REVSTAT – Statistical Journal
Volume 2, Number 1, June 2004

CHANGES OF STRUCTURE IN FINANCIAL TIME SERIES AND THE GARCH MODEL

Authors: Thomas Mikosch*
– Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, and MaPhySto, the Danish Network for Mathematical Physics and Stochastics, Denmark (mikosch@math.ku.dk)

Cătălin Stărică
– Department of Mathematical Statistics, Chalmers University of Technology, S–412 96 Gothenburg, Sweden (starica@math.chalmers.se)

Received: November 2003 Accepted: December 2003

Abstract:

- In this paper we propose a goodness of fit test that checks the resemblance of the spectral density of a GARCH process to that of the log-returns. The asymptotic behavior of the test statistics are given by a functional central limit theorem for the integrated periodogram of the data. A simulation study investigates the small sample behavior, the size and the power of our test. We apply our results to the S&P500 returns and detect changes in the structure of the data related to shifts of the unconditional variance. We show how a long range dependence type behavior in the sample ACF of absolute returns might be induced by these changes.

Key-Words:

- *integrated periodogram; spectral distribution; functional central limit theorem; Kiefer–Müller process; Brownian bridge; sample autocorrelation; change point; GARCH process; long range dependence; IGARCH; non-stationarity.*

AMS Subject Classification:

- Primary: 62P20; Secondary: 60G10 60F17 60B12 60G15 62M15.

*This research supported in part by a grant of the Dutch Science Foundation (NWO), by DYNSTOCH, a research training network under the programme Improving Human Potential financed by The 5th Framework Programme of the European Commission, the Danish Research Council Grant No 21-01-0546, and MaPhySto, the Danish Network for Mathematical Physics and Stochastics, funded by The Danish National Research Foundation.

1. INTRODUCTION

In this paper we introduce a goodness of fit test for the GARCH process. In its simplest form this model is given by

$$\begin{cases} X_t \;=\; \sigma_t\, Z_t\,, \\ \sigma_t^2 \;=\; \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 X_{t-1}^2\,, \end{cases} \qquad t \in \mathbb{Z}\,, \tag{1.1}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid random variables with $EZ_1 = 0,\; EZ_1^2 = 1$. The parameters α_1 and β_1 are non-negative and α_0 is necessarily positive.

Our test decides if the data at hand is a white noise whose squares have a covariance structure which is in agreement with the second order structure of the hypothesized squared GARCH process. The test is related to the classical Grenander–Rosenblatt or Bartlett goodness of fit tests for the spectral distribution of a time series; see for example Priestley [38]. Such tests are analogues to the Kolmogorov–Smirnov test for the distribution of a sample. Other testing procedures exist in the literature. Among them we mention the approach that uses the sequential empirical process for the residuals of an ARCH process; see Horváth et al. [26]. Besides being restricted to the ARCH case, these asymptotic tests present another drawback. The limit distribution of the test statistic depends in general on the distribution of the noise Z_t and the parameters of the model. An advantage of the test proposed in this paper is that the limit distribution of the test statistic is distribution free and, as in the Kolmogorov-Smirnov test, is a function of the Brownian bridge. Moreover, we prove that the limit distribution of our test statistics are insensitive to the replacement of the parameters by their estimators under the null hypothesis.

Although attractive as a model, there is copious empirical evidence in the econometrics literature, coming especially from the analysis of long series of log-returns, that argues against the GARCH$(1,1)$ model. For example, although the squares of a GARCH$(1,1)$ process follow the dynamics of an ARMA process (in particular the ACF goes to zero exponentially fast), the sample ACFs of the absolute values and their squares tend to stabilize around a positive value for larger lags (the so-called long range dependence in absolute returns or in volatility). For longer samples the estimated parameters α_1 and β_1 sum up to values close to 1 (Engle and Bollerslev [19], Mikosch and Stărică [35]). This fact, known as the integrated GARCH finding, implies infinite variance (see Bollerslev [8]) for the returns, a conclusion in strong disagreement with the accepted results of semi-parametric tail analysis that find at least a finite third moment (Embrechts et al. [17]).

The second contribution of the paper is the analysis based on our goodness of fit procedure of a long portion of the S&P500 log-return series (January 1953 to December 1990) which could provide an answer to these critics. For the data under investigation we detect structural changes related to movements of the unconditional variance and show how a long range dependence type behavior in the sample ACF of absolute returns might be induced by these shifts. Our procedure identifies most of the recessions of this period as being structurally different. The major structural change is detected between 1973 and 1975 and seems to correspond to the oil crises. Our analysis seems to indicate that one simple GARCH(1,1) process (which models the first ten years of the data quite well) cannot describe the complicated dynamics of longer, possibly non-stationary log-return time series.

Our paper is organized as follows. In Section 2 we formulate our main theoretical result, a functional central limit theorem for the integrated periodogram of the GARCH process. Then we indicate how this result can be used to build an asymptotic goodness of fit test for the spectral distribution of the GARCH process. We also discuss the behavior of the test statistics under the alternative hypothesis of a different GARCH process. The proofs are rather technical and therefore postponed to Appendix A1. In Section 3 we investigate by means of simulations the small sample properties, the size and the power of our test statistic while in Section 4 we apply our method to the study of a long portion of the S&P log-return series. Some concluding remarks are given in Section 5.

2. LIMIT THEORY FOR THE TWO-PARAMETER INTEGRATED PERIODOGRAM

In fields as diverse as time series analysis and extreme value theory it is generally assumed that the observations or a suitable transformation of them constitute a stationary sequence of random variables. In the context of this section, stationarity is always understood as strict stationarity. One of the aims of this paper is to provide a procedure for testing how good the fit of a stationary GARCH(p, q) model to data is. This section provides the limit theory for a certain two-parameter process which is the basis for the statistical procedure we propose in Section 2.2. This theory is slightly more general than needed for the purposes of this paper. However, the theory for the corresponding one-parameter process (which will be used intensively in the rest of the paper) is essentially the same as for the case of two parameters. The latter case can be used for change point detection in the spectral domain while the former one yields goodness of fit tests. As already mentioned, in the context of this paper, we are mainly interested in test statistics for the goodness of fit of GARCH processes. The statistical procedure will allow us to single out the parts of the data which are not well described by the hypothesized model.

To be precise, we assume that the data come from a stationary generalized autoregressive conditionally heteroscedastic process of order (p, q), for short GARCH(p, q):

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-i}^2 , \qquad t \in \mathbb{Z} , \tag{2.1}$$

where (Z_t) is an iid symmetric sequence with $EZ^2 = 1$, non-negative parameters α_i and β_j, and the stochastic volatility σ_t is independent of Z_t for every fixed t. We also assume that Z_1 has a Lebesgue density on the real line. This ensures that (X_t) is α-mixing with geometric rate; see Boussama [11]. In what follows, we write σ for a generic random variable with the distribution of σ_1, X for a generic random variable with the distribution of X_1, etc.

This kind of model is most popular in the econometrics literature for modeling the log-returns of stock indices, share prices, exchange rates, etc., and has found its way into the practice of forecasting financial time series. See for example Engle [18] for a collection of papers on ARCH. We assume that, for a particular choice of parameters α_i and β_i, the sequence $((X_t, \sigma_t))$ is stationary. Assumptions for stationarity of a GARCH process can be found in Bougerol and Picard [10] for the general GARCH(p, q) case and in Nelson [36] for the GARCH$(1, 1)$ case. For a recent overview on the mathematics of GARCH processes, we refer to Mikosch [32].

Our analysis is based on the spectral properties of the underlying time series. Consider the classical estimator of the spectral density, the periodogram, given by

$$I_{n,X}(\lambda) = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathrm{e}^{-i\lambda t} X_t \right|^2 , \qquad \lambda \in [0, \pi] .$$

Under general conditions, the integrated periodogram or empirical spectral distribution function

$$\frac{1}{2\pi} J_{n,X}(\lambda) = \frac{1}{2\pi} \int_0^{\lambda} I_{n,X}(x)\, dx , \qquad \lambda \in [0, \pi] , \tag{2.2}$$

is a consistent estimator of the spectral distribution function given by

$$F_X(\lambda) = \int_0^{\lambda} f_X(x)\, dx , \qquad \lambda \in [0, \pi] ,$$

provided the spectral density f_X is well defined.

2.1. Main results

As a motivation for our main result, we start by considering the two-parameter process $J_{n,X}(x,\lambda)$ related to (2.2) (see also Appendix A1):

$$
\begin{aligned}
J_{n,X}(x,\lambda) \;&=\; \int_0^\lambda \left(\gamma_{n,[nx],X}(0) \,+\, 2 \sum_{h=1}^{[nx]-1} \gamma_{n,[nx],X}(h)\, \cos(yh) \right) dy \\
&=\; \lambda\, \gamma_{n,[nx],X}(0) \,+\, 2 \sum_{h=1}^{[nx]-1} \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h}\ ,
\end{aligned}
\tag{2.3}
$$

where

$$\gamma_{n,[nx],X}(h) \;=\; \frac{1}{n} \sum_{t=1}^{[nx]-h} X_t\, X_{t+h}\ , \qquad h=0,1,2,...,[nx]-1, \quad x \in [0,1]\ .$$

Clearly,

$$\gamma_{n,X}(h) \;:=\; \gamma_{n,n,X}(h)$$

denotes a version of the sample autocovariance at lag h; the standard version of the sample autocovariance is defined for the centered random variables $X_t - \overline{X}_n$, where $\overline{X}_n$ is the sample mean. We also write

$$\gamma_X(h) = \operatorname{cov}(X_0, X_h) \quad \text{and} \quad v_X(h) = \operatorname{var}(X_0 X_h) = E(X_0 X_h)^2\ , \qquad h \in \mathbb{Z}\ .$$

The processes $\gamma_{n,[n\cdot],X}(h)$ satisfy a fairly general functional central limit theorem (FCLT). Recall that $\mathbb{D}\left([0,1],\mathbb{R}^m\right)$ is the Skorokhod space of $\mathbb{R}^m$-valued cadlag functions on $[0,1]$ (continuous from the right in $[0,1)$, limits exist from the left in $(0,1]$) endowed with the J_1-topology and the corresponding Borel σ-field; see for example Jacod and Shiryaev [27] or Bickel and Wichura [6].

Lemma 2.1. *Consider the* GARCH(p,q) *process* (X_t) *given by* (2.1). *Assume that*

$$E|X|^{4+\delta} < \infty \qquad \text{for some} \quad \delta > 0\ . \tag{2.4}$$

Then for every $m \geq 1$, *as* $n \to \infty$

$$\sqrt{n}\Big(\gamma_{n,[nx],X}(h),\ h=1,...,m\Big)_{x\in[0,1]} \;\xrightarrow{d}\; \Big(v_X^{1/2}(h)\ W_h(x),\ h=1,...,m\Big)_{x\in[0,1]}\ , \tag{2.5}$$

in $\mathbb{D}\left([0,1],\mathbb{R}^m\right)$, *where* $W_h(\cdot)$, $h=1,...,m$, *are iid standard Brownian motions on* $[0,1]$.

The proof of the lemma is given in Appendix A1. A naive argument, based on Lemma 2.1 and the decomposition (2.3), suggests that

$$\sqrt{n}\Big(J_{n,X}(x,\lambda) - \lambda\,\gamma_{n,[nx],X}(0)\Big)_{x\in[0,1],\ \lambda\in[0,\pi]} \xrightarrow{d}$$

$$\xrightarrow{d}\ 2\left(\sum_{h=1}^{\infty} v_X^{1/2}(h)\, W_h(x)\,\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\ \lambda\in[0,\pi]},$$

in $\mathbb{D}\big([0,1]\times[0,\pi]\big)$. This result can be shown to be true; one can follow the lines of the proof of Theorem 2.1 below. However, the two-parameter Gaussian limit field has a distribution that explicitly depends on the covariance structure of (X_t^2), which is not a very desirable property. Indeed, since we are interested in using functionals of the limit process for a goodness of fit procedure, we would like that the asymptotic distribution of those functionals is independent of the null hypothesis we test. In other words, we want a "standard" Gaussian process in the limit since otherwise we would have to evaluate the distributions of its functionals by Monte–Carlo simulations for every choice of parameters of the GARCH(p,q) we consider in the null hypothesis.

A glance at the right-hand side of (2.3) suggests another approach. The dependence of the limiting Gaussian field on the covariance structure of (X_t^2) comes in through the FCLT of Lemma 2.1. However, it is intuitively clear that, if we replaced in (2.3) the processes $\gamma_{n,[n\cdot],X}(h)$ by

$$\widetilde{\gamma}_{n,[n\cdot],X}(h) = \frac{\gamma_{n,[n\cdot],X}(h)}{v_X^{1/2}(h)},$$

the limit process would become independent of the covariance structure of (X_t^2).

Therefore we introduce the following two-parameter process which is a straightforward modification of $J_{n,X}(x,\lambda)$:

$$C_{n,X}(x,\lambda) = \sum_{h=1}^{[nx]-1} \widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}, \qquad x\in[0,1], \quad \lambda\in[0,\pi].$$

Our main result is a FCLT for $C_{n,X}$.

Theorem 2.1. *Let* (X_t) *be a stationary* GARCH(p,q) *process given by* (2.1). *Assume that* (2.4) *holds. Then*

$$\sqrt{n}\big(C_{n,X}(x,\lambda)\big)_{x\in[0,1],\ \lambda\in[0,\pi]} \xrightarrow{d}$$

$$\xrightarrow{d}\ \big(K(x,\lambda)\big)_{x\in[0,1],\ \lambda\in[0,\pi]} = \left(\sum_{h=1}^{\infty} W_h(x)\,\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\ \lambda\in[0,\pi]}, \tag{2.6}$$

in $\mathbb{D}\left([0,1]\times[0,\pi]\right)$ *where* $(W_h(\cdot))_{h=1,\ldots}$ *is a sequence of iid standard Brownian motions on* $[0,1]$*. The infinite series on the right-hand side converges with probability* 1 *and represents a Kiefer–Müller process, i.e., a two-parameter Gaussian field with covariance structure*

$$
\begin{aligned}
E\Big(K(x_1,\lambda_1)\,K(x_2,\lambda_2)\Big) &= \min(x_1,x_2)\sum_{t=1}^{\infty}\frac{\sin(\lambda_1 t)\,\sin(\lambda_2 t)}{t^2}\\
&= 2^{-1}\,\pi^2\min(x_1,x_2)\left(\min\left(\frac{\lambda_1}{\pi},\frac{\lambda_2}{\pi}\right)-\frac{\lambda_1}{\pi}\,\frac{\lambda_2}{\pi}\right)\,.
\end{aligned}
\tag{2.7}
$$

The proof of the theorem is given in Appendix A1.

The series representation of the Kiefer–Müller process can be found in Klüppelberg and Mikosch [29]. This process is known in empirical process theory as the limiting Gaussian field for the sequential empirical process; see Shorack and Wellner [41].

Remark 2.1. The statement of Theorem 2.1 remains valid for wider classes of stationary sequences. In particular the result holds if the conditions in Remark 1.1 are satisfied and in addition, (X_t) is symmetric and $(|X_t|)$ and $(\mathrm{sign}(X_t))$ are independent. The latter conditions are satisfied by any stochastic volatility model of the form $X_t=\sigma_t\,Z_t$, where (Z_t) is a sequence of iid symmetric random variables and the random variables σ_t are adapted to the filtration $\sigma(Z_{t-1},Z_{t-2},\ldots)$, or alternatively, (σ_t) and (Z_t) are independent.

Immediate consequences of Theorem 2.1 and the continuous mapping theorem are limit theorems for continuous functionals of the process $C_{n,X}$ which can be used for the construction of goodness of fit tests and tests for detecting changes in the spectrum of the time series.

Corollary 2.1. *Under the assumptions of Theorem 2.1,*

$$
\sqrt{n}\sup_{x\in[0,1],\ \lambda\in[0,\pi]}\big|C_{n,X}(x,\lambda)\big| \stackrel{d}{\longrightarrow} \sup_{x\in[0,1],\lambda\in[0,\pi]}\big|K(x,\lambda)\big|\,,
$$

$$
n\int_0^1\!\!\int_0^\pi C_{n,X}^2(x,\lambda)\,dx\,d\lambda \stackrel{d}{\longrightarrow} \int_0^1\!\!\int_0^\pi K^2(x,\lambda)\,dx\,d\lambda\,.
$$

For $x=1$, convergence in (2.6) yields

$$
\sqrt{n}\,\widetilde{C}_{n,X}(\cdot) := \sqrt{n}\sum_{h=1}^{n-1}\frac{\gamma_{n,X}(h)}{v_X^{1/2}(h)}\,\frac{\sin(\cdot\,h)}{h} \stackrel{d}{\longrightarrow} B(\cdot) := \sum_{h=1}^{\infty}W_h(1)\,\frac{\sin(\cdot h)}{h}\,,
\tag{2.8}
$$

in $\mathbb{C}\,[0,\pi]$. The series on the right-hand side is the so-called Paley–Wiener representation of a Brownian bridge on $[0,\pi]$; see (2.7) with $x=1$ (see for example Hida [25]).

The one-parameter process $\widetilde{C}_{n,X}$ will be our basic process for testing the goodness of fit of the sample $X_1, ..., X_n$ to a GARCH process. The convergence of the following functionals can be used for constructing Kolmogorov–Smirnov and Cramér–von Mises type goodness of fit tests for a GARCH(p,q) process.

Corollary 2.2. *Under the assumptions of Theorem 2.1,*

$$\widetilde{S}_n := \sqrt{n} \sup_{\lambda\in[0,\pi]} \left|\widetilde{C}_{n,X}(\lambda)\right| \xrightarrow{d} \sup_{\lambda\in[0,\pi]} |B(\lambda)| \,, \tag{2.9}$$

$$n \int_0^\pi \widetilde{C}^2_{n,X}(\lambda)\, d\lambda \xrightarrow{d} \int_0^\pi B^2(\lambda)\, d\lambda \,.$$

2.2. The goodness of fit test

In what follows, we focus on the GARCH$(1,1)$ case but a similar theory can be developed for the general GARCH(p,q) case. The quantities $v_X(h)$ are continuous functions of the GARCH parameters and the fourth moments of the iid noise Z_t. We refer to Appendix A2 where v_X is explicitly given for the GARCH$(1,1)$ case. For an application of the results above it is natural to replace the unknown quantities $v_X(h)$ in the definition of $\widetilde{\gamma}_{n,k,X}(h)$ by their sample versions $\widehat{v}_X(h)$, i.e., the parameters α_i and β_1 are replaced by some estimators $\widehat{\alpha}_i$ and $\widehat{\beta}_1$ and EZ^4 is replaced by the sample mean of the 4$^{\text{th}}$ powers of the residuals $\widehat{EZ^4} = n^{-1}\sum_{i=1}^n \widehat{Z}_t^4$, where $\widehat{Z}_t = X_t/\widehat{\sigma}_t$ and $\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 X_{t-1}^2 + \widehat{\beta}_1 \widehat{\sigma}_{t-1}^2$ and $\widehat{\sigma}_0^2$ and X_0^2 are arbitrarily chosen, but fixed. Denoting by

$$\widehat{\gamma}_{n,[n\cdot],X}(h) \;=\; \frac{\gamma_{n,[n\cdot],X}(h)}{\widehat{v}_X^{1/2}(h)} \,,$$

we produce the straightforward modification of $C_{n,X}(x,\lambda)$:

$$\widehat{C}_{n,X}(x,\lambda) \;=\; \sum_{h=1}^{[nx]-1} \widehat{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \,, \qquad x \in [0,1], \quad \lambda \in [0,\pi] \,,$$

and that of $\widetilde{S}_n$:

$$S_n \;:=\; \sqrt{n} \sup_{\lambda\in[0,\pi]} \left|\widehat{C}_{n,X}(\lambda)\right| \,. \tag{2.10}$$

The following result states that the theory developed in this section remains valid if v_X is replaced by its sample analogue.

Theorem 2.2. *Assume that the parameter estimators $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ based on $X_1,\dots,X_n$ are independent of* $(\operatorname{sign}(X_t))$ *and consistent, i.e.,* $\widehat{\alpha}_1 \stackrel{P}{\to} \alpha_1$ *and* $\widehat{\beta}_1 \stackrel{P}{\to} \beta_1$. *Then Theorem 2.1 and its corollaries remain valid for a* GARCH$(1,1)$ *process if v_X is replaced by its sample analogue $\widehat{v}_X$. In particular*

$$S_n \stackrel{d}{\longrightarrow} \sup_{\lambda\in[0,\pi]} |B(\lambda)| \,. \tag{2.11}$$

Remark 2.2. The Whittle parameter estimators of a GARCH process are consistent if $EX^4<\infty$, and so are the Gaussian quasi maximum likelihood estimators; see Giraitis and Robinson [22] and Mikosch and Straumann [33] for the former case and Berkes et al. [5] for the latter case. Moreover, by their definitions they are calculated from the X_t^2's and σ_t^2's only and therefore they are independent of $(\operatorname{sign}(X_t))$.

The results we presented so far are sufficient for providing a theoretical understanding of the behavior of tests based on functionals of $\widehat{C}_{n,X}$ (for example S_n). These are tests of the null hypothesis that the sample $X_1,\dots,X_n$ comes from a GARCH$(1,1)$ model with given parameters α_i and β_i against the alternative of another GARCH$(1,1)$ model with parameters α_i^a, $i=0,1$, and β_1^a. They reject the null hypothesis if the functional is in a certain region. The rejection region giving the test the right size is constructed based on the quantiles of the appropriate functional of the limit process in Theorem 2.2 (i.e. the supremum of a Brownian bridge in the case of the statistic S_n). As for the power of the test, similar arguments as for the proof of Theorem 2.2 yield under the alternative the following result.

Theorem 2.3. *Assume that (X_t) and (Y_t) are two stationary* GARCH(1,1) *processes* (2.1) *with coefficients α_i, $i=0,1$, β_1, and α_i^a, $i=0,1$, β_1^a respectively. Assume that the parameter estimators $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ (based on the sample $X_1,\dots,X_n$) are independent of* $(\operatorname{sign}(X_t))$ *and consistent, i.e.,* $\widehat{\alpha}_1 \stackrel{P}{\to} \alpha_1$ *and* $\widehat{\beta}_1 \stackrel{P}{\to} \beta_1$. *Define*

$$\widehat{\gamma}^a_{n,[n\cdot]}(h) = \frac{\gamma_{n,[n\cdot],Y}(h)}{\widehat{v}_X^{1/2}(h)}\,,$$

$$\widehat{C}^a_{n,X,Y}(x,\lambda) = \sum_{h=1}^{[nx]-1} \widehat{\gamma}^a_{n,[nx]}(h)\,\frac{\sin(\lambda h)}{h}\,, \qquad x\in[0,1], \quad \lambda\in[0,\pi]\,.$$

Then

$$\sqrt{n}\left(\widehat{C}^a_{n,X,Y}(x,\lambda)\right)_{x\in[0,1],\,\lambda\in[0,\pi]} \stackrel{d}{\longrightarrow} \left(\sum_{h=1}^{\infty} \frac{v_X^{a\,1/2}(h)}{v_X^{1/2}(h)}\, W_h(x)\,\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\,\lambda\in[0,\pi]}, \tag{2.12}$$

in $\mathbb{D}\left([0,1]\times[0,\pi]\right)$ *where* $(W_h(\cdot))$ *is a sequence of iid standard Brownian motions on* $[0,1]$ *while* $v_X^a(h) = E(Y_0Y_h)^2$.

This result yields a theoretical description of the power of tests based on functionals of $\widehat{C}_{n,X}$. It individuates the functional of the Gaussian process on the right hand side of equation (2.12) as an asymptotic equivalent of the desired functional of $\widehat{C}^a_{n,X,Y}$. Note that the distributions of the functionals of the limit processes depend on the parameters of the alternative hypothesis in a rather complicated way. This makes the direct use of the Theorem 2.3 in applications rather difficult. For this reason we will rely on Monte–Carlo based calculations of the distribution of $\widehat{C}^a_{n,X,Y}$. See Section 3 for a simulation study on the size and the power of a test based on the statistics S_n.

Results similar to Theorem 2.3 can be derived for the alternative hypothesis that the sample $X_1, ..., X_n$ consists of subsamples from different GARCH(p, q) processes. Clearly, the asymptotic distribution will be even more complex and the Monte–Carlo approach again inevitable.

3. A SIZE AND POWER MONTE CARLO STUDY

The aim of this section is to investigate the size and the power of a test based on the statistic S_n in (2.10). The set up is relevant to the real data analysis performed in Section 4. There we check the goodness of fit of a GARCH$(1, 1)$ process with parameters estimated on the first 3 years of data (750 observations)

$$\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759, \quad \nu = 5.24 \ , \tag{3.1}$$

to various segments of the data set. Here ν is the number of degrees of freedom of the t-distributed noise sequence (Z_t). The corresponding value of the fourth moment of the estimated residuals of the model 3.1 is $EZ^4 = 7.82$.

The first choice we need to make when applying our test on data is precisely that of the size of the window that guarantees a correct behavior of the statistic S_n. Theoretically, the correct size of the test will be guaranteed by a choice of the rejection region based on the asymptotic behavior of S_n described by Theorem 2.2. These results are only asymptotic and provide the right size if the data window used to calculate the S_n statistic is large. For reasons that are explained in Section 4, we want to keep the length of the window as small as possible. It is by means of simulations that we find the right balance between these opposing requirements on the window size. As a byproduct of the simulation study, we will understand how to adjust the interval provided by Theorem 2.2 in order to maintain the correct size.

The top graph in Figure 1 displays the QQ-plot of 1000 simulated values of $\widetilde{S}_{125}$ (the quantiles on the x-axis), calculated on samples of 125 observations from a GARCH process with Student-t innovations and parameters (3.1) against

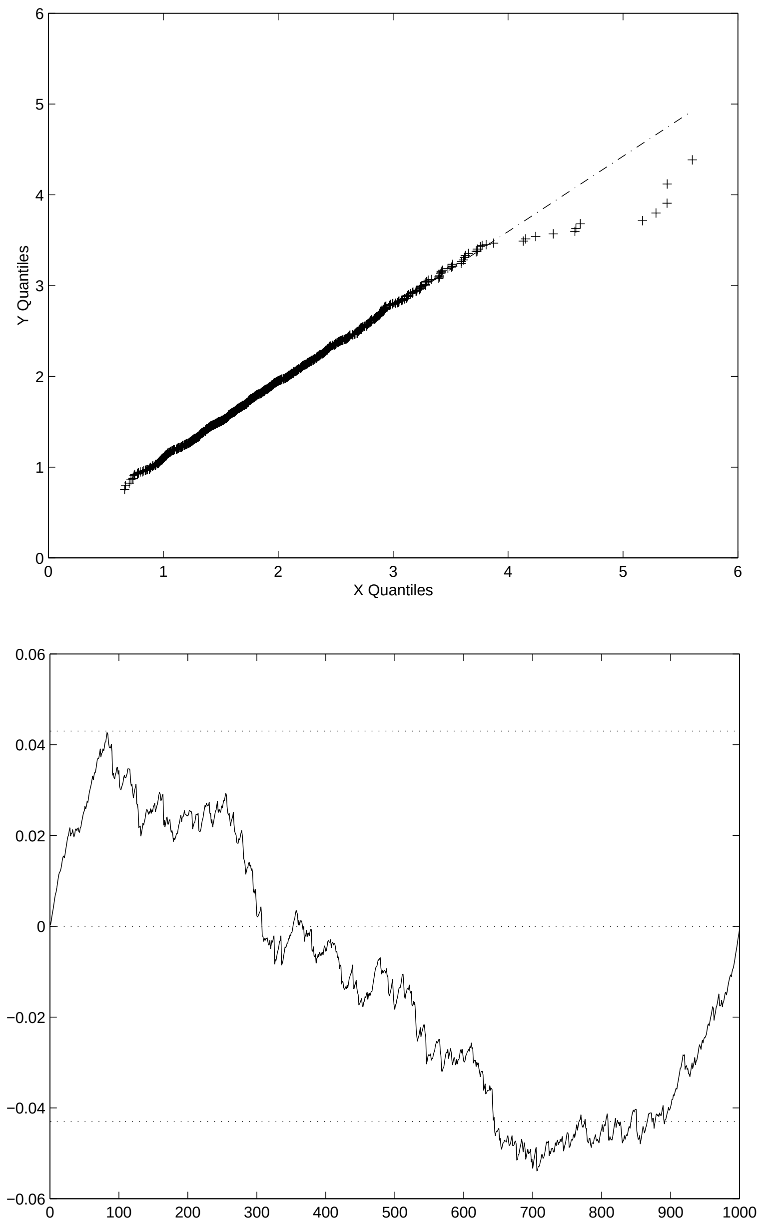


Figure 1: *Top:* QQ-plot of 1000 values of $\widetilde{S}_{125}$ (x-axis) against the quantiles of the supremum of a Brownian bridge (y-axis).
Bottom: The difference between the sample cdf of 1000 simulated values of $\widetilde{S}_{125}$ and the theoretical cdf of the supremum of a Brownian bridge with the Kolmogorov-Smirnov 95% confidence bands.

the quantiles of the supremum of a Brownian bridge (on the y-axis). The bottom graph in the same figure together with the graphs in Figure 2 shows the goodness of fit of the distribution of the supremum of a Brownian bridge to samples of 1000 simulations of $\widetilde{S}_{125}$ (Figure 1), $\widetilde{S}_{500}$ and $\widetilde{S}_{1000}$ (Figure 2) respectively. The statistic

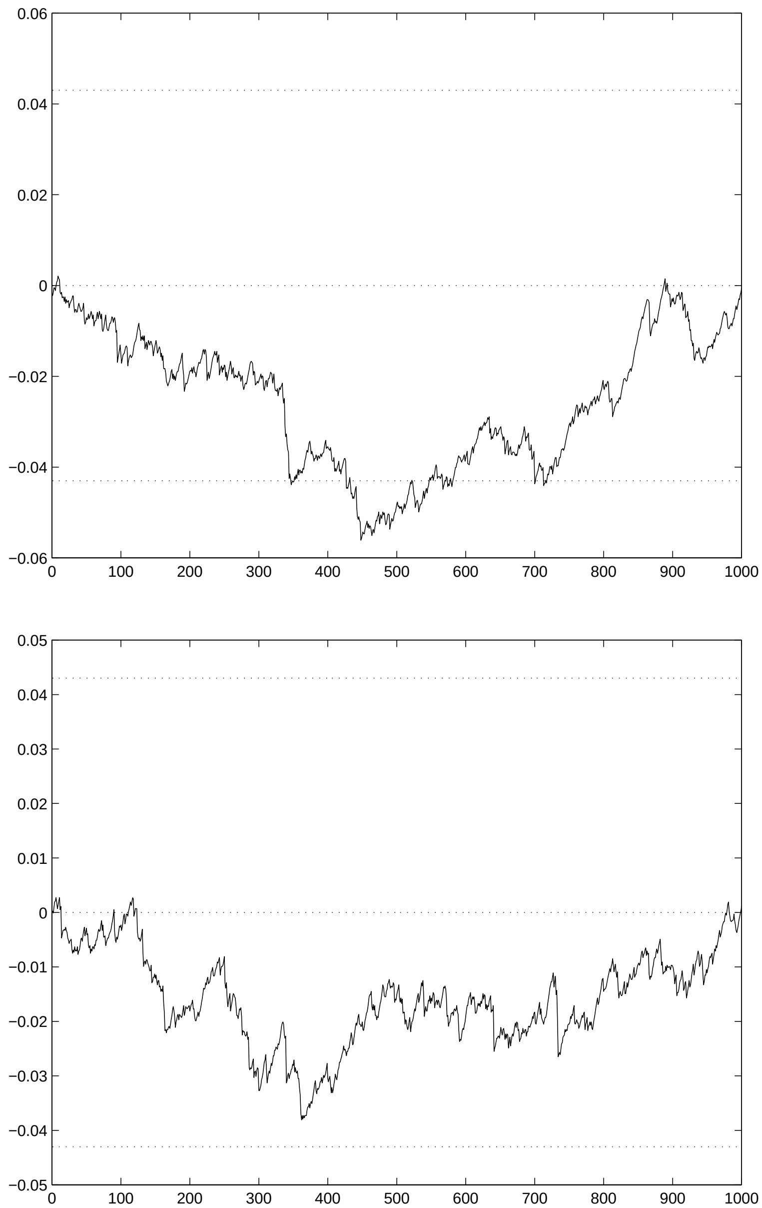


Figure 2: The difference between the sample cdf of the 1000 simulated values of $\widetilde{S}_{500}$ *(Top)* and of $\widetilde{S}_{1000}$ *(Bottom)* and the theoretical cdf of the supremum of a Brownian bridge with the Kolmogorov-Smirnov 95% confidence bands.

$\widetilde{S}$ is calculated using the parameters (3.1). The goodness of fit is based on the Kolmogorov–Smirnov test. The solid line in these graphs represents the difference between the sample cdf and the theoretical cdf of a Brownian bridge while the dotted lines are the 95% confidence intervals stipulated by the Kolmogorov–Smirnov test. This test seems to indicate that the asymptotic behavior fully works for sample sizes of the order 1000 while the qualitative differences between sample sizes of order 125 and 500 are not too big. This observation together with the good fit showed by the QQ-plot in Figure 1 motivate our choice of a window size of 125 data points (or half a business year).

A next issue that we need to clarify is the behavior of S_{125}. Recall that Theorem 2.2 stipulates that the asymptotic behaviors of S_n and $\widetilde{S}_n$ are the same. A verification of this statement is provided in Figure 3 which displays the QQ-plot of 2500 simulated values of S_{125} against 2500 simulated values of $\widetilde{S}_{125}$. In all cases the data generating process is a GARCH model with Student-t innovations and parameters (3.1). To obtain one value of $\widetilde{S}_{125}$, 125 simulated data and the true parameters (3.1) are used, while in the case of S_{125}, 875 data points are simulated, the parameters are estimated on the first 750 data points and the last 125 observations together with the estimated parameters are used to produce the statistic. The two distributions seem indeed very close to each other.

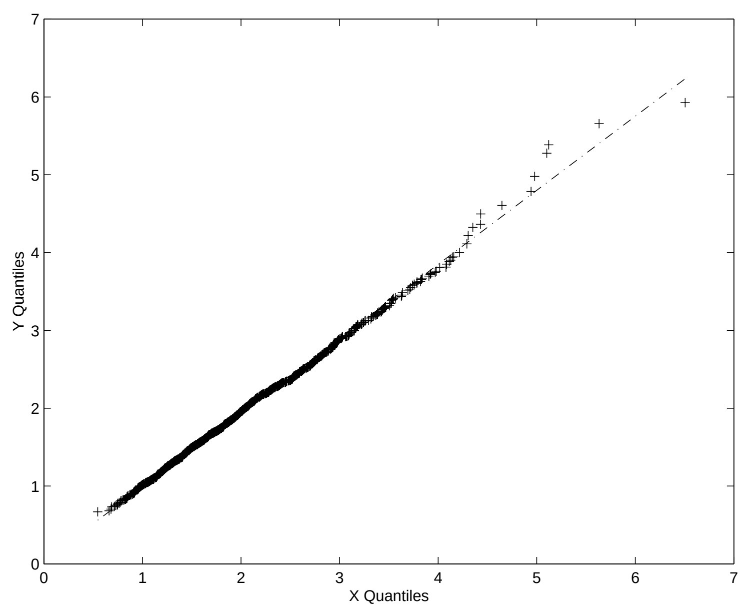


Figure 3: QQ-plot of 2500 values of S_{125} (x-axis) against 2500 values of $\widetilde{S}_{125}$ (y-axis).

This part of the simulation study serves also to define the rejection regions of the test based on the statistic S_{125}. The rejection intervals for a 95% size one-sided, respectively two-sided test for the S_{125} statistic are $(0, 1.01)$ and $(3.32, \infty)$, and $(0, 0.92) \cup (3.8, \infty)$ respectively. The interval that gives a size of 99% to our two-sided test (and that will be used for the data analysis in the next section) is $(0, 0.785) \cup (4.9, \infty)$.

The alternatives we consider in our study are those of GARCH processes with parameters different from those in (3.1). Although Theorem 2.3 gives the theoretical power of our test against various GARCH alternatives, its complicated form renders it of little practical help. For understanding the behavior of the S_n statistic under various GARCH alternatives we again have to turn to a Monte Carlo analysis.

The results are displayed in Figures 4–7. The tests have size 95% and are based on the choice of intervals given by our simulation study as discussed above. In Figures 4 and 5 the parameters α_0 and EZ^4 are kept constant while the other two parameters α_1 and β_1 are made to vary between 0.04 and 0.14 and 0.6 and 0.92 respectively. In Figures 6 and 7 the parameters α_1 and β_1 are kept constant while the other two parameters α_0 and EZ^4 are made to vary between 1.5×10^{-6} and 4.05×10^{-5} and 3 and 9 respectively. For every alternative, 500 simulations were produced. The top and center graphs in Figures 4 and 6 display the power of the one-sided tests (for the top graph, the rejection interval is $(0, 1.01)$, for the center graph $(3.32, \infty)$) while the graphs on the bottom row display the difference between the standard deviation of the alternative models and that of the model with parameters (3.1). The top graphs in Figures 5 and 7 show the power of the two-sided test (rejection region $(0, 0.785) \cup (4.9, \infty)$) while the bottom graphs in the two pictures display the absolute value of the difference between the log of the standard deviation of the alternative models and that of the model with parameters (3.1).

The graphs in Figures 4 and 6 shed light on the relationship between the difference of the unconditional variances and the distribution of the S_{125} statistic under the alternative. They show that the sampling distribution of the statistic S_{125} (calculated with the parameters of the null hypothesis) for GARCH models with lower (higher) unconditional variance is dominated (dominates) the sampling distribution for the null model. Hence rejecting for small (high) values of the statistic S_{125} gives power against alternative models with smaller (larger) unconditional variance. The graphs in Figures 5 and 7 show a strong connection between the power of the test and the absolute value of the difference of the log unconditional variances of the two models. The higher the size of the difference, the higher the power. Even more, Figures 6 and 7 show that the test has equal power against alternatives of equal variance. Note that the variance of the alternative GARCH$(1, 1)$ processes does not depend on the EZ^4 parameter.

As a conclusion, the study motivates the interpretation of the rejection of the null hypothesis not only as signaling the need for another GARCH model but also as a clear sign of a change in the unconditional variance of the time series. More concretely, a rejection on the upper (lower) end of the rejection region also signals an increase (a decrease, respectively) in the unconditional variance of the time series.

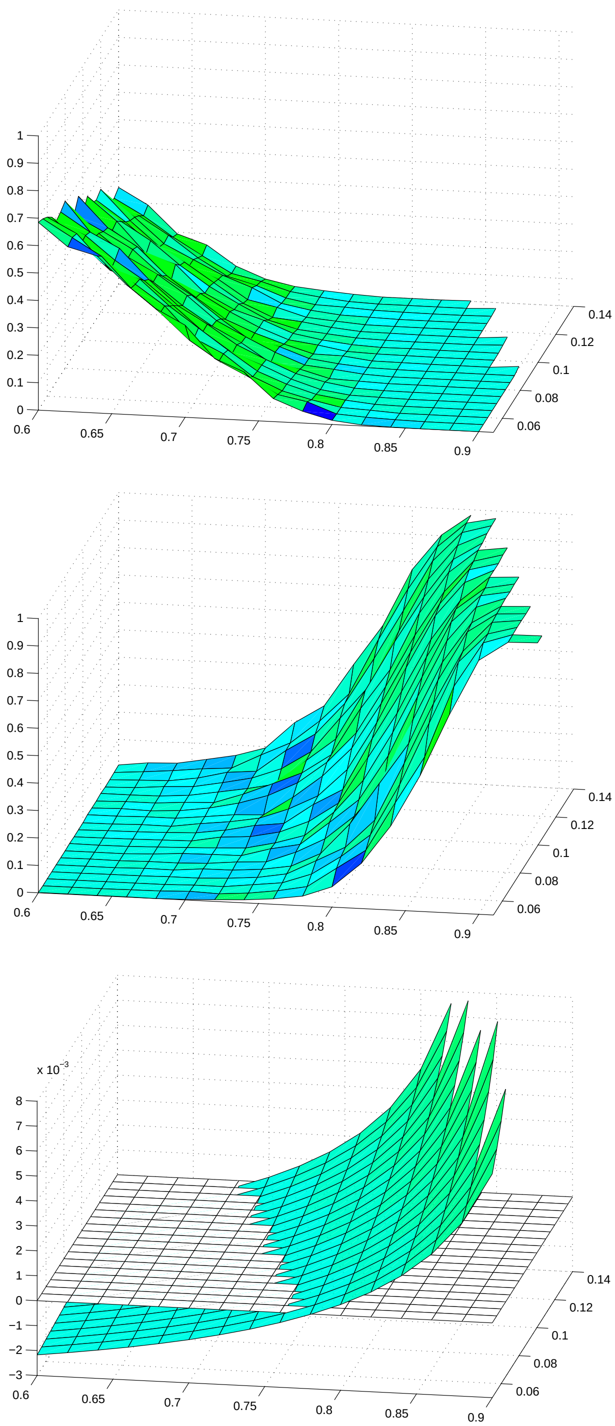


Figure 4: *Top and Center:* Power against GARCH alternatives of a test based on S_{125} and the rejection interval $(0, 1.01)$ *(Top)* and $(3.32, \infty)$ *(Center)*. The parameters α_0^a and ν^a are kept constant and equal to the values in (3.1), i.e., 8.58×10^{-6} and 5.24 respectively. The x- and y-axes show the β_1^a- and α_1^a-values of the alternatives.
Bottom: The difference between the standard deviations of the alternative models and that of the model with parameters (3.1).

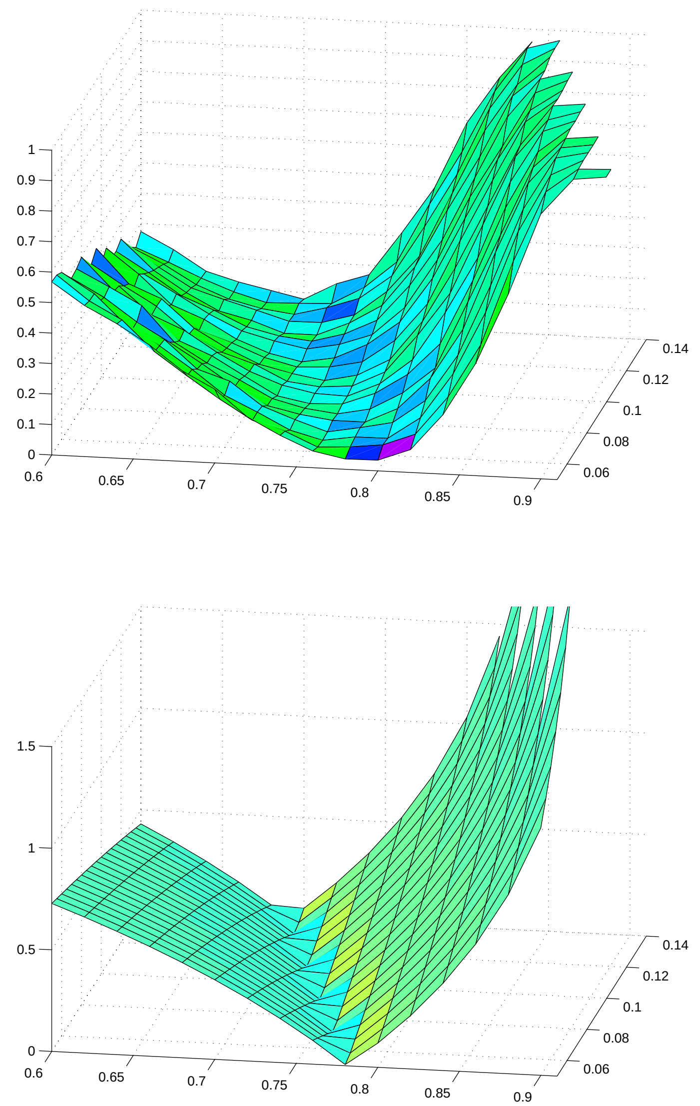


Figure 5: *Top:* Power against GARCH alternatives of a test based on S_{125} and the two-sided rejection region $(0, 0.785) \cup (4.9, \infty)$. The parameters α_0^a and ν^a are kept constant and equal to the values in (3.1), i.e., 8.58×10^{-6} and 5.24, respectively. The x- and y-axes show the β_1^a- and α_1^a-values of the alternatives.
Bottom: The absolute value of the differences between the log of the standard deviations of the alternative models and that of the model with parameters (3.1).

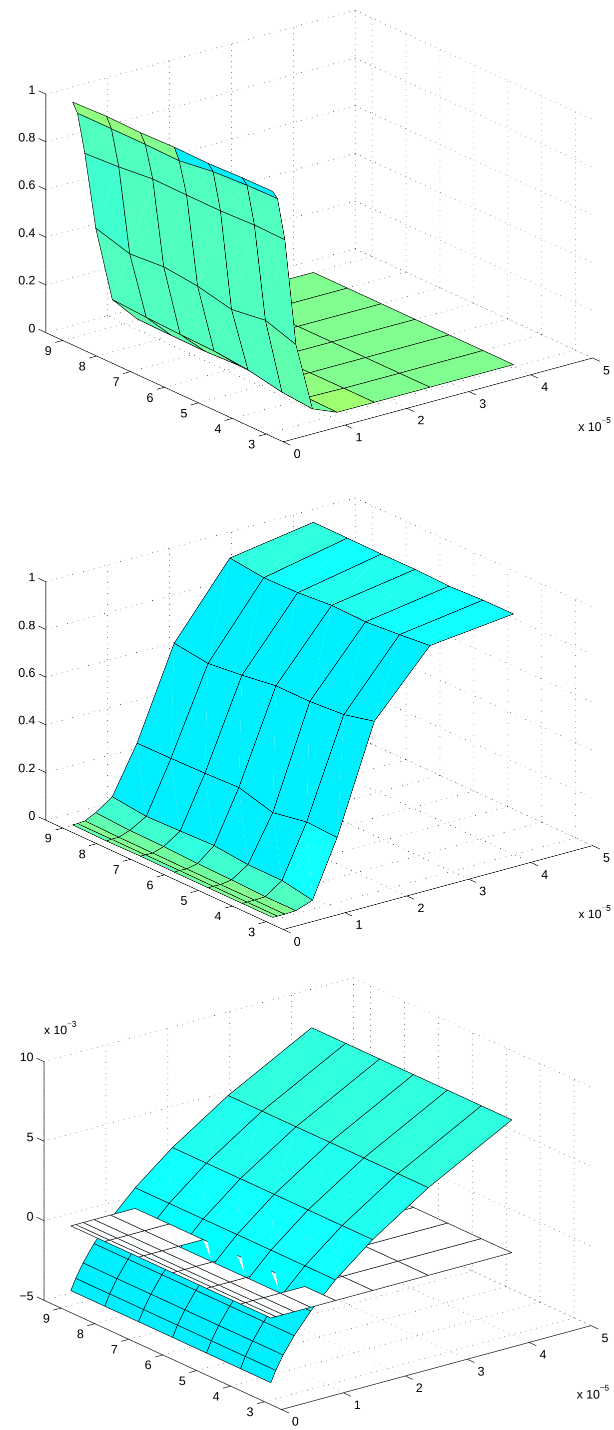


Figure 6: *Top and Center:* Power against GARCH alternatives of a test based on S_{125} and the one-sided rejection intervals $(0, 1.01)$ *(Top)* and $(3.32, \infty)$ *(Center)*. The parameters α_1^a and β_1^a are kept constant to the values in (3.1), i.e., 0.072 and 0.759, respectively. The x- and y-axes show the α_0^a- and ν^a-values of the alternatives.
Bottom: The difference between the standard deviations of the alternative models and that of the model with parameters (3.1).

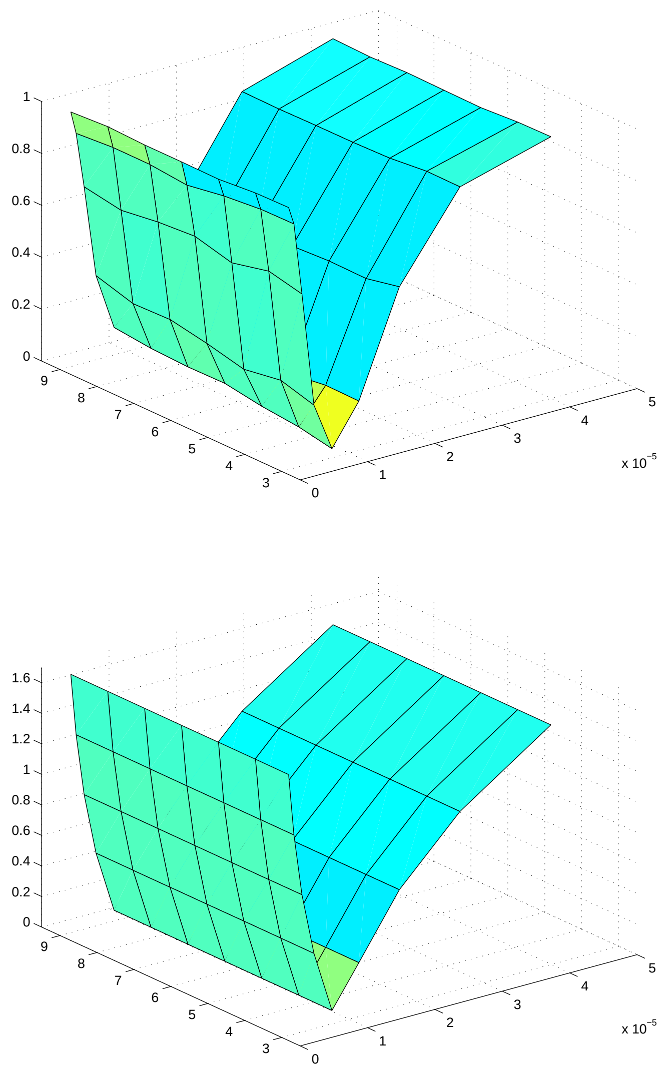


Figure 7: *Top:* Power against GARCH alternatives of a test based on S_{125} and the two-sided rejection interval $(0, 0.785) \cup (4.9, \infty)$ and $(3.32, \infty)$. The parameters α_1^a and β_1^a are kept constant to the values in (3.1), i.e., 0.072 and 0.759, respectively. The x- and y-axes show the α_0^a- and ν^a-values of the alternatives.
Bottom: The absolute value of the difference between the log of the standard deviations of the alternative models and that of the model with parameters (3.1).

4. A STUDY OF THE STANDARD & POOR'S 500 SERIES

We now proceed to analyze a time series that has been previously used to exemplify the presence of LRD in financial log-return series: the Standard 90 and Standard and Poor's 500 composite stock index. This series, covering the period between January 3, 1928, to August 30, 1991, was used in Ding et al. [15], Granger et al. [23], Ding and Granger [14] for an analysis of its autocorrelation structure. It led the authors to the conclusion that the powers of the absolute values of the log-returns are positively correlated over more than 2500 lags, i.e., 10 years. It is hard to believe that this time series is likely to be stationary. It covers the Great Depression, a world war together with the most recent period, marked by major structural changes in the world's economy. In addition, there was a compositional change in the S&P composite index that happened in January 1953 when the Standard 90 was replaced by the broader Standard and Poor's 500 index. Despite all these, Ding et al. [15] conclude the section which describes the data as follows (page 85): "During the Great Depression of 1929 and early 1930s, volatilities are much higher than any other period. There is a sudden drop in prices on Black Monday's stock market crash of 1987, but unlike the Great Depression, the high market volatility did not last very long. *Otherwise, the market is relatively stable.*" Bollerslev and Mikkelsen [9] used the daily returns on the Standard and Poor's 500 composite stock index from January 2, 1953, to December 31, 1990 (a total of 9559 observations) to fit a FIGARCH model under the assumptions of stationarity and LRD. (It is unknown whether the FIGARCH has a stationary version, and if it existed, it had infinite variance marginals, thus the definition of LRD via the ACF would break down. See Giraitis et al. [21] and Mikosch and Stărică [34] for some discussions.)

In the sequel we perform a detailed analysis of the same data set covering the time span from January 2, 1953, to December 31, 1990; see Figure 8. The first goal of the analysis is to check the goodness of fit of a GARCH process with parameters estimated on the first 3 years of data, the period from the beginning in 1953 until the beginning of 1956 (750 observations)

$$\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759, \quad \nu = 5.24 \ , \tag{4.1}$$

to various segments of the data set. Here ν is the number of degrees of freedom of the t-distributed noise sequence (Z_t). The corresponding value of the fourth moment of the estimated residuals of the model (4.1) is $EZ^4 = 7.82$. The analysis verifies if this GARCH$(1,1)$ model which provides a good description to the beginning of the sample can be used to model later periods. In the case of a negative answer we are interested in understanding the type of changes that occurred and, if possible, to pin them to new economic conditions. In other words, the second goal of the analysis is the timing of possible changes in the structure of the data. We try to achieve this goal by evaluating the statistic S_n on a window that moves sequentially through the data. We will chose the window as small as possible to make sure the statistic reacts promptly to possible structural changes. In the end of the section we document the effect which changes in the variance have on the sample ACF. We find that the shape of the sample ACF changes drastically after episodes of increased variance that cannot be properly described by the estimated model.

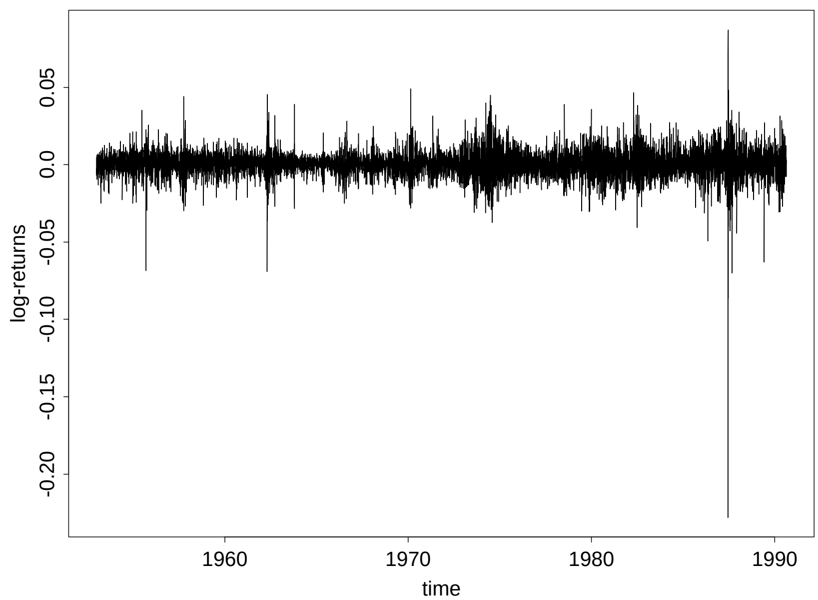


Figure 8: Plot of 9558 *S&P500* daily log-returns.
The year marks indicate the beginning of the calendar year.

The top graph in Figure 9 shows the results of calculating the statistic S_n (see (2.10)) on a weekly basis (i.e., every 5^{th} instant of time) with blocks of $n=125$ past observations, corresponding to approximately 6 months of previous observations. The horizontal lines correspond to the ends of the rejection region of a goodness of fit test of size 99% based on S_{125} statistic as obtained from the simulation study in Section 3. The dotted vertical lines mark the start and the end of economic recessions as determined by the US National Bureau of Economic Research. This graph shows that one simple GARCH$(1,1)$ process (which, according to the S_{125} statistic, models the first ten years of data or so quite well cannot describe the complicated dynamics of longer, possibly non-stationary log-return time series. More precisely, the graph shows that most of the more pronounced violations of the confidence interval are on the upper side. It also shows that most of the recessions of the period under study (apart the one in the beginning of the 60s) are associated with larger than acceptable values of the S_{125} statistic. Recalling the simulation results of Section 3, these two findings also seem to imply that the unconditional variance of the log-returns changes through time and that most of the recessions of the period under study are characterized by higher unconditional variance than the periods of normal economic activity.

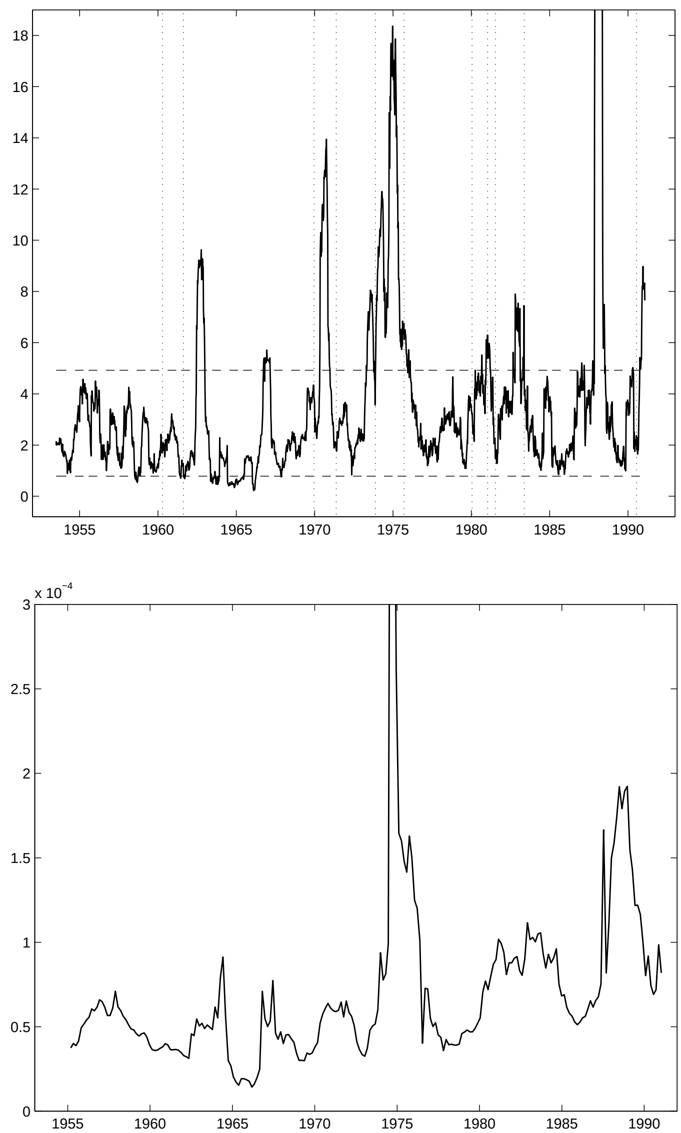


Figure 9: *Top:* The goodness of fit test statistic S_{125} for the *S&P500* data. The horizontal lines are the limits of the 99% confidence interval of S_{125} as obtained from the simulation study in Section 3. The dotted vertical vertical lines mark the start and the end of economic recessions as determined by the National Bureau of Economic Research.
Bottom: The implied GARCH(1,1) unconditional variance of the *S&P500* data. A GARCH(1,1) model is estimated every 2 months using the previous 2 years of data (i.e., 508 observations). The graph displays the variances $\sigma_X^2 = \alpha_0/(1-\alpha_1-\beta_1)$; see (A2.1).

A closer look at the S&P 500 plot in Figure 8 together with the top graph in Figure 9 reveals an almost one-to-one correspondence between the periods of larger absolute log-returns (larger volatility) and the periods when the goodness of fit test statistic S_{125} falls outside and above the confidence region.

If the unconditional variance changes through time, as our analysis seems to indicate, no GARCH$(1,1)$ model could be a good model for the whole period. It is then interesting to verify whether a periodically updated GARCH$(1,1)$ model could account for the more pronounced volatility periods that cannot be explained by the GARCH$(1,1)$ model (4.1). One way to do this is to calculate the implied unconditional GARCH$(1,1)$ variance of a periodically re-estimated GARCH$(1,1)$ model, i.e., one calculates the variance

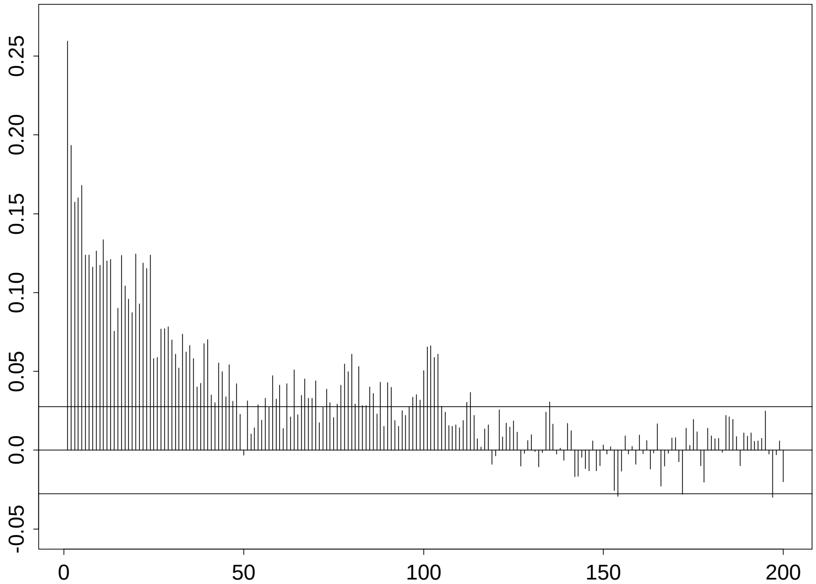
$$\sigma_X^2 = \alpha_0 / \big(1 - (\alpha_1 + \beta_1)\big)$$

based on the periodically re-estimated parameters α_1 and β_1; see (A2.1).

More concretely, we re-estimated a GARCH$(1,1)$ model every 2 months, i.e., every 42 days, on a moving window of 508 past observations, equivalent to roughly two business years of daily log-returns. We then plotted the implied variance σ_X^2. The results of this procedure are displayed in the bottom graph of Figure 9. One notices that the pattern of increased implied unconditional variance is quite similar to the pattern of the excursions of the statistic S_{125} above the 99% quantile threshold. This similarity seems to imply that one can capture the changing patterns of volatility present in the data by periodically updating the GARCH$(1,1)$ model. However a more in-depth study would be needed to substantiate such a statement.

Let us now analyze the impact which these periods of different structural behavior detected by the goodness of fit test statistic S_{125} have on the sample ACF of the time series. The top graph in Figure 9 identifies the period beginning in 1973 and lasting for almost 4 years as the longest and most significant deviation from the hypothesized model. This period is centered around the longest economic recession in the analyzed data. Figure 10 displays the sample ACF of the absolute values $|X_t|$ up to the moment when the change is detected, i.e., beginning of 1973, next to the sample ACF including the 4-year period that followed. The impact of the change in the structure of the time series between 1973 and 1977 on the sample ACF is extremely severe as one sees from the second graph of Figure 10. The graph clearly displays the LRD effect as explained in [34, 35]: exponential decay at small lags followed by almost constant plateau for larger lags together with strictly positive correlations.

Contrary to the belief that the LRD characteristic carries meaningful information about the price generating process, these graphs show that the LRD behavior could be just an artifact due to very plausible structural changes in the log-return data: variations of the unconditional variance due to the business cycle.



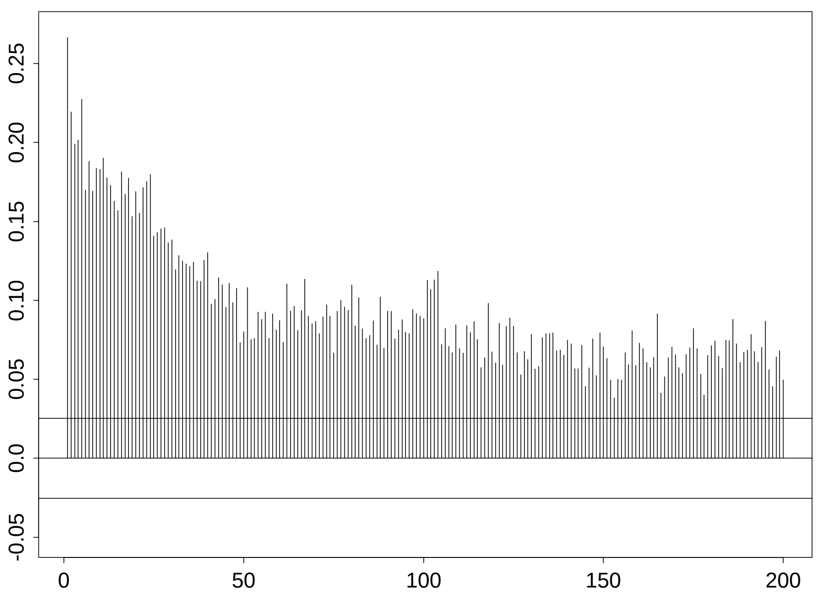


Figure 10: The sample ACF for the absolute log-returns of the first 20 and 24 years *(top and bottom)* of the *S&P500* data.

5. CONCLUDING REMARKS

In this paper we have argued that long financial time series display complicated volatility structures for which the simplifying assumption of constant unconditional variance and constant other moments is too rigid. Modeling the changing unconditional variance (possibly together or instead of the changing conditional one) is an important component of the modelization of long log-returns time series.

A1. APPENDIX

Proof of Lemma 2.1: We have to show the convergence of the finite-dimensional distributions and the tightness in $\mathbb{D}([0,1],\mathbb{R}^m)$. Notice first that for every fixed h,

$$\text{(A1.1)} \qquad \sqrt{n}\Big(\gamma_{n,[nx],X}(h)\Big)_{x\in[0,1]} \xrightarrow{d} \Big(v_X^{1/2}(h)\, W_h(x)\Big)_{x\in[0,1]}\,.$$

in $\mathbb{D}[0,1]$; see Oodaira and Yoshihara [37]; cf. Doukhan [16], Theorem 1 on p. 46. In the latter theorem one has to ensure that $E|X_0X_h|^{2+\epsilon}<\infty$ for some $\epsilon>0$ (this follows from (2.4)) and that the sequence (X_tX_{t+h}) is α-mixing with a sufficiently fast rate for the mixing coefficients; see (A1.2). However, the GARCH(p,q) is strongly mixing with geometric rate since we assume that Z has a Lebesgue density on $\mathbb{R}$ (see Boussama [11]), and so the mixing coefficients converge to zero at an exponential rate, which implies the conditions in the aforementioned theorem.

Thus each of the processes $\sqrt{n}\,\gamma_{n,[n\cdot],X}(h)$ is tight in $\mathbb{D}[0,1]$. Using a generalization of the argument for Lemma 4.4 in Resnick [39], one obtains that the map from $(\mathbb{D}[0,1])^m$ into $\mathbb{D}([0,1],\mathbb{R}^m)$ defined by

$$\Big(x_1,\dots,x_m\Big) \longrightarrow \Big(x_1(t),\dots,x_m(t)\Big)_{t\in[0,1]}$$

is continuous at $(x_1,\dots,x_m)$ in $(\mathbb{C}[0,1])^m$. This and the sample path continuity of the limit process ensure that the processes on the left-hand side of (2.5) are tight in $\mathbb{D}([0,1],\mathbb{R}^m)$.

Notice that the multivariate CLT

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\Big(X_t\,X_{t+1},\,\dots,\,X_t\,X_{t+h}\Big) \xrightarrow{d} \Big(v_X^{1/2}(1)\,W_1(x),\,\dots,\,v_X^{1/2}(h)\,W_h(x)\Big)$$

holds for every fixed x. This is again a consequence of the aforementioned CLT for α-mixing sequences in combination with the Cramér–Wold device. A similar argument for a finite number of x-values yields the convergence of the finite-dimensional distributions. This proves the lemma. □

Remark 1.1. It follows from the argument in the proof of Lemma 2.1 that (2.5) remains valid for stationary strongly mixing sequences (X_t) with $EX=0$, $E|X|^{4+\delta}<\infty$ for some $\delta>0$ and such that $EX_0X_h=0$ for $h\ge 1$, $\mathrm{cov}(X_0X_h,X_0X_l)=0$ for all $h\ne l\ge 1$, and with α-mixing coefficients $\widetilde{\alpha}_i$ satisfying

$$\sum_{i=1}^{\infty}\widetilde{\alpha}_i^{\delta/(2+\delta)}<\infty\ . \tag{A1.2}$$

The latter conditions are needed for the validity of the FCLT in (A1.1); see Oodaira and Yoshihara [37].

Proof of Theorem 2.1: We proceed analogously to Klüppelberg and Mikosch [29]. It follows from Lemma 2.1 and the continuous mapping theorem that, for every fixed $m\ge 1$, in $\mathbb{D}\left([0,1]\times[0,\pi]\right)$

$$\sum_{h=1}^{m}\sqrt{n}\,\widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}\ \stackrel{d}{\longrightarrow}\ \sum_{h=1}^{m}W_h(x)\,\frac{\sin(\lambda h)}{h}\ . \tag{A1.3}$$

According to Theorem 4.2 in Billingsley [7], it remains to show that for every $\epsilon>0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}P\left(\sup_{0\le x\le 1}\ \sup_{0\le\lambda\le\pi}\left|\sum_{h=m+1}^{[nx]-1}\sqrt{n}\,\widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}\right|>\epsilon\right)\ =\ 0\ . \tag{A1.4}$$

Since Z is symmetric the sequences $(r_t)=(\mathrm{sign}(X_t))$ and $(|X_t|)$ are independent. Conditionally on $(|X_t|)$,

$$\sum_{h=m+1}^{k}\sqrt{n}\,\widetilde{\gamma}_{n,k,X}(h)\,\frac{\sin(\cdot h)}{h}\ ,\qquad k=m+1,...,n-1\ ,$$

is a sequence of quadratic forms in the iid Rademacher random variables r_t and with values in the Banach space $\mathbb{C}\,[0,\pi]$ endowed with the sup-norm. Now condition on $(|X_t|)$. Use first a decoupling inequality for Rademacher quadratic forms (e.g. de la Peña and Montgomery–Smith [13], Theorem 1) then the Lévy maximal inequality for sums of iid symmetric random variables, then again the decoupling inequality in reverse order, and finally take expectations with respect to $(|X_t|)$. Then we obtain the inequality

$$P\left(\sup_{0\le x\le 1}\ \sup_{0\le\lambda\le\pi}\left|\sum_{h=m+1}^{[nx]}\sqrt{n}\,\widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}\right|>\epsilon\right)\ \le$$

$$\le\ c_1\,P\left(c_2\sup_{0\le\lambda\le\pi}\left|\sum_{h=m+1}^{n-1}\sqrt{n}\,\widetilde{\gamma}_{nX}(h)\,\frac{\sin(\lambda h)}{h}\right|>\epsilon\right)$$

for certain positive constants c_1, c_2. The right-hand probability can be treated in the same way as the derivation of (6.3) in [28], pp. 1873–1876. Instead of Theorem 3.1 in Rosiński and Woyczyński [40] one can simply use the Cauchy–Schwarz inequality in the first display on p. 1876 in [28] with $\mu = 2$. Then all the calculations for (6.3) remain valid, implying that (A1.4) holds. This concludes the proof of the theorem. □

Remark 1.2. The condition of symmetry of Z is needed only for the application of the Lévy maximal inequality for sums of independent random variables. Alternatively, one can proceed as in the proof of Theorem 3.1 in Klüppelberg and Mikosch [29], p. 980, last display, where instead of the Lévy maximal inequality Doob's 2^{nd} moment maximal inequality for submartingales was used. Then one can follow the lines of the proof of Theorem 1 in Grenander and Rosenblatt [24], Chapter 6.4.

Proof of Theorem 2.2: We start by showing that $\widehat{EZ^4} \xrightarrow{P} EZ^4$. Indeed, consistency of the estimators $\widehat{\alpha}_i$ and $\widehat{\beta}_1$ implies consistency of $\widehat{EZ^4}$. We have by induction, using the definitions of σ_t^2 and $\widehat{\sigma}_t^2$,

$$
\begin{aligned}
&\widehat{EZ^4} - n^{-1}\sum_{i=1}^{n} Z_t^4 = \\
&= n^{-1}\sum_{i=1}^{n} \frac{X_t^4}{\widehat{\sigma}_t^4} - n^{-1}\sum_{i=1}^{n} \frac{X_t^4}{\sigma_t^4} \\
&= n^{-1}\sum_{i=1}^{n} X_t^4 \,\frac{\sigma_t^4 - \widehat{\sigma}_t^4}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= n^{-1}\sum_{i=1}^{n} X_t^4 \,\frac{(\sigma_t^2 - \widehat{\sigma}_t^2)\,(\sigma_t^2 + \widehat{\sigma}_t^2)}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= n^{-1}\sum_{i=1}^{n} X_t^4 \Big[(\alpha_0 - \widehat{\alpha}_0) + (\alpha_1 - \widehat{\alpha}_1)X_{t-1}^2 + (\beta_1 - \widehat{\beta}_1)\sigma_{t-1}^2 + \widehat{\beta}_1(\sigma_{t-1}^2 - \widehat{\sigma}_{t-1}^2)\Big] \times \\
&\qquad\qquad \times \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= (\alpha_0 - \widehat{\alpha}_0)\; n^{-1}\sum_{i=1}^{n} X_t^4 \left(1 + \widehat{\beta}_1 + \widehat{\beta}_1^2 + \cdots + \widehat{\beta}_1^t\right) \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&\quad + (\alpha_1 - \widehat{\alpha}_1)\; n^{-1}\sum_{i=1}^{n} X_t^4 \left(X_{t-1}^2 + \widehat{\beta}_1 X_{t-2}^2 + \cdots + \widehat{\beta}_1^t X_0^2\right) \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&\quad + (\beta_1 - \widehat{\beta}_1)\; n^{-1}\sum_{i=1}^{n} X_t^4 \left(\sigma_{t-1}^2 + \widehat{\beta}_1 \sigma_{t-2}^2 + \cdots + \widehat{\beta}_1^t \sigma_0^2\right) \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&\quad + (\sigma_0^2 - \widehat{\sigma}_0^2)\; n^{-1}\sum_{i=1}^{n} X_t^4\, \widehat{\beta}_1^t\, \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= (\alpha_0 - \widehat{\alpha}_0)\, I_1 + (\alpha_1 - \widehat{\alpha}_1)\, I_2 + (\beta_1 - \widehat{\beta}_1)\, I_3 + (\sigma_0^2 - \widehat{\sigma}_0^2)\, I_4 \;.
\end{aligned}
$$

Notice that, by consistency of the parameter estimators and since $(Z_t^4 \sigma_t^2)$ is ergodic,

$$\begin{aligned} I_1 &\le (1-\widehat{\beta}_1)^{-1}\, n^{-1} \sum_{i=1}^n Z_t^4\, \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\widehat{\sigma}_t^4} \\ &\le (1-\widehat{\beta}_1)^{-1}\, n^{-1} \sum_{i=1}^n Z_t^4 \left[\sigma_t^2\, \widehat{\alpha}_0^{-2} + \widehat{\alpha}_0^{-1}\right] \\ &\stackrel{\text{a.s.}}{\longrightarrow} (1-\beta_1)^{-1}\, EZ^4 \left[E\, \sigma^2\, \alpha_0^{-2} + \alpha_0^{-1}\right] . \end{aligned}$$

By similar arguments, for any $\delta > 0$ and $\epsilon > 0$ such that $\beta_1 + \epsilon < 1$,

$$\begin{aligned} &P\Big(I_2 > \delta\Big) \le \\ &\le P\left(n^{-1} \sum_{i=1}^n X_t^4 \Big(X_{t-1}^2 + (\beta_1+\epsilon) X_{t-2}^2 + \cdots + (\beta_1+\epsilon)^t X_0^2\Big)\, \frac{\sigma_t^2 + \widehat{\sigma}_t^2}{\sigma_t^4\, \widehat{\sigma}_t^4} > \delta\right) \\ &\quad + P\Big(\widehat{\beta}_1 > \beta_1+\epsilon\Big) \\ &\le P\left(n^{-1} \sum_{i=1}^n Z_t^4 \Big(X_{t-1}^2 + (\beta_1+\epsilon) X_{t-2}^2 + \cdots + (\beta_1+\epsilon)^t X_0^2\Big) \left(\sigma_t^2\, \widehat{\alpha}_0^{-2} + \widehat{\alpha}_0^{-1}\right) > \delta\right) \\ &\quad + o(1) . \end{aligned}$$

It is not difficult to see, by an application of the Cauchy–Schwarz inequality, that the first moments of

$$n^{-1} \sum_{i=1}^n Z_t^4 \left(X_{t-1}^2 + (\beta_1+\epsilon)\, X_{t-2}^2 + \cdots + (\beta_1+\epsilon)^t\, X_0^2\right) \sigma_t^2$$

and

$$n^{-1} \sum_{i=1}^n Z_t^4 \left(X_{t-1}^2 + (\beta_1+\epsilon)\, X_{t-2}^2 + \cdots + (\beta_1+\epsilon)^t\, X_0^2\right) ,$$

are bounded, uniformly for n. Therefore I_2 is stochastically bounded, and a similar argument applies to I_3. Finally,

$$P\Big(n\, I_4 > \delta\Big) \le P\left(\sum_{i=1}^n Z_t^4\, (\beta_1+\epsilon)^t \left(\sigma_t^2\, \widehat{\alpha}_0^2 + \widehat{\alpha}_0^{-1}\right) > \delta\right) + P\Big(\widehat{\beta}_1 > \beta_1 + \epsilon\Big) .$$

The second probability vanishes by consistency of $\widehat{\beta}_1$. Moreover,

$$\sum_{i=1}^n Z_t^4\, (\beta_1+\epsilon)^t\, \sigma_t^2 \qquad \text{and} \qquad \sum_{i=1}^n Z_t^4\, (\beta_1+\epsilon)^t ,$$

have bounded first moments. This implies that I_4 is stochastically bounded, and $n^{-1} I_4 \stackrel{P}{\to} 0$. Collecting the bounds for all I_j, we conclude by the law of large numbers that $\widehat{EZ^4} \stackrel{P}{\to} EZ_4$.

For the remaining proof we follow the lines of the proof of Theorem 2.1. Write $\widehat{v}_X(h)$ for the sample version of $v_X(h)$. By consistency of $\widehat{EZ^4}$, $\widehat{\alpha}_i$, $\widehat{\beta}_1$ and the form of $v_X(h)$, see (A2.2), we have $\widehat{v}_X(h) \stackrel{P}{\to} v_X(h)$ for every h. This fact and the continuous mapping theorem immediately yield that (A1.3) remains valid with $v_X(h)$ replaced by $\widehat{v}_X(h)$. So it remains to show

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \frac{\gamma_{n,[nx],X}(h)}{\widehat{v}_X^{1/2}}\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) = 0 \,.$$

Notice that for every $h \ge 1$,

$$\widehat{v}_X(h)^{-1/2} \le \widehat{\sigma}_X^{-4} = \widehat{\alpha}_0^{-4} \big(1-\widehat{\varphi}_1\big)^{-4} \,.$$

See Appendix A2. By the assumptions, the estimators $\widehat{\alpha}_i$, $\widehat{\beta}_1$ are independent of $(\mathrm{sign}(X_t))$, and so is $\widehat{EZ^4}$ by construction of the residuals. Thus, conditionally on $(|X_t|)$,

$$\sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \frac{\gamma_{n,[nx],X}(h)}{\widehat{v}_X^{1/2}}\, \frac{\sin(\lambda h)}{h}$$

is a random quadratic form in the variables $\mathrm{sign}(X_t)$, which, by symmetry of (Z_t) are independent of the coefficients of the quadratic form which only depend on the sequence $(|X_t|)$. An application of the contraction principle for Rademacher quadratic forms (cf. Kwapień and Woyczyński [31]) implies that for some constants $c_1, c_2 > 0$

$$\begin{aligned} &P\left(\sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \frac{\gamma_{n,[nx],X}(h)}{\widehat{v}_X^{1/2}}\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) \le \\ &\quad \le c_1\, P\left(c_2 \max_h \widehat{v}_X^{-1/2} \sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) \\ &\quad \le c_1\, P\left(c_2\, \widehat{\alpha}_0^{-4} \big(1-\widehat{\varphi}_1\big)^{-4} \sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right). \end{aligned}$$

Thus it remains to show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) = 0 \,,$$

which follows along the lines of the proof of Theorem 2.1. □

A2. APPENDIX

Consider a GARCH$(1,1)$ process (X_t) with parameters $\alpha_0, \alpha_1, \beta_1$. We write $\varphi_1 = \alpha_1 + \beta_1$ and assume $EX^4 < \infty$. From the calculations below it follows that the condition

$$1 - \left(\alpha_1^2 \, EZ^4 + \beta_1^2 + 2\,\alpha_1\,\beta_1\right) \; > \; 0$$

must be satisfied. The squared GARCH$(1,1)$ process can be rewritten as an ARMA$(1,1)$ process by using the defining equation (2.1):

$$X_t^2 - \varphi_1 \, X_{t-1}^2 \; = \; \alpha_0 + \nu_t - \beta_1 \, \nu_{t-1} \; ,$$

where $(\nu_t) = (X_t^2 - \sigma_t^2)$ is a white noise sequence. Thus, the covariance structure of

$$U_t = X_t^2 - EX^2 \; , \qquad t \in \mathbb{Z} \; ,$$

is that of a mean-zero ARMA$(1,1)$ process. The values of $\gamma_U(h)$ are given on p. 87 in Brockwell and Davis [12]:

$$\begin{aligned} \gamma_U(0) \; &= \; \sigma_\nu^2 \left[1 + \frac{(\varphi_1 - \beta_1)^2}{1 - \varphi_1^2} \right] \; , \\ \gamma_U(1) \; &= \; \sigma_\nu^2 \left[\varphi_1 - \beta_1 + \frac{(\varphi_1 - \beta_1)^2 \, \varphi_1}{1 - \varphi_1^2} \right] \; , \\ \gamma_U(h) \; &= \; \varphi_1^{h-1} \, \gamma_U(1) \; , \qquad h \geq 2 \; . \end{aligned}$$

Straightforward calculation yields

$$\text{(A2.1)} \qquad \begin{aligned} \sigma_\nu^2 \; &= \; (EZ^4 - 1) \; E\sigma_1^4 \; = \; \frac{1 + \varphi_1}{1 - \varphi_1} \, \frac{\alpha_0^2 (EZ^4 - 1)}{1 - \left(\varphi_1^2 + \alpha_1^2 (EZ^4 - 1)\right)} \; , \\ \sigma_X^2 \; &= \; \frac{\alpha_0}{1 - \varphi_1} \; . \end{aligned}$$

Thus we can calculate the quantities

$$v_X(h) \; = \; E(X_0^2 X_h^2) \; = \; \gamma_U(h) + \sigma_X^4 \; , \qquad h \geq 1 \; ,$$

which occur in the definition of the change point statistics and goodness of fit test statistics of Section 2. We obtain:

$$\text{(A2.2)} \qquad v_X(h) \; = \; \sigma_X^4 \left(\frac{(EZ^4 - 1) \; \alpha_1 \left(1 - \varphi_1^2 + \alpha_1 \, \varphi_1\right)}{1 - \left(\varphi_1^2 + \alpha_1^2 (EZ^4 - 1)\right)} \; \varphi_1^{h-1} + 1 \right) \; , \qquad h \geq 1 \; .$$

ACKNOWLEDGMENTS

We would like to thank the Newton Institute in Cambridge for its stimulating hospitality during the Managing Uncertainty period in June/July 2001.

REFERENCES

[1] Anderson, T.W. (1993). Goodness of fit tests for spectral distributions, *Ann. Statist.*, **21**, 830–847.

[2] Baillie, R.T.; Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **74**, 3–30.

[3] Bartlett, M.S. (1954). Problèmes de l'analyse spectrale des séries temporelles stationnaires, *Publ. Inst. Statist. Univ. Paris*, **III-3**, 119–134.

[4] Beran, J. (1994). *Statistics for Long–Memory Processes*, Monographs on Statistics and Applied Probability, No. 61, Chapman and Hall, New York.

[5] Berkes, I.; Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation, *Bernoulli*, **9**, 201–228.

[6] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications, *Ann. Math. Statist.*, **42**, 1656–1670.

[7] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[8] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**, 307–327.

[9] Bollerslev, T. and Mikkelsen, H.O. (1996). Modeling and pricing long memory in stock market volatility, *J. Econometrics*, **73**, 151–184.

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

[11] Boussama, F. (1998). *Ergodicité, mélange et estimation dans le modelès GARCH*, PhD Thesis, Université 7 Paris.

[12] Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting*, Springer, New York.

[13] de la Peña, V.H. and Montgomery-Smith, S.J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics, *Ann. Probab.*, **23**, 806–816.

[14] Ding, Z. and Granger, C.W.J. (1996). Modeling volatility persistence of speculative returns: A new approach, *J. Econometrics*, **73**, 185–215.

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

[16] Doukhan, P. (1994). *Mixing. Properties and Examples*, Lecture Notes in Statistics, **85**, Springer Verlag, New York.

[17] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

[18] Engle, R.F. (Ed.) (1995). *ARCH Selected Readings*, Oxford University Press, Oxford (U.K.).

[19] Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. With comments and a reply by the authors, *Econometric Rev.*, **5**, 1–87.

[20] Giraitis, L. and Leipus, R. (1992). Testing and estimating in the change-point problem of the spectral function, *Lith. Math. Trans. (Lit. Mat. Sb.)*, **32**, 20–38.

[21] Giraitis, L.; Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: dependence structure and central limit theorem, *J. Econom. Theory*, **16**, 3–22.

[22] Giraitis, L. and Robinson, P.M. (2001). Whittle estimation of ARCH models, *J. Econom. Theory*, **17**, 608–631.

[23] Granger, C.W.J. and Ding, Z. (1996). Varieties of long memory models, *J. Econometrics*, **73**, 61–77.

[24] Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, 2nd edition, Chelsea Publishing Co., New York.

[25] Hida, T. (1980). *Brownian Motion*, Springer, New York.

[26] Horváth, L.; Kokoszka, P. and Teyssiere, G. (2001). Empirical process of squared residuals of an ARCH sequence, *Ann. Statist.*, **29**, 445–469.

[27] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*, Springer, Berlin, New York.

[28] Klüppelberg, C. and Mikosch, T. (1996). The integrated periodogram for stable processes, *Ann. Statist.*, **24**, 1855–1879.

[29] Klüppelberg, C. and Mikosch, T. (1996). Gaussian limit fields for the integrated periodogram, *Ann. Appl. Probab.*, **6**, 969–991.

[30] Kokoszka, P. and Mikosch, T. (1997). The integrated periodogram for long-memory processes with finite or infinite variance, *Stoch. Proc. Appl.*, **66**, 55–78.

[31] Kwapień, S. and Woyczyński, W.A. (1992). *Random Series and Stochastic Integrals: Single and Multiple*, Birkhäuser, Basel.

[32] Mikosch, T. (2003). *Modelling dependence and tails of financial time series*. In: "Extreme Values in Finance, Telecommunications and the Environment" (B. Finkenstädt and H. Rootzén, Eds.), Chapman and Hall, Boca Raton, pp. 185–286.

[33] Mikosch, T. and Straumann, D. (2002). Whittle estimation in a heavy-tailed GARCH(1,1) model, *Stoch. Proc. Appl.*, **100**, 187–222.

[34] Mikosch, T. and Stărică, C. (2003). *Long range dependence effects and ARCH modeling*. In: "Long Range Dependence" (P. Doukhan, G. Oppenheim and M.S. Taqqu, Eds.), Birkhauser, Boston, pp. 439–460.

[35] Mikosch, T. and Stărică, C. (2004). Non-stationarities in financial time series, the long range dependence and the IGARCH effects, *Rev. Economics and Statistics*, **86**, 378–390.

[36] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1, 1) model, *J. Econom. Theory*, **6**, 318–334.

[37] Oodaira, H. and Yoshihara, K. (1972). Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition, *Kōdai Math. Sem. Rep.*, **24**, 259–269.

[38] Priestley, M.B. (1981). *Spectral Analysis and Time Series*, vols. I and II, Academic Press, New York.

[39] Resnick, S.I. (1986). Point processes, regular variation and weak convergence, *Adv. Appl. Probab.*, **18**, 66–138.

[40] Rosiński, J. and Woyczyński, W.A. (1987). Multilinear forms in Pareto–like random variables and product random measures, *Coll. Math.*, **51**, 303–313.

[41] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

[42] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*, Almqvist och Wicksel, Uppsala.