A non-stationary multivariate model for financial returns [1]

Stefano Herzel [2] , Cătălin Stărică [3] and Reha Tütüncü[4]

First draft: November 2001

This version: August 2002

[2]Department of Economics, University of Perugia, Via Pascoli 1, Perugia, 06100, Italy. Email: herzel@unipg.it

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

[4]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh. Email: reha@andrew.cmu.edu

Abstract

A simple non-stationary model for multivariate returns is proposed. Unlike most of the multivariate econometric models for financial returns, this model supposes the volatility to be exogenous. The vector of returns is assumed to follow an AR(1) process whose innovations are independent and have a slowly changing unconditional covariance structure. The methodological frame is that of non-parametric regression with non-random equidistant design points, where the regression function is the evolving unconditional covariance. Special attention is payed to the accurate description of the tails of the innovations. As an application the model is fit to a multivariate data set of returns on three different financial instruments: a foreign exchange rate, an index and an interest rate. The 1-day ahead multivariate distributional forecast performance is evaluated.

Keywords and Phrases: stock returns, volatility, sample autocorrelation, long range dependence, non-parametric regression, Nadaraya-Watson kernel estimator, distributional forecast, heavy tails.

1. Introduction

Describing the evolution of prices by imposing a stationary model on the conditional distribution of returns is probably the most significant recent development in the econometric modeling of financial time series. This approach, that assumes the returns to be strongly stationary, expresses the conventional wisdom that models for financial returns should allow for nonlinear dynamics, seemingly demonstrated by the sample autocorrelation and cross correlation functions (SACF/SCCF) of such time series. As an illustration, consider the tri-variate sample of daily log-returns of the exchange rate Euro/Dollar (EU), the FTSE index and the 10 year US T-bond from January 2, 1990 until September 12, 2001 (the same data set is used in the sequel to exemplify our modeling approach). The SACF/SCCF of the returns (Figure 1.1) show extremely small auto- or cross-correlations at lags greater then 4 between the EU, the FTSE or the 10 year T-bond returns. In contrast to this, the SACF/SCCF of absolute returns in Figure 1.2 show larger correlations in the absolute values.

While the aspect of the displayed SACF/SCCF could be due to stationary non-linear dynamics, a precise interpretation of Figures 1.1 and 1.2 is difficult for several reasons. Since the type of dependency in the data is unknown, the significance of the auto- and cross-correlations in these figures can not be assessed (for this reason no confidence intervals for the correlations are given in the mentioned figures). Another difficulty is related to distinguishing between stationary long memory, i.e. significant correlations at large lags, and non-stationarity. An SACF/SCCF that displays positive correlations at large lags (like that in Figure 1.2), could be a sign of non-stationarities in the second moment structure of the time series as well as a proof of a stationary, non-linear long-range dependence; see Mikosch and Stărică [10].

For the multivariate case, rich classes of stationary models that generalize to more dimensions the successful univariate GARCH-type or stochastic volatility-type models and try to mimic the second moment structure illustrated by Figures 1.1 and 1.2 have been
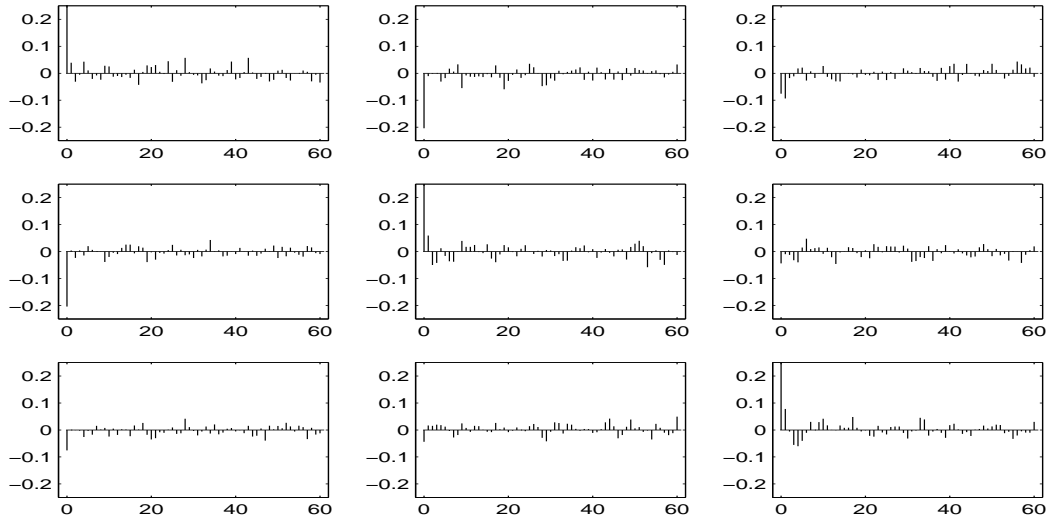
Figure 1.1. *SACF-SCCF of the data (EU returns, the first coordinate, FTSE returns, the second coordinate, the 10 year T-bond returns, the third coordinate respectively). On the diagonal the SACF of the 3 series. Off the diagonal the SCCF of pairs. Since the dependency structure in the data is unknown, no confidence intervals for the correlations are displayed.*

proposed (see the comprehensive surveys by Bollerslev et. al. [2] and Pagan [14]). However, due to their complexity, most of these models are poorly understood both from a probabilistic and a statistical point of view. In particular, they require the estimation of a large number of parameters and do not describe accurately the extremal behavior of the vector of returns (see Engle and Sheppard [6]). In this paper, we describe a simple modeling approach, based on a different interpretation of the SACF/SCCF, that addresses these shortcomings.

Our alternative to the fully parametric, stationary, conditional methodology is a semi-parametric, non-stationary, unconditional modeling approach. The methodological frame is that of non-parametric regression with non-random equidistant design points; see also Drees and Stărică [5]. More concretely, the vector of returns is assumed to follow an AR(1) process with independent innovations whose unconditional covariance matrix changes slowly
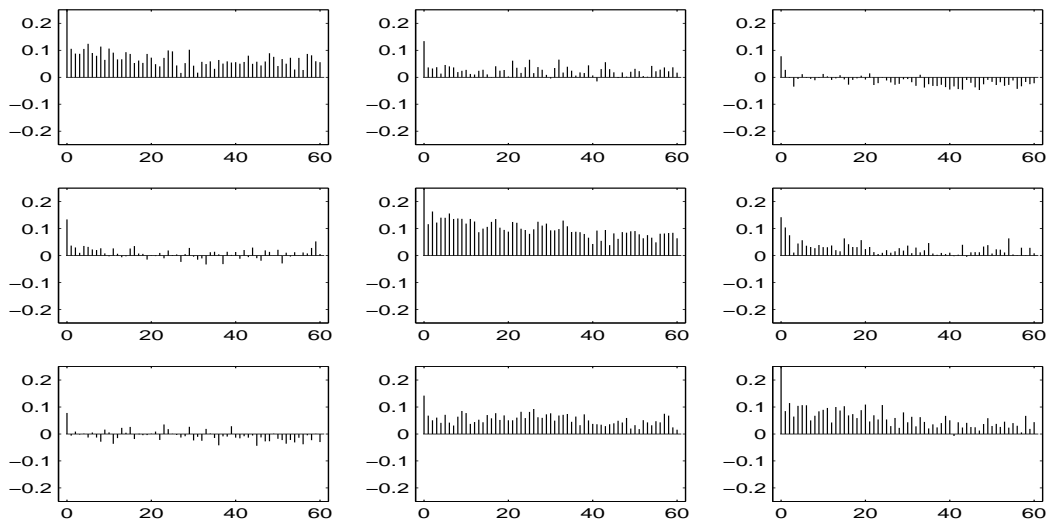
Figure 1.2. *SACF-SCCF of the absolute values of the data (EU returns, the first coordinate, FTSE returns, the second coordinate, the 10 year T-bond returns, the third coordinate respectively). On the diagonal the SACF of the 3 series. Off the diagonal the SCCF of pairs. Since the dependency structure in the data is unknown, no confidence intervals for the correlations are displayed.*

through time and is observed at regularly spaced moments in time (in the case of this paper, daily). The regression function is the changing unconditional covariance matrix and is estimated by a local weighted average. The weights are chosen by evaluating a kernel function at the design points. The vectors of standardized innovations have asymmetric heavy tails and are modeled parametrically. The careful modeling of the extremal behavior of the standardized innovations yields a model suited for precise VaR calculations and for generation of stress-testing scenarios.

The rest of the paper is organized as follows. Section 2 introduces our non-stationary, regression-type model. In Section 3, this model is estimated on a tri-variate sample of returns consisting of the foreign exchange rate Euro/Dollar (EU), the FTSE 100 index, and the 10 year US T-bond (the dimension of the multivariate vector of returns has been intentionally kept low to facilitate an in-depths statistical analysis). Section 4 evaluates the performance of our model in forecasting the distribution of multivariate returns. In Section

5 we comment on the relationship between our modelling approach and the RiskMetrics^{TM} methodology while Section 6 concludes.

2. A simple non-stationary model for multivariate returns

Denote by $\mathbf{r}_t$ the $m \times 1$-dimensional vector of returns at time t. Our discussion focuses on the following class of models with constant mean $\underline{\mu}$,

$$(2.1) \qquad (\mathbf{r}_t - \underline{\mu}) - \Phi(\mathbf{r}_{t-1} - \underline{\mu}) = \mathbf{u}_t$$

where Φ is a diagonal matrix of auto-regressive parameters and $(\mathbf{u}_t)$ is the sequence of innovations.

GARCH-type models assume that $(\mathbf{u}_t)$ is a *stationary, dependent*, white noise sequence with a certain conditional second moment structure. More specifically, the $m \times m$ conditional variance-covariance matrix $\mathbf{H}_t := E(\mathbf{u}_t \mathbf{u}_t' \mid \mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots)$ is assumed to follow a stationary stochastic process defined in terms of past $\mathbf{r}$'s and past $\mathbf{H}$'s. Often, it is assumed that $P(\mathbf{u}_t \in \cdot \mid \mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots) = P(N(0, \mathbf{H}_t) \in \cdot)$. In short, the common assumptions of the GARCH-type models imply that $(\mathbf{r}_t)$ is a strongly stationary sequence. In particular, the unconditional covariance does not change in time.

We assume $(\mathbf{u}_t)$ to be a *non-stationary* sequence of *independent* random vectors. The distribution of $\mathbf{u}_t$ is characterized by a changing covariance structure that is a manifestation of complex market conditions, hence driven by exogenous factors. Since no clear candidates for explanatory exogenous variables are at hand, we choose to model this covariance as a deterministic, smooth function of time. This approach leads to the following regression-type model:

$\mathbf{u}_t = \mathbf{S}(t)\, \underline{\varepsilon}_t, \qquad t = 1, 2, \ldots, n, \quad$ where

$\mathbf{S}(t)$ is an invertible matrix and a smooth, deterministic function of time,

(2.2) $(\underline{\varepsilon}_t)$ is an iid sequence of random vectors with mutually independent coordinates, such that $E\, \underline{\varepsilon}_t = 0$, $Var\, \underline{\varepsilon}_t = I_m$.

The elements of the sequence $(\underline{\varepsilon}_t)$ are called the *standardized innovations*. From (2.1) and (2.2), it follows that

$$E(\mathbf{u}_t \mathbf{u}_t^{'} \mid \mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots) = E(\mathbf{u}_t \mathbf{u}_t^{'}) = \mathbf{S}^2(t), \quad \text{and}$$

$$P(\mathbf{u}_t \in \cdot \mid \mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots) = P(\mathbf{u}_t \in \cdot).$$

Let us emphasize that by modeling the covariance matrix as a deterministic sequence we do not claim that random effects do not play any role in the second moment dynamics. This modeling approach only reflects the belief that the distribution of the vector of future returns incorporates a changing pool of information which is partly expressed in the recent past of the time series and the fact that we are not aware of exogenous variables capable of reliably explaining the dynamics of the volatility.

The preliminary step when fitting this regression-type model to a multivariate time series of returns is obtaining the estimated innovations

$$\hat{\mathbf{u}}_t := (\mathbf{r}_t - \bar{\mathbf{r}}) - \hat{\Phi}(\mathbf{r}_{t-1} - \bar{\mathbf{r}}), \quad t = 1, \ldots, n$$

with $\bar{r}_i := n^{-1} \sum_{t=1}^{n} r_{i,t}$, $i = 1, \ldots, m$, being the natural estimator for the mean $\underline{\mu}$ and $\hat{\Phi}$ the $m \times m$ diagonal matrix whose non-zero entries are the auto-regression coefficients estimated coordinate-wise. Neglecting the estimation error in this first step, the estimated innovations $\hat{\mathbf{u}}_t$ are supposed to be independent with covariance matrix $\mathbf{S}^2(t)$, a smooth function of t. The function $\mathbf{S}^2(t)$ can be estimated by standard non-parametric regression methods for non-random, equidistant design points t using the series $\hat{\mathbf{u}}_t \hat{\mathbf{u}}_t^{'}$, $t = 1, \ldots, n$. In this paper we use the evaluation weighted (Nadaraya-Watson) estimator of $\mathbf{S}^2(t)$ given by

$$(2.3) \qquad \hat{\mathbf{S}}^2(t) := \frac{n^{-1} \sum_{i=1}^{n} K_h(t - i) \hat{\mathbf{u}}_{t-i} \hat{\mathbf{u}}_{t-i}^{'}}{n^{-1} \sum_{i=1}^{n} K_h(t - i)}, \qquad t = 1, 2, \ldots, n,$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$ and K a kernel. In the analysis of the multivariate return time series in Section 3, a Gaussian kernel is used. We note that, according to our experience, an exponential kernel or the LOESS procedure produce very close results. This is in accordance with the established fact that for the equidistant design set-up, the shape of

8

the kernel function makes little difference; see the monographes by Müller [12] and Härdle [7].

The final step is modeling the distribution of the *estimated standardized innovations* defined as

$$(2.4) \qquad \hat{\underline{\varepsilon}}_t := \hat{\mathbf{S}}^{-1}(t)\, \hat{\mathbf{u}}_t, \qquad t = 1, 2, \ldots, n$$

with $\hat{\mathbf{S}}(t)$, the square root of the estimate of $\mathbf{S}^2(t)$ in (2.3). One possibility is to use the empirical cumulative distribution function (cdf) of $\hat{\underline{\varepsilon}}$ as a model for the standardized innovations as done in Barone-Adesi et al. [1]. However, since the estimated standardized innovations are usually heavy tailed (see Section 3.3 for evidence supporting this claim), the use of the empirical cdf will underestimate the probability of extreme standardized innovations and, hence, the risk of extreme returns, with potentially serious consequences for risk managing.

Since we assume the estimated standardized innovations to have independent coordinates, it is sufficient to specify the distributions of $\hat{\varepsilon}_i$, $i = 1, \ldots, m$. A flexible and parsimonious family of distributions that allow for asymmetry between the distributions of positive and negative standardized innovations and, in addition, for arbitrary tail indices can be defined starting from the Pearson type VII distribution with shape parameter m and scale parameter c; see Drees and Stărică [5]. The density of this distribution is

$$(2.5) \qquad f(x; m, c) = \frac{2\Gamma(m)}{c\Gamma(m - 1/2)\pi^{1/2}}\left(1 + \left(\frac{x}{c}\right)^2\right)^{-m}, \qquad x > 0.$$

Note that f is the density of a t-distributed random variable with $\nu = 2m - 1$ degrees of freedom multiplied by the scale parameter $c\nu^{-1/2}$. This family was also used to model the distribution of financial returns in an univariate stochastic volatility framework by Nagahara and Kitagawa [13].

According to our experience, this distribution (concentrated on the positive axis) fits well the positive standardized innovations and the absolute value of the negative ones. Because usually there are about as many positive standardized innovations as there are negative ones, it may be assumed that the cdf of the standardized innovations has median

0. Hence, denoting the densities of the negative and positive standardized innovations by $f_-(\,\cdot\,;m_-,c_-)$ and $f_+(\,\cdot\,;m_+,c_+)$, respectively, the density of the distribution of the coordinates of the standardized innovations is

$$(2.6) \quad f^{VII}(x;m_-,c_-,m_+,c_+) = \frac{1}{2}\Big(f_-(x;m_-,c_-)1_{(-\infty,0)}(x) + f_+(x;m_+,c_+)1_{[0,\infty)}(x)\Big).$$

We refer to the distribution with density (2.6) (that covers the whole real axis) as the asymmetric Pearson type VII and denote its cdf by F^{VII}.

To summarize, for a given coordinate, 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 covariance estimates $\hat{\mathbf{S}}(t)$, the average return $\bar{\mathbf{r}}$, and the diagonal matrix $\hat{\Phi}$, fully specify the distribution of the time series of returns in the model (2.1), (2.2).

3. Modeling multivariate returns. An example.

In this section, we apply the methods described in the previous section to the 2927 observations (from January 2, 1990 until September 12, 2001) of the time series of daily returns of three qualitatively different financial instruments: one foreign exchange rate, the Euro/Dollar (EU), an index, the FTSE 100 and an interest rate, the 10 year US T-bond. The EU and the US T-bond series are available on the site of the US Federal Reserve Board: *http://www.federalreserve.gov/releases/*. To facilitate a graphical display of the empirical analysis, we conduct our study in a trivariate setup. Note that similar modeling results have been achieved with higher dimensional vectors of returns.

In a preliminary step, the sample mean was removed and the series were pre-filtered coordinate-wise with an AR(1) filter.

3.1. **The evolution of the unconditional covariance structure.** The graphs in Figure 3.1 display two estimates of the time-varying standard deviations (sd's) of the three time series. Those in Figure 3.2 show two estimates of the time-dependent correlation between the three pairs of univariate time series (in the top graph, EU and FTSE, in the middle
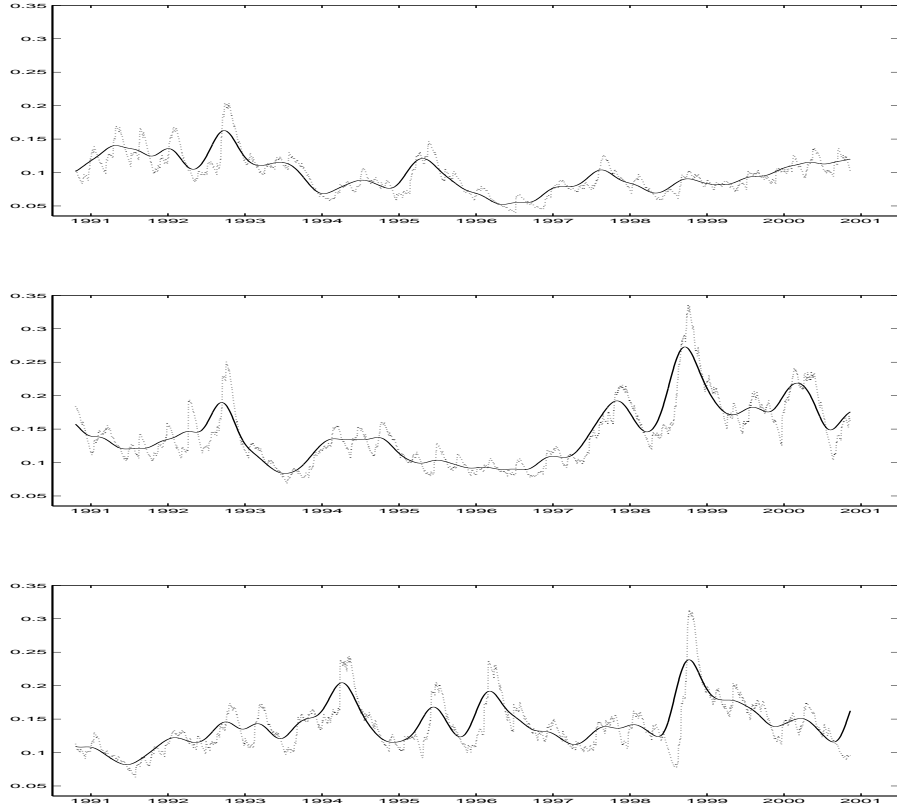
Figure 3.1. *Local estimates ($\hat{\mathbf{S}}^2$, $h = 35$, solid line, $\hat{\mathbf{S}}_1^2$, $h = 25$, dotted line) of the (annualized) standard deviation (sd) of the data: EU, (Top), FTSE (Middle) and the 10-year T-bond (Bottom). The annualized sd is obtained by multiplying the daily sd by a factor of $\sqrt{250}$.*

FTSE and T-bill, in the lower one, EU and T-bill). In all the pictures, the solid line is the estimate obtained using $\hat{\mathbf{S}}^2$, defined in (2.3) with band-width $h = 35$ while the dotted line is obtained using the estimate

$$(3.1) \qquad \hat{\mathbf{S}}_1^2(t) := \frac{n^{-1} \sum_{0 \le i \le t} K_h(t - i) \hat{\mathbf{u}}_{t-i} \hat{\mathbf{u}}_{t-i}'}{n^{-1} \sum_{0 \le i \le t} K_h(t - i)}$$

with band-width $h = 25$. Note that $\hat{\mathbf{S}}_1^2(t)$ uses only the information available at day t. This estimator will be used to produce the forecasting results presented in Section 4.
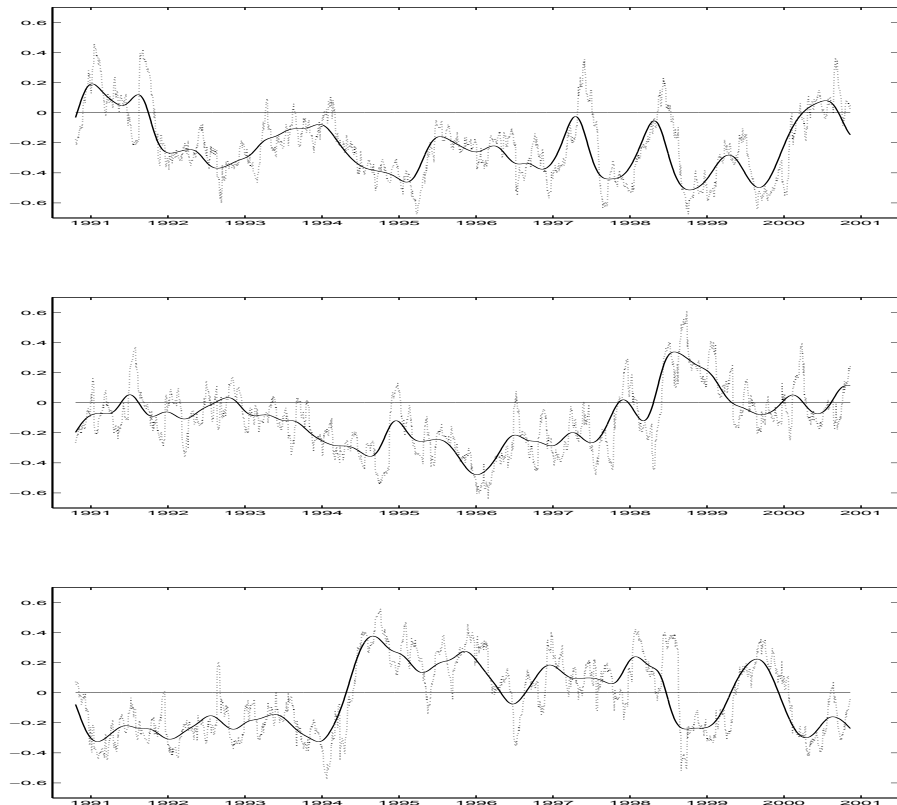
Figure 3.2. *Local estimates ($\hat{\mathbf{S}}^2$, $h = 35$, solid line, $\hat{\mathbf{S}}_1^2$, $h = 25$, dotted line) of the correlations between the data: EU and FTSE, (Top), FTSE and 10-year T-bond (Middle) and 10-year T-bond and EU (Bottom).*

Remark 3.3. A possible explanation for the the wiggly appearance of the dotted lines in Figures 3.1, 3.2 could be the boundary effect that affects the estimator $\hat{\mathbf{S}}_1^2(t)$. Due to the low ratio of signal to noise in the problem at hand, it is not clear that taming this boundary effect would improve the outcome of the forecasting procedure (see Section 4 for a qualitative evaluation). Since the best of such adjustments tend to be rather complicated, evaluating the trade-off between the gains in forecasting accuracy and the increase in the complexity will be considered elsewhere.

The graphs in Figure 3.1 and 3.2 show rather large variations in the estimated standard deviations as well as in the estimated correlation structure of the series. In particular,

the estimated correlation between the EU and the T-bond switched from negative values in the interval (-0.3, -0,2) in the beginning of the 90's, to positive ones around 0.2 in the beginning of the second half of the decade. The largest fluctuations in the estimated sd are displayed by the FTSE with increases from values around 10% in the middle of the decade to a peak of roughly 25% towards the end of the 90's.

3.2. **The dependency structure of the standardized innovations.** In this section we analyze the dependency structure of the estimated standardized innovations $\hat{\underline{\varepsilon}}_t$ defined in (2.4).

A battery of three tests is used to achieve this goal. We begin by verifying that the marginal distributions of the coordinates of the estimated standardized innovations $\hat{\varepsilon}_i$, $i = 1, 2, 3$, do not change through time. Towards this goal, for a given coordinate i, we split the sample $(\hat{\varepsilon}_{i,t})$ in three subsamples of equal length, $(\hat{\varepsilon}_{i,t}^{(1)})$, $(\hat{\varepsilon}_{i,t}^{(2)})$, $(\hat{\varepsilon}_{i,t}^{(3)})$ respectively. Then, we perform a pairwise comparison of the three resulting empirical cdf's using a 2-sample Kolmogorov-Smirnov test, producing three p-values.

For the pair $(\hat{\varepsilon}_i^{(1)}, \hat{\varepsilon}_i^{(2)})$, the working assumptions are that $\hat{\varepsilon}_i^{(1)}$'s and $\hat{\varepsilon}_i^{(2)}$'s are mutually independent (see the independence tests (3.3) in the sequel for evidence supporting this assumption) and that all the observations in the sample $(\hat{\varepsilon}_i^{(1)})$ come from the same continuous population $F_i^{(1)}$, while all the observations in the sample $(\hat{\varepsilon}_i^{(2)})$ come from the same continuous population $F_i^{(2)}$. The null hypothesis is

$$(3.2) \qquad\qquad H_0: \quad F_i^{(1)} \text{ and } F_i^{(2)} \text{ are identical.}$$

Table 3.4 reports the nine p-values (3 for each coordinate) for the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ (left) together with the nine values corresponding to their absolute values $(|\hat{\underline{\varepsilon}}_t|)$ (right).

Table 3.4 supports the hypothesis of stationarity of the coordinates of the sequence of estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$.

In the sequel we use the covariance/autocovariance structure of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ and their absolute values $(|\hat{\underline{\varepsilon}}_t|)$ (see Figures 3.5 and 3.6) to test the
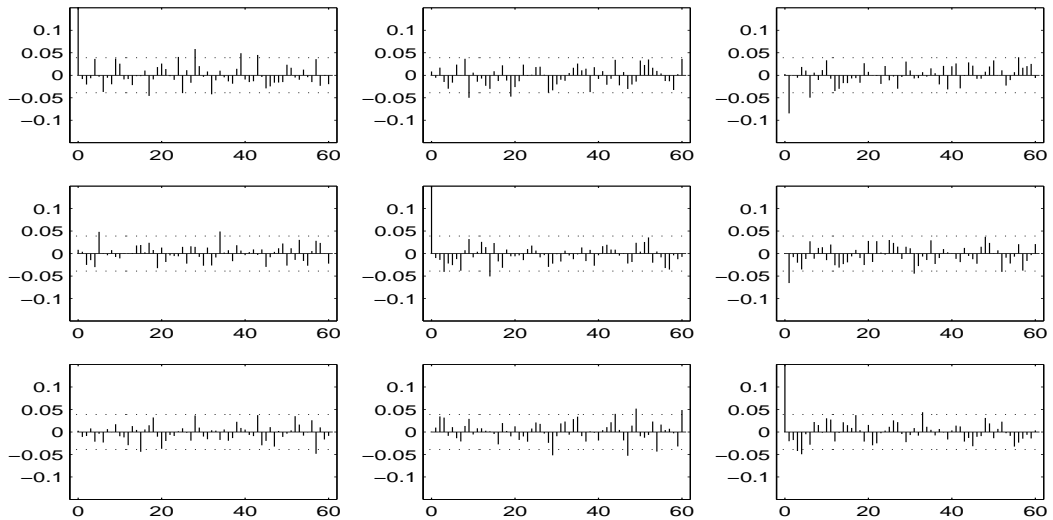
Figure 3.5. *SACF/SCCF of the estimated standardized innovations ($\hat{\underline{\varepsilon}}_t$). The covariance structure was estimated using $\hat{\mathbf{S}}^2$.*

hypothesis

$$(3.3) \qquad H_0 : \hat{\underline{\varepsilon}}_t \text{ are iid vectors with independent coordinates.}$$

The confidence intervals in Figures 3.5 and 3.6 correspond to the null hypothesis (3.3). These figures show that accounting for the changing covariance produces standardized innovations that are practically uncorrelated, removing the long memory look of the SACF of absolute returns in Figure 1.2. They support the choice of modeling the standardized innovations as a sequence of iid vectors with independent coordinates.

	1 and 2	1 and 3	2 and 3		1 and 2	1 and 3	2 and 3
$\hat{\varepsilon}_1$	0.93	0.23	0.05	$|\hat{\varepsilon}_1|$	0.89	0.39	0.21
$\hat{\varepsilon}_2$	0.10	0.46	0.59	$|\hat{\varepsilon}_2|$	0.20	0.43	0.75
$\hat{\varepsilon}_3$	0.23	0.63	0.65	$|\hat{\varepsilon}_3|$	0.09	0.48	0.64

Table 3.4. *The p-values corresponding to the 2-sample Kolmogorov-Smirnov tests on subsamples of estimated standardized innovations ($\hat{\underline{\varepsilon}}_t$) (left) and their absolute values ($|\hat{\underline{\varepsilon}}_t|$) (right). The column labels code the pairs of subsamples.*
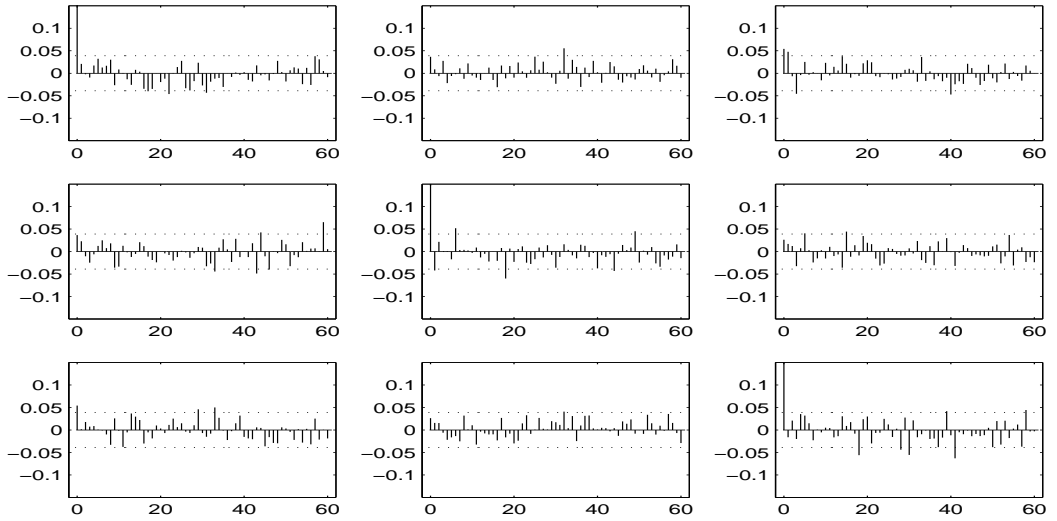
14

Figure 3.6. *SACF/SCCF of the absolute values of the estimated standardized innovations* $(|\hat{\underline{\varepsilon}}_t|)$. *The covariance structure was estimated using* $\hat{\mathbf{S}}^2$.

The visual test of the hypothesis (3.3) is complemented by a Ljung-Box test for the first 25 lags. Table 3.7 gives the p-values for the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ (the left half) and their absolute values (the right half). The value reported at the intersection of the i-th line with the j-th column is the p-value of the Ljung-Box statistic obtained by summing the first 25 values of the SCCF between the coordinate i and past lags of the coordinate j. Besides the pair (1,3), all other p-values do not reject the hypothesis (3.3) at usual statistical significance levels.

	1	2	3	1	2	3
1	0.15	0.17	0.03	0.20	0.89	0.12
2	0.81	0.25	0.11	0.62	0.21	0.16
3	0.70	0.88	0.15	0.25	0.50	0.22

Table 3.7. *The p-values for the Ljung-Box test at lag 25 of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ (left) and their absolute values $(|\hat{\underline{\varepsilon}}_t|)$ (right). The row and column numbers represent the coordinates.*

Finally, the hypothesis that the coordinates of the estimated standardized innovations, $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$, $\hat{\varepsilon}_3$ are pair-wise independent is tested using Kendall's τ distribution-free statistic. Kendall's τ takes values between -1 and 1 (independent variables have τ=0) and provides an alternative measure of dependence between two variables to the usual correlation. While the easy-to-compute correlation is the natural scalar measure of linear dependence, Kendall's τ is a valuable measure of dependency also in the case of non-normality and non-linearity. In large samples, as the sample size n goes to ∞,

$$3\tau\sqrt{\frac{n(n-1)}{2(2n+5)}} \xrightarrow{d} N(0,1).$$

Therefore Kendall's τ can be used as a test statistic for testing the null hypothesis of independent variables. (For more details on Kendall's τ we refer to Kendall and Stuart [8].)

The test is applied to all pairs of coordinates $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ $(i,j=1,2,3, i<j)$ and all pairs of their absolute values. For a given pair (i,j), the assumption is that the n bivariate observations $(\hat{\varepsilon}_{i,t}, \hat{\varepsilon}_{j,t})$, $t=1,\ldots,n$, are mutually independent (see the independence tests (3.3) for evidence supporting this assumption) and come from the same continuous bivariate population. The null hypothesis is

(3.4) $H_0:$ the random variables $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_j$ are independent.

The resulting p-values are given in Table 3.8. For all pairs the hypothesis of independence (3.4) is not rejected at usual statistical levels of significance..

	(1,2)	(1,3)	(2,3)	(1,2)	(1,3)	(2,3)
Kendall	0.34	0.60	0.89	0.98	0.99	0.71

Table 3.8. *The p-values for the Kendall's τ distribution-free test of independence applied to the estimated standardized innovations sequence $(\underline{\hat{\varepsilon}}_t)$ (left) and to the absolute values $(|\underline{\hat{\varepsilon}}_t|)$ (right). The pairs on the top are pairs of coordinates.*

At this point, we conclude that, based on the battery of test described above, the hypothesis that the estimated standardized innovations $(\underline{\hat{\varepsilon}}_t)$ is a stationary sequence of iid

vectors with independent coordinates can not be rejected. Besides providing a goodness of fit check for our model, this fact allows for a careful statistical analysis of the series of estimated standardized innovations.

Remark 3.9. Our approach assumes the standardized innovations to be iid and the coordinates of the standardized innovations independent. According to our experience with other multivariate data sets, significant autocorrelations at lags greater then 0 or significant cross-correlations are occasionally encountered in the SACF/SCCF of the standardized innovations. However, they are usually barely outside the confidence bands. An accurate modeling of these correlations would imply a significant increase in the complexity of the model. The trade-off between the gains in modeling accuracy and the increase in the complexity is subject of further research.

3.3. **The multivariate distribution of the standardized innovations.** In this section we concentrate on modeling the marginal distribution of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$. We begin by presenting some evidence that supports our claim that the marginal distributions of the three coordinate series $(\hat{\varepsilon}_i)$, $i = 1, 2, 3$, are heavy tailed. Figure 3.10 displays the standard normal plots of three coordinate series of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$. The graphs seem to show departures from normality for at least two of the three coordinates (the first and the third) with the right tail apparently heavier than the left one.
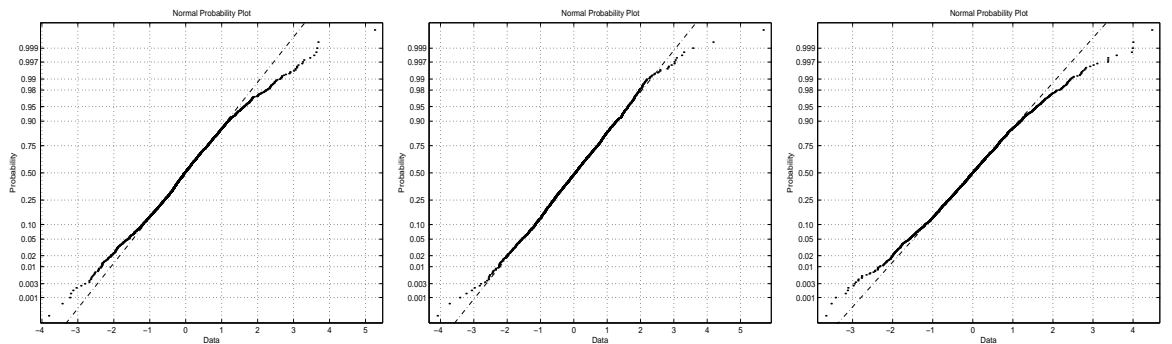

Figure 3.10. *Normal probability plots of the three series of coordinates the of estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$.*

The impression given by Figure 3.10 is confirmed by the p-values of the Kolmogorov-Smirnov (K-S) and Andersen-Darling (A-D) tests (for details on these tests see [16]) applied to the coordinate series $(\hat{\varepsilon}_i)$, $i = 1, 2, 3$ reported in the left half of Table 3.11. The working assumptions for each coordinate i are that $\hat{\varepsilon}_i$'s are mutually independent (see the independence test (3.3) for evidence supporting this assumption) and that all the observations in the sample $(\hat{\varepsilon}_i)$ come from the same continuous population F_i (the results of test of stationarity hypothesis (3.2) legitimates this assumption). The null hypothesis is

(3.5) $\qquad\qquad\qquad H_0 : \quad F_i$ is the standard normal distribution.

(See also the tests of the hypothesis (3.6) $Var(\hat{\underline{\varepsilon}}) = I_m$.)

The K-S and A-D tests are chosen for their complementary nature. It is well known that the Kolmogorov-Smirnov test is sensitive to departures from the hypothesized law affecting the middle of the distribution while the Andersen-Darling test has been proved to be effective in identifying departures that affect the tails. The normality assumption is rejected at the 5% level by the A-D test for all three coordinates, while the K-S rejects it for the first and third coordinate.

H_0: Normal	A-D	K-S	H_0: Pearson VII	A-D	K-S
$\hat{\varepsilon}_1$	0.007	0.038	$\hat{\varepsilon}_1$	0.21	0.20
$\hat{\varepsilon}_2$	0.037	0.108	$\hat{\varepsilon}_2$	0.17	0.48
$\hat{\varepsilon}_3$	0	0.003	$\hat{\varepsilon}_3$	0.10	0.25

Table 3.11. *The p-values for the Andersen-Darling and Kolmogorov-Smirnov tests of normality (left) and of asymmetric VII Pearson (right) applied to the 3 coordinate series of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$.*

Figure 3.10 and the values on the left side of Table 3.11 show that the estimated standardized innovations have tails that are heavier than normal tails.

We continue with the parametric modeling of the marginals of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ as asymmetric Pearson type VII heavy tailed distributions. Table

3.12 contains the estimated parameters obtained by fitting an asymmetric Pearson VII distribution (2.6) to the three coordinates of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$.

	m_-	c_-	m_+	c_+	Left tail	Right tail
$\hat{\varepsilon}_1$	5.94 (1.48)	2.92 (0.47)	3.88 (0.60)	2.24 (0.25)	10.87	6.75
$\hat{\varepsilon}_2$	9.24 (3.71)	3.87 (0.91)	9.84 (4.22)	4.14 (1.03)	17.48	18.67
$\hat{\varepsilon}_3$	6.62 (1.86)	3.16 (0.55)	4.30 (0.75)	2.40 (0.29)	12.23	7.59

Table 3.12. *The parameters of the asymmetric Pearson distribution corresponding to the 3 series of estimated standardized innovations $\hat{\underline{\varepsilon}}_t$ (the standard deviations are provided in parentheses). The tail indices are given by $\nu = 2m - 1$.*

The estimated parameters in Table 3.12 confirm the results of the previous tail analysis: the first and the third coordinates have heavier tails then the second, with the right tail being heavier then the left one. To test the hypothesis

$$(3.6) \qquad\qquad H_0: \quad Var(\hat{\underline{\varepsilon}}_t) = I_m,$$

two estimates of the variances of the coordinates of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ together with the corresponding standard deviations are produced. The working assumption is that $(\hat{\underline{\varepsilon}}_t)$ is an iid sequence of random vectors (see the independence tests (3.3) for evidence supporting this assumption). The first estimate is the sample variance with the standard deviation given by $\sqrt{m_i^4/n}$, $i = 1, 2, 3$, where m_i^4 is the sample fourth moment of $(\hat{\varepsilon}_i)$. The second estimate is the variance of the estimated asymmetric Pearson type VII given by (2.6). Since the variance of any coordinate is a function of the parameters reported in Table 3.12, the standard deviation for this variance estimate is obtained from the covariance matrix of the MLE estimates using the delta method. The three pairs of point estimates together with the standard deviations are reported on the left half of Table 3.13. The right half of the same table reports the sample covariance together with the corresponding standard deviation. According to the values in Table 3.13 the hypothesis that $Var(\hat{\underline{\varepsilon}}_t) = I_m$ is not rejected at the 5% significance level.

	Empirical	Pearson VII		Covariance
$\hat{\varepsilon}_1$	0.971 (0.043)	1.007 (0.10)	$\hat{\varepsilon}_1, \hat{\varepsilon}_2$	0.0080 (0.020)
$\hat{\varepsilon}_2$	0.957 (0.037)	0.997 (0.17)	$\hat{\varepsilon}_1, \hat{\varepsilon}_3$	0.0026 (0.020)
$\hat{\varepsilon}_3$	0.944 (0.041)	1.002 (0.10)	$\hat{\varepsilon}_2, \hat{\varepsilon}_3$	0.0004 (0.019)

Table 3.13. *The estimated variances of the coordinates of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$. The first column reports the sample variance while the second one is the variance of the estimated asymmetric Pearson type VII. The last column reports the sample covariance. The standard deviations are reported in parentheses.*

To verify the goodness of fit of the asymmetric Pearson type VII distribution, the probability plot of the coordinates of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$ using the estimated asymmetric Pearson VII distributions are displayed in Figure 3.14. A good fit of the asymmetric Pearson VII distributions should translate in linear graphs close to the first diagonal. For the coordinate i, the assumptions are that $\hat{\varepsilon}_i$'s are mutually independent (see the test of the hypothesis (3.3) for evidence supporting it) and that all the observations in the sample $(\hat{\varepsilon}_i)$ come from the same continuous population F_i (the test of (3.2) supports this assumption). The null hypothesis is

(3.7) $H_0:$ F_i is the asymmetric Pearson VII distribution with parameters given in
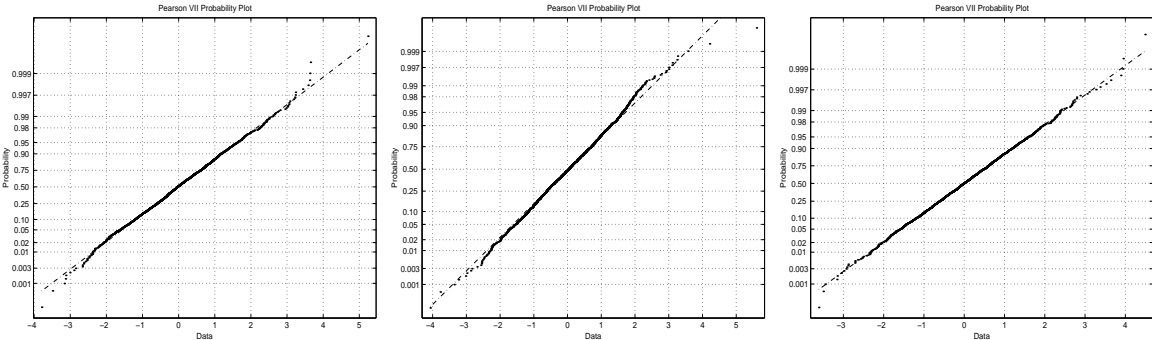
Table 3.11.

Figure 3.14. *The asymmetric Pearson VII probability plots of the three coordinate series of the estimated standardized innovations $(\hat{\varepsilon}_{i,t})$, $i = 1, 2, 3$.*

The straight plots in Figure 3.14 are a confirmation of the good fit of the asymmetric Pearson VII distribution.

The hypothesis (3.7) is formally tested using the Kolmogorov-Smirnov and Andersen-Darling tests. The p-values of these tests are reported on the right in Table 3.11. The hypothesis is not rejected at usual levels of significance.

The plots in Figure 3.14 and the results in Table 3.11 demonstrate that the parametric family described by (2.6) is indeed an appropriate model for the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$.

This concludes the evaluation of the goodness of fit of the model (2.1), (2.2). The statistical analysis has shown that it provides an overall good description of the multivariate data set considered. We now direct our attention towards evaluating the forecasting performance of our model.

4. FORECASTING MULTIVARIATE RETURNS

In this section we discuss aspects related to forecasting multivariate returns using the regression-type model (2.1), (2.2). We begin by specifying the 1-day ahead forecasting methodology. Then we check the quality of our multivariate distributional forecasts. We end the section with a comparison (in the univariate framework) between the forecasting behavior of the industry standard Riskmetrics^{TM} and that of our methodology on randomly generated portfolios containing the three instruments EU, the FTSE, and the US T-bond.

4.1. **The 1-day ahead forecasting methodology.** Denote by $\bar{\mathbf{r}}_t$, the sample mean and by $\hat{\Phi}_t$, the diagonal matrix whose non-zero entries are the coefficients of the component-wise AR(1) regressions, both estimated on the sample up to day t. Let

$$\tilde{\mathbf{u}}_t := (\mathbf{r}_t - \bar{\mathbf{r}}_t) - \hat{\Phi}_t(\mathbf{r}_{t-1} - \bar{\mathbf{r}}_{t-1}), \quad t = 1, \ldots, n$$

be the innovations estimated using only past information.

Given $\hat{\mathbf{S}}_1^2(\cdot)$, an estimate of the unconditional covariance matrix $\mathbf{S}^2(\cdot)$ based only on past information, denote by $\hat{F}_{i,t}^{VII}$, $i = 1, 2, 3$, the asymmetric Pearson type VII distributions

(2.6) with parameters estimated on the coordinates of the series $(\hat{\mathbf{S}}_1^{-1}(1)\,\tilde{\mathbf{u}}_1,\ \hat{\mathbf{S}}_1^{-1}(2)\,\tilde{\mathbf{u}}_2,\dots,$ $\hat{\mathbf{S}}_1^{-1}(t)\,\tilde{\mathbf{u}}_t)$.

Based on the model (2.1), (2.2), the distributional forecast at time t for $\mathbf{r}_{t+1}$ is then given by

$$\mathbf{r}_{t+1}^f = \bar{\mathbf{r}}_t + \hat{\Phi}_t(\mathbf{r}_t - \bar{\mathbf{r}}_t) + \mathbf{u}_{t+1}^f, \quad \text{where} \quad \mathbf{u}_{t+1}^f = \hat{\mathbf{S}}_1(t)\,\underline{\epsilon}_t,$$

(4.1) $\underline{\epsilon}_t$ are independent m-dimensional random vectors,

$\epsilon_{i,t}$ are mutually independent with distributions $\hat{F}_{i,t}^{VII}$, $(i = 1, 2, 3)$.

For our forecasting exercise we use

$$\hat{\mathbf{S}}_1^2(t) = \frac{n^{-1}\sum_{0 \le i \le t} K_h(t-i)\tilde{\mathbf{u}}_{t-i}'\tilde{\mathbf{u}}_{t-i}}{n^{-1}\sum_{0 \le i \le t} K_h(t-i)},$$

the one-sided kernel estimate of the unconditional covariance matrix $\mathbf{S}_t^2$ defined in (3.1). See also Remark 3.3.

4.2. **Multivariate density forecast evaluation.** Evaluating the multivariate distributional forecast (see Diebold et al. [4]) is particularly simple in the case of the model (2.1), (2.2) due to the assumption of independence of the sequence $(\mathbf{u}_t)$. Verifying that the distribution of $\mathbf{r}_{t+1}^f$ defined by (4.1) coincides with that of $\mathbf{r}_{t+1}$ is equivalent to checking that the m-dimensional vectors $(\mathbf{z}_t)$

(4.2) $$z_{i,\,t} := F_{i,t}^{VII}(v_{i,t}), \quad i = 1, 2, 3, \quad \text{where} \quad \mathbf{v}_t = \mathbf{S}_1(t)^{-1}\tilde{\mathbf{u}}_t,$$

are iid, with independent, uniformly (0,1) distributed coordinates.

For evaluating the forecasting performance the sample is split in two: the first 1000 observations are used to produce the initial parameter estimates while the remaining 1926 observations are used to check the goodness of fit of the distribution forecast. The band width used in (3.1) is $h = 25$. The mean, the vector of the AR(1) coefficients and the parameters of the distribution (2.6) are re-estimated every 100 days on a window that starts with the beginning of the sample.
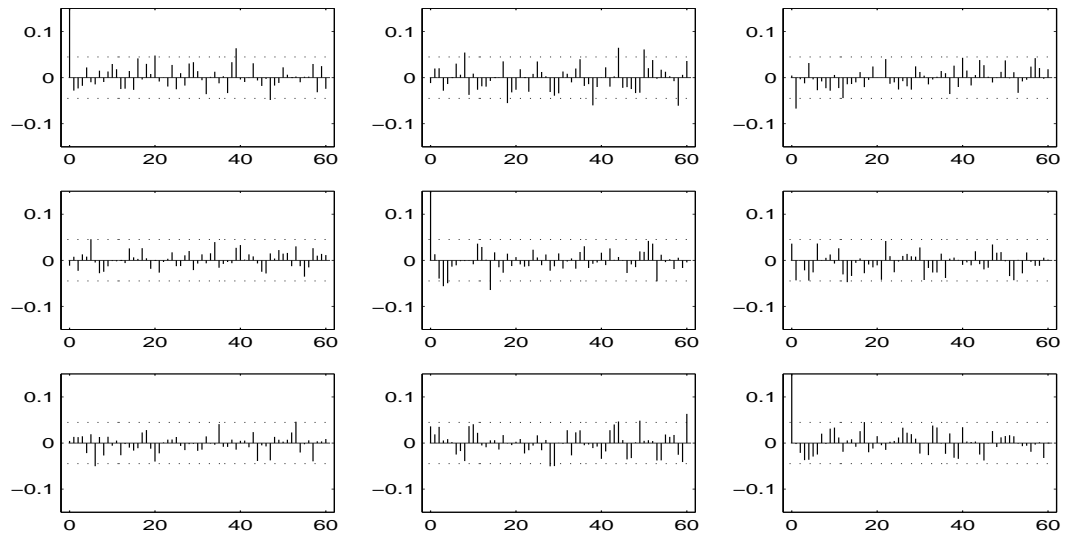
Figure 4.1. *SACF of the sequence* $(\mathbf{z}_t - \bar{\mathbf{z}})$.
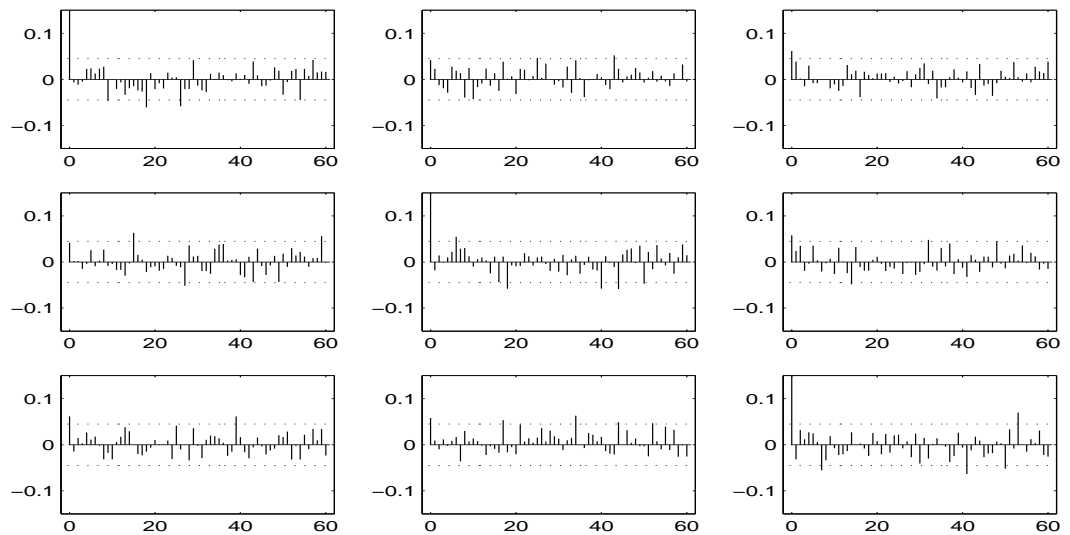

Figure 4.2. *SACF of the absolute values of the sequence* $(\mathbf{z}_t - \bar{\mathbf{z}})$.

A battery of tests similar to the one in Section 3 is employed to verify the hypothesis of iid-ness of the sequence $(\mathbf{z}_t)$ and those of uniformity and mutual independence of the coordinate sequences $(z_{i,t})$, $i = 1, 2, 3$. The precise working assumptions are those of the corresponding tests in Section 3. Figures 4.1 and 4.2 display the SACF/SCCF of the sequence $(\mathbf{z}_t - \bar{\mathbf{z}})$ and that of its absolute values ($\bar{\mathbf{z}}$ is the sample mean). Overall, they seem

to support the hypothesis of iid vectors with independent coordinates for the sequence $(\mathbf{z}_t)$, although small violations of the confidence intervals are observed in the absolute values at the first lag of pairs (1,3) and (2,3).

The visual test of the SACF/SCCF is complemented by the Ljung-Box test for the first 25 lags (the p-values are reported in Table 4.3). The value at the intersection of the row i with column j corresponds to the p-value of the Ljung-Box statistic associated with the SACF/SCCF of the coordinate i and past lags of the coordinate j. The p-values confirm the validity of the assumption of iid vectors with independent coordinates for the sequence $(\mathbf{z}_t)$.

	1	2	3	1	2	3
1	0.53	0.20	0.38	0.50	0.34	0.62
2	0.96	0.20	0.09	0.71	0.51	0.32
3	0.80	0.75	0.67	0.52	0.47	0.49

Table 4.3. *The p-values for the Ljung-Box test at lag 25 of the sequence $(\mathbf{z}_t - \bar{\mathbf{z}})$ (left) and the absolute values $(|\mathbf{z}_t - \bar{\mathbf{z}}|)$(right). The row and column numbers represent the coordinates.*

The hypothesis of pair-wise, mutual independence of the coordinates of the vector $\mathbf{z}$ is tested using the already familiar distribution-free test of Kendall's τ. The p values corresponding to the pairs of coordinates are given in Table 4.4. For all pairs the hypothesis of independent coordinates is not rejected at usual levels of statistical significance.

	(1,2)	(1,3)	(2,3)	(1,2)	(1,3)	(2,3)
Kendall	0.31	0.55	0.95	0.97	0.99	0.98

Table 4.4. *The p values for Kendall's τ distribution-free test of independence applied to the sequence $(\mathbf{z}_t - \bar{\mathbf{z}})$ (left) and to that of absolute values $(|\mathbf{z}_t - \bar{\mathbf{z}}|)$ (right).*

Figure 4.5 displays the uniform probability plots for the three coordinates z_i, $i = 1, 2, 3$. The straight plots in this figure together with the p-values of the Andersen-Darling and Kolmogorov-Smirnov tests of uniformity given in Table 4.6 support the conclusion that the marginal distributions of the three sequences $(z_{i,t})$, $i = 1, 2, 3$ are uniform (0,1).
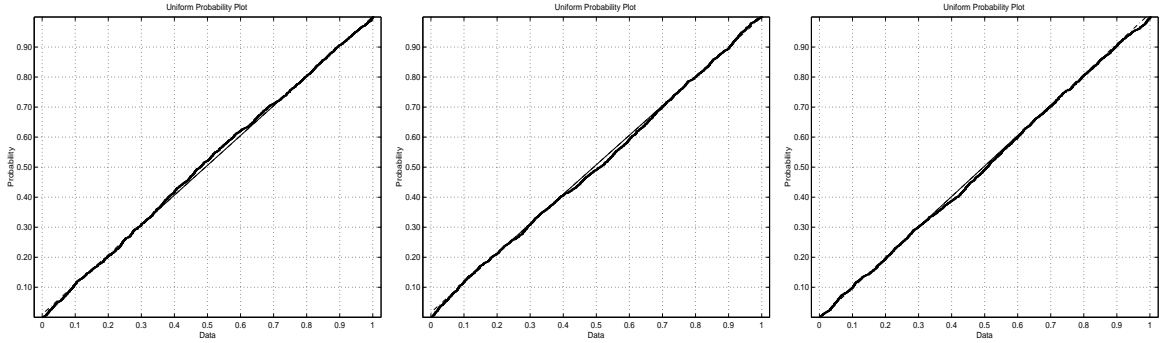
Figure 4.5. *The uniform probability plot of the three series of coordinates z_i, $i = 1, 2, 3$, $t = 1, \ldots, 1926$.*

	1	2	3		1	2	3
A-D	0.14	0.19	0.77	K-S	0.10	0.24	0.58

Table 4.6. *p-values for the Andersen-Darling and Kolmogorov-Smirnov tests of uniformity applied to the coordinates of the sequence $(\mathbf{z}_t)$.*

4.3. Univariate density forecast evaluation. We conclude this section with a distributional forecast comparison in a univariate framework. The comparison is done between the industry standard Riskmetrics^{TM} and the approach described in Section 4.1. Both methodologies are used to produce daily distributional forecasts for the returns of randomly generated portfolios containing the (by now familiar) three financial instruments. More specifically, for a given day t, the two approaches are first used to produce two multivariate (conditional) distributional forecasts for the next day vector of returns. For Riskmetrics^{TM}, the distributional forecast is

$$(4.3) \qquad \mathbf{r}_{t+1}^{f,RM} \overset{d}{=} N(0, \hat{\Sigma}_t^2),$$

where

$$(4.4) \qquad \hat{\Sigma}_t^2 := \sum_{i=0}^{d-1} \lambda^{t-i} \mathbf{r}_{t-i} \mathbf{r}_{t-i}' \Big/ \sum_{i=0}^{d-1} \lambda^{t-i},$$

is the exponential moving average estimate of the conditional covariance matrix Σ_t^2. The parameters are $\lambda = 0.94$ and $d = 74$ as stipulated in the Riskmetrics^{TM} documents [15]. For the regression-type model (2.1),(2.2), the $t+1$ forecast $\mathbf{r}_{t+1}^{f,REG}$ is given by (4.1).

The return at time $t+1$ of a given portfolio $\mathbf{w}$ with weights $\mathbf{w} = (w_1, w_2, w_3)$, is denoted by $r_{t+1}^{(\mathbf{w})}$. The distribution of $r_{t+1}^{(\mathbf{w})}$ forecasted by the Riskmetrics^{TM} methodology, which we denote by F_{t+1}^{RM}, is the distribution of $\mathbf{w}\mathbf{r}_{t+1}^{f,RM}$ (a normal with mean 0 and variance $\mathbf{w}\hat{\Sigma}_t^2\mathbf{w}'$). The distribution forecasted by the regression-type model (2.1), (2.2), denoted by F_{t+1}^{REG}, is that of $\mathbf{w}\mathbf{r}_{t+1}^{f,REG}$.

As explained in Diebold et al. [3], evaluating the correct distributional forecast F_{t+1} at the realized portfolio returns $r_{t+1}^{(\mathbf{w})}$ yields an iid sequence $(F_{t+1}(r_{t+1}^{(\mathbf{w})}))$ of uniform (0,1) random variables. Hence the quality of a distributional forecast G_{t+1} can then be assessed by testing the hypothesis

(4.5) H_0: $(G_{t+1}(r_{t+1}^{(\mathbf{w})}))$ is an iid sequence with uniform (0,1) marginal distribution.

In the sequel we test hypothesis (4.5) for $G_{t+1} = F_{t+1}^{RM}$ and $G_{t+1} = F_{t+1}^{REG}$. More concretely the sequences $(F_{t+1}^{RM}(r_{t+1}^{(\mathbf{w})}))$ and $(F_{t+1}^{REG}(r_{t+1}^{(\mathbf{w})}))$ are tested for variance 1/12 (the variance of a uniform (0,1)), using a test based on the Central Limit Theorem, for uniform (0,1) marginal distribution, employing the Kolmogorov-Smirnov and Andersen-Darling test and for independence, using the Ljung-Box statistic at lag 10. We used the following simulation set-up.

Three thousand portfolios were randomly generated. The weights of each portfolio $\mathbf{w}$ were sampled from a uniform (0,1) distribution then normalized such that they added up to 1. As in Section 4.2, the sample is split into two parts: the first 1000 observations serve to produce the initial parameter estimates for the regression-type model while the remaining 1926 observations are used to compute the sequences $(F_{t+1}^{RM}(r_{t+1}^{(\mathbf{w})}))$ and $(F_{t+1}^{REG}(r_{t+1}^{(\mathbf{w})}))$ for each portfolio $\mathbf{w}$. (We kept the weights of the portfolios constant during the testing period.) For every sequence we produced the p-values corresponding to the four mentioned statistics.

The results of these simulations are summarized in Figure 4.7 where the percentage of p-values smaller than 5% is reported. For a given test, the first bar concerns Riskmetrics^{TM}
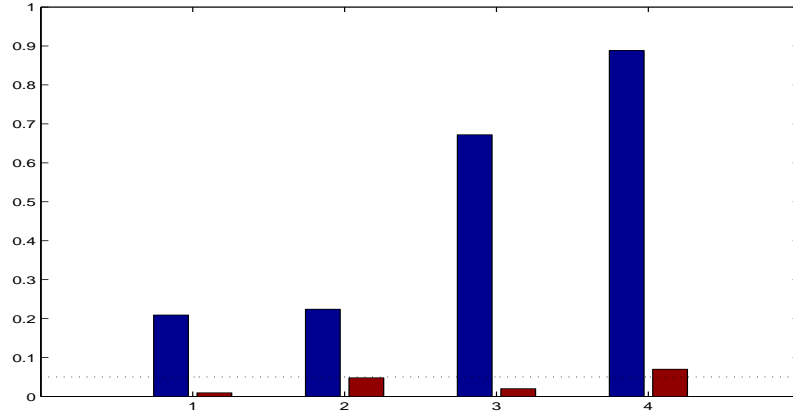
Figure 4.7. *The percentage of the p-values for the K-S (1), A-D (2), the L-B at lag 10 (3) and the variance test (4) that are smaller than 5% . For a given test, the first bar concerns the Riskmetrics^{TM} methodology while the second one refers to the forecasting methodology described in subsection 4.1.*

while the second one refers to the forecasting methodology described in Section 4.1. It is interesting to notice that for almost 90% of the portfolios, Riskmetrics^{TM} fails (at the 5% level) the variance test. This should be compared to the 94% acceptance rate for our methodology. Moreover, 25% of the sequences $(F_{t+1}^{RM}(r_{t+1}^{(\mathbf{w})}))$ fail at least one of the uniformity tests (either K-S or A-D) compared to only 5% of the $(F_{t+1}^{REG}(r_{t+1}^{(\mathbf{w})}))$ sequences. Finally, Riskmetrics^{TM} fails at least one of the four tests in 94% of the cases compared to only 9% for our methodology.

5. THE RELATIONSHIP TO RISKMETRICS[5]

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

---

[5]Reprinted from Drees and Stărică [5]. Although the discussion concentrates on the univariate case, its relevance remains unaltered. The univariate returns are denoted by X. We appologize for the small change of notation.

approach which ignores the mean of the returns and assumes normal innovations, although we also briefly discuss more refined models.

In the RiskMetrics manual [15] it is suggested to forecast the volatility by an exponential filter of the squared returns:

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

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 [15], a conditional, multiplicative model with $\mu = 0$ and normal independent innovations is assumed, i.e.

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

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

$$(5.3) \qquad \sigma^2_t = \lambda \sigma^2_{t-1} + (1-\lambda)X^2_{t-1}.$$

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 [9] and Nelson [11] 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 GARCH(1,1) specification

$$(5.4) \qquad\qquad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

and (5.3) and emphasized by the comparisons in Section 5.2.3 of [15]) is misleading. Instead, the RiskMetrics approach can be motivated by our non-stationary regression model (2.2).

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 (3.1) 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 2 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 [15]. 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 Section 4 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.

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.

6. CONCLUSION

In this paper a simple multivariate model for returns on financial instruments is proposed. The modeling methodology is innovative, replacing the stationary, fully parametric, conditional approach of GARCH- or stochastic volatility-type models with a non-stationary, semi-parametric, unconditional modeling.

Unlike most of the multivariate econometric models for financial returns, our model supposes the volatility to be exogenous. The vector of returns is assumed to follow an AR(1) process whose independent innovations have a changing unconditional covariance structure. The methodological frame is that of non-parametric regression with non-random equidistant design points where the regression function is the evolving unconditional co-variance. The regression function is estimated by a local weighted average with the weights chosen by evaluating a kernel function at the design points. The vectors of standardized innovations have asymmetric heavy tailed and are modeled parametrically.

The model was estimated on a tri-variate sample of returns consisting of the foreign exchange rate Euro/Dollar (EU), the FTSE 100 index, and the 10 year US T-bond. The model fits the data well and provides good multivariate distributional forecasts.

Since the unconditional covariance structure is estimated non-parametrically, we believe that our approach is particularly suited for analyzing large portfolios of assets. We also believe that the careful parametric modeling of the extremal behavior of the standardized

innovations makes it amenable for precise VaR calculations. Evaluating its behavior in these settings is, however, subject of further research.

REFERENCES

[1] BARONE-ADESI, G., GIANNOPOULOS, K., VOSPER, L. (1999) VaR without correlations for portfolio of derivative securities. *Journal of Futures Markets*, **19**, 583–602.

[2] 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.

[3] DIEBOLD, F.X., GUNTHER, T. AND TAY, A. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, **39**, 863–883.

[4] DIEBOLD, F.X., HAHN, J. AND TAY, A. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange. *Review of Economics and Statistics*, **81**, 661–673.

[5] DREES, H. AND STĂRICĂ, C. (2002) A simple non-stationary model for stock returns. Preprint.

[6] ENGLE, R.F. AND SHEPPARD, K. (2002) Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. Preprint. Available at `http://weber.ucsd.edu/~mbacci/engle/index_recent.html`

[7] HÄRDLE, W. (1990) Applied Nonparametric Regression. Econometric Society Monographs No. 19, Cambridge University Press.

[8] KENDALL, M. AND STUART, A. (1979). The advanced theory of statistics. Charles Griffin, London.

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

[10] MIKOSCH, T. AND STĂRICĂ C. (2000) Is it really long memory we see in financial returns? In Extremes and Integrated Risk Management. Ed. P. Embrechts, Risk Books.

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

[12] MÜLLER, H., G. (1988) Nonparametric Regression Analysis of Longitudinal Data. Springer, Berlin.

[13] NAGAHARA, Y., AND KITAGAWA, G. (1999). A Non-Gaussian Stochastic Volatility Model. *The Journal of Computational Finance*, **2**, 33–47.

[14] PAGAN, A. (1996) The Econometrics of Financial Markets. *Journal of Empirical Finance*, **3**, 15–102.

[15] RiskMetrics^{TM}. Technical Document. Available at `http://www.riskmetrics.com/techdoc.html`.

[16] SHORACK, G. R. AND WELLNER J. A. (1986) Empirical processes with applications to statistics. John Wiley & Sons, New York.