Bernoulli **12**(4), 2006, 583–608

Strong approximation for the sums of squares of augmented GARCH sequences

ALEXANDER AUE[1,*], ISTVÁN BERKES[2] and LAJOS HORVÁTH[1,**]

[1]*Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City UT 84112-0090, USA. E-mail: *aue@math.utah.edu; **horvath@math.utah.edu*
[2]*Institut für Statistik, Technische Universität Graz, Steyrergasse 17/IV, A–8010 Graz, Austria. E-mail: berkes@stat.tu-graz.ac.at*

We study so-called augmented GARCH sequences, which include many submodels of considerable interest, such as polynomial and exponential GARCH. To model the returns of speculative assets, it is particularly important to understand the behaviour of the squares of the observations. The main aim of this paper is to present a strong approximation for the sum of the squares. This will be achieved by an approximation of the volatility sequence with a sequence of blockwise independent random variables. Furthermore, we derive a necessary and sufficient condition for the existence of a unique (strictly) stationary solution of the general augmented GARCH equations. Also, necessary and sufficient conditions for the finiteness of moments are provided.

Keywords: augmented GARCH processes; moments; partial sums; stationary solutions; strong approximation

1. Introduction

Starting with the fundamental paper of Engle (1982), autoregressive conditionally heteroscedastic (ARCH) processes and their generalizations have played a central role in finance and macroeconomics. One important class of these generalizations can be subsumed under the notion of augmented GARCH(1,1) processes introduced by Duan (1997). He studied random variables $\{y_k : -\infty < k < \infty\}$ satisfying the equations

$$y_k = \sigma_k \varepsilon_k \tag{1.1}$$

and

$$\Lambda(\sigma_k^2) = c(\varepsilon_{k-1})\Lambda(\sigma_{k-1}^2) + g(\varepsilon_{k-1}), \tag{1.2}$$

where $\Lambda(x)$, $c(x)$ and $g(x)$ are real-valued functions. To solve for σ_k^2, the definitions in (1.1) and (1.2) require that

$$\Lambda^{-1}(x) \text{ exists.} \tag{1.3}$$

Throughout this paper we assume that

$$\{\varepsilon_k : -\infty < k < \infty\} \text{ is a sequence of independent, identically distributed random variables.} \tag{1.4}$$

1350–7265

We start by giving a variety of examples which are included in the framework of the augmented GARCH model. Since all submodels must satisfy (1.1), only the corresponding specific equations (1.2) are stated. Moreover, it can be seen that (1.3) always holds.

Example 1.1 *GARCH*. Bollerslev (1986) introduced the process

$$\begin{aligned}\sigma_k^2 &= \omega + \beta\sigma_{k-1}^2 + \alpha y_{k-1}^2 \\ &= \omega + [\beta + \alpha\varepsilon_{k-1}^2]\sigma_{k-1}^2.\end{aligned}$$

Example 1.2 *AGARCH*. This asymmetric model for volatility was defined by Ding *et al.* (1993) as

$$\begin{aligned}\sigma_k^2 &= \omega + \beta\sigma_{k-1}^2 + \alpha(|y_{k-1}| - \mu y_{k-1})^2 \\ &= \omega + [\beta + \alpha(|\varepsilon_{k-1}| - \mu\varepsilon_{k-1})^2]\sigma_{k-1}^2.\end{aligned}$$

Example 1.3 *VGARCH*. Engle and Ng (1993) used the equations

$$\sigma_k^2 = \omega + \beta\sigma_{k-1}^2 + \alpha(\varepsilon_{k-1} - \mu)^2$$

to study the impact of news on volatility.

Example 1.4 *NGARCH*. This nonlinear asymmetric model was also introduced by Engle and Ng (1993):

$$\sigma_k^2 = \omega + \beta\sigma_{k-1}^2 + \alpha(\varepsilon_{k-1} - \mu)^2\sigma_{k-1}^2.$$

Example 1.5 *GJR-GARCH*. The model of Glosten, Jagannathan and Runke (1993) is given by

$$\begin{aligned}\sigma_k^2 &= \omega + \beta\sigma_{k-1}^2 + \alpha_1 y_{k-1}^2 I\{y_{k-1} < 0\} + \alpha_2 y_{k-1}^2 I\{y_{k-1} \geqslant 0\} \\ &= \omega + \left[\beta + \alpha_1\varepsilon_{k-1}^2 I\{\varepsilon_{k-1} < 0\} + \alpha_2\varepsilon_{k-1}^2 I\{\varepsilon_{k-1} \geqslant 0\}\right]\sigma_{k-1}^2.\end{aligned}$$

Example 1.6 *TSGARCH*. This is a modification of Example 1.1 studied by Taylor (1986) and Schwert (1989):

$$\begin{aligned}\sigma_k &= \omega + \beta\sigma_{k-1} + \alpha_1|y_{k-1}| \\ &= \omega + [\beta + \alpha_1|\varepsilon_{k-1}|]\sigma_{k-1}.\end{aligned}$$

Example 1.7 *PGARCH*. Motivated by a Box–Cox transformation of the conditional variance, Carrasco and Chen (2002) extensively studied the properties of the process satisfying

$$\sigma_k^\delta = \omega + \beta\sigma_{k-1}^\delta + \alpha|y_{k-1}|^\delta$$
$$= \omega + \left[\beta + \alpha|\varepsilon_{k-1}|^\delta\right]\sigma_{k-1}^\delta.$$

Example 1.8 *TGARCH*. Example 1.5 is a non-symmetric version of the GARCH process of Example 1.1. Similarly, a non-symmetric version of Example 1.7 is

$$\sigma_k^\delta = \omega + \beta\sigma_{k-1}^\delta + \alpha_1|y_{k-1}|^\delta I\{y_{k-1} < 0\} + \alpha_2|y_{k-1}|^\delta I\{y_{k-1} \geqslant 0\}$$
$$= \omega + \left[\beta + \alpha_1|\varepsilon_{k-1}|^\delta I\{\varepsilon_{k-1} < 0\} + \alpha_2|\varepsilon_{k-1}|^\delta I\{\varepsilon_{k-1} \geqslant 0\}\right]\sigma_{k-1}^\delta.$$

The special case of $\delta = 1$ was proposed by Taylor (1986) and Schwert (1989) and includes the threshold model of Zakoian (1994).

In Examples 1.1–1.8, $\Lambda(x) = x^\delta$, so these models are usually referred to as polynomial GARCH. Next, we provide two examples of exponential GARCH models.

Example 1.9 *MGARCH*. The multiplicative model of Geweke (1986) is defined as

$$\log\sigma_k^2 = \omega + \beta\log\sigma_{k-1}^2 + \alpha\log y_{k-1}^2$$
$$= \omega + (\alpha + \beta)\log\sigma_{k-1}^2 + \alpha\log\varepsilon_{k-1}^2.$$

Example 1.10 *EGARCH*. The exponential GARCH model of Nelson (1991) is

$$\log\sigma_k^2 = \omega + \beta\log\sigma_{k-1}^2 + \alpha_1\varepsilon_{k-1} + \alpha_2|\varepsilon_{k-1}|.$$

Straumann and Mikosch (2006) obtained a necessary and sufficient condition for the existence and uniqueness of the solution of the recursive equations in Examples 1.1, 1.2 and 1.10. Their method is based on the theory of iterated functions. Carrasco and Chen (2002) studied the existence of solutions in the general framework of (1.1) and (1.2). Since they were also interested in the mixing properties of the sequences $\{y_k\}$ and $\{\sigma_k^2\}$, they assumed that $|c(0)| < 1$, $\mathrm{E}|c(\varepsilon_0)| < \infty$ and $\mathrm{E}|g(\varepsilon_0)| < \infty$. We show that these conditions can be weakened to logarithmic moment conditions. The verification of the mixing property of solutions to (1.1) and (1.2) is based on the theory of hidden Markov models and geometric ergodicity, and it is also assumed that ε_0 has a continuous density which is positive on the whole real line. The exponential mixing proved by Carrasco and Chen (2002) yields the weak convergence as well as the approximation of partial sums of y_i^2 with Brownian motion. However, for these results to hold it should not be necessary to assume either the existence or smoothness of the density of ε_0.

The main aim of this paper is to provide a general method for obtaining strong invariance principles for partial sums of GARCH-type sequences. To this end, we show that $\{\sigma_k^2\}$ can be approximated with a sequence of random variables $\{\tilde{\sigma}_k^2\}$ which are defined such that $\tilde{\sigma}_i^2$ and $\tilde{\sigma}_j^2$ are independent if the distance $|i - j|$ is appropriately large. This is a generalization of the observation in Berkes and Horváth (2001) that the volatility in GARCH(p, q) models can be approximated with random variables which are blockwise independent. It is also known that the general theory of mixing provides a poor bound for the rate of convergence to Brownian motion. We connect this rate of convergence to the finiteness of moments of ε_0. The general blocking method used throughout the proofs yields strong invariance for the partial sums of the random variables $\{y_k^2 - \mathrm{E}y_k^2\}$ with very sharp rates. Our method can, however, also be used to obtain strong approximations for other functions of $y_1, \ldots, y_n$. Additionally, while we are focusing on processes of order (1,1) for the sake of presentation, adaptations of our methods work also for sequences of arbitrary order (p, q).

Strong approximations of partial sums play an important role in, for instance, deriving results in asymptotic statistics. Thus, it has been pointed out by Horváth and Steinebach (2000) that limit results for so-called CUSUM and MOSUM-type test procedures, which are used to detect mean and variance changes, can be based on strong invariance principles. Theorem 2.4 below, in particular, might be used to test for the stability of the observed volatilities. In addition, strong approximations can also be used to develope sequential monitoring procedures (see Berkes *et al.* (2004; Aue and Horváth 2004; Aue *et al.* 2005; and Horváth *et al.* (2006).

The paper is organized as follows. First, we discuss the existence of a unique stationary solution of the augmented GARCH model (1.1) and (1.2). We also find conditions for the existence of the moments of $\Lambda(\sigma_0^2)$ (see Theorems 2.1–2.3 below). These results will be utilized to provide strong approximations for the partial sums of $\{y_k^2\}$ in Theorem 2.4. The proofs of Theorems 2.1–2.3 are given in Section 4, while the strong approximation of Theorem 2.4 is verified in Section 5. In Section 3, we deal with applications of Theorem 2.4 in the field of change-point analysis.

2. Main results

Solving (1.1) and (1.2) backwards, we obtain that, for any $N \geqslant 1$,

$$\Lambda(\sigma_k^2) = \Lambda(\sigma_{k-N}^2) \prod_{1 \leqslant i \leqslant N} c(\varepsilon_{k-i}) + \sum_{1 \leqslant i \leqslant N} g(\varepsilon_{k-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{k-j}). \tag{2.1}$$

This suggests that the general solution of (1.1) and (1.2) is given by

$$\Lambda(\sigma_k^2) = \sum_{1 \leqslant i < \infty} g(\varepsilon_{k-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{k-j}). \tag{2.2}$$

Our first result gives a necessary and sufficient condition for the existence of the random variable

$$X = \sum_{1 \leqslant i < \infty} g(\varepsilon_{-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{-j}). \tag{2.3}$$

Let $\log^+ x = \log(\max\{x, 1\})$.

Theorem 2.1. *We assume that* (1.1)–(1.4) *hold, and that*

$$\mathrm{E}\log^+|g(\varepsilon_0)| < \infty \quad \text{and} \quad \mathrm{E}\log^+|c(\varepsilon_0)| < \infty. \tag{2.4}$$

(i) *If*

$$\mathrm{E}\log|c(\varepsilon_0)| < 0, \tag{2.5}$$

then the sum in (2.3) *is absolutely convergent with probability one.*

(ii) *If*

$$P\{g(\varepsilon_0) = 0\} < 1 \tag{2.6}$$

and the sum in (2.3) *is absolutely convergent with probability one, then* (2.5) *holds.*

We note that in (2.5) we allow that $\mathrm{E}\log|c(\varepsilon_0)| = -\infty$. The result bears some resemblence to Theorem 1.3 in Bougerol and Picard (1992), who concentrate, however, on positive processes and put more emphasis on the problem of irreducibility. The proof of Theorem 2.1 is given in Section 4.

Since under stationarity $\Lambda(\sigma_k^2) \stackrel{\mathcal{D}}{=} X$, the next result gives conditions for the existence of $\mathrm{E}|X|^\nu$ as well as of $\mathrm{E}|\Lambda(\sigma_k^2)|^\nu$.

Theorem 2.2. *We assume that* (1.1)–(1.4) *and* (2.5) *hold,* $\nu > 0$ *and*

$$\mathrm{E}|g(\varepsilon_0)|^\nu < \infty. \tag{2.7}$$

(i) *If*

$$\mathrm{E}|c(\varepsilon_0)|^\nu < 1, \tag{2.8}$$

then

$$\mathrm{E}|X|^\nu < \infty. \tag{2.9}$$

(ii) *If* (2.6) *and* (2.9) *are satisfied, and*

$$c(\varepsilon_0) \geqslant 0 \quad \textit{and} \quad g(\varepsilon_0) \geqslant 0, \tag{2.10}$$

then (2.8) *holds.*

We note that in Examples 1.1–1.8 the left-hand side of the equations defining $\Lambda(\sigma_k^2)$ is always positive, so the right-hand side must also be positive. This requirement restricts the values of possible parameter choices such that (2.10) always holds in these examples. We also point out that Theorem 2.2 is an extension of the results in Nelson (1990) on the existence of the moments of a GARCH(1,1) sequence.

The next result shows that (2.2) is the only solution of (1.1) and (1.2). We say that a strictly stationary solution of (1.1) and (1.2) is non-anticipative if, for any k, the random variable y_k is independent of $\{\varepsilon_i : i > k\}$. According to Brandt (1986), the strictly stationary solution of (1.1) and (1.2) in part (i) of the following theorem is always non-anticipative. The same is true under the assumptions of part (ii) of the same theorem.

Theorem 2.3. *We assume that* (1.1)–(1.4) *and* (2.4) *are satisfied.*

(i) *If* (2.5) *holds, then* (2.2) *is the only stationary solution of* (1.1) *and* (1.2).
(ii) *If, in addition,* (2.6) *and* (2.10) *hold, and* (1.1) *and* (1.2) *have a stationary, non-negative solution, then* (2.5) *must be satisfied.*

Theorem 2.3 gives, in particular, a necessary and sufficient condition for the existence of polynomial GARCH(1,1) and thus generalizes the result of Nelson (1990). The special case of EGARCH introduced in Example 1.10 is also covered by Straumann and Mikosch (2006). Here, stationarity can be characterized as follows. If $|\beta| < 1$ and $\mathrm{E}\log^+|\alpha_1\varepsilon_0 + \alpha_2|\varepsilon_0|| < \infty$, then there is a unique stationary, non-anticipative process satisfying the equation in Example 1.10. If $\beta > 0$ and $P\{\alpha_1\varepsilon_0 + \alpha_2|\varepsilon_0| \geqslant 0\} = 1$, then $\beta < 1$ is necessary and sufficient for the existence of a stationary solution of the EGARCH equations. If we allow anticipative solutions, then (1.1) and (1.2) might have a solution even if (2.5) fails, that is, if $\mathrm{E}\log|c(\varepsilon_0)| \geqslant 0$ (cf. Aue and Horváth 2005).

We now consider approximations for

$$S(k) = \sum_{1\leqslant i\leqslant k} (y_i^2 - \mathrm{E}y_i^2).$$

To this end, we need $\Lambda^{-1}(x)$ to be a smooth function. We make the following assumptions:

$$\Lambda(\sigma_0^2) \geqslant \omega, \qquad \text{for some } \omega > 0; \tag{2.11}$$

and Λ' exists and is non-negative, and there are C, γ such that

$$\left|\frac{1}{\Lambda'(\Lambda^{-1}(x))}\right| \leqslant Cx^{\gamma}, \qquad \text{for all } x \geqslant \omega. \tag{2.12}$$

Assumptions (2.11) and (2.12) hold for the processes in Examples 1.1–1.8, but fail for the remaining Examples 1.9 and 1.10. Thus, more restrictive moment conditions are needed in the exponential GARCH case (see Theorem 2.4(ii) below). Set

$$\overline{\sigma}^2 = \operatorname{var} y_0^2 + 2\sum_{1\leqslant i<\infty} \operatorname{cov}(y_0^2, y_i^2). \tag{2.13}$$

Theorem 2.4. *We assume that* (1.1)–(1.4) *and* (2.5) *hold,* $\overline{\sigma} > 0$, *and*

$$\mathrm{E}|\varepsilon_0|^{8+\delta} < \infty, \qquad \text{for some } \delta > 0. \tag{2.14}$$

(i) *If* (2.11) *and* (2.12) *are satisfied and* (2.7) *holds for some* $\nu > 4(1 + \max\{0, \gamma\})$, *then there is a Wiener process* $\{W(t) : t \geqslant 0\}$ *such that*

$$\sum_{1\leqslant i\leqslant n} (y_i^2 - \mathrm{E}y_i^2) - \overline{\sigma}W(n) = o(n^{5/12+\varepsilon}) \text{ almost surely.} \tag{2.15}$$

for any $\varepsilon > 0$.
(ii) *If* $\Lambda(x) = \log x$, *and we assume that* $\mathrm{E}\exp(t|g(\varepsilon_0)|)$ *exists for some* $t > 4$ *and that there exists*

$$0 < c < 1, \qquad \textit{such that } |c(x)| < c, \tag{2.16}$$

then (2.15) *holds.*

The method of the proof of Theorem 2.4 can be used to establish strong approximations for the sums of functionals of the y_i, too. For example, approximations can be derived for $\sum_{1\leqslant i\leqslant n}(y_i - \mathrm{E}y_i)$, $\sum_{1\leqslant i\leqslant n}(y_i y_{i+k} - \mathrm{E}[y_i y_{i+k}])$ and $\sum_{1\leqslant i\leqslant n}(y_i^2 y_{i+k}^2 - \mathrm{E}[y_i^2 y_{i+k}^2])$, where k is a fixed integer. Examples of strong approximations for dependent sequences other than augmented GARCH processes may be found in Eberlein (1986) and Kuelbs and Philipp (1980). While we are using a blocking argument, the methods in the aforementioned papers are based on examining conditional expectations and variances, and mixing properties, respectively.

3. Applications

In this section, we illustrate the usefulness of Theorem 2.4 with some applications from change-point analysis. One of the main concerns in econometrics is to decide whether the volatility of the underlying variables is stable over time or if it changes in the observation period. While, in general, this allows us to include larger classes of random processes, we will focus on augmented GARCH sequences here. To accommodate the setting, we assume that the volatilities $\{\sigma_k^2\}$ follow a common pattern until an unknown time k^*, called the change-point. After k^* a different regime takes over. It will be demonstrated in the following that, using Theorem 2.4, some of the basic detection methods can be applied to the observed volatilities $\{y_k^2\}$.

Example 3.1. The popular CUSUM procedure is based on the test statistics

$$T_n = \frac{1}{\sqrt{n}\bar{\sigma}} \max_{1\leqslant k\leqslant n} \left| \sum_{1\leqslant i\leqslant k} y_i^2 - \frac{k}{n} \sum_{1\leqslant i\leqslant n} y_i^2 \right|.$$

Under the assumption of no change in the volatility,

$$T_n \xrightarrow{\mathcal{D}} \sup_{0\leqslant t\leqslant 1} |B(t)|, \tag{3.17}$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge. Theorem 2.4 not only gives an alternative way to prove (3.17), but also provides the strong upper bound $\mathcal{O}(n^{-1/12+\varepsilon})$ for the rate of convergence.

Example 3.2. A modification of Example 3.1 is given by the so-called moving-sum or MOSUM procedure. Its basic properties have been presented by Csörgő and Horváth (1997: 181). In our case, the test statistic becomes

$$V_n = \frac{1}{\sqrt{a(n)}\bar{\sigma}} \max_{1\leqslant k\leqslant n-a(n)} \left| \sum_{k<i\leqslant k+a(n)} y_i^2 - \frac{a(n)}{n} \sum_{1\leqslant i\leqslant n} y_i^2 \right|.$$

In addition to the assumptions of Theorem 2.4, let the conditions

$$\frac{n^{5/12+\varepsilon}\sqrt{\log(n/a(n))}}{\sqrt{a(n)}} = \mathcal{O}(1), \qquad \text{for some } \tau > 0, \tag{3.18}$$

and $a(n)/n \to 0$, as $n \to \infty$, be satisfied. Then we obtain the following extreme value asymptotic expression for V_n:

$$\lim_{n\to\infty} P\left\{ A\left(\frac{n}{a(n)}\right) V_n \leq t + D\left(\frac{n}{a(n)}\right) \right\} = \exp(-\mathrm{e}^{-t}) \tag{3.19}$$

for all t, where

$$A(x) = \sqrt{2\log x} \quad \text{and} \quad D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log \pi.$$

We now give a proof of (3.19). Indeed, on using Theorem 2.4, we obtain

$$V_n = \frac{1}{\sqrt{a(n)}} \max_{1 \leq k \leq n-a(n)} \left| W(k + a(n)) - W(k) - \frac{a(n)}{n} W(n) \right| + \mathcal{O}_P\left(\frac{n^{5/12+\varepsilon}}{\sqrt{a(n)}}\right) \tag{3.20}$$

as $n \to \infty$. The moduli of continuity of the Wiener process (see Csörgő and Révész 1981: 30) yield, as $n \to \infty$,

$$\frac{1}{\sqrt{a(n)}} \max_{1 \leq k \leq n-a(n)} |W(k + a(n)) - W(k)| \tag{3.21}$$

$$= \frac{1}{\sqrt{a(n)}} \sup_{0 \leq t \leq n-a(n)} |W(t + a(n)) - W(t)| + \mathcal{O}_P\left(\sqrt{\frac{\log n}{a(n)}}\right).$$

By the scale transformation we obtain for the functional of the Wiener process on the right-hand side of the latter equation

$$\frac{1}{\sqrt{a(n)}} \sup_{0 \leq t \leq n-a(n)} |W(t + a(n)) - W(t)| \stackrel{\mathcal{D}}{=} \sup_{0 \leq s \leq n/a(n)-1} |W(s + 1) - W(s)|. \tag{3.22}$$

Hence, Theorem 7.2.4 of Révész (1990: 72) gives the following extreme value asymptotic expression for the Wiener process:

$$\lim_{n\to\infty} P\left\{ A\left(\frac{n}{a(n)}\right) \sup_{0 \leq s \leq n/a(n)-1} |W(s + 1) - W(s)| \leq t + D\left(\frac{n}{a(n)}\right) \right\} = \exp(\mathrm{e}^{-t}). \tag{3.23}$$

On combining (3.20)–(3.22) with condition (3.18), we arrive at

$$A\left(\frac{n}{a(n)}\right) V_n \stackrel{\mathcal{D}}{=} A\left(\frac{n}{a(n)}\right) \sup_{0 \leq s \leq n/a(n)-1} |W(s + 1) - W(s)| + o_P(1),$$

so that (3.19) follows from (3.23).

Example 3.3. It has been pointed out by Csörgő and Horváth (1997) that weighted versions of the CUSUM statistics have a better power for detecting changes that occur either very early or very late in the observation period. Therefore, T_n from Example 3.1 can be transformed to the weighted statistics

$$T_{n,\alpha} = \frac{1}{\sqrt{n}\overline{\sigma}} \max_{1\leqslant k\leqslant n} \left(\frac{n}{k}\right)^{\alpha} \left| \sum_{1\leqslant i\leqslant k} y_i^2 - \frac{k}{n} \sum_{1\leqslant i\leqslant n} y_i^2 \right|,$$

where $0 \leqslant \alpha < 1/2$. Under the assumptions of Theorem 2.4,

$$T_{n,\alpha} \xrightarrow{\mathcal{D}} \sup_{0<t<1} \frac{|B(t)|}{t^{\alpha}}, \qquad n \to \infty, \tag{3.24}$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ is a Brownian bridge. Theorem 2.4 and the moduli of continuity of the Wiener process (Csörgő and Révész 1981: 30) imply that

$$\sum_{1\leqslant i\leqslant t} (y_i^2 - \mathrm{E}y_0^2) - \overline{\sigma} W(t) = o\left(t^{5/12+\varepsilon}\right) \text{ a.s.} \tag{3.25}$$

as $t \to \infty$. Hence,

$$\frac{1}{\sqrt{n}} \sup_{1/n\leqslant t\leqslant 1} \frac{1}{t^{\alpha}} \left| \sum_{1\leqslant i\leqslant nt} (y_i^2 - \mathrm{E}y_0^2) - \overline{\sigma} W(nt) \right|$$

$$= \mathcal{O}_P\left(\frac{1}{\sqrt{n}} \sup_{1/n\leqslant t\leqslant 1} \frac{(nt)^{5/12+\varepsilon}}{t^{\alpha}} \right) = \begin{cases} \mathcal{O}_P(n^{-1/12+\varepsilon}), & \text{if } \alpha \leqslant 5/12+\varepsilon, \\ \mathcal{O}_P(n^{\alpha-1/2}), & \text{if } 5/12+\varepsilon < \alpha < 1/2, \end{cases}$$

giving that

$$T_{n,\alpha} \stackrel{\mathcal{D}}{=} \sup_{1/n\leqslant t\leqslant 1} \frac{1}{t^{\alpha}} |W(t) - tW(1)| + o_P(1).$$

By the almost sure continuity of $t^{-\alpha}W(t)$, we obtain

$$\sup_{1/n\leqslant t\leqslant 1} \frac{1}{t^{\alpha}} |W(t) - tW(1)| \to \sup_{0<t\leqslant 1} \frac{1}{t^{\alpha}} |W(t) - tW(1)| \text{ a.s.}$$

Since $\{W(t) - tW(1) : 0 \leqslant t \leqslant 1\}$ and $\{B(t) : 0 \leqslant t \leqslant 1\}$ have the same distribution, (3.24) holds.

Example 3.4. Instead of the weighted supremum norm, weighted L_2 functionals can be used. Let

$$Z_n = \frac{1}{n\overline{\sigma}^2} \int_0^1 \frac{1}{t} \left(\sum_{1\leqslant i\leqslant nt} y_i^2 - t \sum_{1\leqslant i\leqslant n} y_i^2 \right)^2 \mathrm{d}t.$$

Under the conditions of Theorem 2.4 we have that

$$Z_n \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{t} \,\mathrm{d}t, \qquad n \to \infty, \tag{3.26}$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge. Using (3.25), we conclude that

$$Z_n = \frac{1}{n}\int_0^t \frac{1}{t}(W(nt) - tW(n))^2 \, dt + \mathcal{O}_P\Big(n^{-1/12+\varepsilon}\Big).$$

Thus, with $B(t) = n^{-1/2}(W(nt) - tW(n))$, we immediately obtain (3.26).

For on-line monitoring procedures designed to detect changes in volatility, see Aue *et al.* (2005). All the applications above contain the parameter $\overline{\sigma}$, which is in general unknown and so has to be estimated from the sample data $y_1, \ldots, y_n$. This can be achieved easily by imposing the method developed by Lo (1991). For further results concerning the estimation of the variance of sums of dependent random variables we refer to Giraitis *et al.* (2003) and Berkes *et al.* (2005).

4. Proofs of Theorems 2.1–2.3

Proof of Theorem 2.1. Part (i) is an immediate consequence of Brandt (1986).

The first step in the proof of (ii) is the observation that by (2.6) there is a constant $a > 0$ such that $P\{|g(\varepsilon_0)| > a\} > 0$. Define the events

$$A_i = \Big\{ |g(\varepsilon_{-i})| \geqslant a, \prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{-j})| \geqslant 1 \Big\}, \qquad i \geqslant 1.$$

We introduce the σ-algebras

$$\mathcal{F}_0 = \{\varnothing, \Omega\},$$
$$\mathcal{F}_i = \sigma(\{\varepsilon_j : -i \leqslant j \leqslant 0\}), \qquad i \geqslant 1. \tag{4.1}$$

Clearly, $A_i \in \mathcal{F}_i$ for all $i \geqslant 1$. Also, by standard arguments,

$$\sum_{1\leqslant i<\infty} P\{A_i|\mathcal{F}_{i-1}\} = P\{|g(\varepsilon_{-i})| \geqslant a\} \sum_{1\leqslant i<\infty} P\Big\{\prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{-j})| \geqslant 1|\mathcal{F}_{i-1}\Big\}$$
$$= P\{|g(\varepsilon_0)| \geqslant a\} \sum_{1\leqslant i<\infty} I\Big\{\sum_{1\leqslant j\leqslant i-1} \log|c(\varepsilon_{-j})| \geqslant 0\Big\}.$$

We claim that

$$\sum_{1\leqslant i<\infty} I\Big\{\sum_{1\leqslant j\leqslant i-1} \log|c(\varepsilon_{-j})| \geqslant 0\Big\} = \infty \text{ a.s.}$$

This follows from the strong law of large numbers if $\mathrm{E}\log|c(\varepsilon_0)| > 0$, while the Chung–Fuchs law (Chow and Teicher 1988: 148) applies if $\mathrm{E}\log|c(\varepsilon_0)| = 0$. Hence, Corollary 3.2 of Durrett (1986, p. 240) gives $P\{A_i \text{ infinitely often}\} = 1$, and therefore

$$P\left\{\lim_{i\to\infty} g(\varepsilon_{-i}) \prod_{1\leqslant j\leqslant i-1} c(\varepsilon_{-j}) = 0\right\} = 0.$$

This contradicts the result that the sum in (2.3) is finite with probability one. □

Proof of Theorem 2.2. If $\nu \geqslant 1$, then by Minkowski's inequality we have

$$\mathrm{E}|X|^\nu \leqslant \left(\sum_{1\leqslant i<\infty}\left[\mathrm{E}\left|g(\varepsilon_{-i})\prod_{1\leqslant j\leqslant i-1} c(\varepsilon_{-j})\right|^\nu\right]^{1/\nu}\right)^\nu$$

$$= \mathrm{E}|g(\varepsilon_0)|^\nu\left(\sum_{1\leqslant i<\infty}[\mathrm{E}|c(\varepsilon_0)|^\nu]^{(i-1)/\nu}\right)^\nu < \infty$$

by assumptions (2.7) and (2.8). If $0 < \nu < 1$, then (see Hardy *et al.* 1959)

$$\mathrm{E}|X|^\nu \leqslant \sum_{1\leqslant i<\infty}\mathrm{E}\left|g(\varepsilon_{-i})\prod_{1\leqslant j\leqslant i-1} c(\varepsilon_{-j})\right|^\nu = \mathrm{E}|g(\varepsilon_0)|^\nu \sum_{1\leqslant i<\infty}[\mathrm{E}|c(\varepsilon_0)|^\nu]^{i-1} < \infty,$$

completing the proof of the first part of Theorem 2.2.

If $\nu \geqslant 1$, then by the lower bound in the Minkowski inequality (Hardy *et al.* 1959) and (2.10) we obtain

$$\mathrm{E}|X|^\nu \geqslant \sum_{1\leqslant i<\infty}\mathrm{E}\left[g(\varepsilon_{-i})\prod_{1\leqslant j\leqslant i-1} c(\varepsilon_{-j})\right]^\nu = \mathrm{E}[g(\varepsilon_0)]^\nu \sum_{1\leqslant i<\infty}(\mathrm{E}[c(\varepsilon_0)]^\nu)^{i-1}.$$

Since by (2.6) and (2.7) we have $0 < \mathrm{E}[g(\varepsilon_0)]^\nu < \infty$, we obtain (2.8).

Similarly, if $0 < \nu < 1$, then

$$\mathrm{E}|X|^\nu \geqslant \left(\sum_{1\leqslant i<\infty}\left(\mathrm{E}\left[g(\varepsilon_{-i})\prod_{1\leqslant j\leqslant i-1} c(\varepsilon_{-j})\right]^\nu\right)^{1/\nu}\right)^\nu$$

$$= \mathrm{E}[g(\varepsilon_0)]^\nu\left(\sum_{1\leqslant i<\infty}(\mathrm{E}[c(\varepsilon_0)]^\nu)^{(i-1)/\nu}\right)^\nu$$

and therefore (2.8) holds. □

Proof of Theorem 2.3. The first part is an immediate consequence of Brandt (1986).

The second part is based on (2.1) with $k = 0$. Since by (2.10) all terms in (2.1) are non-negative, we obtain for all $N \geqslant 1$ that

$$\Lambda(\sigma_0^2) \geqslant \sum_{1 \leqslant i \leqslant N} g(\varepsilon_{-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{-j}). \tag{4.2}$$

If $\mathrm{E} \log c(\varepsilon_0) \geqslant 0$, then the proof of Theorem 2.1(ii) gives that

$$\limsup_{i \to \infty} g(\varepsilon_{-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{-j}) \geqslant a,$$

for some $a > 0$. So by (2.10) we obtain that

$$\lim_{N \to \infty} \sum_{1 \leqslant i \leqslant N} g(\varepsilon_{-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{-j}) = \infty \text{ a.s.},$$

contradicting (4.2). □

5. Proof of Theorem 2.4

Since the proof of Theorem 2.4 is divided into a series of lemmas, we will first outline its structure. The main ideas can be highlighted as follows.

The first goal of the proof is to show that the partial sums of $\{y_k^2\}$ can be approximated by the partial sums of a sequence of random variables $\{\tilde{y}_k^2\}$ which are independent at sufficiently large lags. To do so, we establish exponential inequalities for the distance between the volatilities $\{\sigma_k^2\}$ and a suitably constructed sequence $\{\tilde{\sigma}_k^2\}$. Consequently, letting $\tilde{y}_k = \tilde{\sigma}_k \varepsilon_k$, it suffices to find an approximation for the partial sums $\sum_{k=1}^n (\tilde{y}_k^2 - \mathrm{E}\tilde{y}_k^2)$. Detailed proofs are given in Lemmas 5.1–5.4 below.

Set $\tilde{\xi}_k = \tilde{y}_k^2 - \mathrm{E}\tilde{y}_k^2$. The partial sum $\sum_{k=1}^n \tilde{\xi}_k$ will be blocked into random variables belonging to two subgroups. The random variables associated with each subgroup form an independent sequence which allows the application of the following simplified version of Theorem 4.4 in Strassen (1967: 334).

Theorem A. *Let $\{Z_i\}$ be a sequence of independent centred random variables and let $0 < \kappa < 1$. If*

$$r(N) = \sum_{1 \leqslant i \leqslant N} \mathrm{E} Z_i^2 \to \infty, \qquad N \to \infty,$$

$$\sum_{1 \leqslant N < \infty} r(N)^{-2\kappa} \mathrm{E} Z_N^4 < \infty,$$

then there is a Wiener process $\{W(t) : t \geqslant 0\}$ such that

$$\sum_{1 \leqslant i \leqslant N} Z_i - W(r(N)) = o\left(r(N)^{(\kappa+1)/4} \log r(N)\right) \text{ a.s.}, \qquad N \to \infty.$$

Lemmas 5.5–5.9 help prove that the assumptions of Theorem A are satisfied, while the actual approximations are given in Lemmas 5.10–5.12.

Let $0 < \rho < 1$ and define $\tilde{\sigma}_k^2$ as the solution of

$$\Lambda(\tilde{\sigma}_k^2) = \sum_{1 \leqslant i \leqslant k^\rho} g(\varepsilon_{k-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{k-j}).$$

First, we obtain an exponential inequality for $\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)$.

Lemma 5.1. *If* (1.1)–(1.4), (2.5), (2.7) *and* (2.8) *hold, then there are constants* $C_{1,1}$, $C_{1,2}$ *and* $C_{1,3}$ *such that*

$$P\big\{|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| \geqslant \exp(-C_{1,1} k^\rho)\big\} \leqslant C_{1,2} \exp(-C_{1,3} k^\rho).$$

Proof. Observe that

$$\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2) = \sum_{k^\rho + 1 \leqslant i < \infty} g(\varepsilon_{k-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{k-j}) = \prod_{1 \leqslant i \leqslant k^\rho} c(\varepsilon_{k-i}) \Lambda(\sigma_{k-k^\rho}^2).$$

Using Theorem 2.7 of Petrov (1995) yields

$$P\left\{\prod_{1 \leqslant i \leqslant k^\rho} |c(\varepsilon_{k-i})| \geqslant \exp(-C_{1,4} k^\rho)\right\} = P\left\{\sum_{1 \leqslant i \leqslant k^\rho} (\log|c(\varepsilon_{k-i})| - a) \geqslant -(C_{1,4} + a) k^\rho\right\}$$
$$\leqslant 2 \exp(-C_{1,5} k^\rho),$$

where $\mathrm{E} \log|c(\varepsilon_0)| = a < 0$ and $0 < C_{1,4} < -a$. Theorem 2.2 implies

$$P\left\{|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| \geqslant \exp\left(-\frac{C_{1,4}}{2} k^\rho\right)\right\}$$
$$\leqslant P\left\{\prod_{1 \leqslant i \leqslant k^\rho} |c(\varepsilon_{k-i})| \geqslant \exp(-C_{1,4} k^\rho)\right\} + P\left\{|\Lambda(\sigma_{k-k^\rho}^2)| \geqslant \exp\left(\frac{C_{1,4}}{2} k^\rho\right)\right\}$$
$$\leqslant 2 \exp(-C_{1,5} k^\rho) + \exp\left(-\frac{\nu C_{1,4}}{2} k^\rho\right) \mathrm{E}|\Lambda(\sigma_0^2)|^\nu \leqslant C_{1,6} \exp(-C_{1,7} k^\rho).$$

This completes the proof. □

Lemma 5.2. *If* (1.1)–(1.4), (2.5), (2.7), (2.8), (2.11) *and* (2.12) *hold, then there are constants* $C_{2,1}$, $C_{2,2}$ *and* $C_{2,3}$ *such that*

$$P\big\{|\sigma_k^2 - \tilde{\sigma}_k^2| \geqslant \exp(-C_{2,1} k^\rho)\big\} \leqslant C_{2,2} \exp(-C_{2,3} k^\rho).$$

Proof. By the mean-value theorem,

$$\sigma_k^2 - \tilde{\sigma}_k^2 = \Lambda^{-1}(\Lambda(\sigma_k^2)) - \Lambda^{-1}(\Lambda(\tilde{\sigma}_k^2)) = \frac{1}{\Lambda'(\Lambda^{-1}(\xi_k))} \big(\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)\big),$$

where ξ_k is between $\Lambda(\sigma_k^2)$ and $\Lambda(\tilde{\sigma}_k^2)$. By (2.12),

$$|\sigma_k^2 - \tilde{\sigma}_k^2| \leqslant C\big||\Lambda(\sigma_k^2)|^\gamma + |\Lambda(\tilde{\sigma}_k^2)|^\gamma\big||\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)|. \tag{5.1}$$

If $\gamma \leqslant 0$, then Lemma 5.2 follows from (2.11) in combination with Lemma 5.1. If $\gamma > 0$, then Theorem 2.2 and Lemma 5.1 yield

$$\begin{aligned}
&P\Big\{\big(|\Lambda(\sigma_k^2)|^\gamma + |\Lambda(\tilde{\sigma}_k^2)|^\gamma\big)|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| \geqslant \exp\Big(-\frac{C_{1,1}}{2}k^\rho\Big)\Big\} \\
&\quad\leqslant P\Big\{|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| \geqslant \exp(-C_{1,1}k^\rho)\Big\} + P\Big\{|\Lambda(\sigma_k^2)|^\gamma + |\Lambda(\tilde{\sigma}_k^2)|^\gamma \geqslant \exp\Big(\frac{C_{1,1}}{2}k^\rho\Big)\Big\} \\
&\quad\leqslant 2C_{1,2}\exp(-C_{1,3}k^\rho) + P\Big\{|\Lambda(\sigma_k^2)|^\nu \geqslant \Big(\frac{1}{2}\exp\Big(\frac{C_{1,1}}{2}k^\rho\Big)\Big)^{\nu/\gamma}\Big\} \\
&\qquad + P\Big\{\big(|\Lambda(\sigma_k^2)| + \exp(-C_{1,1}k^\rho)\big)^\nu \geqslant \Big(\frac{1}{2}\exp\Big(\frac{C_{1,1}}{2}k^\rho\Big)\Big)^{\nu/\gamma}\Big\} \\
&\quad\leqslant C_{2,4}\exp(-C_{2,5}k^\rho),
\end{aligned}$$

finishing the proof. □

Lemma 5.3. *If* (1.1)–(1.4), (2.5), (2.7) *and* (2.8) *hold and* $\Lambda(x) = \log x$, *then there are constants* $C_{3,1}$ *and* $C_{3,2}$ *such that*

$$P\big\{|\sigma_k^2 - \tilde{\sigma}_k^2| \geqslant \exp(-C_{3,1}k^\rho)\big\} \leqslant C_{3,2}k^{-\rho\nu}.$$

Proof. By the mean-value theorem, $\sigma_k^2 - \tilde{\sigma}_k^2 = \mathrm{e}^{\log\sigma_k^2} - \mathrm{e}^{\log\tilde{\sigma}_k^2} = \mathrm{e}^{\xi_k}(\log\sigma_k^2 - \log\tilde{\sigma}_k^2)$, where ξ_k is between $\log\sigma_k^2$ and $\log\tilde{\sigma}_k^2$. Hence,

$$|\sigma_k^2 - \tilde{\sigma}_k^2| \leqslant \Big(\mathrm{e}^{\log\sigma_k^2} + \mathrm{e}^{\log\tilde{\sigma}_k^2}\Big)|\log\sigma_k^2 - \log\tilde{\sigma}_k^2|. \tag{5.2}$$

On the set $-1 < \log\sigma_k^2 - \log\tilde{\sigma}_k^2 < 1$, we have $\exp(\log\tilde{\sigma}_k^2) \leqslant 3\exp(\log\sigma_k^2)$, so by Theorem 2.2 and Lemma 5.1, we obtain

$$P\left\{|\sigma_k^2 - \tilde{\sigma}_k^2| \geqslant \exp\left(-\frac{C_{1,1}}{2} k^\rho\right)\right\}$$
$$\leqslant P\left\{3\mathrm{e}^{\log \sigma_k^2} |\log \sigma_k^2 - \log \tilde{\sigma}_k^2| \geqslant \exp\left(-\frac{C_{1,1}}{2} k^\rho\right)\right\} + C_{1,2} \exp(-C_{1,3} k^\rho)$$
$$\leqslant P\{|\log \sigma_k^2 - \log \tilde{\sigma}_k^2| \geqslant \exp(-C_{1,1} k^\rho)\}$$
$$+ P\left\{3\mathrm{e}^{\log \sigma_k^2} \geqslant \exp\left(\frac{C_{1,1}}{2} k^\rho\right)\right\} + C_{1,2} \exp(-C_{1,3} k^\rho)$$
$$\leqslant C_{1,2} \exp(-C_{1,3} k^\rho) + P\left\{\log \sigma_k^2 \geqslant \frac{C_{1,1}}{2} k^\rho - \log 3\right\} + C_{1,2} \exp(-C_{1,3} k^\rho)$$
$$\leqslant C_{1,2} \exp(-C_{1,3} k^\rho) + P\left\{|\log \sigma_k^2|^\nu \geqslant \left(\frac{C_{1,1}}{2} k^\rho - \log 3\right)^\nu\right\} + C_{1,2} \exp(-C_{1,3} k^\rho)$$
$$\leqslant C_{3,3} k^{-\rho\nu}.$$

This completes the proof. □

Let

$$\tilde{y}_i = \tilde{\sigma}_i \varepsilon_i, \qquad -\infty < i < \infty.$$

It is clear that $\tilde{y}_i$ and $\tilde{y}_j$ are independent if $i < j - j^\rho$. The next result shows that it is enough to approximate the sum of the $\tilde{y}_i^2$.

Lemma 5.4. *If the conditions of Theorem 2.4 are satisfied, then*

$$\max_{1 \leqslant k < \infty} \left| \sum_{1 \leqslant i \leqslant k} y_i^2 - \sum_{1 \leqslant i \leqslant k} \tilde{y}_i^2 \right| = \mathcal{O}(1) \ a.s.$$

Proof. Condition (2.14) implies that

$$|\varepsilon_k| = o\left(k^{1/(8+\delta)}\right) \text{ a.s.}$$

as $k \to \infty$. By Lemmas 5.2 and 5.3 we obtain

$$|\sigma_k^2 - \tilde{\sigma}_k^2| = \mathcal{O}(\exp(-C_{4,1} k^\rho)) \text{ a.s.}$$

Thus we conclude that

$$\max_{1 \leqslant k < \infty} \left| \sum_{1 \leqslant i \leqslant k} y_i^2 - \sum_{1 \leqslant i \leqslant k} \tilde{y}_i^2 \right| \leqslant \sum_{1 \leqslant i < \infty} |y_i^2 - \tilde{y}_i^2| = \sum_{1 \leqslant i < \infty} |\sigma_i^2 - \tilde{\sigma}_i^2| \varepsilon_i^2$$
$$\overset{\text{a.s.}}{=} \mathcal{O}(1) \sum_{1 \leqslant i < \infty} i^{2/(8+\delta)} \exp(-C_{4,1} i^\rho) \overset{\text{a.s.}}{=} \mathcal{O}(1),$$

and the proof is complete. □

We note that under the conditions of Theorem 2.4 we have

$$P\{|\sigma_k^2 - \tilde{\sigma}_k^2| \geqslant \exp(-C_{4,2}k^\rho)\} \leqslant C_{4,3}k^{-\mu}, \tag{5.3}$$

where $\mu = \rho\nu$ when $\Lambda(x) = \log x$, while μ can be chosen arbitrarily when (2.11) and (2.12) hold. In both cases μ can be chosen arbitrarily large.

Lemma 5.5. *If the conditions of Theorem 2.4 are satisfied, then*

$$\mathrm{E}(\tilde{\sigma}_k^2)^{4+\delta} = \mathcal{O}(1), \textit{ for some } \delta > 0, \tag{5.4}$$

and

$$\mathrm{E}(\sigma_k^2 - \tilde{\sigma}_k^2)^2 = \mathcal{O}\left(k^{-\mu/2}\right),$$

where μ *is defined in* (5.3).

Proof. We first show that $\mathrm{E}|\sigma_k^2 - \tilde{\sigma}_k^2|^{4+\delta} = O(1)$. Assume first that (2.11) and (2.12) are satisfied. Using (5.1), we obtain in case of $\gamma > 0$ that

$$|\sigma_k^2 - \tilde{\sigma}_k^2|^{4+\delta} \leqslant C\big(|\Lambda(\sigma_k^2)|^{(4+\delta)(1+\gamma)} + |\Lambda(\tilde{\sigma}_k^2)|^{(4+\delta)(1+\gamma)}$$
$$+ |\Lambda(\sigma_k^2)|^{(4+\delta)\gamma}|\Lambda(\tilde{\sigma}_k^2)|^{4+\delta} + |\Lambda(\sigma_k^2)|^{4+\delta}|\Lambda(\tilde{\sigma}_k^2)|^{(4+\delta)\gamma}\big).$$

It follows from our assumptions that $\mathrm{E}|\Lambda(\sigma_k^2)|^{(4+\delta)(1+\gamma)} \leqslant C_{5,1}$. Applying the definition of $\tilde{\sigma}_k^2$, we obtain from the proof of Theorem 2.2 that

$$\mathrm{E}|\Lambda(\tilde{\sigma}_k^2)|^{(4+\delta)(1+\gamma)} \leqslant \mathrm{E}\left(\sum_{1\leqslant i\leqslant k^\rho} |g(\varepsilon_{k-i})| \prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{k-j})|\right)^{(4+\delta)(1+\gamma)}$$
$$\leqslant \mathrm{E}\left(\sum_{1\leqslant i\leqslant \infty} |g(\varepsilon_{k-i})| \prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{k-j})|\right)^{(4+\delta)(1+\gamma)} \leqslant C_{5,2}.$$

Similarly,

$$|\Lambda(\sigma_k^2)|^{(4+\delta)\gamma}|\Lambda(\tilde{\sigma}_k^2)|^{4+\delta}$$
$$\leqslant \left(\sum_{1\leqslant i\leqslant \infty} |g(\varepsilon_{k-i})| \prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{k-j})|\right)^{(4+\delta)\gamma} \left(\sum_{1\leqslant i\leqslant k^\rho} |g(\varepsilon_{k-i})| \prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{k-j})|\right)^{4+\delta}$$
$$\leqslant \left(\sum_{1\leqslant i\leqslant \infty} |g(\varepsilon_{k-i})| \prod_{1\leqslant j\leqslant i-1} |c(\varepsilon_{k-j})|\right)^{(4+\delta)(1+\gamma)}$$

and therefore $\mathrm{E}|\Lambda(\sigma_k^2)|^{(4+\delta)\gamma}|\Lambda(\tilde{\sigma}_k^2)|^{4+\delta} \leqslant C_{5,3}$. The same argument similarly implies that $\mathrm{E}|\Lambda(\sigma_k^2)|^{4+\delta}|\Lambda(\tilde{\sigma}_k^2)|^{(4+\delta)\gamma} \leqslant C_{5,4}$. If $\gamma < 0$, then (5.1) and (2.11) yield

$$|\sigma_k^2 - \tilde{\sigma}_k^2|^{4+\delta} \leqslant C_{5,5}|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)|^{4+\delta}$$

and similarly $\mathrm{E}|\Lambda(\sigma_k^2)|^{4+\delta} \leqslant C_{5,6}$ as well as $\mathrm{E}|\Lambda(\tilde{\sigma}_k^2)|^{4+\delta} \leqslant C_{5,6}$.

If $\Lambda(x) = \log x$, then by (5.2) we have

$$|\sigma_k^2 - \tilde{\sigma}_k^2|^{4+\delta} \leqslant C_{5,7}\left(\mathrm{e}^{(4+\delta)\log \sigma_k^2} + \mathrm{e}^{(4+\delta)\log \tilde{\sigma}_k^2}\right)\left(|\log \sigma_k^2|^{4+\delta} + |\log \tilde{\sigma}_k^2|^{4+\delta}\right).$$

So, it is enough to prove that

$$\mathrm{E}\mathrm{e}^{\eta \log \sigma_k^2} \leqslant C_{5,8} \quad \text{and} \quad \mathrm{E}\mathrm{e}^{\eta \log \tilde{\sigma}_k^2} \leqslant C_{5,8}, \tag{5.5}$$

for some $\eta > 4$ and $C_{5,8} = C_{5,8}(\eta)$. By assumption (2.16),

$$|\Lambda(\sigma_k^2)| \leqslant \left| \sum_{1 \leqslant i < \infty} g(\varepsilon_{k-i}) \prod_{1 \leqslant j \leqslant i-1} c(\varepsilon_{k-j}) \right| \leqslant \sum_{1 \leqslant i < \infty} c^{i-1}|g(\varepsilon_{k-i})|.$$

Now $M(t) = \mathrm{E}\exp(t|g(\varepsilon_0)|)$ exists for some $t > 4$ and therefore

$$\mathrm{E}\exp(t|\Lambda(\sigma_k^2)|) \leqslant \prod_{1 \leqslant i < \infty} M(tc^{i-1}) = \exp\left(\sum_{1 \leqslant i < \infty} \log M(tc^{i-1}) \right).$$

Since $0 < c < 1$, we see that $tc^i \to 0$ as $i \to \infty$. For any $x > 0$, $\log(1 + x) \leqslant x$ and $M(x) = 1 + \mathcal{O}(x)$ as $x \to 0$, so we obtain the first part of (5.5). A similar argument applies to the second statement.

For $k \geqslant 1$, define the events $A_k = \left\{|\sigma_k^2 - \tilde{\sigma}_k^2| \leqslant \exp(-C_{5,9}k^\rho)\right\}$ and write

$$(\sigma_k^2 - \tilde{\sigma}_k^2)^2 = (\sigma_k^2 - \tilde{\sigma}_k^2)^2 I\{A_k\} + (\sigma_k^2 - \tilde{\sigma}_k^2)^2 I\{A_k^c\}.$$

Clearly,

$$\mathrm{E}[(\sigma_k^2 - \tilde{\sigma}_k^2)^2 I\{A_k\}] = \mathcal{O}(\exp(-C_{5,9}k^\rho)).$$

Applying (5.3), (5.4) and the Cauchy–Schwarz inequality, we arrive at

$$\mathrm{E}\left[(\sigma_k^2 - \tilde{\sigma}_k^2)^2 I\{A_k^c\}\right] \leqslant \left(\mathrm{E}|\sigma_k^2 - \tilde{\sigma}_k^2|^4\right)^{1/2} P(A_k^c)^{1/2} = \mathcal{O}\left(k^{-\mu/2}\right),$$

completing the proof of Lemma 5.5. □

Lemma 5.6. *If the conditions of Theorem 2.4 are satisfied, then*

$$\sum_{1 \leqslant k < \infty} \mathrm{cov}(y_0^2, y_k^2)$$

is absolutely convergent.

Proof. Since y_0 and $\tilde{y}_k$ are independent for all $k > 1$, we obtain that

$$\begin{aligned}E(y_0^2 - Ey_0^2)(y_k^2 - Ey_k^2) &= E(y_0^2 - Ey_0^2)(y_k^2 - \tilde{y}_k^2 + \tilde{y}_k^2 - Ey_k^2)\\ &= E(y_0^2 - Ey_0^2)(\tilde{y}_k^2 - Ey_k^2) + E(y_0^2 - Ey_0^2)(y_k^2 - \tilde{y}_k^2)\\ &= E(y_0^2(y_k^2 - \tilde{y}_k^2)) - Ey_0^2 E(y_k^2 - \tilde{y}_k^2).\end{aligned}$$

Since $y_k^2 - \tilde{y}_k^2 = (\sigma_k^2 - \tilde{\sigma}_k^2)\varepsilon_k^2$ and the random variables $\sigma_k^2 - \tilde{\sigma}_k^2$, ε_k^2 are independent, Lemma 5.5 and Cauchy's inequality give $|\mathrm{cov}(y_0^2, y_k^2)| = \mathcal{O}(k^{-\mu/4})$, so Lemma 5.6 is proved on account of $\mu/4 > 1$. □

Lemma 5.7. *If the conditions of Theorem 2.4 are satisfied, then there is a constant* $C_{7,1}$ *such that*

$$E\left(\sum_{n<k\leqslant n+m} (\tilde{y}_k^2 - E\tilde{y}_k^2)\right)^2 \leqslant C_{7,1}\, m$$

for all $n, m \geqslant 1$.

Proof. Set $\xi_k = y_k^2 - Ey_k^2$ and $\tilde{\xi}_k = \tilde{y}_k^2 - E\tilde{y}_k^2$. It is easy to see that

$$E\left(\sum_{n<k\leqslant n+m} \tilde{\xi}_k\right)^2 = \sum_{n<k,l\leqslant n+m} E\xi_k\xi_l + \sum_{n<k,l\leqslant n+m} E\tilde{\xi}_k(\tilde{\xi}_l - \xi_l) + \sum_{n<k,l\leqslant n+m} E\xi_l(\tilde{\xi}_k - \xi_k).$$

Lemma 5.5 and Cauchy's inequality imply that

$$\begin{aligned}\sum_{n<k,l\leqslant n+m} E\left|\tilde{\xi}_k(\tilde{\xi}_l - \xi_l)\right| &\leqslant \sum_{n<k,l\leqslant n+m} \left(E\tilde{\xi}_k^2 E(\tilde{\xi}_1 - \xi_1)^2\right)^{1/2}\\ &\leqslant \sum_{n<k,l\leqslant n+m} \left(E\tilde{\sigma}_k^4 E\varepsilon_k^4\right)^{1/2}\left(E\varepsilon_l^4 E(\sigma_l^2 - \tilde{\sigma}_l^2)^2\right)^{1/2}\\ &= \mathcal{O}(1) \sum_{n<k,l\leqslant n+m} l^{-\mu/4} = \mathcal{O}(1)m \sum_{n<l<\infty} l^{-\mu/4} = \mathcal{O}(1)m.\end{aligned}$$

Similarly,

$$\sum_{n\leqslant k,l\leqslant n+m} E|\xi_l(\tilde{\xi}_k - \xi_k)| = \mathcal{O}(1)m.$$

By the stationarity of $\{\xi_k : -\infty < k < \infty\}$, we have

$$\sum_{n\leqslant k,l\leqslant n+m} E\xi_k\xi_l = \sum_{n\leqslant k,l\leqslant n+m} E\xi_0\xi_{|k-l|} = \mathcal{O}(1)m$$

by an application of Lemma 5.6. □

Lemma 5.8. *If the conditions of Theorem 2.4 are satisfied, then there are constants* n_0, m_0 *and* $C_{8,1} > 0$ *such that*

$$\mathrm{E}\left(\sum_{n<k\leqslant n+m}(\tilde{y}_k^2 - \mathrm{E}\tilde{y}_k^2)\right)^2 \geqslant C_{8,1}m$$

for all $n \geqslant n_0$ *and* $m \geqslant m_0$.

Proof. Following the proof of Lemma 5.7, we obtain

$$\left|\mathrm{E}\left(\sum_{n<k\leqslant n+m}\tilde{\xi}_k\right)^2 - \sum_{n<k,l\leqslant n+m}\mathrm{E}\xi_k\xi_l\right| \leqslant C_{8,2}m\sum_{n<l<\infty} l^{-\mu/4}.$$

Since

$$\frac{1}{m}\sum_{n<k,l\leqslant n+m}\mathrm{E}\xi_k\xi_l = \frac{1}{m}\sum_{n<k,l\leqslant n+m}\mathrm{E}\xi_0\xi_{|k-l|} \to \operatorname{var} y_0^2 + 2\sum_{1\leqslant k<\infty}\operatorname{cov}(y_0^2, y_k^2) = \overline{\sigma} > 0,$$

the proof is complete. □

Lemma 5.9. *If the conditions of Theorem 2.4 are satisfied, then there is a constant* $C_{9,1}$ *such that*

$$\mathrm{E}\left(\sum_{n<k\leqslant n+m}(\tilde{y}_k^2 - \mathrm{E}\tilde{y}_k^2)\right)^4 \leqslant C_{9,1}m^2 \tag{5.6}$$

for all n, $m \geqslant 1$.

Proof. We prove (5.6) using induction. If $m = 1$, then by (5.4) we have $\mathrm{E}(\tilde{y}_k^2 - \mathrm{E}\tilde{y}_k^2)^4 \leqslant C_{9,2}$, for all $k \geqslant 1$. We assume that (5.6) holds for m and for all n. We have to show that

$$\mathrm{E}\left(\sum_{n<k\leqslant n+m+1}\tilde{\xi}_k\right)^4 \leqslant C_{9,1}(m+1)^2, \qquad \text{for all } n. \tag{5.7}$$

To this end, let $\lfloor\cdot\rfloor$ denote integer part, $m_\pm(\rho) = m/2 \pm 2m^\rho$ and set

$$S_1(n, m) = \sum_{k=n+1}^{n+m_-(\rho)}\tilde{\xi}_k, \qquad S_2(n, m) = \sum_{k=n+m_+(\rho)+1}^{n+m+1}\tilde{\xi}_k, \qquad S_3(n, m) = \sum_{k=n+m_-(\rho)}^{n+m_+(\rho)}\tilde{\xi}_k.$$

By the triangle inequality, we have

$$\left(\mathrm{E}\left[\sum_{1\leqslant j\leqslant 3}S_j(n, m)\right]^4\right)^{1/4} \leqslant \left(\mathrm{E}[S_1(n, m) + S_2(n, m)]^4\right)^{1/4} + \left(\mathrm{E}[S_3(n, m)]^4\right)^{1/4}.$$

Also, by (5.4),

$$\left(\mathrm{E}[S_3(n,\ m)]^4\right)^{1/4} \leqslant C_{9,3}m^{\rho}.$$

Let $A_k = \left\{|\sigma_k^2 - \tilde{\sigma}_k^2| \leqslant \exp(-C_{9,4}k^{\rho})\right\}$ and write $\tilde{\xi}_k = \xi_k + (\tilde{\xi}_k - \xi_k)I\{A_k\} + (\tilde{\xi}_k - \xi_k)I\{A_k^c\}$. Let $U = \{n+1,\ n+2,\ \ldots,\ n + \lfloor m_-(\rho)\rfloor\} \cup \{n + \lfloor m_+(\rho)\rfloor,\ \ldots,\ n+m+1\}$. Then,

$$\left(\mathrm{E}[S_1(n,\ m) + S_2(n,\ m)]^4\right)^{1/4}$$

$$\leqslant \left(\mathrm{E}\left[\sum_{k\in U}\xi_k\right]^4\right)^{1/4} + \left(\mathrm{E}\left[\sum_{k\in U}(\tilde{\xi}_k - \xi_k)I\{A_k\}\right]^4\right)^{1/4} + \left(\mathrm{E}\left[\sum_{k\in U}(\tilde{\xi}_k - \xi_k)I\{A_k^c\}\right]^4\right)^{1/4}.$$

It follows from the definition of the A_k that

$$\mathrm{E}\left[\sum_{k\in U}(\tilde{\xi}_k - \xi_k)I\{A_k\}\right]^4 \leqslant C_{9,5}.$$

Moreover, by (5.4), (5.3) and Minowski's and Hölder's inequalities we obtain that

$$\left(\mathrm{E}\left(\sum_{k\in U}(\tilde{\xi}_k - \xi_k)I\{A_k^c\}\right)^4\right)^{1/4} \leqslant \sum_{k\in U}\mathrm{E}\left((\tilde{\xi}_k - \xi_k)^4 I\{A_k^c\}\right)^{1/4}$$

$$\leqslant \sum_{k\in U}(\mathrm{E}(\tilde{\xi}_k - \xi_k)^{4+\delta})^{1/(4+\delta)}(P\{A_k^c\})^{\delta/(4+\delta)} \leqslant C_{9,6}\sum_{1\leqslant k<\infty}k^{-\mu\delta/(4+\delta)} \leqslant C_{9,7}.$$

Observe that by the stationarity of the ξ_k,

$$\mathrm{E}\left[\sum_{k\in U}\xi_k\right]^4 = \mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)}\xi_k + \sum_{m_+(\rho)\leqslant k\leqslant m+1}\xi_k\right]^4.$$

Arguing as before, we estimate

$$\left(\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)}\xi_k + \sum_{m_+(\rho)\leqslant k\leqslant m+1}\xi_k\right]^4\right)^{1/4} \leqslant \left(\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_+(\rho)}\tilde{\xi}_k + \sum_{m_+(\rho)\leqslant k\leqslant m+1}\xi_k\right]^4\right)^{1/4} + C_{9,8}.$$

Since the partial sums $\sum_{1\leqslant k\leqslant m_-(\rho)}\tilde{\xi}_k$ and $\sum_{m_+(\rho)\leqslant k\leqslant m+1}\tilde{\xi}_k$ are independent, if $m \geqslant m_0$ for some m_0 and $\mathrm{E}\tilde{\xi}_k = 0$, we obtain that

$$\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)}\tilde{\xi}_k + \sum_{m_+(\rho)\leqslant k\leqslant m+1}\tilde{\xi}_k\right]^4$$

$$= \mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)}\tilde{\xi}_k\right]^4 + \mathrm{E}\left[\sum_{m_+(\rho)\leqslant k\leqslant m+1}\tilde{\xi}_k\right]^4 + 6\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)}\tilde{\xi}_k\right]^2\mathrm{E}\left[\sum_{m_+(\rho)\leqslant k\leqslant m+1}\tilde{\xi}_k\right]^2.$$

Using the induction assumption we obtain

$$\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)} \tilde{\xi}_k\right]^4 \leqslant C_{9,1}\frac{m^2}{4} \quad \text{and} \quad \mathrm{E}\left[\sum_{m_+(\rho)\leqslant k\leqslant m+1} \tilde{\xi}_k\right]^4 \leqslant C_{9,1}\frac{m^2}{4}.$$

Lemma 5.7 yields

$$6\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)} \tilde{\xi}_k\right]^2 \mathrm{E}\left[\sum_{m_+(\rho)\leqslant k\leqslant m+1} \tilde{\xi}_k\right]^2 \leqslant 6C_{7,1}^2\frac{m^2}{4}.$$

Since we can assume that $C_{9,1} \geqslant 6C_{7,1}^2$, we conclude that

$$\mathrm{E}\left[\sum_{1\leqslant k\leqslant m_-(\rho)} \tilde{\xi}_k + \sum_{m_+(\rho)\leqslant k\leqslant m+1} \tilde{\xi}_k\right]^4 \leqslant \frac{3}{4}C_{9,1}m^2,$$

thus completing the proof. □

Lemmas 5.7–5.9 contain upper and lower bounds for the second and fourth moments of the increments of the partial sums of the $\tilde{y}_i^2 - \mathrm{E}\tilde{y}_i^2$. As we will see, these inequalities, together with the independence of the $\tilde{y}_j$ and $\tilde{y}_i$, where $i < j - j^\rho$, are sufficient to establish the strong approximation in Theorem 2.4.

According to Lemma 5.4 it suffices to approximate $\sum(\tilde{y}_i^2 - \mathrm{E}\tilde{y}_i^2)$. Let $\eta > 1+\rho$, $\eta\rho < \tau < \eta - 1$, and set

$$X_k = \sum_{k^\eta\leqslant i\leqslant (k+1)^\eta-(k+2)^\tau} \tilde{\xi}_i \quad \text{and} \quad Y_k = \sum_{(k+1)^\eta-(k+2)^\tau < i < (k+1)^\eta} \tilde{\xi}_i,$$

where $\tilde{\xi}_i = \tilde{y}_i^2 - \mathrm{E}\tilde{y}_i^2$. Then $\{X_k\}$ and $\{Y_k\}$ are sequences each consisting of independent random variables. To show that X_k and X_{k+1} are independent, consider the closest lag, that is, the index difference of the first $\tilde{\xi}_i$ in X_{k+1} and of the last $\tilde{\xi}_i$ in X_k. We obtain that $(k+1)^\eta - [(k+1)^\eta - (k+2)^\tau] > (k+1)^\tau > (k+1)^{\eta\rho}$ which implies the assertion. Similar arguments apply to $\{Y_k\}$.

Finally, let

$$r(N) = \mathrm{E}\left[\sum_{1\leqslant k\leqslant N} X_k\right]^2 \quad \text{and} \quad r_*(N) = \mathrm{E}\left[\sum_{1\leqslant k\leqslant N} Y_k\right]^2. \tag{5.8}$$

Lemma 5.10. *If the conditions of Theorem 2.4 are satisfied, then there is a Wiener process* $\{W_1(t) : t \geqslant 0\}$ *such that*

$$\sum_{1\leqslant k\leqslant N} X_k - W_1(r(N)) = o\Big(r(N)^{(\kappa+1)/4}\log r(N)\Big) \ \textit{a.s.}$$

for any $(2\eta-1)/(2\eta) < \kappa < 1$.

Proof. It follows from the definition of $\tilde{\xi}_i$ that $X_1, X_2, \ldots$ are independent random variables with mean 0. By Lemmas 5.7 and 5.8, there are $C_{10,1}$ and $C_{10,2}$ such that

$$\mathrm{E}X_k^2 \leqslant C_{7,1}((k+1)^\eta - (k+2)^\tau - k^\eta) \leqslant C_{10,1} k^{\eta-1}$$

and

$$\mathrm{E}X_k^2 \geqslant C_{8,1}((k+1)^\eta - (k+2)^\tau - k^\eta) \geqslant C_{10,2} k^{\eta-1}. \tag{5.9}$$

Using Lemma 5.9, we can find $C_{10,3}$ such that

$$\mathrm{E}X_k^4 \leqslant C_{9,1}((k+1)^\eta - (k+2)^\tau - k^\eta)^2 \leqslant C_{10,3} k^{2(\eta-1)}. \tag{5.10}$$

Now (5.9) and (5.10) imply

$$\sum_{1\leqslant k<\infty} \left(\sum_{1\leqslant i\leqslant k} \mathrm{E}X_i^2 \right)^{-2\kappa} \mathrm{E}X_k^4 \leqslant C_{10,4} \sum_{1\leqslant k<\infty} k^{2(\eta-1)-2\kappa\eta} < \infty$$

and the result follows from Theorem A. □

Lemma 5.11. *If the conditions of Theorem 2.4 are satisfied, then there is a Wiener process* $\{W_2(t) : t \geqslant 0\}$ *such that*

$$\sum_{1\leqslant k\leqslant N} Y_k - W_2(r_*(N)) = o\left(r_*(N)^{(\kappa_*+1)/4} \log r_*(N) \right) \ a.s.$$

for any $(2\tau+1)/(2(1+\tau)) < \kappa^* < 1$.

Proof. Following the proof of Lemma 5.10, we have $\mathrm{E}Y_k^2 \geqslant C_{11,1} k^\tau$ and $\mathrm{E}Y_k^4 \leqslant C_{11,2} k^{2\tau}$. Hence,

$$\sum_{1\leqslant k<\infty} \left(\sum_{1\leqslant i\leqslant k} \mathrm{E}Y_i^2 \right)^{-2\kappa^*} \mathrm{E}Y_k^4 \leqslant C_{11,3} \sum_{1\leqslant k<\infty} k^{2\tau-2(\tau+1)\kappa^*} < \infty,$$

so the assertion follows from Theorem A. □

The next lemma gives an upper bound for the increments of $\sum_{i\leqslant n}(\tilde{y}_i^2 - \mathrm{E}\tilde{y}_i^2)$.

Lemma 5.12. *If the conditions of Theorem 2.4 are satisfied, then*

$$\max_{k^\eta \leqslant j \leqslant (k+1)^\eta} \left| \sum_{j\leqslant i\leqslant (k+1)^\eta} \tilde{\xi}_i \right| = \mathcal{O}\left(r(k)^{(\kappa+1)/4} \log r(k) \right) \ a.s.,$$

where κ *is specified in Lemma 5.10.*

Proof. On account of Lemma 5.9 we can apply Theorem 12.2 of Billingsley (1968: 94), resulting in

$$P\left\{\max_{k^\eta \leqslant j \leqslant (k+1)^\eta}\left|\sum_{j\leqslant i\leqslant (k+1)^\eta}\tilde{\xi}_i\right| \geqslant \lambda\right\} \leqslant \frac{C_{12,1}}{\lambda^4}((k+1)^\eta - k^\eta)^2 \leqslant \frac{C_{12,2}}{\lambda^4}k^{2(\eta-1)}.$$

So,

$$P\left\{\max_{k^\eta \leqslant j \leqslant (k+1)^\eta}\left|\sum_{j\leqslant i\leqslant (k+1)^\eta}\tilde{\xi}_i\right| \geqslant k^{\eta(\kappa+1)/4}\log k\right\} \leqslant C_{12,2}k^{2(\eta-1)-\eta(\kappa+1)}.$$

Since $(\eta-1)/\eta < (2\eta-1)/(2\eta) < \kappa$, we obtain that

$$\sum_{1\leqslant k<\infty} P\left\{\max_{k^\eta \leqslant j \leqslant (k+1)^\eta}\left|\sum_{j\leqslant i\leqslant (k+1)^\eta}\tilde{\xi}_i\right| \geqslant k^{\eta(\kappa+1)/4}\log k\right\} < \infty$$

and therefore, using the Borel–Cantelli lemma, the proof is complete. □

For any real $n \geqslant 1$ (not necessarily an integer), we introduce the notation

$$T_n = \mathrm{E}\left[\sum_{1\leqslant k\leqslant n}\xi_k\right]^2 \quad \text{and} \quad \tilde{T}_n = \mathrm{E}\left[\sum_{1\leqslant k\leqslant n}\tilde{\xi}_k\right]^2.$$

The final lemma contains certain properties of the quantities T_n and $\tilde{T}_n$.

Lemma 5.13. *Let $N_k = (k+1)^\eta$. Then we have, for some constant $C > 0$,*

(i) $|\tilde{T}_{N_k} - r(k)| \leqslant Cr(k)^{1/2}r_*(k)^{1/2}$,
(ii) $|\tilde{T}_n - \tilde{T}_{N_k}| \leqslant Ck^{(2\eta-1)/2}$, *for* $N_k < n \leqslant N_{k+1}$,
(iii) $|T_n - \tilde{T}_n| \leqslant C\sqrt{n}$, *for* $n \geqslant 1$.

Proof. By Minkowski's inequality, we have $|\tilde{T}_{N_k}^{1/2} - r(k)^{1/2}| \leqslant r_*(k)^{1/2}$ and thus

$$|\tilde{T}_{N_k} - r(k)| \leqslant r_*(k)^{1/2}(\tilde{T}_{N_k}^{1/2} + r(k)^{1/2}) \leqslant r_*(k)^{1/2}(2r(k)^{1/2} + r_*(k)^{1/2}),$$

proving (i). The proof of (ii) is similar, using Lemma 5,7. Finally, to prove (iii) we apply Minkowski's inequality again to get

$$|T_n^{1/2} - \tilde{T}_n^{1/2}| \leqslant \sum_{k=1}^n\left(\mathrm{E}[\xi_k - \tilde{\xi}_k]^2\right)^{1/2}$$

and, using the independence of ε_k and $\sigma_k^2 - \tilde{\sigma}_k^2$ and Lemma 5.5, we obtain

$$\left(\mathrm{E}[\xi_k - \tilde{\xi}_k]^2\right)^{1/2} \leqslant 2\left(\mathrm{E}[\sigma_k^2 - \tilde{\sigma}_k^2]^2\mathrm{E}[\varepsilon_k^4]\right)^{1/2} = \mathcal{O}\left(k^{-\mu/4}\right).$$

As μ can be chosen arbitrarily large, we get $T_n^{1/2} - \tilde{T}_n^{1/2} = \mathcal{O}(1)$, which implies (iii), since $\tilde{T}_n = \mathcal{O}(n)$ by Lemma 5.7. □

We are at last in a position to give the proof of Theorem 2.4.

Proof of Theorem 2.4. Putting together Lemmas 5.10, 5.11 and the law of the iterated logarithm for W_2 yields

$$\sum_{1\leqslant i\leqslant (N+1)^\eta} \tilde{\xi}_i - W_1(r(N)) = \mathcal{O}\Big(r(N)^{(\kappa+1)/4}\log r(N)\Big) \text{ a.s.,} \tag{5.11}$$

provided the constants involved are chosen according to the side conditions

$$\eta(\kappa+1) > 2(\tau+1) \quad \text{and} \quad \frac{2\eta-1}{2\eta} < \kappa < 1, \tag{5.12}$$

which ensure that the sum of the Y_k is of smaller order of magnitude than the remainder term in Lemma 5.10. Observe that $r(N) \approx N^\eta$ and $r_*(N) \approx N^{\tau+1}$, where $\approx$ means that the ratio of the two sides is between positive constants. Letting $S_t^* = \tilde{\xi}_1 + \ldots + \tilde{\xi}_{\lfloor t\rfloor}$, $t \geqslant 0$, we obtain for $N_{k-1} < n \leqslant N_k$, using Lemmas 5.12, 5.13, relation (5.11) and standard estimates for the increments of Wiener processes (see, for example, Theorem 1.2.1 of Csörgő and Révész 1981: 30),

$$\begin{aligned}
|S_n^* - W_1(T_n)| \leqslant{}& |S_n^* - S_{N_k}^*| + |S_{N_k}^* - W_1(r(k))| \\
&+ |W_1(r(k)) - W_1(\tilde{T}_{N_k})| + |W_1(\tilde{T}_{N_k}) - W_1(\tilde{T}_n)| + |W_1(\tilde{T}_n) - W_1(T_n)| \\
={}& \mathcal{O}\Big(r(k)^{(\kappa+1)/4}\log r(k)\Big) + \mathcal{O}\Big(r(k)^{(\kappa+1)/4}\log r(k)\Big) \\
&+ \mathcal{O}\Big(r(k)^{1/4} r_*(k)^{1/4}\log k\Big) + \mathcal{O}\Big(k^{(2\eta-1)/4}\log k\Big) + \mathcal{O}\Big(n^{1/4}\log n\Big) \\
={}& \mathcal{O}\Big(n^{(\kappa+1)/4}\log n\Big) + \mathcal{O}\Big(n^{(\eta+\tau+1)/4\eta}\log n\Big) \\
&+ \mathcal{O}\Big(n^{(2\eta-1)/4\eta}\log n\Big) + \mathcal{O}\Big(n^{1/4}\log n\Big).
\end{aligned} \tag{5.13}$$

In the last step we expressed the remainder terms in n, using the fact that $k \approx n^{1/\eta}$. Choose $\eta = 3/2$, τ near 0 and κ very near to, but larger than, 2/3. In this case the relations (5.12) are satisfied and hence (5.13) yields

$$S_n^* - W_1(T_n) = \mathcal{O}\Big(n^{5/12+\varepsilon}\Big) \text{ a.s.} \tag{5.14}$$

for any $\varepsilon > 0$. Now $\{\xi_k\}$ is a stationary sequence with zero mean and covariances $\mathrm{E}(\xi_0\xi_k) = \mathcal{O}(k^{-\mu})$ for any $\mu > 0$, from which it easily follows that

$$T_n = \mathrm{E}\Bigg[\sum_{1\leqslant k\leqslant n}\xi_k\Bigg]^2 = n\overline{\sigma}^2 + \mathcal{O}(1),$$

where $\overline{\sigma}^2$ is defined in (2.13). Thus, on applying Theorem 1.2.1 of Csörgő and Révész (1981, p .30), we arrive at $W_1(T_n) = W_1(n\overline{\sigma}^2) + \mathcal{O}(\log n)$ a.s., so Theorem 2.4 is readily proved on setting $W(t) = \overline{\sigma}^{-1}W_1(\overline{\sigma}^2 t)$. □

Acknowledgements

The research for this paper was partially supported by NATO grant PST.EAP.CLG 980599 and NSF–OTKA grant INT-0223262.

References

Aue, A. and Horváth, L. (2004) Delay time in sequential detection of change. *Statist. Probab. Lett.*, **67**, 221–231.

Aue, A. and Horváth, L. (2005) A note on the existence of a stationary augmented GARCH process. Preprint, University of Utah.

Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2005) Change-point monitoring in linear models with conditionally heteroskedastic errors. Preprint, University of Utah.

Berkes, I. and Horváth, L. (2001) Strong approximation for the empirical process of a GARCH sequence. *Ann. Appl. Probab.*, **11**, 789–809.

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004) Sequential change-point detection in GARCH(p,q) models. *Econometric Theory*, **20**, 1140–1167.

Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.-M. (2005) On the almost sure convergence of the Bartlett estimator. *Period. Math. Hungar.*, **51**, 11–25.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

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

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

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

Carrasco, M. and Chen, X. (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17–39.

Chow, Y.S. and Teicher, H. (1988) *Probability Theory* (2nd edn). New York: Springer-Verlag.

Csörgő, M. and Horváth, L. (1997) *Limit Theorems in Change-Point Analysis*. Chichester: Wiley.

Csörgő, M. and Révész, P. (1981) *Strong Approximations in Probability and Statistics*. New York: Academic Press.

Ding, Z., Granger, C.W.J. and Engle, R. (1993) A long memory property of stock market returns and a new model. *J. Empir. Finance*, **1**, 83–106.

Duan, J.-C. (1997) Augmented GARCH(p,q) process and its diffusion limit. *J. Econometrics*, **79**, 97–127.

Durrett, R. (1986) *Probability: Theory and Examples*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Eberlein, E. (1986) On strong invariance principles under dependence assumptions. *Ann. Probab.*, **14**, 260–270.

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

Engle, R. and Ng, V. (1993) Measuring and testing impact of news on volatility. *J. Finance*, **48**, 1749–1778.

Geweke, J. (1986) Modeling persistence of conditional variances: a comment. *Econometric Rev.*, **5**, 57–61.

Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2003) Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics*, **112**, 265–294.

Glosten, L., Jagannathan, R. and Dunke, D. (1993) Relationship between expected value and the volatility of the nominal excess return on stocks. *J. Finance*, **48**, 1779–1801.

Hardy, G.H., Littlewood, J.E. and Pólya, G. (1959) *Inequalities* (2nd edn). Cambridge: Cambridge University Press.

Horváth, L. and Steinebach, J. (2000) Testing for changes in the mean or variance of a stochastic process under weak invariance. *J. Statist. Plann. Inference*, **91**, 365–376.

Horváth, L., Kokoszka, P. and Zhang, A. (2006) Monitoring constancy of variance in conditionally heteroscedastic time series. *Econometric Theory*. To appear.

Kuelbs, J. and Philipp, W. (1980) Almost sure invariance principles for partial sums of mixing *B*-valued random variables. *Ann. Probab.*, **8**, 1003–1036.

Lo, A. (1991) Long-run memory in stock market prices. *Econometrica*, **59**, 1279–1313.

Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, **6**, 318–334.

Nelson, D.B. (1991) Conditional heteroscedasticity in asset returns: a new approach. *Econometrica*, **59**, 347–370.

Petrov, V.V. (1995) *Limit Theorems of Probability Theory*. Oxford: Clarendon Press.

Révész, P. (1990) *Random Walk in Random and Non-random Environments*. Singapore: World Scientific.

Schwert, G.W. (1989) Why does stock market volatility change over time? *J. Finance*, **44**, 1115–1153.

Strassen, V. (1967) Almost sure behaviour of sums of independent random variables and martingales. In L. Le Cam and J. Neyman (eds), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 315–344. Berkeley: University of California Press.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in heteroscedastic time series: a stochastic recurrence equations approach. *Ann. Statist.* To appear.

Taylor, S. (1986) *Modelling Financial Time Series*. New York: Wiley.

Zakoian, J.-M. (1994) Threshold heteroscedastic models. *J. Economic Dynamics and Control*, **18**, 931–955.

Received April 2005 and revised October 2005