A simple non-stationary model for stock returns [1]

Holger Drees [2] and Cătălin Stărică [3]

First draft: February 2002

This version: September 2002

[1]This work would not have been possible without the generous financial support of The Bank of Sweden Tercentenary Foundation and of the Department of Applied Mathematics of the University of Heidelberg. The hospitality of the Department of Applied Mathematics of the Heidelberg University (especially Rainer Dahlhaus) and of ECARES, Universitè Libre de Bruxelles (especially Marc Hallin) are warmly acknowledged. Holger Drees was partly supported by the Heisenberg program of the DFG. We also thank Enno Mammen (Heidelberg University) for fruitful discussions and Bas Werker (Tilburg University) for valuable remarks.

[2]Department of Mathematics, Saarland University, Saarbrücken, Germany. Email: drees@num.uni-sb.de

[3]Department of Mathematical Statistics, Chalmers University of Technology, Gothenburg, Sweden. Email: starica@math.chalmers.se. Web page: www.math.chalmers.se/˜starica

Abstract

The aim of the present paper is to show by the example of the S&P 500 return series that a simple non-stationary model seem to fit the data significantly better than conventional GARCH-type models outperforming them also in forecasting the distribution of tomorrow's return. Instead of a complex endogenous specification of the conditional variance, we assume that the volatility dynamics is exogenous. Since no obvious candidates for explanatory exogenous variables are at hand, we model the volatility as deterministic. This approach leads to a structurally simple regression-type model. Special attention is paid to the accurate description of the tails of the innovations.

Keywords and Phrases: distributional forecasts, GARCH process, non-parametric regression, non-stationarity, stock returns, volatility, heavy tails, sample autocorrelation.

1. Introduction

Accurate modelling of time series of stock returns is essential for the risk management of financial investments as well as for the pricing of derivatives. For at least two decades, the econometric models have expressed the common wisdom that appropriate modelling of financial returns should allow for non-linear dynamics. Usually the sample autocorrelation functions (SACF) of such time series are brought as evidence of their non-linear nature. Consider, for example the daily log returns of the closing prices of the Standard & Poor's 500 stock index from the January 2, 1990 to February 21, 2002, resulting in a time series of 3062 observations (Figure 3.1) (the same data set is used in the sequel to exemplify our modelling approach). While the returns show almost no autocorrelation at all lags, the absolute returns seem to have higher correlations over several hundred lags (the so-called long memory in volatility) (see Figure 1.1). In fact, after declining relatively fast for the first couple of dozen lags, the SACF of the absolute returns remains almost constant, seemingly evidence of long-range dependence in the time series of absolute returns.

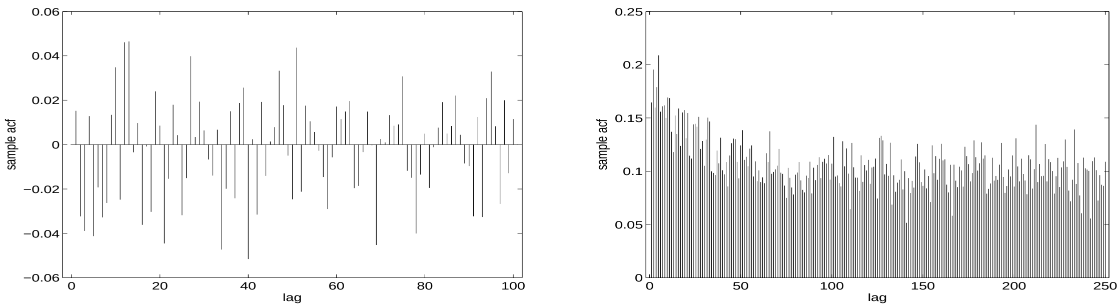

Figure 1.1. *Sample ACF for S&P 500 returns (*Left*) and absolute returns (*Right*). Since the dependency structure in the data is unknown, no confidence intervals for the correlations are displayed.*

While the aspect of the displayed SACFs could be due to stationary, non-linear dynamics, a precise interpretation of Figure 1.1 is difficult. Since the type of dependency in the data is unknown, the significance of the auto-correlations in these figures can not be assessed (for this reason no confidence intervals for the correlations are given in the mentioned figure). Another difficulty is related to distinguishing between stationary long memory, i.e.

significant correlations at large lags, and non-stationarity. A SACF that displays positive correlations at large lags (like that in Figure 1.1) could be evidence of stationary, non-linear long-range dependence as well as a sign of non-stationarities in the second moment structure of the time series; see Mikosch and Stărică [20],[22].

Among the non-linear models that try to mimic the second moment structure illustrated by Figure 1.1, the class of GARCH-type models has become extremely popular. Starting with the seminal papers by Engle [10] and Bollerslev [2] which introduced the relatively simple ARCH(p) and GARCH(p,q) models, respectively, more and more complicated modifications were proposed to take into account various stylized facts about returns. While elaborated GARCH-type models fit short return time series reasonably well, inconsistencies arise when one tries to model daily returns over longer periods (decades for example). For instance, the typical outcome of GARCH(1,1) quasi-maximum likelihood estimation on a long financial time series is a IGARCH(1,1) model, which implies an infinite variance of the observed random variables. This clearly contradicts the findings of a direct tail analysis which indicate that daily returns have a finite second moment (see for example [12]).

It has been known for a while that both the long range dependence effect and the IGARCH effect could be explained by non-stationary changes in the time series; see e.g. Diebold [7], Lamoreux and Lastrapes [19], Hidalgo and Robinson [15], Granger and Hyung [11], Diebold and Inoue [9] and Mikosch and Stărică [20], [21], to mention just a few articles from the extensive econometric literature on this topic. Models built under the hypothesis of structural breaks in the volatility of the time series of returns have been proposed in the econometric literature (Hamilton and Susmel [13], Cai [4]). However, even when switches between different regimes of low and high volatility are assumed, the hypothesis of stationarity of the whole time series is, in general, preserved. A common working hypothesis for this type of models is that the switches between various levels of volatility are described by a stationary, Markovian mechanism. The pattern of change is, in this way, supposed to remain unchanged through decades of economic activity ([13]). Given the fast pace of changes in the financial markets as well as in the economic environment, one may seriously question whether stationarity is a reasonable assumption for long financial time series.

In this paper, we describe a modelling approach, based on a different interpretation of the SACF in Figure 1.1 that seems to address these shortcomings. We show by the example of the above mentioned S&P 500 time series that a simple non-stationary model fits the data well and produces forecasts of the distribution of tomorrow's return that seem to outperform those obtained from conventional GARCH-type models, alleviating at the same time the somewhat philosophical discomfort caused by the use of a stationary paradigm in modelling the particularly dynamic environment of the financial markets.

The paper is organized as follows: Section 2 presents our modelling approach and introduces our simple non-stationary model. In Section 3 this model is estimated on the above mentioned S&P 500 time series and the goodness of fit is carefully checked. Section 4 evaluates the performance of our model in forecasting conditional distributions of returns over various horizons and compares it with classical models for financial returns. In Section 5 we comment on the relationship between our modelling approach and the RiskMetrics^{TM} methodology while Section 6 concludes.

2. A regression model for returns

Denote by X_t the return at time t. We focus on multiplicative models with constant mean:

$$X_t = \mu + \sigma_t \varepsilon_t, \qquad t = 1, 2, \ldots, n, \tag{2.1}$$

with innovations ε_t satisfying $E(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = 1$ and volatility process (σ_t). In GARCH-type models, (σ_t) is a stationary stochastic process. More concretely, a function of σ_t is modelled as a linear combination of certain functions of past volatilities, $\sigma_{t-1}, \sigma_{t-2}, \ldots$ and past returns $X_{t-1}, X_{t-2}, \ldots$. For instance, the GARCH$(1,1)$ model is given by the relation

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2.2}$$

while an exponential GARCH$(1,1)$ (or EGARCH(1,1), see [23]) model is specified by

$$\log \sigma_t^2 = \alpha_0 + \alpha_1 \frac{|X_{t-1}|}{\sigma_{t-1}} + \gamma_1 \frac{X_{t-1}}{\sigma_{t-1}} + \beta_1 \log \sigma_{t-1}^2. \tag{2.3}$$

Note that by treating positive and negative returns differently, the EGARCH$(1,1)$ model allows for a leverage effect, that is, large negative returns seemingly have a significantly stronger impact on the future returns than large positive returns.

In order to improve the fit of this type of models, more and more complicated processes were used to specify the volatility dynamics (often calling for sophisticated fitting procedures) producing an enormous number of models (for an overview see [3]). In our opinion a possible explanation for the need for an increasing complexity in the modelling of volatility could simply be that a simple endogenous specification of the volatility dynamics does not exist. If this was the case, the fit could be improved not by a more complex endogenous specification of the volatility process, but through a change of the working hypothesis. More concretely, in our approach the volatility is thought to be exogenous to the process of returns. The evolution of prices is seen as a manifestation of complex market conditions, hence driven by exogenous factors. Since no obvious candidates for explanatory exogenous variables are at hand, we model the volatility as deterministic. This approach leads to the following structurally simple regression-type model:

$$\begin{aligned} & X_t = \mu + \sigma(t)\varepsilon_t, \qquad t = 1, 2, \ldots, n, \\ & \varepsilon_t \text{ iid}, \quad E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = 1, \\ & \sigma(t), \quad t = 1, 2, \ldots, n, \qquad \text{a smooth, deterministic function of time.} \end{aligned} \tag{2.4}$$

It should be emphasized that by modelling the volatilities as a deterministic sequence we do not claim that random effects do not play any role in the volatility dynamics. This modelling approach merely reflects the belief that both recent past and close future returns (with respect to some moment t) are manifestations of the same unspecified, exogenous economic factors about which we are only willing to hypothesize a gradually changing nature. In our model these economic factors express themselves mainly in the level of the unconditional variance. Our methodology quantifies the expression of these economic factors in the recent past (with respect to the moment t) prices, i.e. evaluates the current level of unconditional variance, and uses it to forecast the close future returns since they are supposed to be the manifestation of (almost) the same constellation of exogenous variables.

To fit this regression-type model to a time series of returns we first remove the mean by subtracting the average return over the whole period under consideration producing the de-meaned returns

$$R_t = X_t - \bar{X}_n, \tag{2.5}$$

where $\bar{X}_n := n^{-1} \sum_{t=1}^{n} X_t$ is the natural estimator for the mean μ. Neglecting the estimation error in this first step, the squared de-meaned returns R_t^2 are independent with mean $\sigma^2(t)$, that is supposed to be a smooth function of time. Note that the setup is now that of a non-parametric regression with deterministic equispaced design points t and the squared volatilities are estimated in the next step by standard non-parametric regression estimators applied to the sequence (R_t^2), $t = 1, \ldots, n$. In our data analysis of the returns on the S&P 500 index, we will use the evaluation weighted (Nadaraya-Watson) estimator with a normal kernel. However, local linear kernel estimators like the LOESS procedure work equally well.

Finally, one must model the distribution of the innovations ε_t. To this end, define the *estimated innovations* as

$$\hat{\varepsilon}_t := \frac{R_t}{\hat{\sigma}(t)}, \tag{2.6}$$

with $\hat{\sigma}(t)$ denoting the square root of the estimates of $\sigma^2(t)$ obtained in the second step. In order to avoid further model assumptions, one might be tempted to use the empirical cumulative distribution function (cdf) of the innovations as an estimate of the cdf of the innovations. However, very often the distribution of the innovations is heavy tailed. Thus using the empirical cdf would underestimate the risk of extreme innovations and hence the risk of extreme returns, with potentially serious consequences when this approach is used in managing risk.

Our experience shows that a one-sided Pearson type VII distribution (concentrated on the positive half of the real line) with shape parameter m and scale parameter c having density

$$f(x;m,c) = \frac{2\Gamma(m)}{c\Gamma(m-1/2)\pi^{1/2}}\Big(1+\Big(\frac{x}{c}\Big)^2\Big)^{-m}, \qquad x>0, \tag{2.7}$$

fits the positive innovations and the absolute value of the negative returns very well. Note that f is the conditional density of a t-distributed random variable with $\nu = 2m-1$ degrees of freedom and scale parameter $c\nu^{-1/2}$, given that it is positive. Denote the densities of the negative and positive standardized innovations by $f_-(\,\cdot\,; m_-, c_-)$ and $f_+(\,\cdot\,; m_+, c_+)$, respectively. Because usually there are about as many positive innovations as there are negative ones, it may be assumed that the cdf of the innovations has median 0. Therefore, we propose modelling the innovations using the following simple but flexible family of distributions that allows for asymmetry between the distributions of positive and negative innovations and, in addition, for arbitrary tail indices for both tails:

$$f^{VII}(x; m_-, c_-, m_+, c_+) = \frac{1}{2}\Big(f_-(x; m_-, c_-)1_{(-\infty,0)}(x) + f_+(x; m_+, c_+)1_{[0,\infty)}(x)\Big). \tag{2.8}$$

We refer to the distribution with density (2.8) (that covers the whole real axis) as the asymmetric Pearson type VII and denote its cdf by F^{VII}.

To summarize, f^{VII} is determined by 4 parameters m_-, c_-, m_+, and c_+, with (m_-, c_-) and (m_+, c_+) being estimated separately by fitting a one-sided Pearson type VII distribution to the absolute values of the negative and positive standardized innovations, respectively, e.g. by maximum likelihood. These parameters, together with the variance estimates $\hat{\sigma}^2(t)$ and the average return $\bar{X}_n$, fully specify the distribution of the time series of returns in the model (2.4).

In some instances, in particular when analyzing returns over several decades, the assumption that all innovations are identically distributed is rather bold. Instead one may assume that the distribution changes slowly over time. In that case one would fit the asymmetric Pearson type VII distribution not to all innovations at once but only to those in a moving window of a given length (e.g. 1000 consecutive innovations). Anyway, as we will see in the next section, such a refinement of model (2.4) is not needed for the time series of S&P 500 returns examined here.

3. Modelling returns of the S&P 500 index

In this section we apply the methods described in the previous section to the time series of daily returns of the S&P 500 index from the January 2, 1990 to February 21, 2002 (3062 observations) displayed in Figure 3.1.

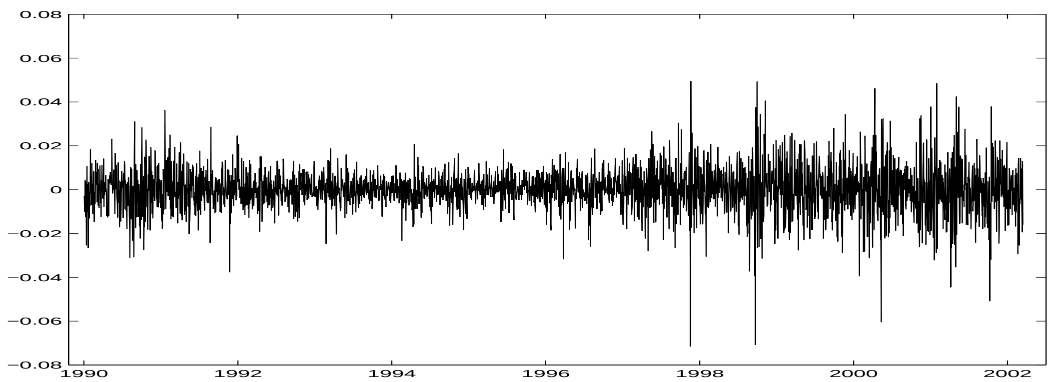

Figure 3.1. *Returns of S&P 500 from 01/02/1990 to 02/21/2002.*

In Figure 3.2 we display two estimates of the annualized time-varying standard deviations (s.d.'s) $\sqrt{250}\sigma(t)$ of the S&P 500 time series. The solid line is the estimate obtained by using a two-sided evaluation weighted (Nadaraya-Watson) estimator:

$$\hat{\sigma}^2(t) := \frac{\sum_{i=1}^{n} K_h(i-t) R_i^2}{\sum_{i=1}^{n} K_h(i-t)}, \tag{3.1}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$ and K is the normal kernel. The band width parameter is $h = 40$. For practical reasons, the weights outside a window of length 300 centered at the return whose variance is being estimated were put to 0, thus obtaining estimates for the volatility at $t = 151, \ldots, n-150$. (Note that by a minor modification using boundary kernels one may also estimate $\sigma^2(t)$ for $t = 1, \ldots, 150$ and $t = n-149, \ldots, n$.)

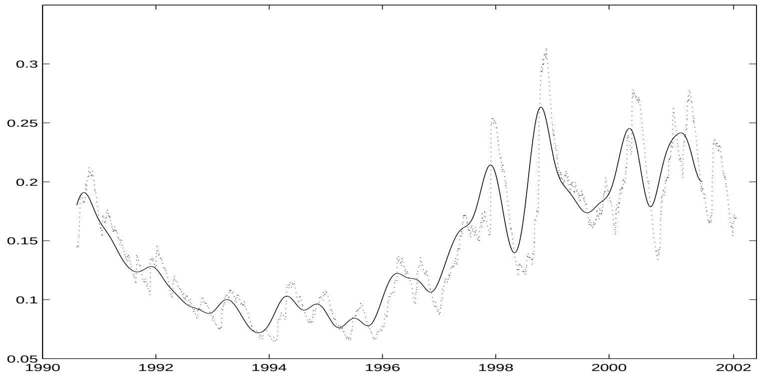

Figure 3.2. *Annualized estimated volatilities of S&P 500 returns using two-sided (3.1) (*solid line*) and one-sided (3.2) (*dotted line*) smoothing.*

The dotted line is the estimate obtained using the one-sided evaluation weighted estimator:

$$\hat{\sigma}_1^2(t) := \frac{\sum_{1 \leq i \leq t} K_h(i-t)\tilde{R}_i^2}{\sum_{1 \leq i \leq t} K_h(i-t)}. \tag{3.2}$$

with bandwidth $h = 25$, where $\tilde{R}_t$ are defined in (4.1). Note that $\hat{\sigma}_1^2(t)$ uses only the information available at day t. This estimator (with various values for h) will be used to produce the forecasting results presented in Section 4.

The annualized volatility estimates shown in Figure 3.2 (i.e. $\sqrt{250}\,\hat{\sigma}(t)$ and $\sqrt{250}\,\hat{\sigma}_1(t)$, respectively) clearly reflect the three obvious consecutive periods of high, low and high volatilities, respectively. In addition, they also exhibit strong fluctuations of the volatilities in the last period which are less obvious from Figure 3.1.

To finish fitting the model, the asymmetric Pearson type VII distribution F^{VII} (2.8) is estimated on the sequence of the estimated innovations (2.6) using the maximum likelihood for iid data (a careful check that $\hat{\varepsilon}_t$, $t = 151, \ldots, n-150$, are approximately iid follows below). The resulting estimates of the four parameters are (the s.d. in parentheses)

(3.3)

$$\hat{m}_- = 3.27 \quad (0.28), \quad \hat{c}_- = 1.88 \quad (0.14), \quad \hat{m}_+ = 6.65 \quad (1.32), \quad \hat{c}_+ = 3.23 \quad (0.40),$$

implying a tail index point estimate of 5.53 for the left tail and of 12.31 for the right tail. The fact that the left tail seems heavier is in accordance with the observation that extreme negative stock returns are usually larger in absolute value than the largest positive return, thus underpinning the need for a model that allows for tail asymmetry.

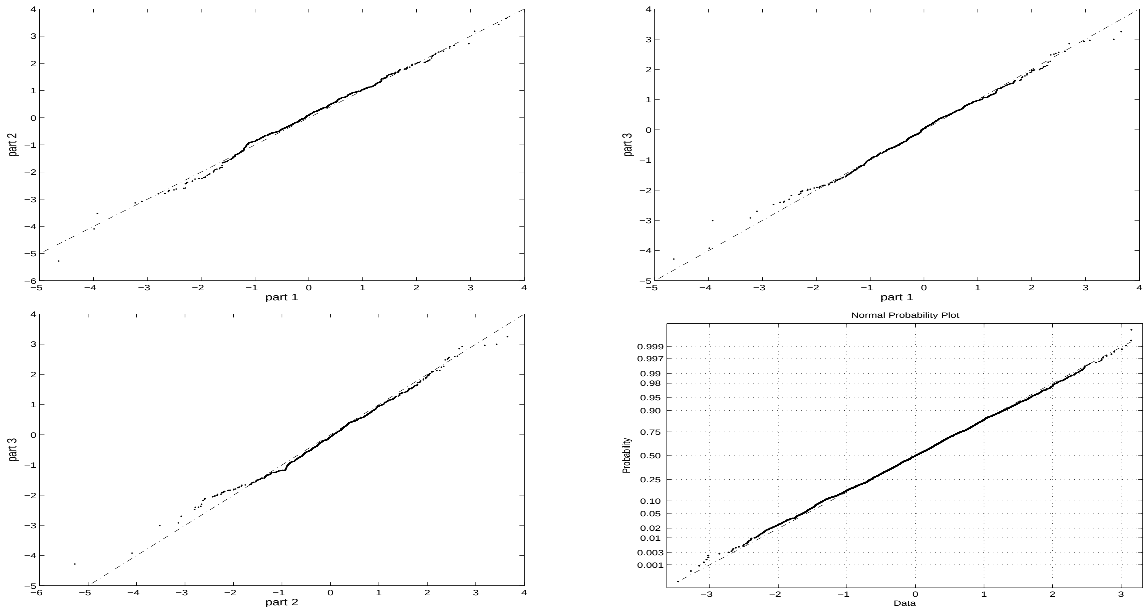

Figure 3.3. *QQ-plots of the pairs of 3 periods of innovations (*Top and Bottom Left*). Normal probability plot of the innovations transformed with estimated asymmetric Pearson type VII cdf,* $\hat{F}^{VII}$ *with the data on the* x*-axis and the normal quantile function on the* y*-axis (*Bottom Right*).*

Next we check whether our model assumptions are satisfied. First, we test whether the estimated innovations are identically distributed. For this, we divide the sample $(\hat{\varepsilon}_t)$ in three subsamples of equal length (920 observations each), $(\hat{\varepsilon}_t^{(1)})$, $(\hat{\varepsilon}_t^{(2)})$, $(\hat{\varepsilon}_t^{(3)})$ respectively. Figure 3.3 displays the QQ-plots for each pair of blocks. The fact that all 3 QQ-plots are close to the diagonal supports the hypothesis that the sample cdf of the innovations is approximately the same in all three periods. The visual check is complemented by a pairwise comparison of the three empirical cdf's using a two-sample Kolmogorov-Smirnov test.

For a given pair, say $(\hat{\varepsilon}^{(1)}, \hat{\varepsilon}^{(2)})$, the working assumptions are that $\hat{\varepsilon}^{(1)}$'s and $\hat{\varepsilon}^{(2)}$'s are mutually independent (see the independence tests in the sequel for evidence supporting this assumption) and that all the observations in the sample $(\hat{\varepsilon}^{(1)})$ come from the same continuous population $F^{(1)}$, while all the observations in the sample $(\hat{\varepsilon}^{(2)})$ come from the same continuous population $F^{(2)}$. The null hypothesis is

$$H_0: \quad F^{(1)} \text{ and } F^{(2)} \text{ are identical.} \tag{3.4}$$

The two-sample Kolmogorov-Smirnov test does not reject the hypothesis of identical distributions for all periods at the 5%-level (p-values of 0.11, 0.59 and 0.24, respectively).

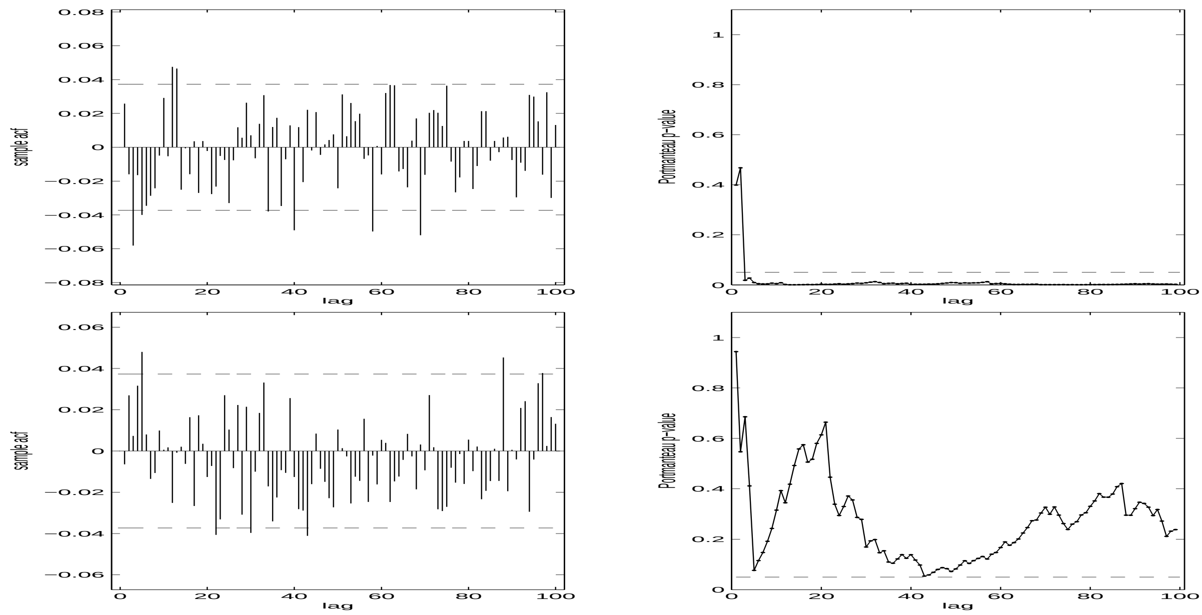

Figure 3.4. *Sample ACF (*Left*) and Portmanteau test (*Right*) for the estimated innovations $(\hat{\varepsilon}_t)$ of the regression model (2.4) fit to the series of S&P 500 returns (*Top*) and of their absolute values (*Bottom*).*

The hypothesis

$$H_0: \hat{\varepsilon}_t \text{ are iid random variables.} \tag{3.5}$$

is tested using the autocorrelation structure of the estimated innovations $(\hat{\varepsilon}_t)$ and their absolute values $(|\hat{\varepsilon}_t|)$. The left plots of Figure 3.4 show the SACF of the innovations (*Top*) and the absolute innovations (*Bottom*). The dotted lines indicate the 95%-confidence interval under the hypothesis of independence (3.5). In accordance with the model assumptions, both SACFs stay within the boundary of the confidence interval for most lags. However, the Portmanteau test for the innovations, that checks at lag h whether the sum of the first h squared sample autocorrelations multiplied with the number of innovations exceeds the 0.95-quantile of the χ^2-distribution with h degrees of freedom, rejects the independence of the innovations for most lags (*Right Top*), while the Portmanteau test based on the SACF of the absolute innovations accepts the null hypothesis (3.5) for the first 100 lags. At a first glance, this seems very peculiar since in Figure 1.1 the autocorrelation between the absolute returns appears to be much stronger than the autocorrelation between the returns. This puzzle can possibly be solved as follows. The apparently larger autocorrelations in the absolute returns could be due to non-stationarities in the second moment structure of the time series. The standardization with the changing volatility displayed in Figure 3.2, which addresses precisely this type of non-stationarity, changes only the absolute values and not the signs of the returns. As a result, it removes the aspect of long range dependence present in the SACF of absolute returns while producing a SACF of the innovations very similar to that of the returns in Figure 1.1. (The SACF of the innovations (or that of the returns) may indicate either weak dependence in the signs of the returns or non-stationarities in the mean of the time series. Neither of them is taken into account by our model or by GARCH-type models.) Up to this minor inaccuracy, we conclude that the innovations seem close to independence.

To test the hypothesis

$$H_0 : E(\hat{\varepsilon}_t) = 0, \quad Var(\hat{\varepsilon}_t) = 1, \tag{3.6}$$

two estimates of the mean and the variance of the estimated innovations $(\hat{\varepsilon}_t)$ together with the corresponding standard deviations are produced. The working assumption is that $(\hat{\varepsilon}_t)$ is an iid sequence of random variables (see the independence tests (3.5) for evidence supporting this assumption).

Under the hypothesis (3.6), according to the Central Limit Theorem, the following test statistics

$$S_{1,n} := \frac{1}{n-300}\sum_{t=151}^{n-150} \hat{\varepsilon}_t, \qquad S_{2,n} := \frac{\sum_{t=151}^{n-150}(\hat{\varepsilon}_t^2 - 1)}{\sqrt{\sum_{t=151}^{n-150}\hat{\varepsilon}_t^4}}, \tag{3.7}$$

should be approximately standard normal. The corresponding two-sided tests give p-values of 0.31 for $S_{1,n}$ and 0.41 for $S_{2,n}$, respectively, thus confirming the hypothesis (3.6).

In addition, we check whether the fitted asymmetric Pearson type VII distribution $\hat{F}^{VII}$ has the correct mean and variance. Point estimates of these two quantities, which are simple functions of the parameters, are obtained using the values in (3.3). The delta method and the large-sample approximation to the covariance matrix of the maximum likelihood estimators of the Pearson type VII model given in Section 27.6 of [16] are used to calculate the standard deviations of these estimates. This approach yields an estimated mean of 0.04 with s.d. 0.08 and variance of 1.01 with s.d. of 0.03. The pertaining two-sided parametric tests for mean 0 and variance 1 have p-values 0.55 and 0.93, respectively.

Finally, we check the goodness of fit of the asymmetric Pearson type VII distribution $\hat{F}^{VII}$ to the estimated innovations $(\hat{\varepsilon}_t)$. We do it by testing the null hypothesis

$$H_0 : \text{The marginal distribution of } (\Phi^{-1}(\hat{F}^{VII}(\hat{\varepsilon}_t))) \text{ is a standard normal, } \Phi(\cdot). \tag{3.8}$$

The bottom-right graph in Figure 3.3 displays the normal probability plot of the sequence $(\Phi^{-1}(\hat{F}^{VII}(\hat{\varepsilon}_t)))$. The resulting plot is very close to a straight line providing evidence that the parametric family (2.8) is indeed an appropriate model for the innovations. In addition, we also test the transformed innovations for normality using the two-sided Kolmogorov-Smirnov (K-S), Shapiro-Wilks (S-W) and Jarque-Bera (J-B) tests. While the first is a general non-parametric goodness of fit test, the Shapiro-Wilks and the Jarque-Bera tests are specifically designed to test normality. The Shapiro-Wilks test ([26]), based on a quadratic statistic which measures the distance of the normal probability plot from a straight line is an omnibus test for normality. In contrast, the Jarque-Bera test, commonly used in econometric applications due to its intuitive appeal (see e.g. [17]), checks only whether the skewness and kurtosis are approximately equal 0 and 3, respectively, and thus it detects only specific alternatives. The p-values of these three tests are reported in the last line of Table 1.

For comparison, a GARCH$(1,1)$ and a EGARCH$(1,1)$ model were fit to the same data. For the innovations we chose the most common distributions encountered in the literature. For the GARCH$(1,1)$ model this was the Student's t with ν degrees of freedom, while for the EGARCH$(1,1)$ model we used the Generalized Error Distribution (GED) with shape parameter ν. The models were fit using the ML estimation based on the UCSD GARCH toolbox for Matlab (http://econ.ucsd.edu/∼ksheppar/research.htm) (some errors were corrected). In our preliminary studies, the performance of this estimation procedures was superior to that of the GARCH module of the commercial software S-plus. The estimated parameters (the s.d. in parenthesis) were

$$\alpha_0 = 2.81 \cdot 10^{-7}\ (2.48 \cdot 10^{-6}),\ \alpha_1 = 0.045\ (0.002),\ \beta_1 = 0.953\ (6.85 \cdot 10^{-4}),\ \nu = 6.14\ (0.54),$$

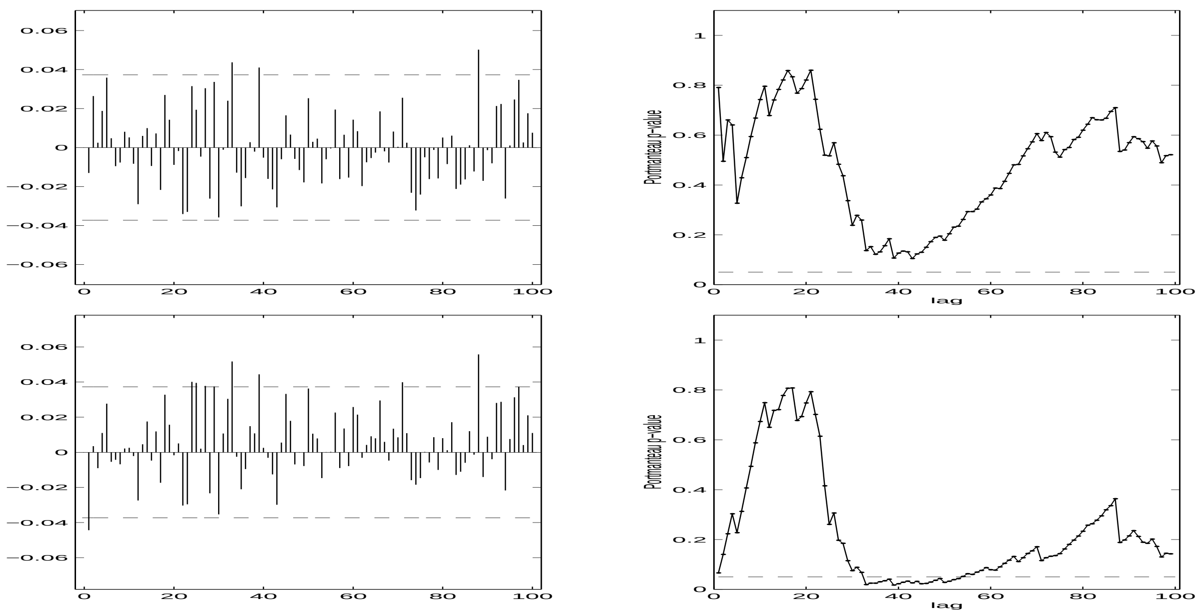

Figure 3.5. *Sample ACF (*Left*) and Portmanteau test (*Right*) for the absolute values of the estimated innovations ($\hat{\epsilon}_t$) of the Student's t* GARCH$(1,1)$ *(*Top*) and of the GED* EGARCH$(1,1)$ *(*Bottom*) fit to the series of S&P 500 returns.*

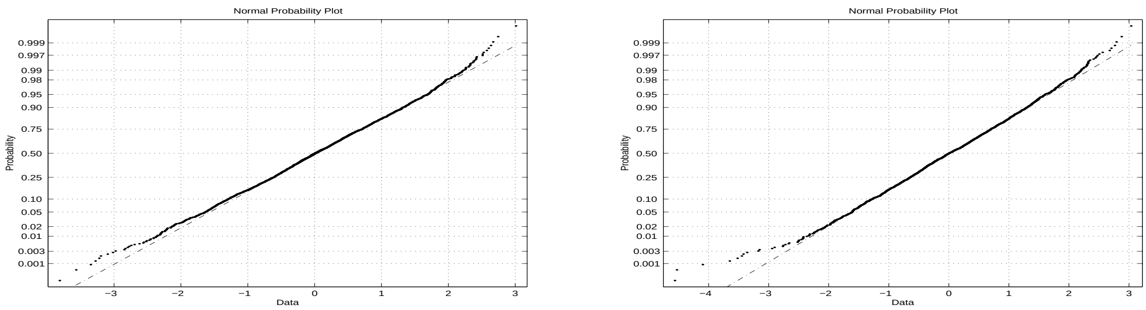

Figure 3.6. *Normal probability plot of the innovations of the Student's t* GARCH$(1,1)$ *transformed with estimated t cdf and the inverse of the standard normal cdf (*Left*) and of the innovations of the GED* EGARCH$(1,1)$ *transformed with estimated GED cdf and the inverse of the standard normal cdf (*Right*) with the data on the x-axis and the normal quantile function on the y-axis (See (3.9).*

for the GARCH$(1,1)$ model and

$$\alpha_0 = -0.226 \quad (0.004), \quad \alpha_1 = 0.083 \quad (0.011), \quad \gamma_1 = 0.122 \quad (0.004),$$
$$\beta_1 = 0.986 \quad (2.52 \cdot 10^{-5}), \quad \nu = 1.39 \quad (0.04),$$

for the EGARCH$(1,1)$ respectively. The SACF of the GARCH$(1,1)$ and EGARCH$(1,1)$ estimated innovations are very similar to the one on Figure 3.4 (*Top*) and hence we do not reproduce them. The SACF of the absolute values of the GARCH$(1,1)$ and EGARCH$(1,1)$ estimated innovations are displayed in Figure 3.5. The GARCH$(1,1)$ innovations pass the Portmanteau test for any number of lags up to 100 while the EGARCH$(1,1)$ innovations fail if the number of lags used in building the statistic is between 30 and 55.

As a goodness of fit of the marginal distribution of the innovations, we display in Figure 3.6 the normal probability plot of

$$\Phi^{-1}(U_\nu(\hat{\epsilon}_t)), \quad \hat{\epsilon}_t = R_t/\hat{\sigma}_t, \tag{3.9}$$

where $\hat{\sigma}_t$ is the GARCH$(1,1)$ (EGARCH$(1,1)$) estimated volatility and U_ν is the cdf of the Student's t with $\nu = 6.14$ degrees of freedom (the cdf of the GED distribution with shape parameter $\nu = 1.39$). The visual impression of a poor fit of the two marginal distributions

Model/Test	K-S	S-W	J-B
Student's t GARCH$(1,1)$	0.19	$6.73 \cdot 10^{-5}$	$1.74 \cdot 10^{-5}$
GED EGARCH$(1,1)$	0.50	$9.70 \cdot 10^{-6}$	$1.35 \cdot 10^{-5}$
Our model	0.70	0.42	0.84

Table 1. p-values for the Kolmogorov-Smirnov, Shapiro-Wilks and Jarque-Bera tests of normality applied to the transformed estimated innovations of the three models.

to the innovations is confirmed by the p-values of the Shapiro-Wilks and Jarque-Bera tests of normality applied to the transformed estimated innovations (3.9) given in Table 1. This table together with Figure 3.5 shows that the fit of our model is superior to that of the two GARCH-type models.

4. Forecasting comparison

Although GARCH-type models are believed to produce accurate conditional variance forecasts for 1-day returns ([3]), evaluating their performance (as forecasters of conditional variances) is difficult due the un-observed nature of the forecasted quantity (see [1]). By contrast, comparing the accuracy of various forecasts of the whole conditional distribution of future returns (and not only of the conditional variance) is, to some extent, easier. It is known, for example (see [5], [8]) that transforming the observed d-days returns with their true conditional cdf produces a $(d-1)$-dependent sequence of uniform observations. The quality of various forecasters can be then compared by assessing how close the d-days returns transformed with the conditional cdf hypothesized by various models resemble a $(d-1)$-dependent, marginally uniform sequence. The more the transformed returns resemble such a sequence, the closer is the conditional distribution hypothesized by the model to the true conditional distribution of the data. Hence our forecasting comparisons will focus on assessing the quality of the distributional forecast of future returns. Furthermore, forecasting only the volatilities is not sufficient when risk measures (like the *Value at Risk* for example) are to be calculated. This task becomes trivial when accurate distributional forecasts are available.

Most of the studies in the literature (see [8], [6] and the references therein) focus on model evaluation based on the 1-day-ahead conditional distribution forecast. Little is however known about the quality of the longer-than-1-day horizon conditional distribution forecasts of classic models like (2.2) and (2.3). The focus of this section is two-fold: to compare the forecasting performance of the two models which specify the volatility endogenously ((2.2) and (2.3)) with that of our model (which treats the volatility as being exogenous) and to evaluate qualitatively the conditional distribution forecasts over longer horizons of 20 days and 40 days of the mentioned classic models. The comparison concerns the forecasted conditional distributions of returns over 1 day, 20 days and 40 days.

Besides the belief that differences between modelling approaches will be the more pronounced the longer the forecast period, it is the interest in an accurate description of longer term risk associated with a portfolio that prompted our study of the distributional forecasts over longer horizons.

Define

$$\tilde{R}_t = X_t - \bar{X}_{t-1}, \tag{4.1}$$

with $\bar{X}_t := n^{-1}\sum_{i=1}^{t} X_i$ being the natural estimator for the mean μ based on returns up to day t. Given $\hat{\sigma}_1^2(\cdot)$, an estimate of the unconditional variance $\sigma^2(\cdot)$ based only on past information, denote by $\hat{F}_t^{VII}$, the asymmetric Pearson type VII distributions (2.8) with MLE parameters estimated on the series $(\tilde{R}_1/\hat{\sigma}_1(1),\ \tilde{R}_2/\hat{\sigma}_1(2),\ldots,\ \tilde{R}_t/\hat{\sigma}_1(t))$.

Based on the model (2.4), the distributional forecast at time t of the d-day return $X_{t+1}+\ldots+X_{t+d}$ is then given by

$$X_t^{(d)} = d\bar{X}_t + \hat{\sigma}_1(t)\ (\epsilon_1+\ldots+\epsilon_d),$$

$$\epsilon_i\ ,\ \text{i=1},\ldots d, \text{ are iid random variables with distributions } \hat{F}_t^{VII}. \tag{4.2}$$

Hence the distributional forecast for model (2.4) is that of a scaled (by an estimate of the standard deviation $\sigma(t)$) and centered (by an estimate of the mean μ) convolution of d iid random variables with cdf $\hat{F}_t^{VII}$, thus implicitly assuming that over the forecasting horizon the volatility does not change much. In practice, since a tractable theoretical description of these convolutions is not available, we used simulations: d iid observations from the distribution $\hat{F}_t^{VII}$ are summed and then scaled by the current level of the volatility $\hat{\sigma}_1(t)$ and centered by the mean estimate $\bar{X}_t$. The operation is repeated 10,000 times.

For our forecasting exercise we use

$$\hat{\sigma}_1^2(t) = \frac{\sum_{1\le i\le t} K_h(i-t)\tilde{R}_i^2}{\sum_{1\le i\le t} K_h(i-t)},$$

the one-sided kernel estimate of the unconditional covariance $\sigma^2(t)$ defined in (3.2).

The bandwidth h depends on the forecasting horizon: for the 1-day, $h_1 = 25$, for the 20-days forecast $h_{20} = 60$, while for the 40-days forecast we used $h_{40} = 100$. The weights outside a window of length 150 (400, 500 respectively) at the left of the return whose variance is being estimated were put to 0 for the 1-day (20-day, 40-day respectively) forecast. Note that it is intuitively plausible that forecasters over longer horizons use a larger bandwidth because they need to predict an average volatility level over the forecasting horizon (unlike the 1-day-ahead forecasters which produce instantaneous volatility readings).

For the GARCH-type models, the 1-day conditional distributions are given by (2.1) and the volatility dynamics postulated by (2.2) or (2.3). For $d \ge 2$ the distributions cannot be described in a simple closed form and simulations are needed. The conditional distributions are used to sequentially simulate next day return for a sequence of days of the desired length d. Then the d simulated daily returns are summed up and centered by d times an estimate of μ to produce one return over a d-day period. Again 10,000 simulations were made. The

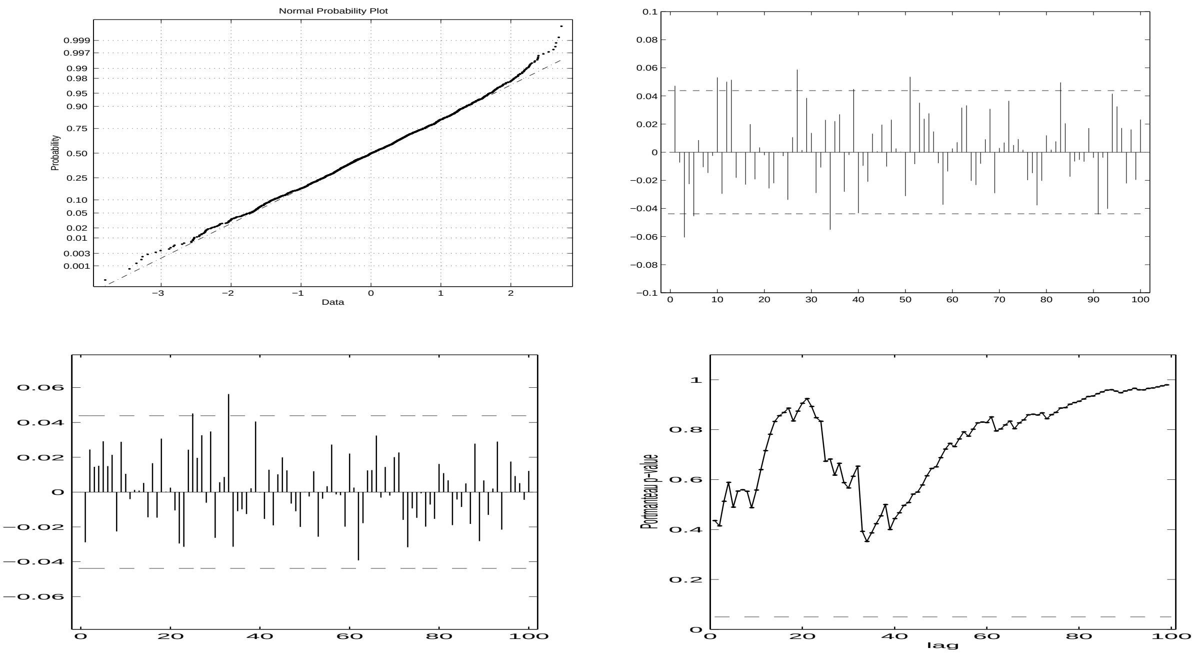

Figure 4.1. *Goodness of fit of the 1-day-ahead conditional distribution of Student's t* GARCH(1, 1) *model. The data analyzed are* $\tilde{z}_t^{(1)}$ *(4.4), i.e. daily returns transformed with the conditional cdf specified by (2.1) and (2.2) and the inverse of the standard normal cdf.* Top: *The normal probability plot (*Left*). The sample ACF of the data (*Right*).* Bottom*: The sample ACF of the absolute values (*Left*). The Portmanteau test for the absolute values (*Right*).*

performance of the models is judged based on the quality of the distributional forecast. More concretely, we are interested in evaluating the ability of a forecaster based on past observations $G_t^{(d)} = G_t^{(d)}(X_t, X_{t-1}, \ldots)$ to describe the conditional cdf of the returns over a d-days period

$$F_t^{(d)}(x) = P\Big(\sum_{i=1}^{d} X_{t+i} \leq x \,\Big|\, X_t, X_{t-1}, \ldots\Big). \tag{4.3}$$

If $G_t^{(d)} = F_t^{(d)}$, i.e. if the forecaster perfectly describes the conditional distribution of the returns over a d-days period, transforming the observed d-days returns with $G_t^{(d)}$ will produce a $(d-1)$-dependent sequence $z_t^{(d,G)} = G_t^{(d)}(\sum_{i=1}^{d} X_{t+i})$ of uniform observations (see [5], [8]). Equivalently,

$$\tilde{z}_t^{(d,G)} = \Phi^{-1}\Big(G_t^{(d)}\big(\sum_{i=1}^{d} X_{t+i}\big)\Big) \tag{4.4}$$

(where Φ is the normal cdf) will be a $(d-1)$-dependent sequence of normal observations. (We prefer the transformation to normals for the sake of commodity and also for a better emphasize on the tail behavior.) The closer the stochastic process $\tilde{z}_t^{(d,G)}$ to a marginally normal, $(d-1)$-dependent sequence, the more accurate the forecaster $G_t^{(d)}$. (In the sequel the index (d, G) in (4.4) will be dropped whenever the context makes clear the forecaster in discussion.)

Before discussing in detail the results of the forecasting of the conditional distribution, we need to say a few words about the concrete model specifications used in our comparisons. For 1-day (20-day and 40-day) forecasts and for all the models, the first 1000 (1300) observations were used for mean and parameter estimation while the rest of 2062 (1762) were used to check the quality of the forecasts (the increase in the number of observations dedicated to the preliminary estimation step accommodated the longer window used in the estimation of the local variance for the 20-day and 40-day forecasting). Always an estimate of the mean was subtracted from the data before fitting the multiplicative models. For the GARCH$(1,1)$ and EGARCH$(1,1)$ models we investigated two possible setups. In both of them the models were re-estimated periodically every 100 days. (This is done to save on computing time but simulations with re-estimation every 10 days yield essentially the same results.) In the first approach (called in the sequel Setup 1), the re-estimation of the mean and the model parameters was done on all the data from the beginning of the sample to the day of re-estimation, while in the second approach (called from now on Setup 2), a moving window of most recent 1000 observations was employed. The mean of the window of data used for the re-estimation was always subtracted from the data before re-estimating the models.

To summarize, in a given day t, the forecasters we evaluate are the conditional distributions of returns over d days given by model (2.1) with (2.2) and (2.3) and (2.4), respectively,

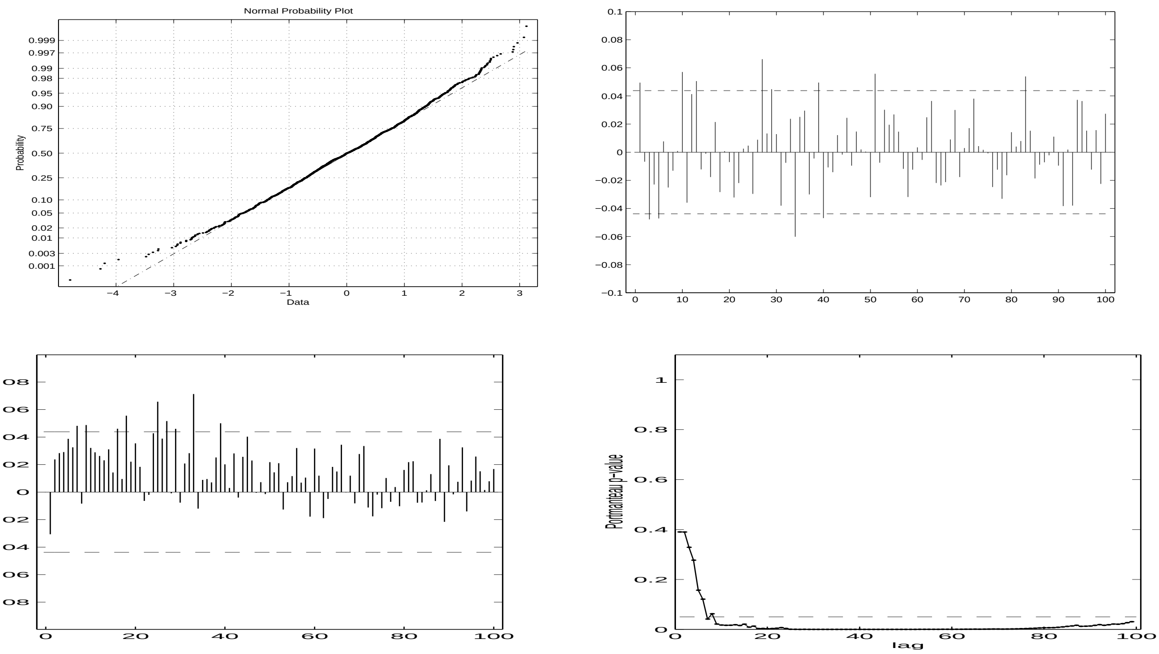

Figure 4.2. *Goodness of fit of the 1-day-ahead conditional distribution of GED-*EGARCH$(1,1)$ *model. The data analyzed are* $\tilde{z}_t^{(1)}$ *(4.4), i.e. daily returns transformed with the conditional cdf specified by (2.1) and (2.3) and the inverse of the standard normal cdf.* Top: *The normal probability plot (*Left*). The sample ACF of the data (*Right*).* Bottom*: The sample ACF of the absolute values (*Left*). The Portmanteau test for the absolute values (*Right*).*

with the most recently estimated parameters. The evaluation consists in testing the hypothesis:

$$H_0: \ (\tilde{z}_t^{(d,G)}) \text{ is a } (d-1)\text{-dependent sequence of } \mathcal{N}(0{,}1) \text{ random variables.} \tag{4.5}$$

For the 1-day ahead forecasts the test is facilitated by the simpler form of the hypothesis

$$H_0: \ (\tilde{z}_t^{(1,G)}) \text{ is an iid sequence of } \mathcal{N}(0{,}1) \text{ random variables.} \tag{4.6}$$

We start now presenting the results of the comparison of the 1-day ahead forecasts. Since the differences between Setup 1 and 2 are in this case minimal we will only present the results of the analysis under the Setup 2. To test the null of normality, the top left graphs in

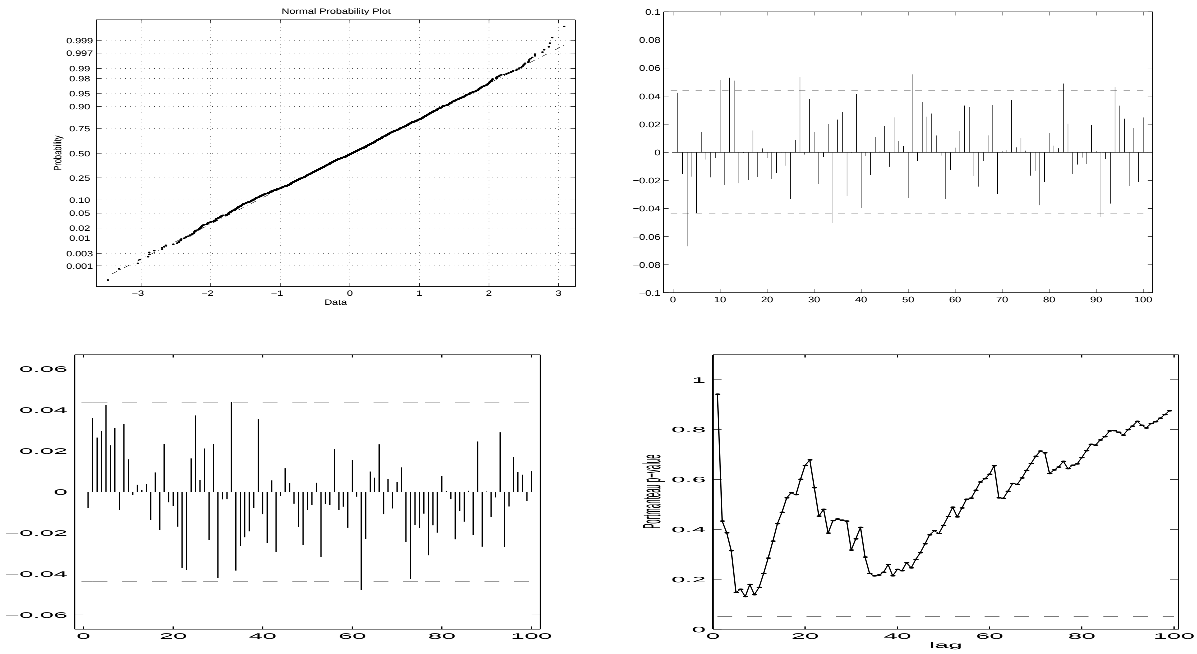

Figure 4.3. *Goodness of fit of the 1-day-ahead conditional distribution of the regression model. The data analyzed are $\tilde{z}_t^{(1)}$ (4.4), i.e. daily returns transformed with the conditional cdf specified by (2.4) and the inverse of the standard normal cdf.* Top: *The normal probability plot (*Left*). The sample ACF of the data (*Right*).* Bottom*: The sample ACF of the absolute values (*Left*). The Portmanteau test for the absolute values (*Right*).*

Figures 4.1–4.3 display the normal probability plot of the time series $(\tilde{z}_t^{(1)})$ based on the 1-day-ahead forecasts for the GARCH(1,1) model (2.1), (2.2) with t-distributed innovations, the EGARCH(1,1) model (2.1), (2.3) with GED-innovations and the regression-type model (2.4), respectively. The sequences $(\tilde{z}_t^{(1)})$ associated with the GARCH-type models are poorly fitted by a normal distribution displaying strong deviations in the tails. This visual impression is confirmed by the p-values of the Kolmogorov-Smirnov, Shapiro-Wilks and Jarque-Bera tests given in the Table 2. Indeed, for both GARCH-type models the Shapiro-Wilks and Jarque-Bera strongly reject the hypothesis of normality and the EGARCH(1,1) model even fails the Kolmogorov-Smirnov test. In contrast, the normal probability plot of the standardized returns $(\tilde{z}_t^{(1)})$ corresponding to our model (2.4) is very close to a straight line. This sequence comfortably passes all the three tests of normality.

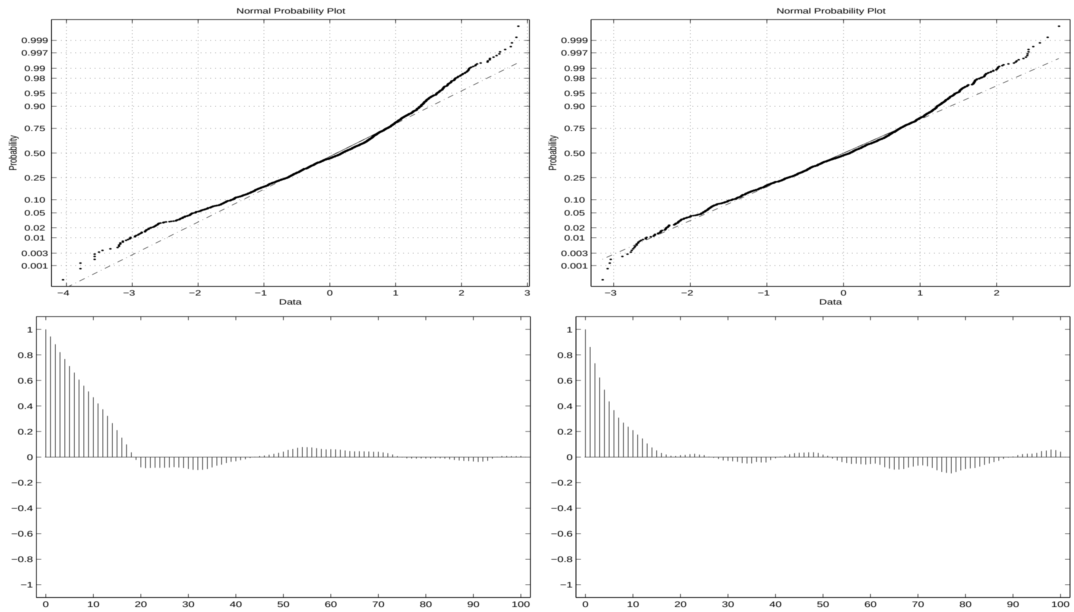

Figure 4.4. *Goodness of fit of the 20-days-ahead conditional distribution of Student's t* GARCH$(1,1)$ *model. The data analyzed are* $\tilde{z}_t^{(20)}$*, the 20-day returns transformed with the conditional cdf specified by (2.1) and (2.2) and the inverse of the standard normal cdf.* Top: *The normal plot of the data obtained in Setup 1 (*Left*) and Setup 2 (*Right*).* Bottom*: The sample ACF of data (Setup 2)(*Left*) and absolute values (*Right*).*

To check the independence of the standardized return sequences $(\tilde{z}_t^{(1)})$ implied by the three models, we display the SACF of $(\tilde{z}_t^{(1)})$ (*Top Right*) and of $(|\tilde{z}_t^{(1)}|)$ (*Bottom Left*) and the Portmanteau p-values for $(|\tilde{z}_t^{(1)}|)$ (*Bottom Right*) (Figures 4.1–4.3). The SACFs of the sequences $(\tilde{z}_t^{(1)})$ are very similar for all three models, and they closely resemble a SACF of an iid sequence. However, they are affected by the same problem as the SACF of the returns (X_t) and of the estimated innovations $(\hat{\varepsilon}_t)$. That is, despite the small values of the SACFs, a Portmanteau test rejects the hypothesis of independence of the sequence $(\tilde{z}_t^{(1)})$ for all three models (see the discussion in Section 3). In contrast, the Portmanteau test based on the absolute standardized returns $(|\tilde{z}_t^{(1)}|)$ rejects the hypothesis of independence for the EGARCH(1,1) model only (its SACF is positive for most of the first 50 lags!), while the sequences $(|\tilde{z}_t^{(1)}|)$ seem independent for the GARCH(1,1) model and our regression model.

To summarize the evaluation of the 1-day forecasts, the sequences $(\tilde{z}_t^{(1)})$ are approximately iid standard normal for the regression-type model while the normality is rejected for both GARCH-type models, and independence is rejected for the EGARCH(1,1) model, which seems least suited for 1-day-ahead forecasts.

Next we turn to evaluating the 20-days-ahead (Figures 4.4–4.6) and 40-days-ahead (Figures 4.7–4.9) forecasts. For the GARCH-type models, we consider both Setup 1, using the whole past in the re-estimation step, and Setup 2, where only the last 1000 observations are used to re-fit the model. The figures display in the top panels the normal probability plots, while the Setup 2 SACFs of $(\tilde{z}_t^{(d)})$ and $(|\tilde{z}_t^{(d)}|)$, for d=20 and 40, are displayed in the bottom panels (the SACFs under Setup 1 are very similar and are not displayed for space reasons). Since the dependency structure in the data is unknown, no confidence intervals for the autocorrelations are displayed.

The evaluation of the three models consists in a discussion of the hypothesis (4.5) and has a more qualitative flavor than the analysis performed for the 1-day forecasts. The significance of the departures from a straight line of the normal probability plots as well

Model/Test	K-S	S-W	J-B
Student's t GARCH$(1,1)$	0.06	$4.9 \cdot 10^{-4}$	$3.4 \cdot 10^{-3}$
GED EGARCH$(1,1)$	0.01	$6.5 \cdot 10^{-5}$	$6.0 \cdot 10^{-5}$
Our model	0.29	0.27	0.25

TABLE 2. p-values for the Kolmogorov-Smirnov (K-S), Shapiro-Wilks (S-W) and Jarque-Bera (J-B) tests of normality applied to the sequences $(\tilde{z}_t^{(1)})$ associated with the three models: (2.1) with (2.2) and (2.3), (2.4), respectively.

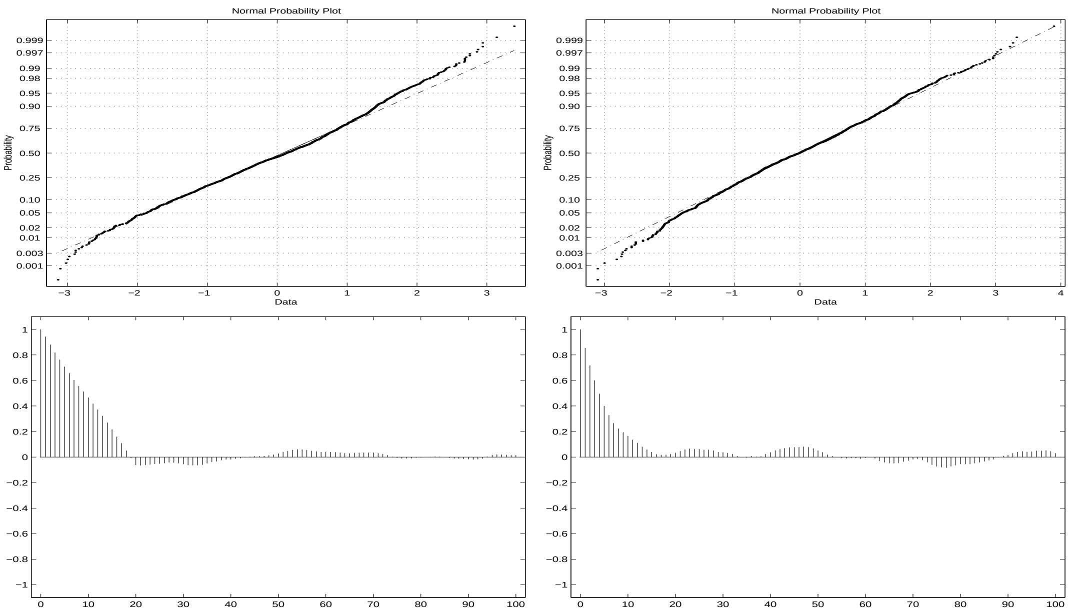

Figure 4.5. *Goodness of fit of the 20-days-ahead conditional distribution of GED* EGARCH(1, 1) *model. The data analyzed are* $\tilde{z}_t^{(20)}$*, the 20-day returns transformed with the conditional cdf specified by (2.1) and (2.3) and the inverse of the standard normal cdf.* Top: *The normal plot of the data obtained in Setup 1 (*Left*) and Setup 2 (*Right*).* Bottom*: The sample ACF of data (Setup 2)(*Left*) and absolute values (*Right*).*

as the significance of the auto-correlations in Figures 4.4–4.9 are not easy to assess due to the $(d-1)$-dependence of the series. It is a well-known fact that dependency can be mistaken for departure from normality, if one does not take it into account while testing. Classical time series theory also shows that the presence of positive dependence widens the confidence intervals for the autocorrelation.

Bearing in mind this caveat, let us note that for the GARCH(1,1) model the normal fit seems better in the Setup 2 when only the last 1000 returns are used to re-fit the model. For the EGARCH(1,1) model this finding is less obvious with a slightly straighter plot for the Setup 2. The regression model produces a sequence $(\tilde{z}_t)$ whose distribution seems very close to a standard normal. The SACFs of the sequences $(\tilde{z}_t)$ and $(|\tilde{z}_t|)$ in Setup 2 (bottom panels) are rather small for lags greater than d and resemble closely the

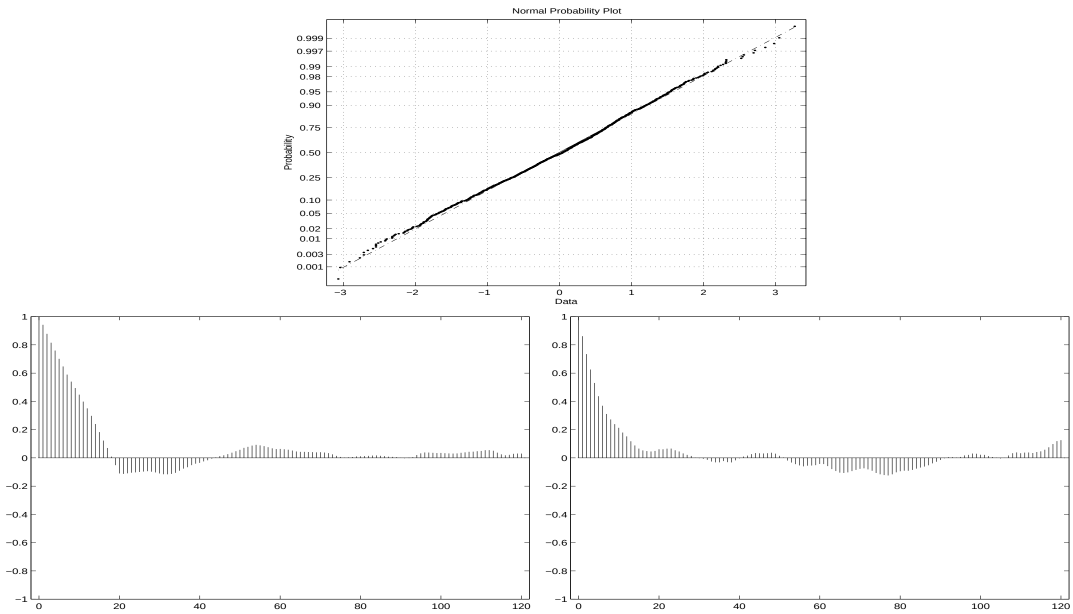

Figure 4.6. *Goodness of fit of the 20-days-ahead conditional distribution of the regression model. The data analyzed are* $\tilde{z}_t^{(20)}$*, the 20-day returns transformed with the conditional cdf specified by (2.4) and the inverse of the standard normal cdf.* Top: *The normal plot of the data.* Bottom*: The sample ACF of data (Setup 2)(*Left*) and absolute values (*Right*).*

dependency structure of a MA(d) process for all models. The SACFs do not seem to indicate violations of the null of $d-1$ dependence.

Lacking a viable methodology, we do not formally test either the normality of the standardized d-day returns or their d-dependent structure. However, we will complement the graphical evaluation of the normality hypothesis for the sequences $(\tilde{z}_t^{(20)})$ based on the normal probability plots in Figures 4.4–4.6 with two qualitative investigations. (The very strong and far-reaching dependence and the small sample size are good arguments for not extending the qualitative analysis to the 40-day forecasts.)

As mentioned, the classical tests employed previously cannot be used directly to assess the null of standard normal marginals, since the correct critical values are unknown. In [5] and [8] it is suggested to split the whole time series into d sub-samples $(\tilde{z}_{i+td}^{(d)}))$, $1 \leq i \leq d$, $t = 0, 1, \ldots$, of iid random variables and to construct a test of size less than or equal to α of

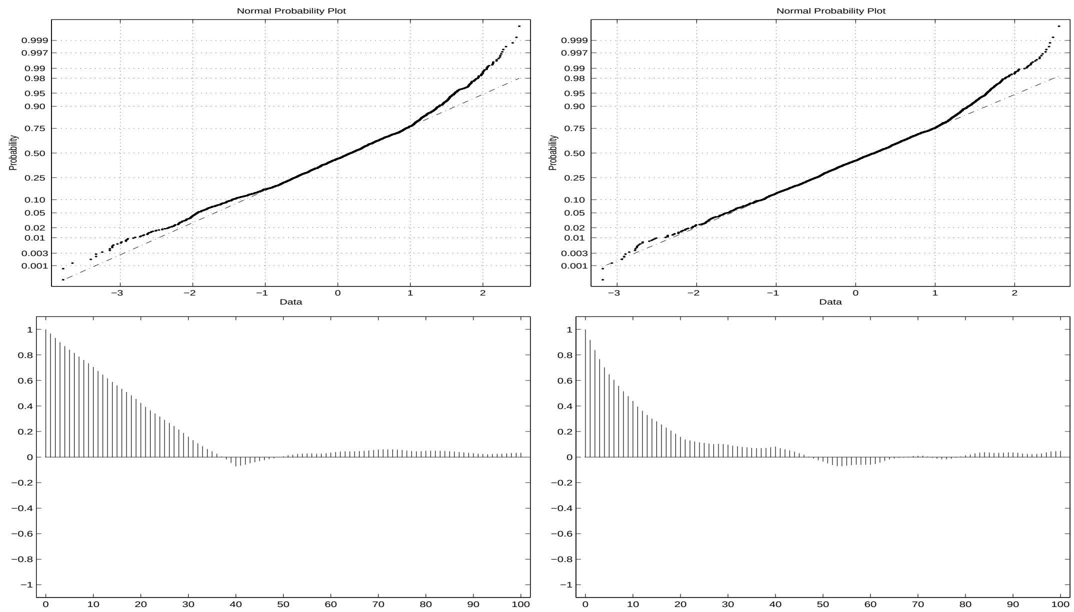

Figure 4.7. *Goodness of fit of the 40-days-ahead conditional distribution of Student's t* GARCH(1, 1) *model. The data analyzed are* $\tilde{z}_t^{(40)}$*, the 40-day returns transformed with the conditional cdf specified by (2.1) and (2.2) and the inverse of the standard normal cdf.* Top: *The normal plot of the data obtained in Setup 1 (*Left*) and Setup 2 (*Right*).* Bottom*: The sample ACF of data (Setup 2)(*Left*) and absolute values (*Left*).*

the null (4.5) by performing d tests of size α/d of the hypothesis (4.6) on the sub-samples. The test is formulated as follows:

(4.7) reject the null (4.5) for the whole sample if at least one of the d test rejects.

This approach is probably not sound for time series displaying rather strong positive dependency like in our case. While the Bonferroni inequality guarantees a size of at most α, the presence of the strong positive dependency between the sub-samples can make the real size of the test be much smaller than α. However, the method could offer a qualitative insight highlighting the relative behavior of the three models.

In Table 3 we provide the minimum $p_{\min}$ of the twenty p-values for the Shapiro-Wilks test of normality applied to the sub-samples of iid random variables of the sequences $(\tilde{z}_t^{(20)})$ associated with the three models. Since $p_{\min}$ exceeds 0.05/20=0.0025, the assumption of

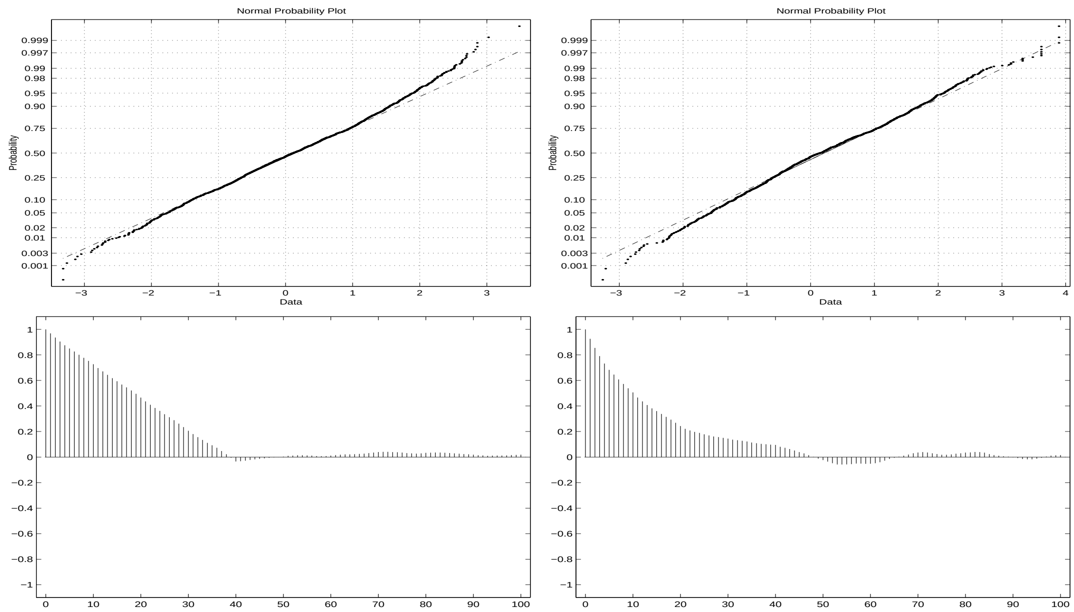

Figure 4.8. *Goodness of fit of the 40-days-ahead conditional distribution of GED* EGARCH(1, 1) *model. The data analyzed are* $\tilde{z}_t^{(40)}$*, the 40-day returns transformed with the conditional cdf specified by (2.1) and (2.3) and the inverse of the standard normal cdf.* Top: *The normal plot of the data obtained in Setup 1 (*Left*) and Setup 2 (*Right*).* Bottom*: The sample ACF of data (Setup 2)(*Left*) and absolute values (*Right*).*

normality cannot be rejected at the 5%-level for any of the three models. The second line of the table gives the p-value of the test (4.7) if the sub-samples were independent, i.e. $1-(1-p_{\min})^{20}$. The size of EGARCH(1, 1) model comes the closest to the 5% level while the other two sizes seem comfortably far from it.

The second qualitative analysis consists in simulating the asymptotic distribution of the three statistics, Kolmogorov-Smirnov, Shapiro-Wilks and Jarque-Bera for MA(20) processes with normal innovations and coefficients and variance of the residuals estimated on the three series $(\tilde{z}_t^{(20)})$. (We note that the three sets of estimated coefficients are very close. In fact, they do not differ statistically.) The approximation of the dependency structure of the sequences $(\tilde{z}_t^{(20)})$ by a MA(20) (although justified by a good fit), together with the use of estimated parameters in generating the asymptotic distributions of the three mentioned statistics might have distorted to some extent the p-values in Table 4. It is, hence, without giving the weight of a formal test to these results, that we state our conclusions.

	GARCH(1,1)	EGARCH(1,1)	Our model
$p_{\min}$	0.021	0.006	0.044
p-value (if indep.)	0.35	0.11	0.59

TABLE 3. *First line:* The smallest p-values for the Shapiro-Wilks test of normality applied to the 20 sub-samples of iid random variables of the the sequences $(\tilde{z}_t^{(20)})$ associated with the three models: (2.1) with (2.2) and (2.3), (2.4), respectively. *Second line:* The p-value of the test (4.7) if the sub-samples were independent $(1-(1-p_{\min})^{20})$.

Model/Test	K-S	S-W	J-B
Student's t GARCH(1,1)	0.68	0.02	0.04
GED EGARCH(1,1)	0.65	0.69	0.56
Our model	0.74	0.75	0.48

TABLE 4. p-values for the Kolmogorov-Smirnov (K-S), Shapiro-Wilks (S-W) and Jarque-Bera (J-B) tests of normality for MA(20) data applied to the sequences $(\tilde{z}_t^{(20)})$ associated with the three models: (2.1) with (2.2) and (2.3), (2.4), respectively.

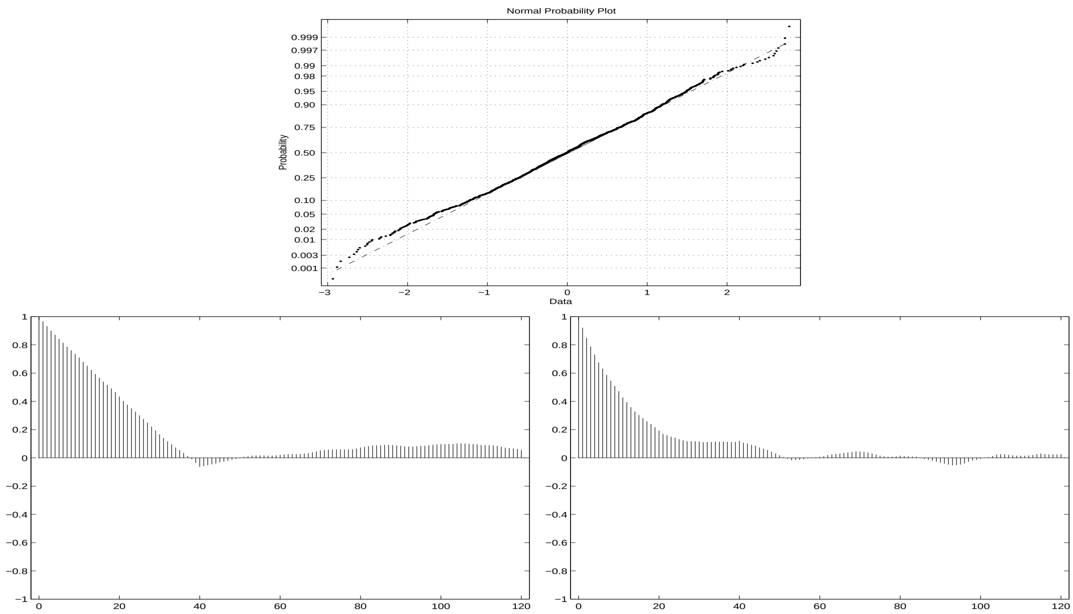

Figure 4.9. *Goodness of fit of the 40-days-ahead conditional distribution of the regression model. The data analyzed are* $\tilde{z}_t^{(40)}$*, the 40-day returns transformed with the conditional cdf specified by (2.4) and the inverse of the standard normal cdf.* Top: *The normal plot of the data.* Bottom*: The sample ACF of data (*Left*) and absolute values (*Right*).*

The Jarque-Bera and the Shapiro-Wilks tests reject (at the statistical level of 5%) the normality of the sequence $(\tilde{z}_t^{(20)})$ associated with the Students's t GARCH$(1,1)$ model while seemingly not rejecting it for the ones obtained using the GED EGARCH$(1,1)$ and the regression model. The Kolmogorov-Smirnov test does not reject the normality of $(\tilde{z}_t^{(20)})$ for any of the models confirming in this way the impression of being the least discriminating test among the three.

To summarize, the qualitative results of the two tests we applied to the dependent sequences $(\tilde{z}_t^{(20)})$ corresponding to the three models confirm the visual impression given by the normal probability plots in Figures 4.4–4.6 that the regression model seems to behave better in forecasting future returns over 20-day intervals then the other classic GARCH-type models.

We conclude that, also for longer horizons, the distributional forecasts based on our regression model seem to outperform the forecasts given by the GARCH-type models considered.

5. The Relationship to RiskMetrics

In the last few years the RiskMetrics^{TM} methodology developed by J.P.Morgan has become a kind of un-official industry standard for forecasting volatilities and assessing the risk of financial investments. In this section we discuss similarities and differences between this methodology and our approach. To this end, we focus on the standard RiskMetrics approach which ignores the mean of the returns and assumes normal innovations, although we also briefly discuss more refined models.

In the RiskMetrics manual [25] it is suggested to forecast the volatility by an exponential filter of the squared returns:

$$\hat{\sigma}^2_{RM,t} := \frac{\sum_{i=1}^{m} \lambda^i X^2_{t-i}}{\sum_{i=1}^{m} \lambda^i}, \tag{5.1}$$

where for a one-day-ahead forecast the values $\lambda = 0.94$ and $m = 74$ are recommended. In the next step, the standardized returns $X_t/\hat{\sigma}_{RM,t}$ are treated as (approximately) independent normal random variables when the Value at Risk and related quantities are calculated.

Unfortunately, the probabilistic model that forms the basis of this statistical procedure is, according to our reading of the RiskMetrics manual, faulty. In the conclusion of the fourth chapter, "Statistical and probability foundations", section 4.6 titled "RiskMetrics model of financial returns: A modified random walk", the model receives the following description: "The variance of each return, $\sigma^2_{i,t}$ and the correlation between returns, $\rho_{i,t}$, (*the index i and the presence of ρ are for the description of the multivariate case, n.n*) are a function of time. The property that the distribution of returns is normal given a time dependent mean and correlation matrix assumes that returns follow a conditional normal distribution-conditional on time."

In formulae, as specified on page 73 of [25], a conditional, multiplicative model of the type (2.1) with $\mu = 0$ and normal independent innovations is assumed, i.e.

$$X_t = \sigma_t \, \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0,1). \tag{5.2}$$

The vague description of σ_t as "a function of time" is made precise eight chapters later, i.e. in Section B.2.1 of the Appendix B as

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) X_{t-1}^2. \tag{5.3}$$

This specification is, up to a constant term, that of a IGARCH process (explaining why in the literature the RiskMetrics model is often thought of as being an IGARCH model). However, seemingly small in appearance, this difference is very big in substance. Results by Kesten [18] and Nelson [24] imply that a time series evolving according to the dynamics (5.2) and (5.3) will tend to 0 almost surely!

From our point of view, the claimed close relationship between the RiskMetrics methodology and GARCH-type models (prompted by the deceiving formal analogy between the specification (2.2) and (5.3) and emphasized by the comparisons in Section 5.2.3 of [25]) is misleading. Instead, the RiskMetrics approach can be motivated by our non-stationary regression model (2.4).

Indeed, if one assumes a zero mean (i.e. $X_t = R_t$), then the forecast (5.1) is just a kernel estimator of the type (3.2) with an exponential kernel $K_{exp}(x) = a^x 1_{[-1,1]}(x)$, $a = \lambda^m$ and $h = m$. Our experience shows that replacing the normal kernel with the exponential leads to results very similar to the ones reported in Section 4. This finding is in line with the well-known fact that the choice of the bandwidth h affects the performance of a kernel regression estimator much more strongly than the choice of the kernel. In fact, in the Sections 3 and 4 we have deliberately chosen the normal kernel instead of the exponential filter (more common in time series analysis) to demonstrate that the choice of the kernel does not matter much.

While the volatility forecasts by the RiskMetrics methodology are similar to ours, the assumption of normal innovations is too restrictive to yield accurate forecasts of the distribution of future returns. This has also been observed in [25]. In Appendix B of the RiskMetrics document normal mixture models or GED models for the innovations are proposed. However, these alternative models lack two features that turned out to be essential for a successful fit of many real data sets: they do not allow for asymmetry of the distribution of innovations and they assume densities with exponentially decaying tails, thus excluding heavy tails. Simulations as well as the results in Herzel et al. [14] show that the reason for the inferiority of distributional forecasts based on RiskMetrics (like for the GARCH-type models discussed in the previous sections) is mainly due to this inflexibility of the model for the innovations.

A second, though minor, difference between our approach and the RiskMetrics procedure is that we explicitly include a non-vanishing mean in our model. Although the standard deviation of the S&P 500 returns is about 27 times bigger than their mean, ignoring the mean altogether leads to a substantial loss in model and forecasting accuracy, while the simple assumption of a constant mean suffices to obtain a good model fit and accurate forecasts as well.

To summarize this discussion, our non-stationary approach of modelling the financial returns as independent observations with a changing unconditional variance provides the needed theoretical background for the RiskMetrics methodology. At the same time, our more flexible modelling of innovations significantly improves the fit and the forecasting performance. (For evidence in the multivariate setup we refer the reader to Herzel et al. [14].)

6. Conclusions

In this paper a simple model for stock returns was introduced. Unlike most of the econometric models for financial returns in our approach the volatility is thought to be exogenous to the process of returns. The evolution of prices is seen as a manifestation of complex market conditions, hence driven by exogenous factors. Since no clear candidates for explanatory exogenous variables are at hand, we model the volatility as a deterministic function of time $\sigma(t)$. The methodological frame is that of nonparametric regression with non-random equidistant design points where the regression function is the evolving unconditional variance. The regression function is estimated by an evaluation weighted (Nadaraya-Watson) estimator with a normal kernel, although the particular choice of the kernel is of minor importance. The innovations are iid random variables with asymmetric heavy tailed and are modelled parametrically.

The model, together with an Student's t GARCH$(1,1)$ and Generalized Error Distribution EGARCH$(1,1)$ were estimated on the last 12 years of daily log returns of the closing prices of the Standard & Poor's 500 stock index. Our model apparently fits the data significantly better than conventional GARCH-type models.

In the forecast setup, our approach models future returns as an iid sequence with a variance estimated on the very recent past returns and a distribution estimated on a long series of past innovations. A comparison of the forecasted distribution of returns over various horizons by our model and the classic GARCH-type of models mentioned earlier revealed a significantly better performance of the regression-type model.

References

[1] Andersen, T. and Bollerslev, T. (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts," *International Economic Review*, **39**, 885–905.

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

[3] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). GARCH models. In: Engle, R.F., and McFadden, D.L. (eds.) *Handbook of Econometrics*, **Vol. 4**, Elsevier, 2961–3038.

[4] Cai, J. (1994) A Markov model of unconditional variance in ARCH. *J. Business and Economic Statist.* **12**, 309–316.

[5] Christoffersen, P. F. (1998) Evaluating interval forecasts. *International Economic Review*, **39**, 841–862.

[6] Clements, M., Smith J. (2000) Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting*, **19**, 255–276.

[7] Diebold, F.X. (1986) Modelling the persistence of the conditional variances: a comment. *Econometric Reviews* **5**, 51–56.

[8] Diebold, F.X., Gunther, T. and Tay, A. (1998) Evaluating density forecasts, with applications to financial risk management, *International Economic Review*, **39**, 863–883.

[9] Diebold, F.X. and Inoue, A. (2001) Long memory and regime switching. *J. Econometrics* **105**, 131–159.

[10] Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**, 987–1008.

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

[12] de Haan, L., Jansen, D., Koedijk, K. and de Vries, C. (1994) Safety first portfolio selection, extreme value theory and long run asset risks In Galambos, J.(ed.), *Extreme Value Theory and Applications*, Kluwer, 471-487

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

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

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

[16] Johnson, N.L. and Kotz, S. (1970) *Continuous Univariate Distributions.* Houghton Mifflin, Boston.

[17] Judge, G.G., Lee, T.C. and Hill, R.C. (1988). *Introduction to the Theory and Practice of Econometrics, 2nd ed.* Wiley, New York.

[18] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

[19] 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.

[20] Mikosch, T. and Stărică, C. (2002) Long range dependence effects and ARCH modelling. In: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.) *Long Range Dependence.* Birkhauser, Boston. To appear.

[21] Mikosch, T. and Stărică, C. (2002) Changes of structure in financial time series and the GARCH model *Preprint.* Available at www.math.chalmers.se/˜starica

[22] Mikosch, T. and Stărică, C. (2002) Non-stationarities in financial time series, the long range dependence and the IGARCH effects. *Preprint.* Available at www.math.chalmers.se/˜starica

[23] Nelson, D.B. (1989) Modelling stock market volatility changes. *ASA 1989 Proceedings of the Business and Economics Statistics Section*, 93–98,

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

[25] J.P.Morgan and Reuters (1996). *Riskmetric*^{TM}. Technical Document, available at www.jpmorgan.com/RiskManagement/RiskMetrics/RiskMetrics.html.

[26] Shapiro, S.S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.