Non-stationarities in stock returns[1]

Cătălin Stărică[2] and Clive Granger[3]

First draft: May 2000

This version: October 2004

[1]This research has been supported by The Bank of Sweden Tercentenary Foundation and by a grant from Jan Wallanders och Tom Hedelius Stiftelse.

[2]Department of Mathematical Statistics, Chalmers University of Technology and Department of Economics, Göteborg University, Gothenburg, S–412 96, Sweden. Email: starica@math.chalmers.se. Web page: www.math.chalmers.se/˜starica

[3]Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Email: cgranger@weber.ucsd.edu

Abstract

The paper outlines a methodology for analyzing daily stock returns that relinquishes the assumption of global stationarity. Giving up this common working hypothesis reflects our belief that fundamental features of the financial markets are continuously and significantly changing. Our approach approximates locally the non-stationary data by stationary models. The methodology is applied to the *S&P 500* series of returns covering a period of over seventy years of market activity. We find most of the dynamics of this time series to be concentrated in shifts of the unconditional variance. The forecasts based on our non-stationary unconditional modeling were found to be superior to those obtained in a stationary long memory framework or to those based on a stationary Garch(1,1) data generating process.

JEL classification: C14, C22, C52, C53.

Keywords and Phrases: stock returns, non-stationarities, locally stationary processes, volatility, sample autocorrelation, long range dependence, Garch(1,1) data generating process.

1. Introduction

This paper reports the results of a non-stationary analysis of the time series properties of daily returns of the *S&P 500* index between January 1928 and May 2000, i.e. more than seventy years of financial market history. Non-stationary modeling has a long tradition in the econometric literature that focuses on modeling financial returns predating the currently prevalent *stationary, conditional* paradigm (of which the autoregressive conditionally heteroscedastic (ARCH) -type processes and stochastic volatility models are outstanding examples) (see, for example, Officer (1976) or Hsu, Miller and Wichern (1974)). Our endeavor is motivated by growing evidence of instability in the stochastic features of stock returns[4] as well as by an increased awareness of the severe negative impact of assuming stationarity when it is not a good modeling approximation (see Stărică (2003) and Herzel et al. (2004)). The project addresses three central questions.

The first question is methodological: How can one analyze index returns in the non-stationary conceptual framework? Our approach is to approximate locally the non-stationary data by stationary models. The changing nature of the probabilistic features of the data, i.e. marginal distribution and dependency structure imposes a periodic updating of the

[4]See, for example, Stock and Watson (1996). Recently, a growing body of econometric literature (Diebold (1986), Lamoureux and Lastrapes (1990), Simonato (1992), Cai (1994), Lobato and Savin (1998), Mikosch and Stărică (2004) among others) argues that most of the features of return series that puzzled through their omni-presence, the so called "stylized facts", including the ARCH effects, the slowly decaying sample ACF for absolute returns and the IGARCH effect (for definitions and details see Mikosch and Stărică (2003)) are manifestations of non-stationary changes in the dynamic of returns.

approximating stationary model. The goal of our methodology is identifying the *intervals of homogeneity*, i.e. the intervals where a certain estimated stationary model describes well the data. On an interval of homogeneity the parameters of the data generating process do not vary much relative to the estimation error of the parameters of the stationary model used as an approximation[5] (see Härdle et al. (1999)). The main tool in identifying the homogeneity intervals is a goodness of fit test for linear models in the spectral domain, related to the ones proposed in Picard (1985) and Klüppelberg and Mikosch (1996) (see Section 2 for a detailed description of the methodology and Section 3 for the relevant statistical results concerning the goodness of fit test).

A second related question is: What type of (major) non-stationarities affect the *S&P 500* returns? The in-depth analysis in Sections 4 to 6 as well as the forecasting results in Sections 7 and 8 indicate the time-varying second unconditional moment as the main source of non-stationarity of returns on the *S&P 500* index.[6]

[5]For example, if the data generating process of returns is an independent sequence of random variables with time-changing unconditional variance, an interval of homogeneity is a period of time when one has reasons to believe that the variance is almost constant (more precisely, that the change in variance cannot be distinguished from estimation error). On the intervals of homogeneity, one approximates the (slowly) changing unconditional variance of returns with a constant. Hence, in the end, the changing pattern of unconditional variance will be approximated by a step function. The resulting model is a process with piecewise constant variance.

[6]Our findings and the modeling methodology that they motivate follow on the steps of Officer (1976) and Hsu, Miller and Wichern (1974). The former, using a non-parametric approach to volatility estimation, reports evidence of time-varying second moment for the time series of returns on the *S&P 500* index and

The third question is conceptual: How should we interpret the slow decay of the sample *autocorrelation function* (ACF) of absolute returns? Should we take it at face value, supposing that events that happened a number of years ago bear an impact on the present dynamics of returns? Or, are the non-stationarities[7] in the returns responsible for its presence? The answer to this question has important implications for estimation and forecasting. In the case of the first alternative, a long history of the time series would carry significant information while in the second case only a short past will be of most use in forecasting. A commonly held belief in the econometric community is that taking the slow decay of the sample ACF at face value (even though it might be caused by non-stationary changes in the unconditional variance) is a meaningful way of making use of the past in forecasting the future. In other words, estimating long memory *stationary* models (based on the slow decay of the sample ACF) and using them in forecasting exploits *in a meaningful* way the patterns of change observed in the past. In Section 7 we investigate the relevance of this assumption by means of a comparison of the forecasting performance of a stationary long memory process estimated on the series of absolute returns with that of a model based on the paradigm of changing unconditional variance. The results seem to show the superiority of the second method, supporting the hypothesis that the changes of the unconditional variance are the source of long memory in absolute stock returns.

industrial production. The later modeled the returns as a non-stationary process with discrete shifts in the unconditional variance.

[7]The list of related relevant references includes Hidalgo and Robinson (1996), Lobato and Savin (1998), Granger and Hyung (1999), Granger and Teräsvirta (1999), Diebold and Inoue (2001), Mikosch and Stărică (2004).

As the paper addresses the issue of volatility and proposes a novel modeling paradigm, a comparison with the Garch framework, the current market leader in volatility modeling, is inevitable. In Section 8 we present the results of a forecasting comparison of our methodology with that based on a Garch(1,1) model. We find that the non-stationary unconditional approach produces significantly better volatility forecasts at longer time horizons, i.e. between 10 and 250 days.

2. Delimitation of intervals of homogeneity

From a traditional time series point of view, the information contained in the time series of daily returns can be split in two components: the sign of the returns and their size. Empirical evidence shows that the sign of daily returns is not predictable. Hence in what follows we will concentrate on studying the time series of absolute returns. More precisely, due to the presence of heavy tails in the absolute returns, we analyze the *logarithm of the absolute values of daily returns*[8], $X_t := \log(|r_t|)$.

In what follows we assume that X_t follow a locally stationary process in the sense of Dahlhaus (1997). In words, we assume that the stochastic features of the data generating process of the absolute values of daily returns, i.e. the marginal distribution and the dependency structure, evolve slowly and smoothly through time as a result of continuous changes in financial markets.

[8]The implications of this analysis on the model choice for the series of returns (r_t) are also discussed in Section 5 and 6.

Our methodology consists in *locally* approximating the dynamics of the data X_t with stationary linear models. Assume, for example, that X_t are generated by an ARMA process with parameters that are smooth functions of time

$$\Phi(t,B)(X_t-\mu(t))=\Theta(t,B)Z_t,\quad Z_t=\sigma(t)\epsilon_t, \tag{2.1}$$

where ϵ_t are iid with $E\epsilon_t=0$, $E\epsilon_t^2$=1, while B denotes the back-shift operator.[9] Our approach yields an approximation of the functions $\Phi(t,B)$, $\Theta(t,B)$, $\mu(t)$, $\sigma^2(t)$ with step functions that are constant on appropriately defined *homogeneity intervals*. This section describes our approach to identifying the homogeneity intervals of the series X_t.

A linear process X_t with unconditional mean μ is defined as

$$X_t-\mu=\sum_{j=-\infty}^{\infty}\psi_j Z_{t-j}=\psi(B)Z_t,\quad t\in\mathbb{Z}, \tag{2.2}$$

where the innovations (Z_t) are a sequence of iid random variables with mean 0 and finite variance σ^2, and

$$\psi(z)=\sum_{j=-\infty}^{\infty}\psi_j z^j.$$

The assumption

$$\sum_{j=-\infty}^{\infty}|\psi_j|j<\infty$$

ensures that X_t is properly defined as an a.s. absolutely converging series. The linear process (2.2) has the following spectral density function

$$f_\psi(\lambda)=|\psi(e^{-i\lambda})|^2\sigma^2/(2\pi),\quad \lambda\in[0,\pi] \tag{2.3}$$

[9]This type of stochastic process is a locally stationary process in the sense of Dahlhaus (1997).

while the spectral distribution function is given by

$$F_\psi(\lambda) = \int_0^\lambda f_\psi(x)\, dx = \sigma^2/(2\pi) \int_0^\lambda |\psi(e^{-ix})|^2\, dx\,, \quad \lambda \in [0,\pi]\,.$$

In the sequel, the linear process (2.2) with mean μ, variance of the noise σ^2, and spectral density f_ψ will be parcimoniously denoted by $\mathcal{M}_{\mu,\,\sigma^2,\,f_\psi}$.

The intervals of homogeneity are constructed by monitoring the changes in the spectral distribution function of X_t as follows.[10] Assume we know that the subsample X_{m_1}, X_{m_1+1}, $\ldots$, X_{m_2} is well described by $\mathcal{M}_{\mu,\,\sigma^2,\,f_\psi}$, a linear parametric model with mean μ, variance of the noise σ^2, and spectral density f_ψ. In other words, assume that the interval of homogeneity under construction contains at least the observations m_1 up to m_2. We want to decide whether the following p observations, $X_{m_2+1}, \ldots, X_{m_2+p}$ also belong to the interval. To accomplish this, we test the hypothesis that the linear model $\mathcal{M}_{\mu,\,\sigma^2,\,f_\psi}$ fits well the subsample $X_{m_2+p-s}, \ldots, X_{m_2+p}$ that contains the new p data points (the size s of the subsample on which the test is conducted is kept constant). A test statistic $T(n, X, \mathcal{M}_{\mu,\sigma^2,f_\psi})$ with a known asymptotic distribution to be specified in the sequel, (see (3.3)) is calculated. As the notation emphasizes, the test statistic $T(n, X, \mathcal{M}_{\mu,\sigma^2,f_\psi})$ is a function of the subsample $X_{m_2+p-s}, \ldots, X_{m_2+p}$ as well as of the model $\mathcal{M}_{\mu,\,\sigma^2,\,f_\psi}$. The value obtained is compared with the asymptotic distribution of the test statistic under the null

[10]The method is related to the one proposed in Picard (1985) for detecting changes in the spectral distribution function of a time series and further developed for various linear processes under mild assumptions on the moments of X and the coefficients of the process by Giraitis and Leipus (1992) and Klüppelberg and Mikosch (1996).

hypothesis that the subsample $X_{m_2+p-s}, \ldots, X_{m_2+p}$ is a stationary sequence from the model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$. If the value of the statistic falls within the asymptotic confidence interval, the homogeneity interval is extended to include the observations $X_{m_2+1}, \ldots, X_{m_2+p}$. Otherwise a new homogeneity interval commences with the block $X_{m_2+p-s}, \ldots, X_{m_2+p}$. The model to describe the data dynamic on the new interval of homogeneity is estimated on this block. The estimated parameters are then assumed to be the true parameters of the new data generating process and the procedure of building a homogeneity interval is reiterated.[11]

3. A goodness of fit test based on Bartlett's weighted integrated periodogram

We concentrate now on the statistical aspects of the goodness of fit test central to the methodology presented above. In this section we define the test statistic $T(n, X, \mathcal{M}_{\mu,\sigma^2,f_\psi})$ and specify its asymptotic distribution. Let

$$\gamma_{n,Y}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(Y_t-\mu)(Y_{t+h}-\mu)\,, \quad h = 0, 1, 2, \ldots n-1,$$

[11]Note that we choose to neglect the estimation error implicit in taking the estimated values of the model's parameters for the real ones. This approximation is acceptable in view of the following two facts. First, a simulation study (see Section 4 for a brief discussion of the results of the study) indicates that the finite sample distribution of the test statistic constructed using the true parameters is very close to that of the test statistic obtained using the estimated parameter values. More important, our empirical experience shows that allowing for large deviations of the asymptotic variance of the test statistics from the values prescribed by Corollary 3.2 does not change at all the results of the analysis.

denote the *sample autocovariance function* of the stationary sequence Y_t centered with the true mean μ. Recall that the *periodogram* defined as

$$(3.1) \qquad I_{n,Y}(\lambda) \;=\; \frac{1}{2\pi}\left(\gamma_{n,Y}(0)+2\sum_{h=1}^{n-1}\gamma_{n,Y}(h)\cos(h\lambda)\right), \quad \lambda\in[0,\pi],$$

is the natural (method of moment) estimator of the spectral density f_Y of the stationary sequence (Y_t); see Brockwell and Davis (1991) or Priestley (1981).

If the linear model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$ defined in (2.2) is the true data generating process for the subsample $X_{m_2+p-s},\ldots,X_{m_2+p}$, the covariances observed in the data ought to match up[12] to the covariances implied by the linear process. The information on the covariance structure of the data is conveniently summarized by the *integrated periodogram* or *empirical spectral distribution function*

$$(3.2) \qquad J_{n,X}(\lambda) := \int_0^\lambda I_{n,X}(y)dy = \frac{1}{2\pi}\left(\lambda\gamma_{n,X}(0)+2\sum_{h=1}^{n-1}\gamma_{n,X}(h)\frac{\sin(\lambda h)}{h}\right),$$

for $\lambda\in[0,\pi]$. Under general conditions, the integrated periodogram is a consistent estimator of the *spectral distribution function* given by

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

provided the density f_X is well defined.

The test statistic is defined as

$$(3.3) \qquad T(n,X,\mathcal{M}_{\mu,\sigma^2,f_\psi}) := \sup_{\lambda\in[0,\pi]}\left|\int_{-\pi}^{\lambda}\left(\frac{I_{n,X}(y)}{f_\psi}-\frac{\widehat{\sigma}^2}{\sigma^2}\right)dy\right|$$

[12]If they do not match up, we conclude that a regime shift has occurred and we re-estimate the model.

where

$$\widehat{\sigma}^2 := \int_{-\pi}^{\pi} \frac{I_{n,X}(z)}{|\psi(e^{-iz})|^2} dz \tag{3.4}$$

is an estimate of σ^2. The test statistic is a function of the data, through the periodogram $I_{n,X}(y)$, as well as of the hypothesized model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$.

Note that the test statistic (3.3) does not involve directly the integrated periodogram as defined by (3.2). Instead, Bartlett's weighted form of the integrated periodogram (Bartlett (1954); cf. Priestley (1981))

$$J_{n,X,f_\psi}(\lambda) := \int_0^{\lambda} \frac{I_{n,X}(y)}{f_\psi(y)}\, dy, \quad \lambda \in [0,\pi] \tag{3.5}$$

has been used as a building block. Bartlett's weighted form of the integrated periodogram was preferred to the empirical spectral distribution function in constructing the test statistic (3.3) for the following statistical reason.

Given a finite 4th moment for X and supposing that (X_t) is the linear process (2.2), the limit of $\sqrt{n}(J_{n,X} - F_\psi)$ in $\mathbb{C}\,[0,\pi]$, the space of continuous functions on $[0,\pi]$ endowed with the uniform topology, is an unfamiliar Gaussian process with a covariance structure that *depends on the spectral density* f_ψ; see for example Anderson (1993) or Mikosch (1998). Hence, a goodness of fit test based on the asymptotic distribution of the integrated periodogram[13] is impractical as it would require tabulating a distribution for every null hypothesis to be tested.

Dividing the periodogram by the spectral density to produce the test statistic (3.3) makes the limit process independent of the spectral density. More concretely, the process $J_{n,X,f_\psi}(\lambda)$, properly centered and scaled, converges in distribution in the Skorokhod space $\mathbb{D}([0,\pi])$ to a Brownian bridge; see Shorack and Wellner (1986) as well as Theorem 3.1 in the sequel.

Let us recall now the main result which yields the asymptotic distribution of the test statistic (3.3).

Theorem 3.1 (Klüppelberg and Mikosch (1996))**.** *Assume that $EZ = 0$, $EZ^4 < \infty$ and denote $var(Z) = \sigma^2$. Let X_t denote the linear processes (2.2) and $\widehat{\sigma}^2$ the estimate of σ^2 defined in (3.4). Then, the following holds*

$$\sqrt{n}\int_{-\pi}^{\lambda}\left(\frac{I_{n,X}(y)}{f_\psi(y)} - \frac{\widehat{\sigma}^2}{\sigma^2}\right) dy \xrightarrow{d} \pi B(\lambda/\pi),$$

in $\mathbb{D}([0,\pi])$ where the function f_ψ is defined in (2.3) and $B(\cdot)$ is a Brownian bridge.[14]

[13]The null hypothesis of such a goodness of fit test is: data X_t is generated by a linear process (2.2) with spectral density f_ψ (spectral distribution function F_ψ). The test statistic is (for example) $\sup_{\lambda\in[0,\pi]} |J_{n,X}(\lambda) - F_\psi(\lambda)|$. Not rejecting the null confirms a good fit of the hypothesized linear model with spectral density f_ψ (spectral distribution function F_ψ) to the sample. Note that the asymptotic distribution of the test statistic changes with the null hypothesis (see also Grenander and Rosenblatt (1984)).

[14]A Brownian Bridge on [0,1] is defined as $B(\lambda) := W(\lambda) - \lambda W(1)$ where W is a standard Brownian motion.

The following corollary yields the critical values for the hypothesis testing central to the methodology explained above.

Corollary 3.2. *Under the hypothesis and with the notation of Theorem 3.1 it follows that*

a. If X_t is the linear processes (2.2) then

$$\sqrt{n} \sup_{\lambda \in [0,\pi]} \left| \int_{-\pi}^{\lambda} \left(\frac{I_{n,X}(y)}{f_\psi(y)} - \frac{\widehat{\sigma}^2}{\sigma^2} \right) dy \right| \xrightarrow{d} \pi \sup_{\lambda \in [0,\pi]} |B(\lambda/\pi)| \tag{3.6}$$

b. Denote $\widetilde{\sigma}^2 := n^{-1} \sum_{t=1}^{n} Z_t^2$. If $(X_t) = (Z_t)$ the following holds

$$\sqrt{n} \int_{-\pi}^{\lambda} \left(\frac{I_{n,X}(y)}{\sigma^2/2\pi} - \frac{\widetilde{\sigma}^2}{\sigma^2} \right) dy \xrightarrow{d} \pi B(\lambda/\pi),$$

$$\sqrt{n} \sup_{\lambda \in [0,\pi]} \left| \int_{-\pi}^{\lambda} \left(\frac{I_{n,X}(y)}{\sigma^2/2\pi} - \frac{\widetilde{\sigma}^2}{\sigma^2} \right) dy \right| \xrightarrow{d} \pi \sup_{\lambda \in [0,\pi]} |B(\lambda/\pi)|, \tag{3.7}$$

in $\mathbb{D}([0,\pi])$.

The distribution of the random variable $\sup_{\lambda \in [0,\pi]} |B(\lambda)|$ is known.[15]

We would like to emphasize the generality of Theorem 3.1 which imposes only low moment restrictions[16] and no distributional restrictions on the innovations Z_t.

We turn now to our detailed data analysis.

[15]Although its distribution function $F(x) = 1 + 2\sum_{k=1}^{\infty} (-1)^k exp(-2k^2x^2)$ involves an infinite sum, the series is extremely rapidly converging. Usually a few terms suffice for very high accuracy. The limiting distribution was tabulated in Massey (1951), (1952). For example, the 90%, 95% and 99% quantiles are 1.225, 1.359 and 1.628 respectively.

[16]Although a Central Limit Theorem for the integrated periodogram (3.2) holds true provided $\sigma^2 < \infty$, we are not aware if a similar result holds for the Bartlett's weighted periodogram (3.5). As noted in Klüppelberg and Mikosch (1996), the results in Theorem 3.1 are sensitive to large fluctuations in the innovations. This could imply that a goodness of fit test based on the statistic (3.3) has power

4. Analysis of the local dependency structure

We begin our data analysis by an investigation of the local dependency structure of the logarithm of the absolute returns X_t. The aim of this section is two-fold. We want first, to measure the strength of the (local) dependency in the log-absolute returns and second, to evaluate its time evolution. Towards these goals we locally estimate AR(1), MA(1) and ARMA(1,1) processes using the methodology outlined in Section 2. Consecutively, we use the time-varying estimated parameters of these processes to shed light on the two issues of interest.

The data on which our analysis is based are the daily returns of *S&P 500* index, $r_t := \log P_t - \log P_{t-1}$, where P_t is the daily closing level of the index between January 3, 1928 and May 25, 2000. There were 390 zeros among 19261 log-returns (2% of the observations), unevenly distributed through the sample. The subsample beginning in 1928 and ending with the introduction of the current definition[17] of the S&P500 index in March 1957 contains 4% (or 319) zeros, while the subsample between 1957 to 2000 contains 0.75% (or 71) zeros. When taking the logarithm of the absolute returns to produce the sequence $X_t = log(|r_t|)$, the zeros have been removed from the sample.[18]

against an alternative of a linear model with infinite variance innovations. Since the result in Theorem 3.1 fundamentally assumes iid innovations, we expect the test to have power against alternative hypothesis that violate the iid assumption of the sequence (Z_t).

[17]In 1957, the *S&P 90* was expanded to 500 stocks and became the *S&P 500* Index. The 500 stocks contained exactly 425 industrials, 25 railroads, and 50 utility firms. This requirement was relaxed in 1988.

[18]The results of the analysis remain unchanged if the zeros are replaced with e^{-6}.

To successfully implement the approach described above we need to specify the statistical elements of the procedure which we do in the sequel.

4.1. **The test statistic for ARMA(1,1) processes.** The test statistic we use to built the homogeneity interval is (3.3). For an ARMA(1,1) process defined as

$$\phi(B)(X_t - \mu) = \theta(B)Z_t, \quad \phi(z) = 1 - \phi_1 z, \quad \theta(z) = 1 + \theta_1 z, \tag{4.1}$$

the filter ψ and the spectral density function are

$$\psi(z) = \frac{\theta(z)}{\phi(z)}, \quad f_\psi(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\psi(e^{-i\lambda})}\right|^2. \tag{4.2}$$

These quantities, related to the model, together with the data-based periodogram (3.1) are all we need to construct the test statistic according to (3.3). The method of estimation of the ARMA parameters was that of quasi-maximum likelihood.

The size of the subsamples used in the goodness of fit test was $s = 250$ (roughly a business year) days.[19] The homogeneity interval was extended with 20 observations at a time, i.e. $p = 20$ (approximately one business month).

[19]This choice strikes a balance between the need for a sample that is large enough for the asymptotics of Theorem 3.1 to work and, in the same time, short enough for the hypothesis of local stationarity to hold. The size s has been chosen empirically. Simulation studies were conducted to evaluate the asymptotic behavior of the test statistic (3.3).

4.2. **Local approximation by AR(1) and MA(1) processes.** We begin by reporting the results of fitting locally stationary AR(1) and MA(1) processes defined by three parameters (μ, σ^2, ϕ_1), $(\mu, \sigma^2, \theta_1)$ respectively, to the logarithm of absolute returns. The findings were very similar and for this reason we concentrate on AR(1) approximation.

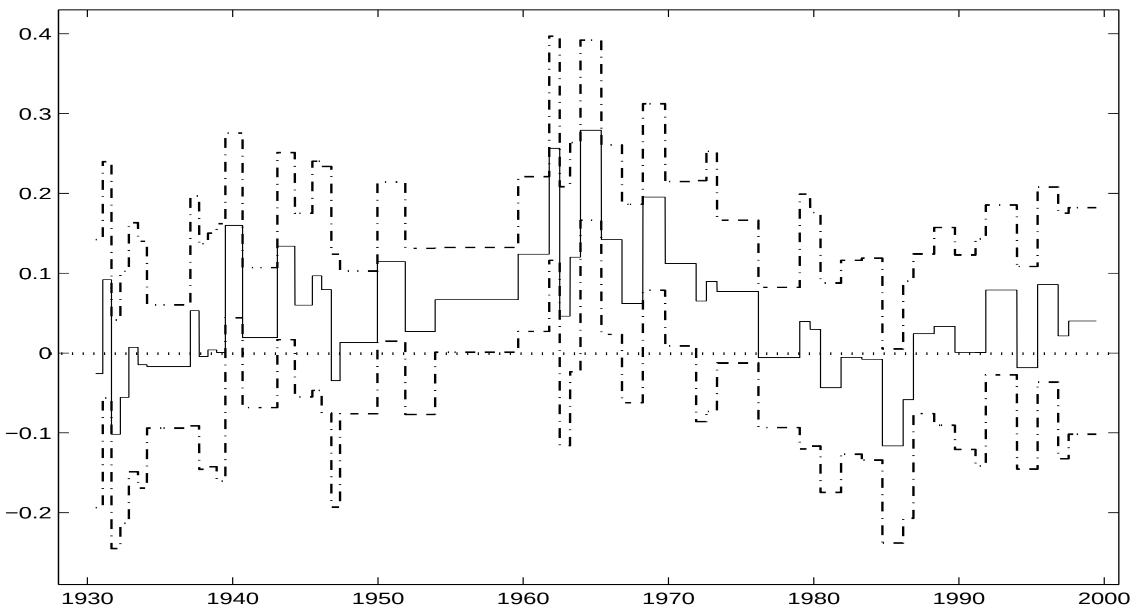

Figure 4.1. The time-varying AR coefficient, ϕ_1 (with 95% confidence intervals) estimated on the intervals of homogeneity of the log-absolute returns of the *S&P500*. The intervals of homogeneity correspond to the AR(1) local approximation and are built using the test statistic (3.3) with $\theta_1 = 0$. Zero is most of the time covered by the interval.

Figure 4.1 displays the time-varying AR(1) coefficient estimated on intervals of homogeneity built using the methodology described in Section 2. In calculating the test statistic (3.3), $\theta_1 = 0$ was used in (4.2). One notices that the AR(1) coefficients are, most of the time, not significant. The periods when the coefficient is significant are short and the value of the AR coefficient is, in absolute value, almost always smaller than 0.15.

The situation is identical when MA(1) processes are used as local approximations, i.e. when the test statistic (3.3) is built setting $\phi_1 = 0$ in (4.2) and MA(1) processes are estimated on the intervals of homogeneity so built: the MA(1) parameter is almost always smaller than 0.15 and most of the time not significantly different from zero. Moreover, the values θ_1 takes are almost identical with those taken by ϕ_1 and displayed in Figure 4.1.[20]

From the graph in Figure 4.1 we conclude that, based on the measures given by the AR(1) and MA(1) local approximations, there is almost no local linear dependency in the log-absolute returns. Besides short episodes when the linear dependency is low, the data is uncorrelated.

4.3. **Local approximation by ARMA(1,1) processes.** The over-all picture on local dependency does not change when one uses ARMA(1,1) processes as local stationary approximations of the dynamics of log-absolute returns. Figure 4.2 displays AR and MA

[20]If the AR(1) (MA(1)) coefficient is not significant, the representation $X_t = \phi_1 X_{t-1} + Z_t$ does not differ from $X_t = \phi_1 Z_{t-1} + Z_t$.

coefficients (together with confidence intervals) estimated on intervals of homogeneity defined using a test statistic (3.3) corresponding to an ARMA(1,1) process. For the sake of visual clarity, the figure only displays the decade 1970-1980.[21]

[21]The rest of the sample shows the same behavior.

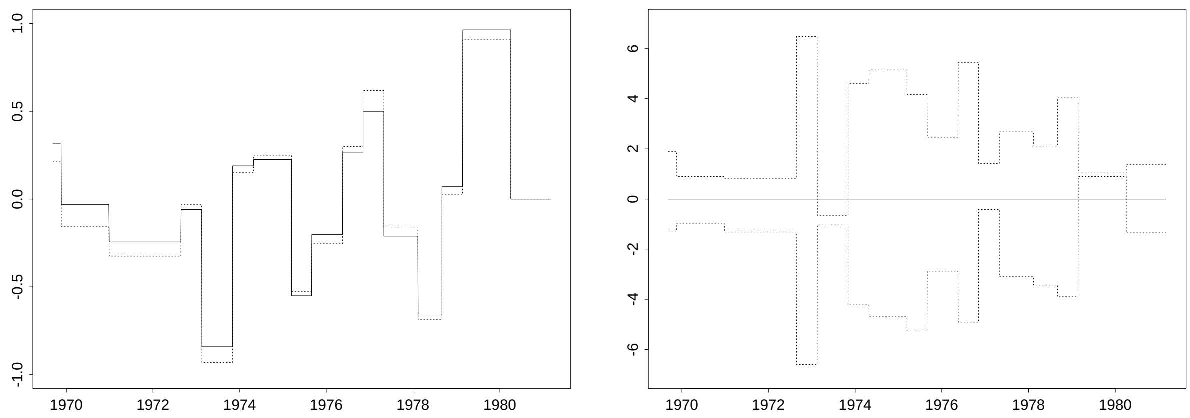

Figure 4.2. Left: The time-varying AR coefficient, ϕ_1 (*continuous line*) and the minus MA coefficient, $-\theta_1$ (*dotted line*) estimated on the intervals of homogeneity of the log-absolute returns of the *S&P500*. The intervals of homogeneity correspond to the ARMA(1,1) local approximation scheme. Right: The 95% confidence intervals for the AR coefficient ϕ_1 displayed on the left. Zero is most of the time covered by the interval.

Figure 4.2 summarizes two remarkable findings of the local dependency analysis based on approximating the true data generating process with ARMA(1,1) stationary processes. *First*, the estimated AR coefficient, ϕ_1 and the minus MA coefficient, $-\theta_1$, although taking a wide range of values, are always very close. This situation corresponds to the following particular form of equation (4.1)

$$(1-\phi_1 B)(X_t-\mu) = (1-\phi_1 B)Z_t, \quad \text{i.e.} \quad X_t-\mu = Z_t.$$

Second, the 95% confidence intervals contain 0 most of the time. This situation is typical of fitting ARMA(1,1) models to white-noise data. As when using AR(1) and MA(1) processes as local approximations, we find evidence of the absence of local linear dependency in the sequence of log-absolute returns.

As a conclusion, in a flexible modeling framework which nests time-varying linear dependency structure and time-varying mean and variance, the data chooses as most appropriate a simple model with no linear dependency but with significant changes in the mean and in the variance of the time series. *Figures 4.1 and 4.2 suggests that piecewise, on the intervals of homogeneity, the data is, approximately, a white noise.*

4.4. **The asymptotic distribution of $T(n, X,\ \mathcal{M}_{\widehat{\mu},\, \widehat{\sigma}^2,\, \widehat{f}_\psi})$.** In constructing the homogeneity intervals that lead to the results presented in Figure 4.2, we took the ARMA(1,1) linear model estimated in the initial part of an interval of homogeneity for the real data generating process of the rest of the interval. In other words, we exchanged $T(n, X,\ \mathcal{M}_{\mu,\, \sigma^2,\, f_\psi})$ for $T(n, X,\ \mathcal{M}_{\widehat{\mu},\, \widehat{\sigma}^2,\, \widehat{f}_\psi})$ while we continued to use the asymptotic distribution of the first to decide on the extension of the homogeneity interval. The issue of the relationship between the distribution of the relevant test statistic $T(n, X,\ \mathcal{M}_{\widehat{\mu},\,\widehat{\sigma}^2,\,\widehat{f}_\psi})$ and the theoretical distribution of $T(n, X,\ \mathcal{M}_{\mu,\,\sigma^2,\,f_\psi})$ has been addressed through a small simulation study. ARMA(1,1) linear models with parameters (μ, σ^2, f_ψ) similar to the ones obtained from the estimation procedure that produced the results in Figure 4.2 were simulated. For every model 1000 samples were generated. The length of the sample was 250 (the same with the sample length used in the previous analysis). For every sample, both $T(n, X,\ \mathcal{M}_{\mu,\,\sigma^2,\,f_\psi})$ and $T(n, X,\ \mathcal{M}_{\widehat{\mu},\,\widehat{\sigma}^2,\,\widehat{f}_\psi})$ were calculated ($\widehat{\mu}$, $\widehat{\sigma}^2$ and $\widehat{f_\psi}$ are the sample estimates of the true parameters μ, σ^2 and f_ψ). Figure 4.3 displays the simulation results for the models $\phi_1 = -0.8$, $\theta_1 = -0.8$ (left), $\phi_1 = 0.01$, $\theta_1 = 0.01$ (center) and $\phi_1 = 0.8$, $\theta_1 = 0.8$ (right) ($\mu = -6$, $\sigma^2 = 1.4$ in all three cases). Each graph displays the qqplot of the sample (of size 1000) of $T(n, X,\ \mathcal{M}_{\mu,\,\sigma^2,\,f_\psi})$ (on the x-axis) vs the sample (of the same length) of $T(n, X,\ \mathcal{M}_{\widehat{\mu},\,\widehat{\sigma}^2,\,\widehat{f}_\psi})$ (on the y-axis). The graphs show that, for the purpose of testing at statistically common levels of confidence (say 95%), the distribution of the two statistics, $T(n, X,\ \mathcal{M}_{\mu,\,\sigma^2,\,f_\psi})$ and $T(n, X,\ \mathcal{M}_{\widehat{\mu},\,\widehat{\sigma}^2,\,\widehat{f}_\psi})$ are practically identical.[22]

[22]Removing the last 25 more extreme pairs in any of the graphs, i.e. 2.5% of the sample, leaves us with an almost straight line indicating a very good correspondence of the quantiles (up to the 0.975-th one) of the two distributions.

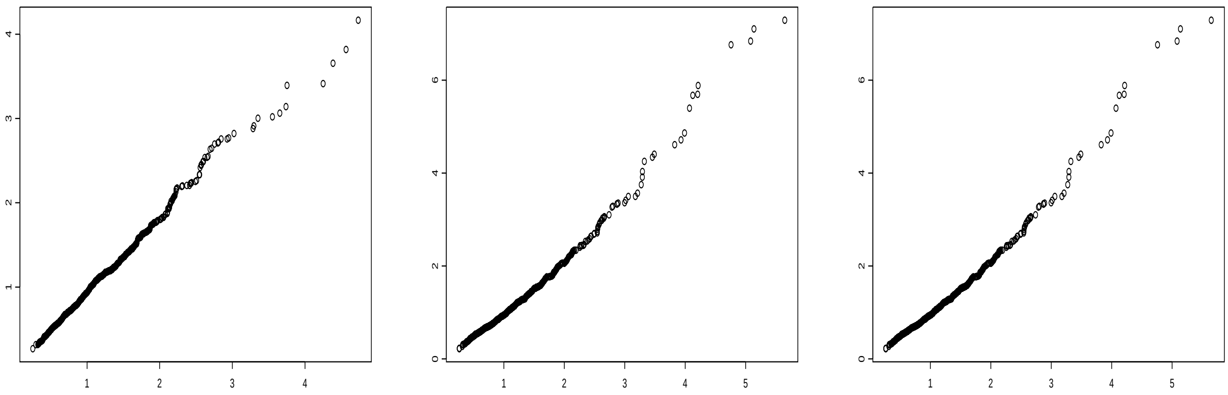

Figure 4.3. QQplot of $T(n, X,\ \mathcal{M}_{\mu,\, \sigma^2,\, f_\psi})$ (on the *x*-axis) vs $T(n, X,\ \mathcal{M}_{\widehat{\mu},\, \widehat{\sigma}^2,\, \widehat{f}_\psi})$ (on the *y*-axis) corresponding to ARMA(1,1) models $\phi_1 = -0.8$, $\theta_1 = -0.8$ (*left*), $\phi_1 = 0.01$, $\theta_1 = 0.01$ (*center*) and $\phi_1 = 0.8$, $\theta_1 = 0.8$ (*right*) ($\mu = -6$, $\sigma^2 = 1.4$ in all three cases). $\widehat{\mu}$, $\widehat{\sigma}^2$ and $\widehat{f_\psi}$ are the sample estimates of the true parameters μ, σ^2 and f_ψ. The length of the sample on which the test statistics were calculated was 250. 1000 samples of each model were generated.

5. Local approximation by iid processes

The results in the previous section suggest that a simple local approximation by processes free of second order structure, characterized simply by mean μ and variance σ^2 might be appropriate. In this section we present the results of an analysis that uses iid processes with changing unconditional mean and variance to locally describe the movements of the log-absolute returns. We estimate the time-varying first and second unconditional moments, we bring some evidence supporting the choice of the simple local approximation under discussion and briefly discuss its implications on describing the dynamics of the returns. Our analysis implies the following dynamics of the data suggested by the results in the previous section.

5.1. **A model for returns.** The results displayed in Figures 4.1, 4.2 indicate that our data, $X_t = \log(|r_t|)$ can be modeled as independent[23] and suggest the following simple model

$$X_t = \mu(t) + \sigma(t)\varepsilon_t, \tag{5.1}$$

where ε_t are iid with $E\varepsilon_t = 0$, $E\varepsilon_t^2$=1 and $\mu(t)$, the unconditional mean, and $\sigma^2(t)$, the unconditional variance, are functions of t. This model yields the following ones for the absolute returns and returns

$$|r_t| = h(t)^{1/2}\,\epsilon_t, \quad r_t = h(t)^{1/2}\,\epsilon_t\,S_t, \tag{5.2}$$

[23]Further evidence supporting the choice of locally approximating the dynamics of the logarithm of the absolute return data by *iid* sequences are presented in Section 6.

where $h(t) := e^{2\mu(t)}E(e^{2\sigma(t)\varepsilon_t})$ is the time-varying unconditional variance function, $\epsilon_t := e^{\sigma(t)\varepsilon_t}/(E(e^{2\sigma(t)\varepsilon_t}))^{1/2}$ are independent innovations with a time-dependent distribution, $E\epsilon_t^2 = 1$. The sequence (S_t) is iid, $S_t = -1, 1$ with probability 0.5.

In words, the returns are modeled as *independent random variables* with a time-varying unconditional variance[24] (if σ^2 is also changing through time, they might have other time-varying unconditional probabilistic characteristics). In particular, the returns are a non-stationary sequence of random variables. Note that if $\sigma(t)$ can be assumed close to constant, the innovation sequence (ϵ_t) can be modeled as iid and the only time-varying feature of the returns is the unconditional variance.

To conduct the local analysis based on approximations with iid processes, we need to implement the steps described in Section 2. In particular, we need to make precise the test statistic to be used.

5.2. **Test statistic for iid processes.** Our earlier findings motivate a simpler test statistic than (3.3), namely

$$\widetilde{T}(n,\, X,\ \mu,\ \sigma^2) := \sup_{\lambda\in[0,\pi]} \left| \int_{-\pi}^{\lambda} \left(\frac{I_{n,X}(y)}{\sigma^2/2\pi} - \frac{\widetilde{\sigma}^2}{\sigma^2} \right) dy \right| \tag{5.3}$$

to help construct the homogeneity intervals. The asymptotic distribution which provides the critical values needed for hypothesis testing is given in Corollary 3.2.

5.3. **Estimation of the time-varying unconditional mean and variance.** In approximating the movements of log-absolute returns by iid processes, the dynamics of the

[24]The time-varying second moment is responsible for the structure present in the sample ACF of absolute returns.

data is concentrated in changes of the unconditional first two moments. Goodness-of-fit test based on the statistic (5.3) were used as part of the methodology described in Section 2 to produce homogeneity intervals on which the unconditional mean and variance were estimated.

Figure 5.1 displays the estimated unconditional mean $\widehat{\mu}(t)$ and the estimated unconditional variance $\widehat{\sigma}^2(t)$ of the logarithm of absolute values of daily returns. The confidence intervals are those given by the Central Limit Theorem (CLT) applied to X_t and $(X_t - E(X_t))^2$, respectively.

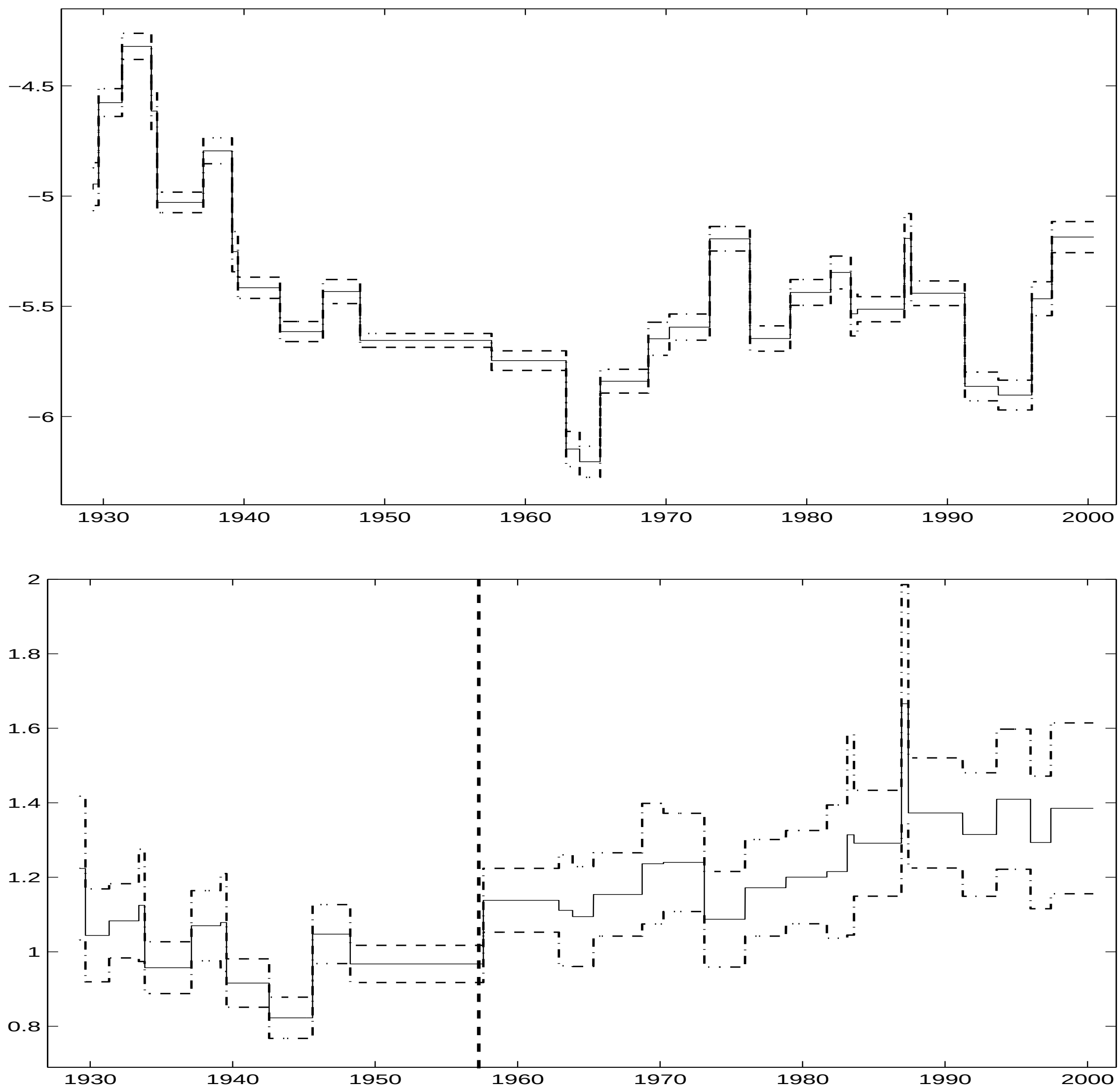

Figure 5.1. The estimated time-varying unconditional mean $\widehat{\mu}$ *(left)* and variance $\widehat{\sigma}^2$ *(right)* of the logarithm of absolute returns on *S&P 500*, X_t. The dotted lines are the 95% confidence intervals based on the CLT. The vertical line in the second graph marks the date when the definition of the *S&P 500* index changes (see Note 17).

The top graph in Figure 5.1 shows a very volatile decade between 1928-1938 (the high-mean period for log-absolute returns ended rather abruptly around the beginning of the Second World War) followed by a mild downwards trend until the middle of the '60s (the post-war economic boom) and a general upwards trend from then to the end of the sample. One can possibly see a certain connection between the higher levels of the mean of log-absolute returns (hence higher unconditional variance of returns) and the 1973 oil crisis and the economic recessions in the beginning of the '80s and '90s. The strong market movements around the 1987 stock market crash are also visible. After a period of low mean in the middle of the '90s, the end of the recent past period of economic expansion that spanned the '90s is also characterized by a higher level of the mean of log-absolute returns, i.e. higher unconditional variance of returns.

The bottom graph in Figure 5.1 shows a significantly lower variance of the log-absolute returns before the middle of 1950s followed by a slight upwards trend. The vertical line corresponds to the date when the definition of the *S&P 500* index changed.

Figure 5.2 display some evidence supporting the choice of iid sequences as local approximations of the dynamic of log-absolute returns.

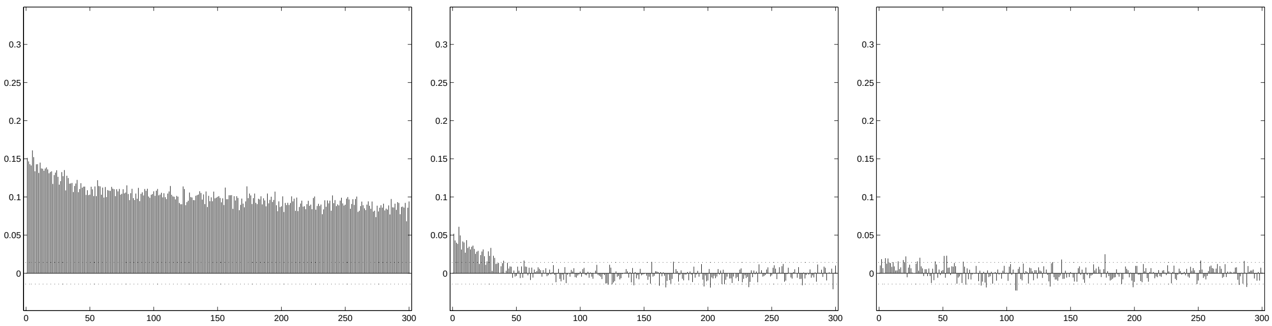

Figure 5.2. Sample ACFs of X_t, the logarithm of the absolute returns on the *S&P 500*, before *(left)* and after subtracting the estimated mean $\widehat{\mu}$ and standardizing with the estimated standard deviation $\widehat{\sigma}$ *(center)*. The sample ACF of the absolute values of centered and standardized data, $|X_t - \widehat{\mu}(t)|/\widehat{\sigma}(t)$ *(right)*.

The first two graphs in Figure 5.2 display the sample ACF for the logarithm of absolute values of daily returns before and after the data was centered by the mean $\widehat{\mu}$ and standardized by the standard deviation $\widehat{\sigma}$ estimated by our methodology. They show a strong reduction of the dependency present in the sample ACF (the residual dependency is extremely small, less than 0.05, and reduced to the first circa 30 lags).[25] The last graph of Figure 5.2 presents the sample ACF of the absolute values of centered and standardized data, $|X_t - \widehat{\mu}(t)|/\widehat{\sigma}(t)$ and shows no linear dependency. The two sample ACF corresponding to the estimated residuals, i.e. middle and right graphs in Figure 5.2 and showing almost no linear dependence suggest that *independent* sequences indeed provide good local approximation to the dynamics of the data.

5.4. **The** $S\&P500$ **volatility from 1928 to 2000.** The intervals of homogeneity for the logarithms of absolute returns translate into intervals of homogeneity for the absolute returns.[26]

[25]A better approximation (than our rough step function approximation) of the changing mean removes completely the linear dependence still present in the sample ACF of log-absolute returns. For an analysis conducted on returns, see Drees and Stărică (2002) for the univariate case and Herzel et al. (2002) for the multivariate case.

[26]The methodology described in Section 2 is applicable as long as the data has a forth finite moment; see Theorem 3.1. This condition is barely satisfied by the absolute returns which have a negative tail index close to four. Although strictly speaking theoretically feasible, our methodology does not produce meaningful results when applied directly on absolute returns.

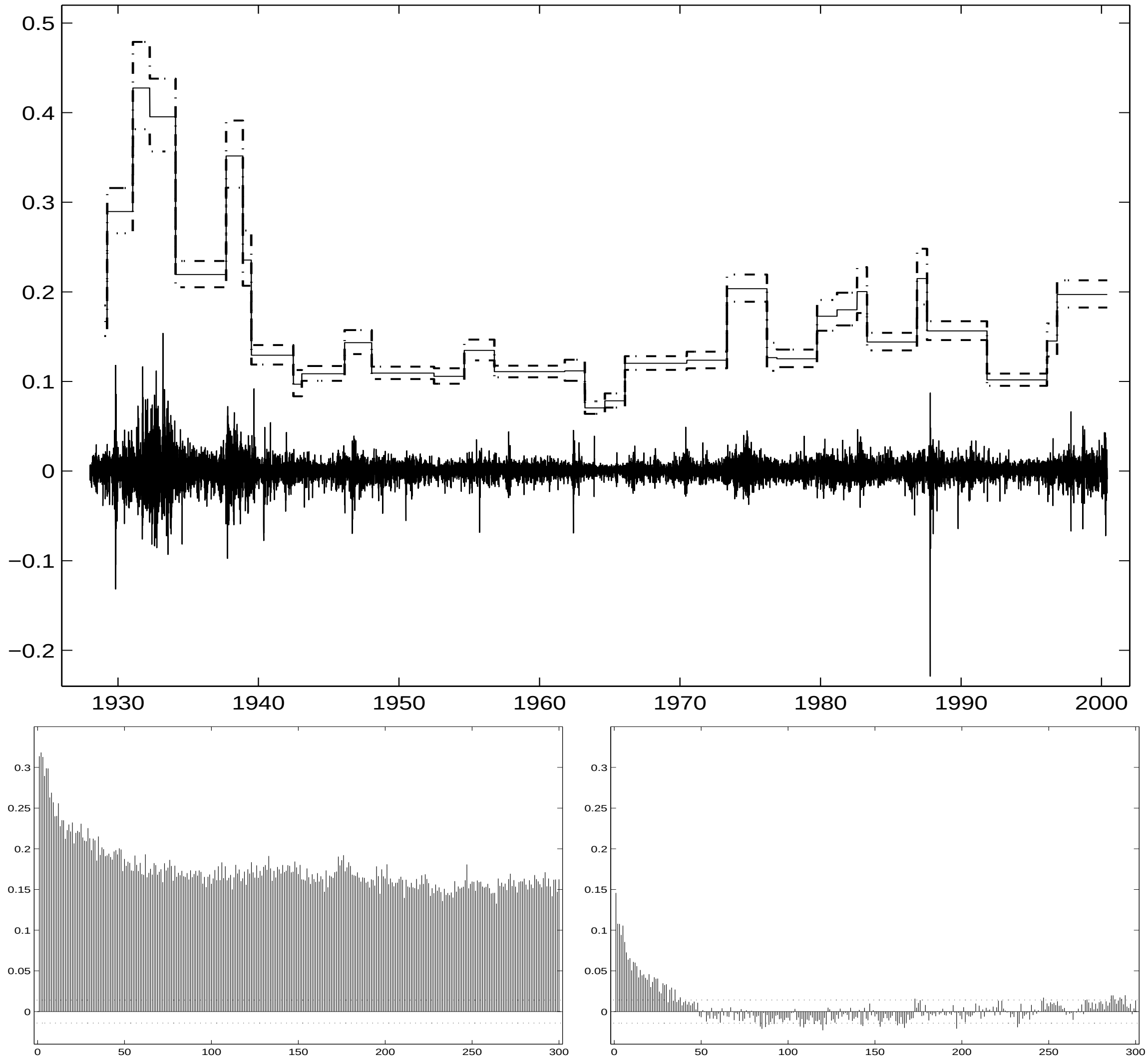

Figure 5.3. Top: Estimated time-varying unconditional standard deviation (annualized) with 95% confidence intervals together with the returns on the *S&P 500*. Bottom: Sample ACFs of the absolute values of returns on the *S&P 500* before *(left)* and after *(right)* scaling with the standard deviation in Top.

Figure 5.3 (Top) displays the unconditional, time-varying annualized standard deviation the returns of the *S&P 500* together with the returns themselves. Figure 5.3 (Bottom) displays the sample ACF for the absolute values of daily returns before and after the data was scaled by the standard deviation in the graph on top. The two sample ACF graphs show a strong reduction of the dependency. The remaining dependency is small, less than 0.15, and it is reduced to the first circa 50 lags.

5.5. **Intervals of homogeneity based on the Central Limit Theorem.** We end this section with a discussion on the performance of our procedure when replacing the goodness of fit test based on the integrated periodogram with a very simple goodness of fit based on the Central Limit Theorem. The null hypothesis is that locally the data are independent with the mean μ and variance σ^2. The test statistic is simply

$$\overline{T}(n, X, \mu, \sigma) := \frac{\bar{X} - \mu}{\sigma}. \tag{5.4}$$

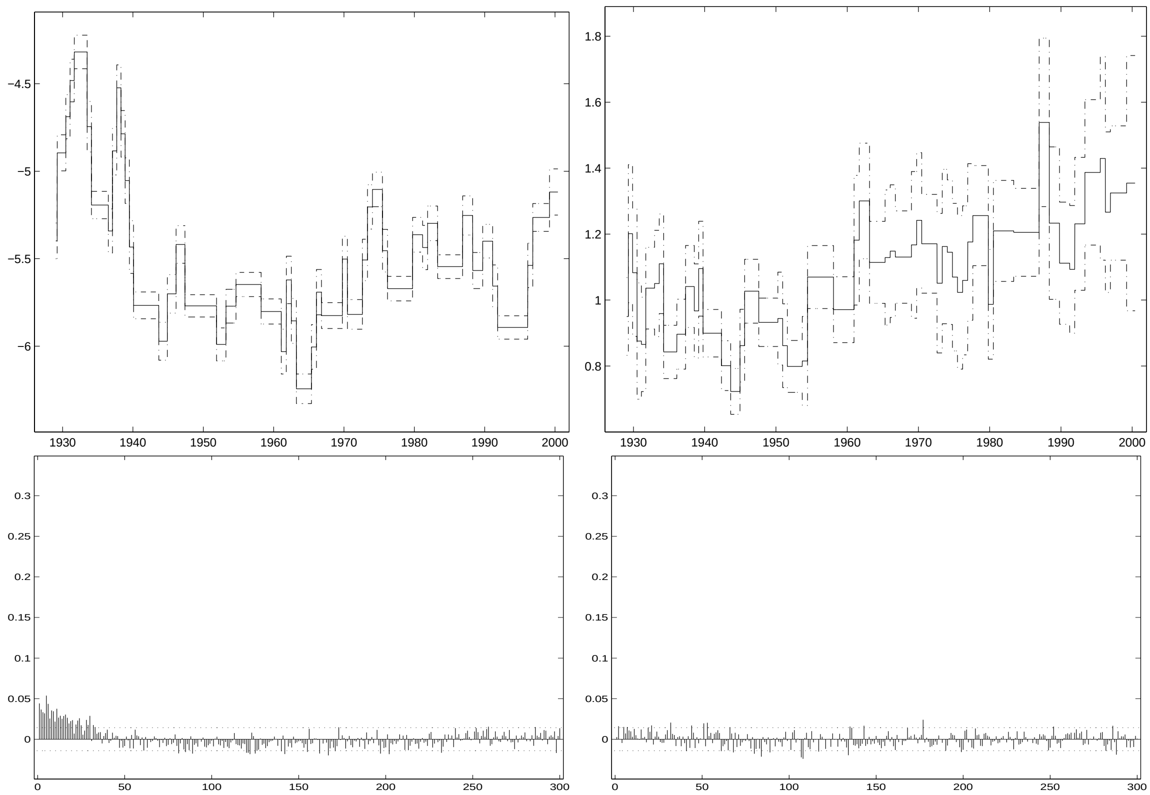

Figure 5.4. Top: The estimated time-varying unconditional mean $\widehat{\mu}(t)$ *(left)* and variance $\sigma^2(t)$ *(right)* of the logarithm of the absolute returns on the *S&P 500*. The intervals of homogeneity were constructed using the test statistic $\overline{T}(n, X, \mu, \sigma)$ defined in (5.4). The dotted lines are the 95% confidence intervals based on the CLT. Bottom: Sample ACFs of the estimated residuals $(X_t - \widehat{\mu}(t))/\widehat{\sigma}(t)$ *(left)* and their absolute values *(right)*.

The methodology outlined in Section 2 and the test statistic (5.3) were used to produce the homogeneity intervals on which the two first unconditional moments were estimated. The estimation results are displayed in Figure 5.4.

While the tests based on the statistic $\overline{T}(n, X, \mu, \sigma)$ found more changes than the integrated periodogram, the overall pattern of change is the same. The overall amount of dependency in the absolute returns explained by the shifts in the first two unconditional moments (as measured by the residual correlation in sample ACF of the standardized time series) is practically the same.

As a conclusion, the integrated periodogram approach offers a simpler overall picture of the pattern of changes in the long time series and, more importantly, serves to motivate the assumption of locally independent log-absolute returns on which the use of the test statistic $\overline{T}(n, X, \mu, \sigma)$ is based.

6. A simple model for the period 1957-2000.

In this section a simple model for the log-absolute returns and returns covering the period between 1957 and 2000 is discussed. A great deal of attention is devoted to checking the goodness of fit of the model to the data. The results confirm our choice of modeling the return data as locally *iid* sequences.

6.1. **The model for returns revisited.** The bottom graph in Figure 5.1 shows a change in the estimated variance of the log-absolute returns occurring in the middle of the '50s. The estimated change coincide in date with the change in the definition of the index. Before mid'50s the variance σ^2 of the log-absolute returns was lower. Note also that after 1960 the value of σ^2 stayed roughly constant around the value of 1.2 until the end of the '80s and around 1.4 since then. Moreover, the confidence intervals indicate that this increase might be not significant. This motivates the assumption of constant σ for the period after 1957. We will see in the sequel that this assumption provides a good approximation to the data dynamic.

Hence, for the period between 1957 and the end of the sample the model for log-absolute returns (5.1) could be further simplified to

$$X_t = \mu(t) + \sigma\varepsilon_t, \tag{6.1}$$

where ε_t are iid with $E\varepsilon_t = 0$ and $E\varepsilon_t^2$=1.

This yields the following model for the absolute returns and returns

$$|r_t| = h(t)^{1/2}\,\widetilde{\varepsilon}_t, \quad r_t = h(t)^{1/2}\,\widetilde{\varepsilon}_t S_t, \tag{6.2}$$

with $h(t) := e^{2\mu(t)}E(e^{2\sigma\,\varepsilon_t})$, $\widetilde{\varepsilon}_t := e^{\sigma\,\varepsilon_t}/(E(e^{2\sigma\,\varepsilon_t}))^{1/2}$, $E\widetilde{\varepsilon}_t^2 = 1$, $S_t = -1, 1$ with probability 0.5, $(\widetilde{\varepsilon}_t)$ and (S_t) iid sequences.

In words, the returns are modeled as *independent random variables* with a time-varying unconditional variance. They form a non-stationary sequence, free of any dependency[27] but with a marginal distribution that evolves through time. Moreover, the only changing probabilistic feature of the marginal distribution is the unconditional variance.

[27]Independent non-stationary sequences can display significant sample ACF. In particular, the long memory in volatility effect can occur for independent sequences with a time-varying unconditional variance. For more details on this issue, see Mikosch and Stărică (2004).

The reduction in the complexity of the model implies also a simplification of the statistical procedure of constructing the homogeneity intervals.

6.2. **The test statistic revisited.** In the test statistic (5.3), the ratio $\widetilde{\sigma}^2/\sigma^2$ is replaced by one. In this way, only the changes in the mean can produce extreme values of the test statistic and hence cause an homogeneity interval to end. The result of estimating the function $\mu(t)$ is practically identical to the one in Figure 5.1 and hence we do not re-display it.

In the sequel we evaluate carefully how well the model (6.1) fits the log-absolute returns between 1957 and 2000.[28] The assessment of the goodness of fit will also yield a validation of the choice of modeling the return data as locally *iid*.

6.3. **Goodness-of-fit analysis.** The goodness-of-fit analysis is based on the sequence of *estimated residuals*

$$\widehat{\varepsilon}_t := X_t - \widehat{\mu}(t), \tag{6.3}$$

where the time-varying first unconditional moment $\mu(t)$ is estimated on the homogeneity intervals constructed using the test statistic described above.

We begin with an evaluation of the assumption that ε_t are independent. Towards this end we asses the dependence in the sequences $\widehat{\varepsilon}_t$ and $|\widehat{\varepsilon}_t|$. The linear dependence is measured via the sample ACF, while the non-linear dependence is evaluated by means of copulas, a notion to be described shortly.

[28]Moving the beginning of the subsample to 1960 as the estimated variance in Figure 5.4 could suggest does not change the nature of the results in any way. Hence we preferred the longer subsample.

6.3.1. *Independent estimated residuals.* The top-left graph in Figure 6.1 displays a plot of the estimated residuals. The visual inspection shows no signs of dependence. The apparent pattern in the lower part of the graph is due to the discrete nature of the prices (the minimum price increment was big in the beginning of the period compared to its end). The top-right graph displays the histogram of the estimated returns showing a distribution skewed to the right. The left-bottom graph is the sample ACF of the logarithm of absolute returns between 1957 and 2000. It displays the so-called long-memory effect in volatility. The last two graphs are the sample ACFs of the sequence (6.3) and their absolute values.

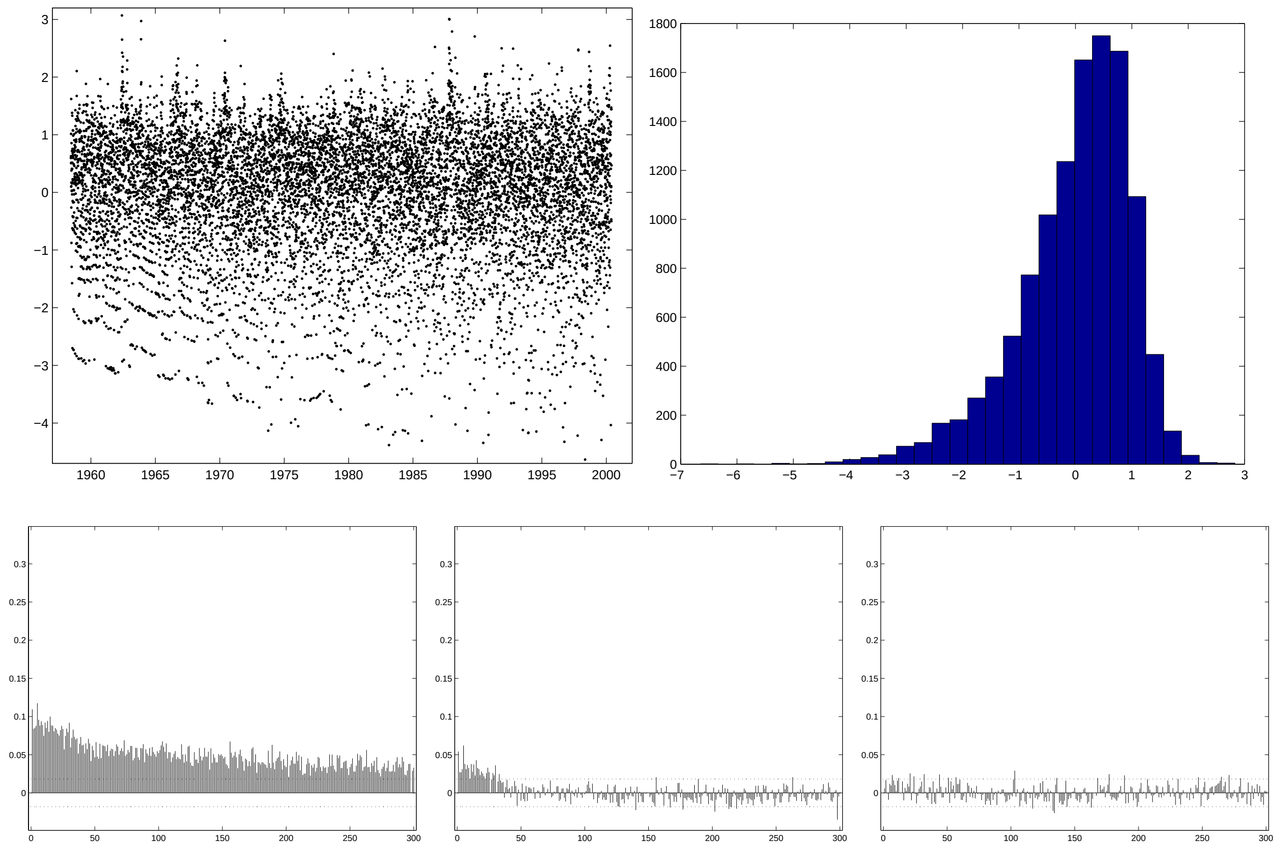

Figure 6.1. Top: Plot *(left)* and histogram (*right*) of the estimated residuals $\widehat{\varepsilon}_t$ based on the model (6.1) fitted to the period 1957-2000. Bottom: Sample ACFs of X_t, the logarithm of the absolute returns of the *S&P* 500, period 1957-2000 *(left)*. Sample ACF for the estimated residuals $\widehat{\varepsilon}_t$ *(center)*. Due to the rough approximation of the mean of the time series by a step function, the first circa 25 lags are slightly significant. Sample ACF of the absolute values of residuals $|\widehat{\varepsilon}_t|$ *(right)*. The graph shows no properties in the variance of the residuals.

Both sample ACFs of the estimated residuals $\widehat{\varepsilon}_t$ as well as their absolute values $|\widehat{\varepsilon}_t|$ shown on the bottom line of Figure 6.1 (center and right), are close to being statistically not significant indicating that almost no linear dependence remains in the time series of residuals and in their absolute values.

To search for possible patterns of non-linear dependence, it is most useful to have a look at copulas of estimated residuals paired with lagged estimated residuals $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+i})$ and the absolute values of the pairs $(|\widehat{\varepsilon}_t|, |\widehat{\varepsilon}_{t+i}|)$, $i = 1, 2, \ldots$ Before showing the results of this assessment let us say a few words about the notion of copula (for more details see Nelsen (1999)).

The joint distribution of a pair of random variables (U, V) is uniquely determined by the marginal distribution of the coordinates F_U and F_V and by their copula, i.e. the distribution on the unit square of $(F_U(U), F_V(V))$. Hence, it is the copula that provides the complete description of the dependency structure between the marginal random variables. Moreover, U and V are independent if and only if their copula is the uniform copula. Graphically, this corresponds to an uniform filling of the unit square by the pairs $(F_U(U), F_V(V))$. Hence, a simple but very informative way of assessing the independence of the coordinates of a bivariate random vector is looking at realizations of its copula. The appearance of an uniformly covered unit square supports the assumption of independence while the presence of patterns indicates dependency.

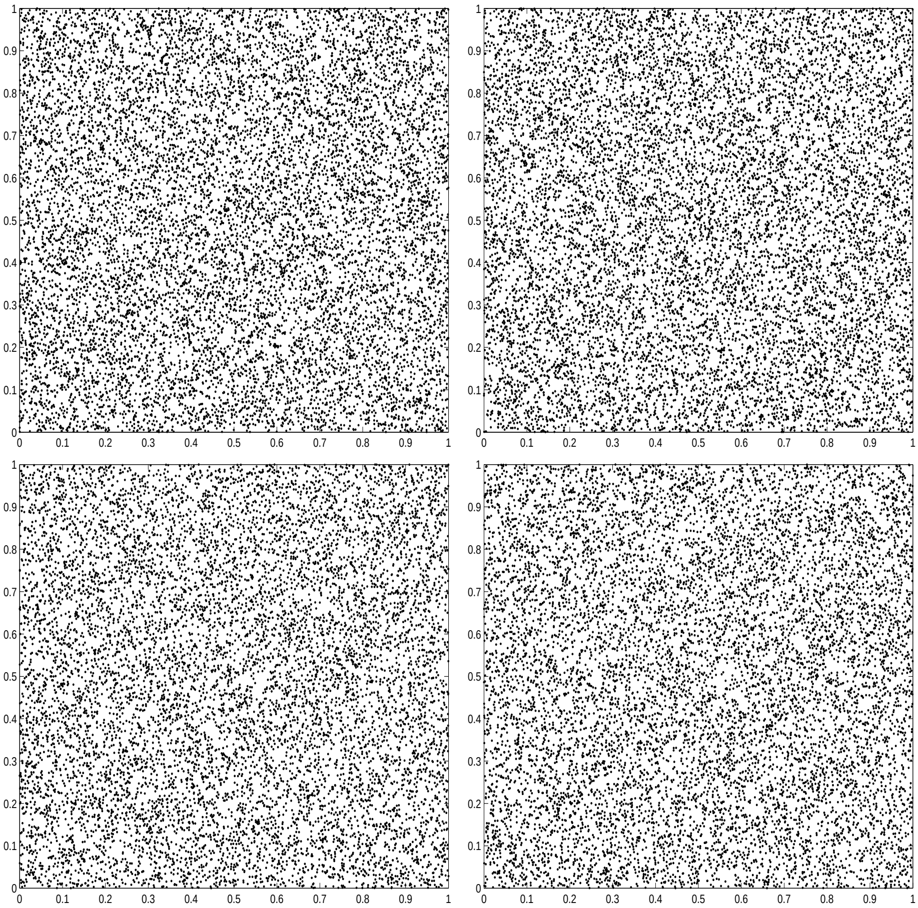

Figure 6.2. Scatter plots $(Y_t,\, Y_{t+1})$ (*Top*) and $(Y_t,\, Y_{t+2})$ (*Bottom*) for $Y_t = \widehat{F}_{\widehat{\varepsilon}}(\widehat{\varepsilon}_t)$ *(left)* and $Y_t = \widehat{F}_{|\widehat{\varepsilon}|}(|\widehat{\varepsilon}_t|)$ *(right)*. The empirical distribution function estimated on the sample $\widehat{F}_{\widehat{\varepsilon}}$ $(\widehat{F}_{|\widehat{\varepsilon}|})$ was used to transform the marginal distribution of the data to uniform. An uniform filling of the unit square is interpreted as evidence of independent components.

To obtain the copula associated with the mentioned bivariate random vectors, we transformed first the residuals and their absolute values into uniform random variables using the empirical distribution functions $\widehat{F}_{\widehat{\varepsilon}}$ and $\widehat{F}_{|\widehat{\varepsilon}|}$. Then we produced the scatter plots $(Y_t,\, Y_{t+i})$ for $Y_t = \widehat{F}_{\widehat{\varepsilon}}(\widehat{\varepsilon}_t)$ (Figure 6.2, left) and $Y_t = \widehat{F}_{|\widehat{\varepsilon}|}(|\widehat{\varepsilon}_t|)$ (Figure 6.2, right) and $i = 1$ (Figure 6.2, top) and $i = 2$ (Figure 6.2, bottom). As mentioned, an uniform filling of the unit square is interpreted as evidence of *independent* components. The graphs in Figure 6.2 reveal a uniform covering of the unit square for transformed residuals paired with their first and second lagged values. The same behavior is apparent for higher lags i. This finding together with the previous evidence on the linear dependence in the time series of estimated residuals confirm that the assumption of *independent innovations* provides a reasonable approximation for the dynamics of the log-absolute data under scrutiny.

6.3.2. *Identically distributed residuals.* To check the relevance of the assumption that the innovations are identically distributed, we divided the sample of residuals in three subsamples of equal length, the first corresponding roughly to the period 1958-1972, the second to 1972-1986 and the third to 1986-2000. Due to the discrete nature of the observations, only the residuals bigger than (-2) were included in the subsamples. Figure 6.3 displays the qqplots corresponding to the three pairs of subsamples. Note that we are comparing, on the three subsamples, the conditional distribution of the residual sequence given that a value bigger than (-2) was taken. The three graphs support the hypothesis of identically distributed innovations of the model (6.1).

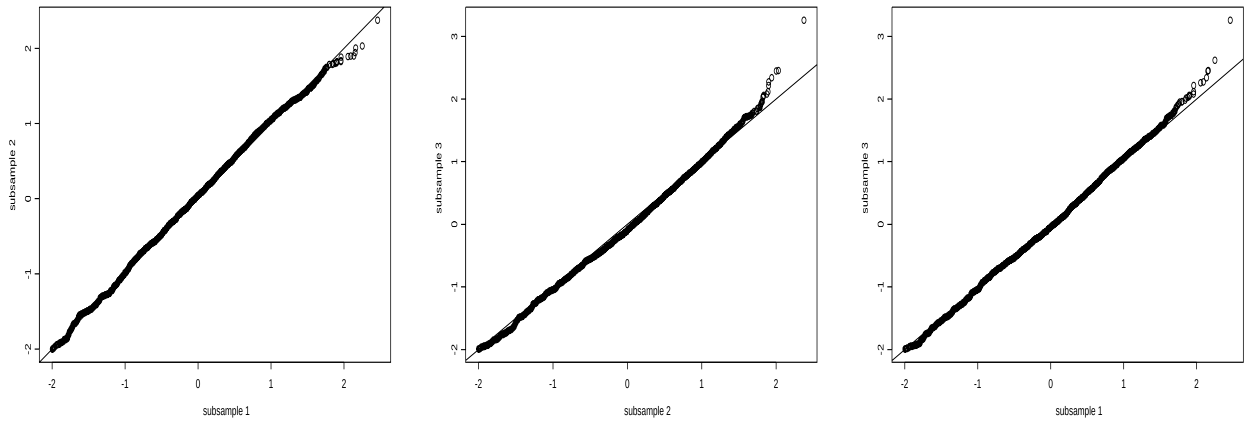

Figure 6.3. Pair-wise qqplots of three subsamples of equal length roughly corresponding to the period 1958-1972, the first, to 1972-1986, the second and to 1986-2000, the third. Due to the discrete nature of the data, only the returns bigger than (-2) were considered.

To summarize, the simple model (6.1) that assumes the returns to be independent with a time-varying unconditional variance describes well the data from the period 1957-2000. Due to the rather coarse estimation method which approximates smooth functions with piecewise constant ones, the residuals present slight traces of linear dependence. The absolute values of the residuals seem independent. Practically speaking, all dynamics of the log-absolute return (return) time series seem to be concentrated in shifts of the unconditional mean (variance).

7. Forecasting comparison: Non-stationary vs. Stationary and Long memory

In the stationary framework, a sample ACF behavior as shown in the left graphs in Figure 5.2 and Figure 5.3 will be interpreted as evidence of long memory. Hence we are facing a modeling choice for $X_t = \log(|r_t|)$. The choice is between a stationary long memory model and a non-stationary model with the dynamics mainly concentrated in changes of the mean. One possible way of solving this dilemma is to compare the forecasting behavior of two paradigms on the data at hand. Since our approach is to describe the volatility directly by analyzing the sequence of absolute returns, a natural choice for a long memory stationary model is the fractionally ARIMA class introduced by Granger and Joyeux (1981) and Hosking (1980).

The process (X_t) is said to be a $FARIMA(p,d,q)$ with $d \in (0, 0.5)$ if (X_t) is stationary and satisfies the difference equation

$$\phi(B) \bigtriangledown^d X_t = \theta(B) Z_t, \tag{7.4}$$

where (Z_t) is white noise and ϕ, θ are polynomials of degree p, q respectively. The operator $\bigtriangledown^d$ is defined by

$$\bigtriangledown^d := (1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j, \tag{7.5}$$

where

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(1-d)} = \Pi_{0<k\leq j} \frac{k-1-d}{k}, \quad j = 1, 2, \ldots.$$

The data used in the out-of-sample comparison are the logarithm of the absolute values of daily returns in the interval 1961-2000 (the data from 1957 to 1960 were used for initial estimation of the models). A FARIMA(1,d,1) model (LM) was estimated on the first 1000 observations (corresponding roughly to the period 1957-1960) and re-estimated every month (i.e. every 20 observations) using all the past observations.

7.1. **Comparison of forecasts of daily log-absolute returns.** With the estimated long memory model, predictions (f^{LM}) for the future values of $X_t = \log(|r_t|)$ were made every month (i.e. every 20 observations). The maximal forecasting horizon p was 200 days ahead. The other model used which will be referred as the shifts-in-the-mean model (SM) is described by equation (6.1). The observations anterior to the date when a forecast was made were used for determining the (then) current interval of homogeneity. The forecasts for the future values of X_t, (f^{SM}), were simply the estimated mean on this homogeneity interval and hence do not change with the horizon.

One way of comparing the two forecasts would be by assessing the orthogonality of one forecast error (at horizon p) to the other forecast. Concretely one can test something less general, i.e whether one forecast error is uncorrelated with the other forecast. This would be accomplished by means of a regression. For example, to test if the SM forecasts and the LM forecast errors are uncorrelated, one would test whether $\alpha = 0$, $\beta_1 = 0$ in the regression

$$X_{t+p} - f^{LM}_{t+p} = \alpha + \beta_1 f^{SM}_{t+p} + \varepsilon_t. \tag{7.6}$$

However, given the possibly non-stationary nature of the time series, the assumption of ergodic stationarity of the regressors and dependent variables needed for the well-functioning of the GMM machinery, is likely to be violated (f^{SM}_{t+p} is close to a piecewise constant function).

To address this possible problem we reformulate our test. Testing whether $\alpha = 0$, $\beta_1 = 0$ in (7.6) is equivalent to testing whether

$$\alpha = 0, \quad \beta_1 = 0, \quad \beta_2 = 1 \tag{7.7}$$

in the following regression

$$X_{t+p} - X_t = \alpha + \beta_1 (f^{SM}_{t+p} - X_t) + \beta_2 (f^{LM}_{t+p} - X_t) + \varepsilon_t, \tag{7.8}$$

Notice that testing for

$$\alpha = 0, \quad \beta_1 = 1, \quad \beta_2 = 0 \tag{7.9}$$

in the same regression would be equivalent to verifying that the SM forecast error $X_{t+p} - f_{t+p}^{SM}$ is uncorrelated with the FARIMA forecast f_{t+p}^{LM}. For this regression the violations of the assumption of ergodic stationarity of the regressors and dependent variables are likely to be less severe. Indeed, the vector $(X_{t+p} - X_t, f_{t+p}^{SM} - X_t, f_{t+p}^{LM} - X_t)$ is stationary and ergodic on every interval on which the volatility process $\sigma(t)$ is constant, i.e. on any interval of homogeneity. Because under the null hypothesis (7.9) the error term ε_t in regression (7.8) equals the forecast error $X_{t+p} - f_{t+p}^{SM}$ orthogonal to anything known at date t including $f_t^{(i)} - X_t$, $i = 1, 2$, the regressors are guaranteed to be orthogonal to the error term (a similar statement holds under the null hypothesis (7.7)).

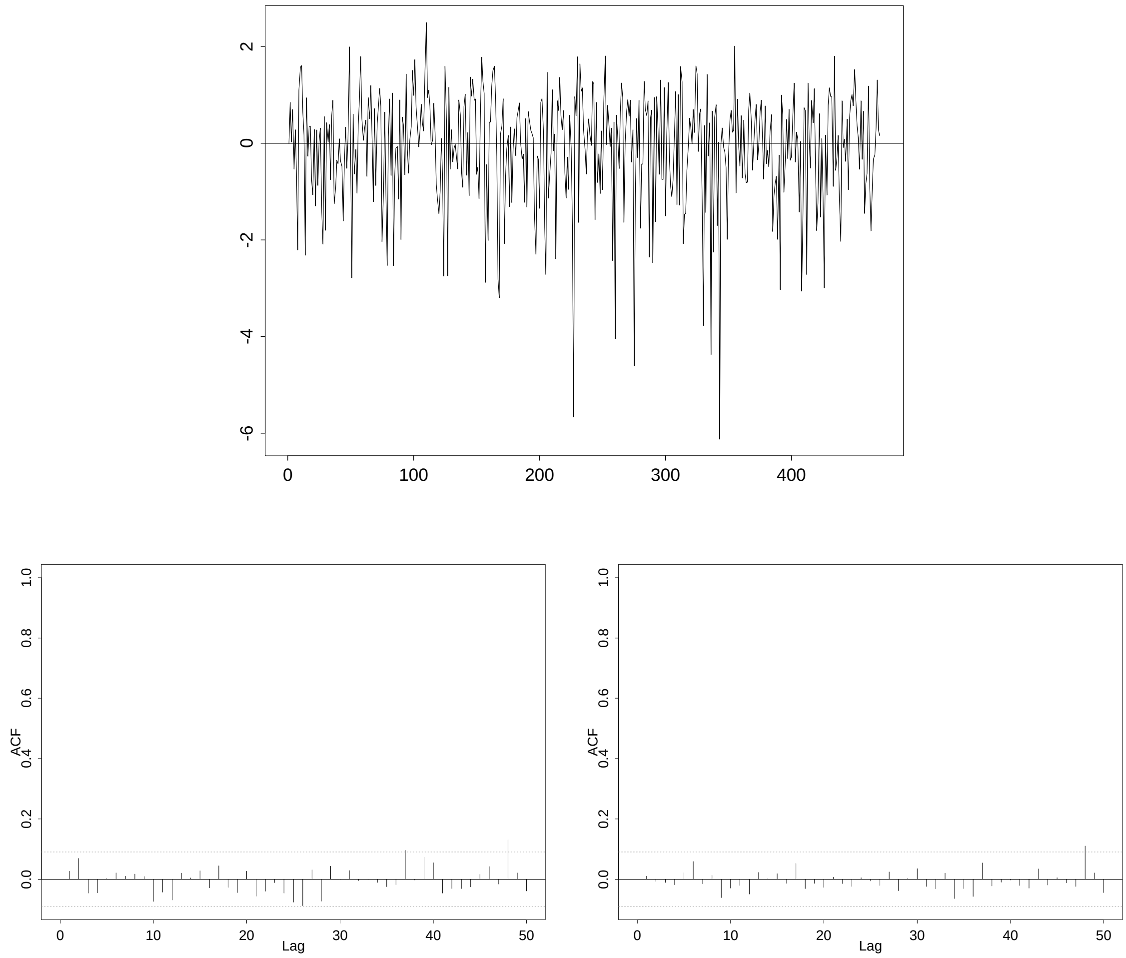

Figure 7.1. Top: The forecast errors $X_{t+p} - f_{t+p}^{SM}$ based on the model (6.1) corresponding to the period 1961-present (the period 1957-1960 is used for the preliminary estimation of the LM model. Bottom: *(left)* Sample ACF for the forecast errors $X_{t+p} - f_{t+p}^{SM}$. *(right)* Sample ACF of absolute forecast errors $|X_{t+p} - f_{t+p}^{SM}|$. The graphs suggest that the forecast errors are uncorrelated and homoscedastic.

Even more, Figure 7.1 supports the hypothesis of uncorrelated forecasting errors (the forecast errors $X_{t+p} - f_{t+p}^{LM}$ have a similar behavior), and hence it appears that an OLS estimate would suffice. Note that the regression (7.8) is closely related to the so-called *forecast encompassing equation*

$$X_{t+p} = \alpha + \beta_1 f_{t+p}^{SM} + \beta_2 f_{t+p}^{LM} + \varepsilon_t \tag{7.10}$$

which cannot be employed due to the possibly non-stationary nature both of the regressors and dependent variables.

Horizon p (days)	P-value of the Wald statistic for $H_0: \alpha=0, \beta_1=0, \beta_2=1$, i.e. $f_{t,p}^{SM} \perp e_{t,p}^{LM}$	$H_0: \alpha=0, \beta_1=1, \beta_2=0$, i.e. $f_{t,p}^{LM} \perp e_{t,p}^{SM}$	Horizon p (days)	P-value of the Wald statistic for $H_0: \alpha=0, \beta_1=0, \beta_2=1$, i.e. $f_{t,p}^{SM} \perp e_{t,p}^{LM}$	$H_0: \alpha=0, \beta_1=1, \beta_2=0$, i.e. $f_{t,p}^{LM} \perp e_{t,p}^{SM}$
10	0.00	0.21	110	0.03	0.22
20	0.03	0.27	120	0.06	0.27
30	0.00	0.31	130	0.00	0.06
40	0.05	0.67	140	0.06	0.40
50	0.01	0.49	150	0.02	0.27
60	0.09	0.23	160	0.03	0.08
70	0.01	0.32	170	0.00	0.48
80	0.09	0.39	180	0.06	0.16
90	0.01	0.44	190	0.00	0.59
100	0.01	0.12	200	0.06	0.23

TABLE 1. Comparison of forecasting performance between the LM model and SM model. The LM process is re-estimated every 20 days using all past observations. The Wald statistic of the F-ratio test is calculated under the two alternatives and the p-values are reported. A small p-value is a signal of the failure of the null. Overall the table shows a better performance of the SM model in forecasting. $e_{t,p}^{LM} := X_{t+p} - f_{t+p}^{LM}$ (the forecast error of the long memory model at horizon p) and $e_{t,p}^{SM} := X_{t+p} - f_{t+p}^{SM}$ (the forecast error of the SM model at horizon p).

The p-values of the F-test Wald statistic corresponding to $H_0 : \alpha = 0$, $\beta_1 = 0$ and $\beta_2 = 1$ and $H_0 : \alpha = 0$, $\beta_1 = 1$ and $\beta_2 = 0$, respectively are reported in Table 1. For most of the forecast horizons the hypothesis of orthogonality of the SM forecast on the LM forecast errors is rejected while the hypothesis of orthogonality of the LM forecast on the SM forecast errors remains unchallenged.

A possible critique of the previous analysis could be that we have used daily data as a measure against which to compare the performance of the two methodologies. As Figure 6.1 shows, the daily data contain a large amount of idiosyncratic noise added to the signal of interest for us, the mean. In other words, daily data might not provide such a good check as they are a poor measure of the mean.

7.2. **Comparison of forecasts of the log absolute returns over various time intervals.** To answer this possible critique and in order to get a more complete picture, aggregated data at various horizons p was also used as an alternative control measure. Define

$$X_{t,p} = \sum_{i=1}^{p} X_{t+i}, \quad \overline{f}_{t,p}^{\,SM} = \sum_{i=1}^{p} f_{t+i}^{SM}, \quad \overline{f}_{t,p}^{\,LM} = \sum_{i=1}^{p} f_{t+i}^{LM}. \tag{7.11}$$

Note that $\overline{f}_{t,p}$ are forecasts of the mean of $X_{t,p}$. For $p > 1$, through averaging some of the idiosyncratic noise in the daily data is canceled yielding a better measure against which to check the quality of the two forecasts.

We calculated and compared the MSE of the two methods defined as

$$MSE^{*}(p) := \sum_{t=1}^{n} (X_{t,p} - \overline{f}_{t,p}^{\,*})^2 \tag{7.12}$$

with $*$ standing for SM or LM.

Horizon	$MSE^{SM}(p)/MSE^{LM}(p)$		Horizon	$MSE^{SM}(p)/MSE^{LM}(p)$	
p (days)	S&P 500	Nasdaq	p (days)	S&P 500	Nasdaq
10	0.99	1.02	110	0.91	0.56
20	0.99	0.88	120	0.91	0.54
30	0.98	0.82	130	0.91	0.53
40	0.96	0.75	140	0.91	0.52
50	0.95	0.70	150	0.91	0.51
60	0.93	0.65	160	0.91	0.51
70	0.93	0.63	170	0.91	0.51
80	0.92	0.61	180	0.91	0.51
90	0.91	0.59	190	0.91	0.51
100	0.91	0.57	200	0.91	0.51

TABLE 2. Comparison of forecasting performance between the LM model and SM model. The LM process is re-estimated every 20 days using all the previous observations in the sample. The ratio $MSE^{SM}(p)/MSE^{LM}(p)$ is reported. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the LM model. The figure shows an over-all better longer-horizon volatility forecast performance of the SM model.

The results are presented in Table 2 where the ratio[29]

$$MSE^{SM}(p)/MSE^{LM}(p)$$

is given for $p = 10, 20, \ldots, 200$. Besides the *S&P 500* data, the MSE analysis has been performed also on the returns of the Nasdaq index between October 12, 1984 and November 8, 2002. The results for both series seem to support the conclusion that the SM model over-performs the LM model in forecasting.

8. Forecasting comparison: Unconditional and Non-stationary vs. Conditional and Stationary

As the paper addresses the issue of volatility and proposes a novel modeling paradigm, a comparison with the Garch framework, the current market leader in volatility modeling, is inevitable. In this section, we present the results of a forecasting comparison between our model (6.2) and a Garch(1,1) model.[30]

[29]Although the ratios in Table 2 are sensibly different from 1 (especially in the case of Nasdaq index), their statistical significance is difficult to assess. Tests of statistical significance for the differences of MSE's have been developed in the literature (see Diebold and Mariano (1995), West (1996), Harvey et al. (1997)among others). The null hypothesis is that $Ed_{t,p} = 0$, where $d_{t,p} := SFE^{LM}_{t,p} - SFE^{SM}_{t,p}$ is the loss-differential series, $SFE^{LM}_{t,p} := (X_{t,p} - \overline{f}^{LM}_{t,p})^2$ and $SFE^{SM}_{t,p} := (X_{t,p} - \overline{f}^{SM}_{t,p})^2$. However, to the best of our knowledge, all tests assume that the loss-differential series are *stationarity and short memory.* Neither of these assumptions hold in the case at hand. Recall that we are trying to distinguish between a null of long memory and an alternative of non-stationarity (shifts in the mean).

[30]Both a Student-t Garch, estimated with an exact maximum likelihood approach, as well as a general Garch(1,1) model with a non-specified conditional distribution, estimated with a quasi-maximum likelihood

In the stationary, conditional framework, the working assumption is that *a stationary Garch(1,1) process*

$$r_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \, r_{t-1}^2 + \beta_1 \, \sigma_{t-1}^2, \tag{8.13}$$

where the innovations (ε_t) are iid, mean zero, variance one, (Student-t distributed in the case of the Student-t Garch) is a good approximation of the data generating process. In the non-stationary, unconditional set-up, we assume that *the model (6.2) provides a good description of the daily returns.*

We begin with an evaluation of the relevance of the assumption of a stationary Student-t Garch(1,1) data generating process (DGP) for the *S&P 500* return data. The relevance of a general Garch(1,1) process, without a concrete specification of the conditional distribution, as DGP for the *S&P 500* return series is discussed in detail in Stărică (2003). [31]

8.1. **Student-t Garch(1,1) model as data generating process for returns.** The evaluation is based on the time evolution of the estimated parameters of the model. More concretely, a Student-t Garch(1,1) process was periodically re-estimated (every 100 days) using maximum likelihood in three different set-ups. In the first set-up, the estimation was

approach, was used. With this comparison we aim at shading some light on the the relevance of the non-stationary, unconditional approach as compared with the stationary, conditional modeling paradigm. We chose to specifically include the Student-t Garch(1,1) model following the wide consensus that seems to exist in the financial econometric literature around the fact that this model provides an adequate description of return data.

[31]There both the analysis of the estimated model parameters as well as the forecasting performance clearly show the inadequacy of the Garch(1,1) model.

done on a sample which included 1000 past observations, in the second set-up the sample contained 2000 past returns while all past observations from the beginning of the sample were used in the third set-up.

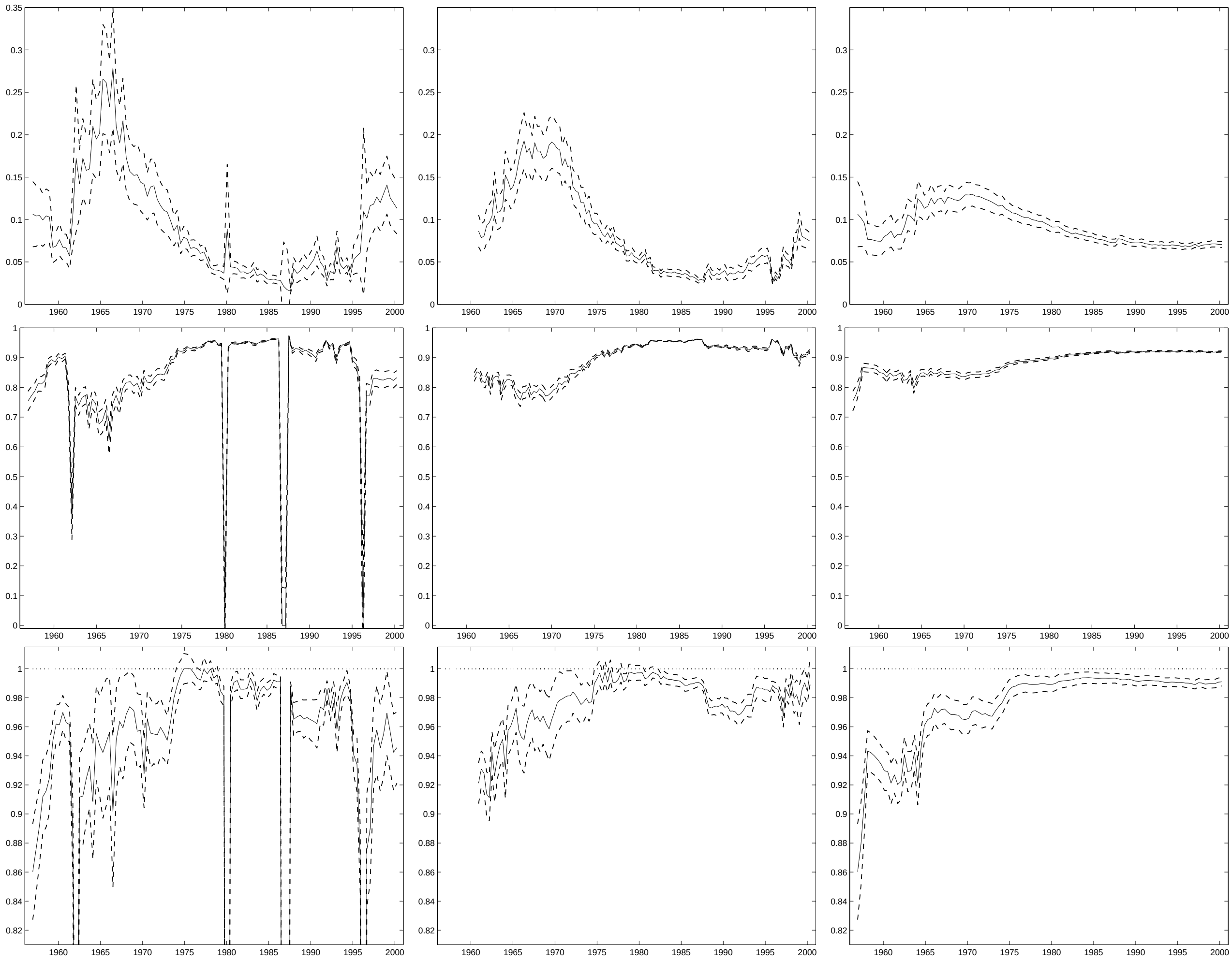

Figure 8.1. The estimated α_1 (Top line), β_1 (Middle line) and $\alpha_1 + \beta_1$ (Bottom line) coefficients in (8.13) together with 95% confidence intervals. The model was re-estimated every 100 days on a sample containing 1000 past observations (*left* column), 2000 past observations (*center* column), all the past observations (*right* column). The graphs reject the hypothesis of a stationary Student-t Garch(1,1) data generating process.

Figure 8.1 displays the results of the estimation. The estimated parameters are plotted together with the 95% confidence intervals from the maximum likelihood estimation procedure[32].

If the data generating process for the *S&P 500* return sample under scrutiny was a Student-t Garch(1,1) model, the parameters estimated on a moving window of constant length (Figure 8.1, all lines, left and center) should not vary significantly through time. Even more, the estimated values obtained using the whole sample (Figure 8.1, all lines, right) should not differ significantly from the ones based only on a part of the sample.

The plots on the bottom line of Figure 8.1 show the so-called IGARCH effect as defined and discussed in Mikosch and Stărică (2004) , i.e. the sum $\alpha_1 + \beta_1$ stays away from 1 when estimated on shorter samples while approaches 1 when the sample size grows. In that paper the authors argue that a possible reason for this behavior is the non-stationary nature of longer samples.

The plots in Figures 8.1 show clear signs of instability of the model parameters rejecting the assumption of a stationary Student-t Garch(1,1) data generating process. This finding is consistent with the results in Stărică (2003) where a Garch(1,1) is fitted to the *S&P 500* returns series between 1957 and 2003 using the quasi-maximum likelihood. In contrast with the estimation yielding the parameters in Figure 8.1, the quasi-maximum likelihoods assumes, possibly erroneously, a normal conditional distribution and accounts for the possible misspecification by corrected, larger confidence intervals (see Mikosch and

[32]Note that the exact ML confidence intervals are much narrower than the quasi-likelihood confidence intervals which account for possible misspecification of the conditional distribution. See Stărică (2003).

Straumann (2003)). The estimated quasi-maximum likelihood parameters also display a clear time evolution incompatible with the assumption of a Garch(1,1) data generating process.

8.2. **Comparison of longer-horizon volatility forecasts.** Finally, we compare the volatility forecasting results of the two methodologies. As in the previous comparison of MSEs, aggregated data at various horizons p is used as a more reliable measure against which the performances are evaluated.

In more detail, assuming a Garch(1,1) data generating process (8.13) that also satisfy $\alpha_1 + \beta_1 < 1$,[33] it follows that the minimum Mean Square Error (MSE) forecast for Er^2_{t+p}, the return variance p-steps ahead, is

$$\sigma^{2,\,Garch}_{t+p} := E_t r^2_{t+p} = \sigma^2_{Garch} + (\alpha_1 + \beta_1)^p (\sigma^2_t - \sigma^2_{Garch}), \tag{8.14}$$

where $\sigma^2_{Garch} := \alpha_0/(1-\alpha_1-\beta_1)$ is the unconditional variance. The minimum MSE forecast for $E(r_{t+1} + \ldots + r_{t+p})^2$, the variance of the next p aggregated returns is then given by

$$\overline{\sigma}^{2,\,Garch}_{t,p} := E_t(r_{t+1} + \ldots + r_{t+p})^2 = \sigma^{2,\,Garch}_{t+1} + \cdots + \sigma^{2,\,Garch}_{t+p}. \tag{8.15}$$

Under the model (6.2), the forecast for Er^2_{t+p} is given by

$$\sigma^{2,\,SM}_{t+p} := Er^2_t = h(t), \tag{8.16}$$

[33] If this condition is not fulfilled, the Garch(1,1) process, although possibly strongly stationary, has infinite variance.

with $h(t)$ as defined by (6.2) while the forecast for $E(r_{t+1} + \ldots + r_{t+p})^2$, the variance of the next p aggregated returns is simply

$$\overline{\sigma}_{t,p}^{2,\,SM} := pEr_t^2. \tag{8.17}$$

Define the following measure of the realized volatility in the interval [t+1, t+p]

$$\overline{r}_{t,p}^{\,2} := \sum_{i=1}^{p} r_{t+i}^2, \tag{8.18}$$

We calculated and compared the following MSE

$$MSE^*(p) := \sum_{t=1}^{n} (\overline{r}_{t,p}^{\,2} - \overline{\sigma}_{t,p}^{2,\,*})^2 \tag{8.19}$$

with $*$ standing for SM or Garch. The MSE (8.19) is preferred to the simpler MSE

$$\sum_{t=1}^{n} (r_{t+p}^2 - \sigma_{t+p}^{2,\,*})^2$$

since this last one uses a poor measure of the realized return volatility.[34] Through averaging some of the idiosyncratic noise in the daily squared return data is canceled yielding (8.18), a better measure against which to check the quality of the two forecasts.

Besides the *S&P 500* data, the MSE analysis has been performed also on the returns of the Nasdaq index between October 12, 1984 and November 8, 2002. The data used in the out-of-sample variance forecasting comparison are the daily returns in the interval 1965-2000 (the data from 1957 to 1964 were used for initial estimation of the models) for

[34] It is well known (see Andersen and Bollerslev (1998)) that the realized square returns are poor estimators of the day-by-day movements in volatility, as the idiosyncratic component of daily returns is large.

the *S&P 500* returns and 1993-2002 for the Nasdaq returns (with the interval 1984-1992 for the initial estimation). The two models were re-estimated every 20 days and forecasts were made based on the most recent information available. Given the fact that its parameters are significantly varying through time, the Garch model, initially estimated on a sample of length 2000, was re-estimated every 20 days on 2000 past observations.[35]

[35] The choice of the re-estimation sample size of 2000 was based on the precision of the parameter estimation. In the case of a Garch(1,1) process, this sample size guarantees acceptable standard errors for the quasi-maximum estimators. See Straumann (2003).

Horizon	$MSE^{SM}(p)/MSE^{Garch}(p)$		Horizon	$MSE^{SM}(p)/MSE^{Garch}(p)$	
p (days)	S&P 500	Nasdaq	p (days)	S&P 500	Nasdaq
10	0.26	0.94	110	0.51	0.60
20	0.50	0.93	120	0.51	0.58
30	0.41	0.83	130	0.53	0.56
40	0.44	0.83	140	0.54	0.54
50	0.42	0.78	150	0.56	0.52
60	0.44	0.76	160	0.56	0.49
70	0.44	0.71	170	0.58	0.47
80	0.46	0.70	180	0.59	0.44
90	0.46	0.66	190	0.61	0.43
100	0.48	0.64	200	0.62	0.41

TABLE 3. Comparison of volatility forecasting performance between a GARCH(1,1) model and the SM model at longer horizons. The two models are re-estimated every 20 observations using all the past observations. The ratio $MSE^{SM}(p)/MSE^{Garch}(p)$ is reported. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the Garch(1,1) model. The figure shows an over-all better longer-horizon volatility forecast performance of the SM model.

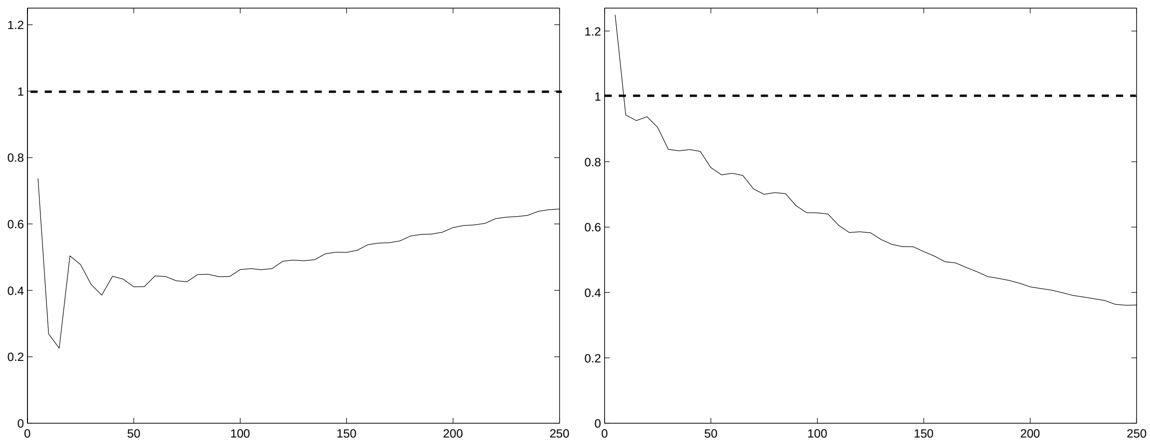

Figure 8.2. The ratio $MSE^{SM}(p)/MSE^{Garch}(p)$ for the $S\&P500$ (left) and Nasdaq (right) returns. On the x-axis, the horizon p. The Garch(1,1) model was re-estimated every 20 observations using 2000 past observations. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the Garch(1,1) model. The figure shows an over-all better longer-horizon volatility forecast performance of the SM model.

The results[36] of the comparison are given in Table 3 and in Figure 8.2. Table 3 reports the ratio[37]

$$MSE^{SM}(p)/MSE^{Garch}(p)$$

for $p = 10, 20, \ldots, 200$. Figure 8.2 gives more ample information on the ratio of mean square errors. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the Garch(1,1) model.

The results in Table 3 and Figure 8.2 show a superior volatility forecasting performance of the non-stationary model (6.2) over the Garch(1,1) model for the time series considered. They are consistent with the findings in Stărică (2003) where a Garch(1,1) process is fitted to the *S&P 500* returns series between 1995 and 2003 using the quasi-maximum likelihood. There it is shown that, in a set-up close to the one described above, a simple locally

[36]Using a misspecified maximum likelihood (Student-t could be one of them) to estimate the Garch coefficients can yield unconsistent estimators (see Straumann (2003)). To guard against the possibly serious consequences of this type of misspecification, the Garch(1,1) coefficients were estimated using the quasi-maximum likelihood method. The results of the forecasting exercise are qualitatively the same if a Student-t maximum likelihood is used for estimation of the model.

[37]The statistical significance of the results in Table 3 has been assessed using the test of statistical significance for the differences of MSE's of Diebold and Mariano (1995) (see also West (1996), Harvey et al. (1997)). The null hypothesis is that $Ed_{t,p} = 0$, where $d_{t,p} := SFE^{Garch}_{t,p} - SFE^{SM}_{t,p}$ is the loss-differential series, $SFE^{Garch}_{t,p} := (r^2_{t+p} - \sigma^{2,\,Garch}_{t+p})^2$ and $SFE^{SM}_{t,p} := (r^2_{t+p} - \sigma^{2,\,SM}_{t+p})^2$. The assumptions on the loss-differential series $d_{t,p}$ are those of *stationarity and short memory.* These assumptions are likely to hold when working under the null hypothesis of a stationary Student-t Garch(1,1) data generating process. All differences in Table 3 are significant at a 95% significance level.

constant volatility model closely related to (6.2) performed significantly better than the Garch(1,1) model in forecasting volatility over all horizons from one day to one business year. The same over-all picture emerges from the detailed analysis of a large number of series of returns on various financial indexes in Herzel et al. (2004).

The results presented in Table 3 and Figure 8.1 convincingly show that a stationary Garch(1,1) is a less appealing choice of data generating process for long series of daily returns on *S&P 500* under discussion than the non-stationary model (6.2).

9. Conclusions

In this paper an analysis of the *S&P 500* absolute returns is conducted giving up the usual assumption of global stationarity. We approximate locally the non-stationary data generating process by stationary models and identify the intervals on which stationary processes provide a good approximation. This is done using a goodness of fit test based on the integrated periodogram (Picard (1985), Klüppelberg and Mikosch (1996)). Our approach leads to modeling the returns as a sequence of independent variables with a piecewise constant variance function. More concretely, the *S&P 500* returns, r_t can be described by the following

$$r_t = h(t)^{1/2}\widetilde{\epsilon}_t, \quad t = 0, 1, \dots$$

where $(\widetilde{\epsilon}_t)$ is an iid sequence and $h(t)$ a function of time which can be well approximated by a step function, yielding a model with piecewise constant variance. We show that even a rough approximation of the variance dynamics by a step function is enough to explain most of the dependency structure present in the sample ACF of long absolute return series, providing an explanation for the so called "long memory in volatility" phenomenon.

We compared the forecasting implications of our non-stationary, unconditional modeling with first, a stationary, long memory paradigm and second, a stationary, conditional methodology. Both comparisons show the superiority of the non-stationary, unconditional approach.

Acknowledgment. Part of the research presented in this paper was carried out while the first author visited the Department of Economics of the University of California at San Diego. He would like to thank his colleagues there, in particular to Rob Engle, for their warm hospitality and generous financial support that made this work possible. The idea of trying the goodness of fit test based on the Central Limit Theorem came from Olivier Perrin. The paper gained in clarity due to the careful reading of Holger Drees, of the two referees and of John Campbell, the associated editor.

References

[1] Andersen, T, Bollerslev, T. (1998) Answering the critics: Yes, ARCH models do provide good volatility forecasts. *International Economic Review* **39**, 885–905.
[2] Anderson, T.W. (1993) Goodness of fit tests for spectral distributions. *Ann. Statist.* **21**, 830–847.
[3] Bartlett, M.S. (1954). Problemes de l'analyse spectrale des séries temporelles stationnaires. *Publ. Inst. Statist. Univ. Paris.* **III-3**, 119–134.
[4] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, New York.
[5] Cai, J. (1994) A Markov model of unconditional variance in ARCH. *J. Business and Economic Statist.* **12**, 309–316.
[6] Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37
[7] Diebold, F.X. (1986) Modeling the persistence of the conditional variances: a comment. *Econometric Reviews* **5**, 51–56.
[8] Diebold, F.X. and Inoue, A. (2001) Long memory and regime switching. *J. Econometrics* **105**, 131–159.

[9] Diebold, F.X. and Mariano, A. (1995) Comparing predictive accuracy. *J. Business and Economic Statist.* **13**, 253–263.

[10] Drees, H. and Stărică, C. (2002) A simple non-stationary model for stock returns. Available at www.math.chalmers.se/∼starica.

[11] 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.

[12] Granger, C.W.J., Hyung, N. (1999) Occasional structural breaks and long-memory. Discussion paper 99-14, University of California, San Diego.

[13] Granger, C., W. and Joyeux, R. (1980) An introduction to long-memory time series models and fractional differencing. *J. Time Series Analysis* **1**, 15–29.

[14] Granger, C., W., Spear, S., Ding, Z-X. (2000) Stylized facts on the temporal and distributional properties of absolute returns: an update. In Chan, W.-S., Li, W., K. and Tong, H., eds. *Statistics and Finance* Imperial College Press.

[15] Granger, C., W., Teräsvirta, T. (1999) A simple non-linear time series model with misleading linear properties. *Economics Letters* **62**, 161–165.

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

[17] Hamilton, J. and Susmel, R. (1994) Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics*, **64**, 307–333.

[18] Harvey D., Leybourne, S. and Newbold P. Testing the equality of prediction mean squared errors. *International J. of Forecasting*, **13**, 281–291.

[19] Herzel, S., Stărică, C. and Tütüncu, R. (2002) A non-stationary multivariate model for financial returns. *Working paper.* Available at www.math.chalmers.se/∼starica.

[20] Herzel, S., Nord, T., and Stărică, C. (2004) Why does the Garch(1,1) model often fail to produce reasonable longer-horizon forecasts? *Working paper.* Available at www.math.chalmers.se/∼starica.

[21] Hidalgo, J. and Robinson, P.M. (1996) Testing for structural change in a long–memory environment. *J. Econometrics* **70**, 159 – 174.

[22] Hosking, J. R. (1981) Fractional differencing . *Biometrika* **68**, 165–176.

[23] Hsu, D.A., Miller, R., and Wichern, D. (1974) On the stable Paretian behavior of stock-market prices. *J. of American Statistical Association* **69**, 108–113.

[24] Härdle, W., Spokoiny, V., Teyssière, G. (2000) Adaptative estimation for a time inhomogeneous stochastic volatility model. *Technical Report.* Available http://sfb.wiwi.hu-berlin.de.

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

[26] Lobato, I. N. and Savin, N. E (1998). Real and Spurious Long-Memory Properties of Stock-Market Data. *Journal of Business & Economic Statistics* **16**, 261–268.

[27] Lamoureux, C.G. and Lastrapes, W.D. (1990) Persistence in variance, structural change and the GARCH model. *J. Business and Economic Statist.* **8**, 225–234.

[28] Massey, F.J. Jr. (1951) The distribution of the maximum deviation between two sample cumulative step functions. *AMS* **22**, 125–128.

[29] Massey, F.J. Jr. (1952) Distribution table for the deviation between two sample cumulatives. *AMS* **23**, 435–441.

[30] Mikosch, T. (1998) Periodogram estimates from heavy–tailed data. In: R. Adler, R. Feldman and M.S. Taqqu (eds.) *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy–Tailed Distributions,* pp. 241–258. Birkhäuser, Boston.

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

[32] MIKOSCH, T. AND STĂRICĂ, C. (2003) Long range dependence and ARCH modeling. In: Doukhan, P., Oppenheim, G. and Taqqu, M. (eds) *Theory and Applications of Long Range Dependence,* 439–460. Birkhäuser, Boston. Available at www.math.chalmers.se/˜starica

[33] NELSEN, R.B. (1999) An Introduction to Copulas. *Lecture Notes in Statistics*, 139, Springer, NY.

[34] OFFICER, R. (1971) A time series examination of the market factor of the New York Stock Exchange. Ph.D. Dissertation. University of Chicago.

[35] PICARD, D. (1985) Testing and estimating change-points in time series. *Adv. Appl. Probab.* **17**, 841–867.

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

[37] SIMONATO, J., G. (1992) Estimation of GARCH processes in the presence of structural change. *Economic Letters* **40**, 155–158.

[38] SHORACK, G.R. AND WELLNER, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York.

[39] STOCK, J. AND WATSON, M. (1996) Evidence on structural instability in macroeconomic time series relations. *J. Business and Economic Statist.* **14**, 11–30.

[40] STĂRICĂ, C. (2003) Is Garch(1,1) as good as a model as the accolades of the Nobel prize would imply? *Working paper.* Available at www.math.chalmers.se/˜starica

[41] STRAUMANN, D. (2003) Estimation in conditionally heteroscedastic time series models. *PhD Thesis.* Laboratory of Actuarial Mathematics, University of Copenhagen.

[42] STRAUMANN, D. AND MIKOSCH, T. (2003) Quasi-MLE in heteroscedastic times series: a stochastic recurrence equations approach. *Working paper.* Available at http://www.math.ku.dk/ mikosch

[43] WEST, K., D. (1996) Asymptotic inference about predictive ability, *Econometrica*, **64**, 1067–1084.