

A non-stationary paradigm for the dynamics of multivariate financial returns¹

Stefano Herzel² , Cătălin Stărică³ and Reha Tütüncü⁴

This version: May 2004

¹This research has been supported by The Bank of Sweden Tercentenary Foundation, by the NSF through grant CCR-9875559 and by the Consiglio Nazionale delle Ricerche.

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

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

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

Abstract

A simple non-stationary paradigm for the dynamics of multivariate returns is discussed. Unlike most of the multivariate econometric models for financial returns, our approach supposes the volatility to be exogenous and non-stationary. The vectors of returns are assumed to be animated by a slowly changing *unconditional* covariance structure. The methodological frame is that of non-parametric regression with fixed, equidistant design points. The regression function is the time evolving unconditional covariance. Special attention is paid to the accurate description of the extremal dependence of the vector of returns. The non-stationary paradigm is first applied to describe the changing dynamics of a multivariate data set of returns on three financial risk factors: a foreign exchange rate, an index and an interest rate. Then, its one-day-ahead multivariate distributional forecast performance is evaluated. We show through an out-of sample simulation experiment that our methodology is superior to the plain-vanilla specification of the industry standard RiskMetrics in forecasting the distribution of returns on portfolios of the three risk factors over horizons of one day, ten days and twenty days.

JEL classification: C14, C16, C32.

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

1. INTRODUCTION

This paper discusses a non-stationary, unconditional approach to understanding the dynamic of multivariate financial returns. Non-stationary modeling has a long tradition in financial econometric literature predating the currently prevalent *stationary, conditional* paradigm of which the autoregressive conditionally heteroscedastic (ARCH) -type processes and stochastic volatility models are outstanding examples (see for example, Officer [27] or Hsu, Miller and Wichern [17]). Our work is motivated by growing evidence of instability in the stochastic features of stock returns. More concretely, a growing body of econometric literature (Diebold [4], Lamoureux and Lastrapes [21], Simonato [33], Cai [3], Lobato and Savin [20], Mikosch and Stărică [22], [23] among others) argues that most of the features of return series that puzzle through their omni-presence, the so called “stylized facts”, including the ARCH effects, the slowly decaying sample ACF for absolute returns and the IGARCH effect (for definitions and details see Mikosch and Stărică [23]) could be manifestations of non-stationary changes in the second moment dynamic of returns (see also Stock and Watson [36]). We illustrate our methodology through a detailed analysis of a tri-variate sample of daily log-returns consisting of the foreign exchange rate Euro/Dollar (EU), the FTSE 100 index, and the 10 year US T-bond. The three series are common examples of risk factors⁵.

The paper concentrates on answering the following methodological question: How can one analyze the multivariate dynamic of returns in the non-stationary conceptual framework? We argue that a possible adequate set-up could be that of classical non-parametric regression with fixed equidistant design points (see Campbell *et al.* [2] or Wand and Jones

⁵For a definition and examples of the importance of modeling the joint dynamic of risk factors, see for example the RiskMetrics document [31]. Briefly, a common current approach to modeling the joint dynamic of large portfolios of financial instruments consists in reducing the size of the model by relating the movements of a large number of the instruments in the portfolio to a relatively small number of so called risk factors (market indices, foreign exchange rates, interest rates). The modeling then concentrates on describing the dynamics of the risk factors.

[37]). More concretely, the vectors of financial returns are assumed have a time-varying *unconditional* covariance matrix that evolves smoothly. Its dynamics is estimated by a local weighted average or local smoothing. The vectors of standardized innovations are assumed to have asymmetric heavy tails and are modeled parametrically. The careful description of the extremal behavior of the standardized innovations yields a model suited for precise VaR calculations and for generation of stress-testing scenarios.

A closely related issue to the methodological question discussed is: What type of non-stationarities might affect the multivariate dynamic of financial returns? The in-depth analysis in Section 5 as well as the forecasting results in Section 6 indicate the time-varying second unconditional moment as a possible main source of non-stationarity of returns on the three financial instruments we use to exemplify our approach⁶.

An important aspect of the methodology we propose is related to answering the following: How should we interpret the slow decay of the sample autocorrelation function (SACF) of absolute returns (see Figures 5.1 and 5.2)? Should we take it at face value, supposing that events that happened a number of months (or years) ago bear a strong impact on the present dynamics of returns? Or are the non-stationarities in the returns responsible for its presence as a number of authors have argued lately (the list of related relevant references includes Hidalgo and Robinson [16], Lobato and Savin [20], Granger and Hyung [14], Granger and Teräsvirta [15], Diebold and Inoue [5], Mikosch and Stărică [22], [23])? In a recent paper, Stărică and Granger [35] have documented the superiority of the paradigm of time-varying unconditional variance over some specifications of stationary long memory and stationary conditional autoregressive heteroscedastic methodology in

⁶Our findings and the modeling methodology that they motivate extend to the multivariate framework the work of Officer [27] and Hsu, Miller and Wichern [17]. The former, using a non-parametric approach to volatility estimation, reports evidence of time-varying second moment for the time series of returns on the *S&P 500* index and industrial production. The later modeled the returns as a non-stationary process with discrete shifts in the unconditional variance. Note also that, although the paper only reports the detailed results of an analysis of three risk factors, qualitatively similar results are obtained for a large number of other risk factors.

longer horizon volatility forecasting. Our approach is based on interpreting the slow decay of the SACF/SCCF of absolute returns as a sign of the presence of non-stationarities in the second moment structure.

Our primary goal is to propose an approach that, while capable of explaining the multivariate dynamics of financial data adequately, is simple and easy to implement. For this reason, at each step of our modeling and estimation approach, we deliberately choose simple and well known methodologies rather than complex estimation techniques. Our empirical study, to which a substantial portion of this article is devoted, demonstrates that the non-stationary paradigm is capable of fitting multivariate data accurately and that it outperforms the plain-vanilla specification of the industry standard Riskmetrics in a simulation study of distributional forecasts.

Non-parametric techniques have been extensively used in the econometric literature on financial and macro-economic time series. For example, Rodríguez-Poo and Linton [28] use kernel-based inference technique to estimate the time-dependent volatility structure of residuals of an VAR process. They apply their methodology to macro-economic time series. Fan *et al.* [9] also use kernel regression to estimate time-dependent parametric models for means and covariances in a Gaussian setting. These models are time-dependent generalizations of the time-homogeneous, stationary models discussed in Fan and Yao [10]. Unlike these studies, we focus on the the dynamic modeling of the *full distribution* of multivariate returns and not only on particular features of it (like mean or second moment structure). We emphasize a non-Gaussian, heavy-tailed modeling of the standardized innovations for an accurate description of the extremal behavior of the multivariate distribution of returns.

The rest of the paper is organized as follows. Section 2 introduces our non-stationary paradigm, Section 3 collects the relevant results from the statistical literature on non-parametric curve estimation. Section 4 discusses a heavy-tail parametric model for the innovation series. In Section 5, the non-stationary paradigm described in Section 2 is used to analyse the dynamics on a tri-variate sample of returns on 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 6 evaluates the performance of our model in forecasting the distribution of multivariate returns. In Section 7 we comment on the relationship between our modeling approach and the RiskMetrics methodology while Section 8 concludes.

2. A SIMPLE NON-STATIONARY PARADIGM FOR MULTIVARIATE RETURN MODELING

Denote by \mathbf{r}_k the $d \times 1$ -dimensional vector of returns $k = 1, 2, \dots, n$. ARCH-type models assume that (\mathbf{r}_k) is a *stationary, dependent*, white noise sequence with a certain conditional second moment structure. More specifically, the $d \times d$ conditional variance-covariance matrix $\mathbf{H}_k := E(\mathbf{r}_k \mathbf{r}_k' \mid \mathbf{r}_{k-1}, \mathbf{r}_{k-2}, \dots)$ 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 $P(\mathbf{r}_k \in \cdot \mid \mathbf{r}_{k-1}, \mathbf{r}_{k-2}, \dots) = P(N(0, \mathbf{H}_k) \in \cdot)$. The common assumptions of the ARCH-type models imply that (\mathbf{r}_k) is a strongly stationary sequence. In particular, the unconditional covariance does not change in time (see Stărică [34] for a discussion on the implications of this assumption on modeling and forecasting univariate index returns).

Our alternative approach assumes (\mathbf{r}_k) to be a *non-stationary* sequence of *independent* random vectors. More concretely, the distribution of \mathbf{r}_k is characterized by a changing unconditional covariance structure that is a manifestation of complex market conditions. The covariance dynamics is hence driven by exogenous factors. We emphasize that, in our approach, the presence of autocorrelation structure in absolute (square) returns is explained by a non-stationary covariance structure⁷. To acknowledge the slow nature of the changes, i.e. the persistence in the second moment structure, the covariance is modeled

⁷Sequences of independent observations will display spurious autocorrelation structure if there is a break in the unconditional variance. In other words, the presence of autocorrelation structure is not incompatible with the assumption of independence. See Diebold and Inoue [5] and Mikosch and Stărică [23].

as a *smooth* function of time. This approach leads to the following regression-type model⁸:

$$\begin{aligned} \mathbf{r}_k &= \mathbf{S}(t_{k,n}) \underline{\varepsilon}_{k,n}, \quad k = 1, 2, \dots, n, \quad \text{where} \quad t_{k,n} := k/n, \quad t_k \in [0, 1] \\ (2.1) \quad \mathbf{S}(t) : [0, 1] &\rightarrow \mathbb{R}^{d \times d} \text{ is an invertible matrix and a smooth function of time,} \\ (\underline{\varepsilon}_{k,n}) &\text{ is an iid sequence of random vectors with mutually independent coordinates,} \\ \text{such that } E \underline{\varepsilon}_{k,n} &= 0, \quad Var \underline{\varepsilon}_{k,n} = I_d. \end{aligned}$$

(The notation is that of the classical non-parametric regression set-up and is motivated by the specific nature of the asymptotic results⁹. We will omit indices n whenever feasible.)

The precise smoothness assumptions on $\mathbf{S}(t)$ are discussed in the sequel. The elements of the sequence $(\underline{\varepsilon}_{k,n})$ are called the *standardized innovations*. From (2.1), it follows that

$$E(\mathbf{r}_k \mathbf{r}_k' \mid \mathbf{r}_{k-1}, \mathbf{r}_{k-2}, \dots) = E(\mathbf{r}_k \mathbf{r}_k') = \mathbf{S}(t_{k,n}) \mathbf{S}'(t_{k,n}) := \boldsymbol{\Sigma}(t_{k,n}), \quad \text{and}$$

$$P(\mathbf{r}_k \in \cdot \mid \mathbf{r}_{k-1}, \mathbf{r}_{k-2}, \dots) = P(\mathbf{r}_k \in \cdot), \quad k = 1, 2, \dots, n.$$

This modeling approach 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. In other words, our uncertainty about the form of the model is manifestly expressed in the choice of the non-parametric regression approach.

⁸A mean term could be included in model (2.1). Denoting by $\mathbf{u}_k := \mathbf{r}_k - E\mathbf{r}_k$, $k = 1, \dots, n$, the model would then assume \mathbf{u}_t to be independent with covariance matrix $\mathbf{S}(t)\mathbf{S}'(t)$, a smooth function of t . We have implemented both procedures, i.e. with and without removing of a mean estimate in a preliminary step, and obtained qualitatively equal results. Hence, in the sequel, we work under the simplifying assumption of a negligible mean of the return series.

⁹Unlike in other fields of statistics, the asymptotic results involve not only an increasing number of observations but also an increase in the frequency with which the unknown function is observed. To attain this goal, the observations are indexed between 0 and 1. In this way an increase in the sample size implies also an increase in the frequency with which we observe the regression function.

Furthermore, we will assume the existence of a smooth¹⁰ function $\mathbf{V}(t) : [0, 1] \rightarrow \mathbb{R}^{d \times d}$ such that $\text{Var } r_{i,k} r_{j,k} = v_{ij}(t_k)$ (in short, $\text{Var } \mathbf{r}_k \mathbf{r}_k' = \mathbf{V}(t_k)$) where $r_{i,k}$ is the i -th coordinate of \mathbf{r}_k and $t_k = k/n$ ¹¹. With this notation

$$(2.2) \quad r_{i,k} r_{j,k} = \Sigma_{ij}(t_k) + v_{ij}^{1/2}(t_k) \tilde{\varepsilon}_k^{ij}, \quad k = 1, 2, \dots, n, \quad i, j = 1, 2, \dots, d,$$

where the errors $\tilde{\varepsilon}_k^{ij}$ are iid vectors with independent coordinates, such that $E \tilde{\varepsilon}_k^{ij} = 0$, $\text{Var } \tilde{\varepsilon}_k^{ij} = 1$. Hence the function $\Sigma(t)$ can be estimated by standard non-parametric heteroscedastic regression methods for non-random, equidistant design points using the series $\mathbf{r}_k \mathbf{r}_k'$, $k = 1, \dots, n$.

The non-stationary paradigm that we have introduced above can be used both for describing the dynamics of multivariate data as well as for short horizon forecasting. The methodological difference between applying it for data description or for forecasting will become clear in the next section.

3. NON-PARAMETRIC SMOOTHING

Our main reference in the context of non-parametric regression is Müller and Stadtmüller [25] on kernel curve estimation in the heteroscedastic regression model

$$(3.1) \quad y_{k,n} = \mu(t_{k,n}) + \sigma(t_{k,n}) \epsilon_{k,n}, \quad k = 1, 2, \dots, n.$$

The random variables y_k are observations of the unknown regression function $\mu(t) : [0, 1] \rightarrow \mathbb{R}$, perturbed by heteroscedastic errors $\sigma(t_k) \epsilon_k$. The standardized errors ϵ_k are iid with mean zero and unit variance not necessarily Gaussian. The functions $\mu : [0, 1] \rightarrow \mathbb{R}$ and $\sigma : [0, 1] \rightarrow \mathbb{R}_+$ are assumed smooth (the smoothness requirements will be made precise in the sequel).

Our analysis uses kernel regression smoothing. For an introduction on smoothing estimator and in particular, on kernel estimators, see Section 12.3 of Campbell *et al.* [2] or

¹⁰The precise smoothness assumptions on $\mathbf{V}(t)$ are discussed in the sequel.

¹¹In words, we assume that the covariance structure and the variance of the covariance are evolving smoothly through time.

Wand and Jones [37]. The following kernel estimator will be used in the various steps of mean and variance estimation in the heteroscedastic regression model (3.1)

$$(3.2) \quad \hat{f}(t; h) = \sum_{k=1}^n W_k(t; h) U_k,$$

where U_k stand for $\tilde{\sigma}^2(t_k)$, preliminary variance estimates, in the estimation of $f := \sigma^2(t)$ and for y_k in the estimation of $f := \mu(t)$. The weights $W_k(t)$ satisfy

$$(3.3) \quad W_k(t; h) = W_{k,n}(t; h) = \frac{1}{h} \int_{s_{k-1}}^{s_k} K\left(\frac{t-u}{h}\right) du, \quad s_k = \frac{t_{k-1} + t_k}{2}.$$

The quantity $h > 0$ is the bandwidth of the estimator and the kernel function K on $[-1, 1]$ satisfies the basic condition $\int K(u) du = 1$ and some further assumptions. These are satisfied by the Gaussian kernel density function when it is first truncated at $[-3, 3]$, then rescaled to $[-1, 1]$ and finally made Lipschitz continuous such that $K(-1) = K(1) = 0$ by changing the kernel appropriately in $[-1, 1] \setminus [-1 + \delta, 1 - \delta]$ for $\delta = 0.01$. This is the kernel used in Section 5. We note that such estimates use past and future information. A modified kernel, only based on the past, will be introduced later.

3.1. Estimation of the variance. Let us summarize now some of the necessary theory for the estimation of $\hat{\sigma}$ in the heteroscedastic model (3.1). The kernel estimator of $\sigma(t)$ in the heteroscedastic regression model (3.1) is defined in two steps.

- (1) First, a preliminary smoothing removes the mean function μ in (3.1) in some neighborhood of t_k . The preliminary estimator of the variance at an inner point t_k in $[0, 1]$ is given by

$$(3.4) \quad \tilde{\sigma}^2(t_k) = \left(\sum_{j=-m_1}^{m_2} w_j y_{j+k} \right)^2,$$

with the weights w_j satisfying $\sum_{j=-m_1}^{m_2} w_j = 0$ and $\sum_{j=-m_1}^{m_2} w_j^2 = 1$ for some fixed $m_1, m_2 \geq 0$.

- (2) Second, we view the preliminary estimates of the variance, $\tilde{\sigma}^2(t_k)$ as measurements from the following regression model:

$$(3.5) \quad \tilde{\sigma}^2(t_k) = \sigma^2(t_k) + \tilde{\epsilon}_k, \quad 1 \leq k \leq n,$$

where the errors $\tilde{\epsilon}_k$ form an $m_1 + m_2$ -dependent sequence, $E\tilde{\epsilon}_k = 0$.

The estimator of the variance is then given by

$$(3.6) \quad \hat{\sigma}^2(t) := \hat{\sigma}^2(t; h_{\sigma^2}) = \sum_{k=1}^n W_k(t; h_{\sigma^2}) \tilde{\sigma}^2(t_k),$$

where the weights $W_k(t; h)$ are defined in (3.3).

In the sequel we assume that σ^2 is twice differentiable with a continuous second derivative, μ is Lipschitz continuous of order $\alpha \geq 0.25$ and $E|\epsilon_i|^{5+\delta} < \infty$ for some $\delta > 0$. Then the following statements can be derived from Theorem 3.1 and Remark at the bottom of p. 622 in Müller and Stadtmüller [25]:

- (1) The estimated variance $\hat{\sigma}^2(t)$ satisfies

$$|\hat{\sigma}^2(t) - \sigma^2(t)| \leq c(h_{\sigma^2}^2 + (\log n/nh_{\sigma^2})^{1/2}),$$

almost surely, for some unspecified positive constant c , uniformly on any compact of the interval $(0,1)$, if the bandwidth h_{σ^2} satisfies $\liminf n^{1/5+\delta'} h_{\sigma^2}/\log n > 0$, $\liminf nh_{\sigma^2}^2 > 0$, where $0 < \delta' < \delta$.

- (2) The expected value $E\hat{\sigma}^2(t)$ satisfies

$$|E\hat{\sigma}^2(t) - \sigma^2(t)| \leq c(h_{\sigma^2}^2 + n^{-1})$$

for some unspecified positive constant c , uniformly on any compact of the interval $(0,1)$.

3.2. Estimation of the mean in the heteroscedastic regression model. If moreover, μ is twice differentiable with continuous second derivative, Lemma 5.3 of Müller and Stadtmüller [25] gives the following results for $\hat{\mu}_{\text{He}}(t; h_\mu)$, the estimator given by (3.2) with $f := \mu$:

- (1) The expected value $E\hat{\mu}_{\text{He}}(t)$ satisfies, as $n \rightarrow \infty$ and $h_\mu := h_{\mu,n} \rightarrow 0$, $nh_\mu \rightarrow \infty$

(i)

$$(3.7) \quad E\hat{\mu}_{\text{He}}(t) - \mu(t) = \mu''(t)h_\mu^2 B + o(h_\mu^2) + O(n^{-1}),$$

where $B = \int K(u)u^2 du/2$,

(ii)

$$|E\hat{\mu}_{\text{He}}(t) - \mu(t)| \leq c(h_\mu^2 + n^{-1}),$$

for some unspecified positive constant c , uniformly for $t \in [\delta, 1 - \delta]$, any fixed $\delta \in (0, 1)$.

(2) The variance of $\hat{\mu}_{\text{He}}(t)$ satisfies for every t , as $n \rightarrow \infty$ and $h_\mu := h_{\mu,n} \rightarrow 0$, $nh_\mu \rightarrow \infty$

$$(3.8) \quad \text{Var}(\hat{\mu}_{\text{He}}(t)) = \frac{\sigma^2(t)}{nh_\mu} U(1 + o(1)),$$

where $U = \int K^2(u)du = 0.84$ for the normal kernel used in our analysis. Note that the bandwidths h_μ in the estimation of μ and h_{σ^2} in that of σ^2 are in general very different.

These results apply to the concrete heteroscedastic regressions of interest (2.2) as follows. The estimator of $\Sigma(t)$ as given by the heteroscedastic approach described in subsection 3.2 is

$$(3.9) \quad \hat{\Sigma}(t; h) := \sum_{k=1}^n W_{k,n}(t; h) \mathbf{r}_k \mathbf{r}_k',$$

where the weights $W_{k,n}$ are defined in (3.3). Note that the matrix $\hat{\Sigma}(t; h)$ is positive definite by construction.

The estimator of $\mathbf{V}(t)$ in (2.2) given by the methodology described in subsection 3.1 is

$$(3.10) \quad \hat{\mathbf{V}}(t; \tilde{h}) := \sum_{k=1}^n W_{k,n}(t; \tilde{h}) \left(\sum_{l=-m_1}^{m_2} w_l \mathbf{r}_{l+k} \mathbf{r}_{l+k}' \right)^2,$$

where the square operation has to be intended component-wise. The weights w_l satisfying $\sum_{l=-m_1}^{m_2} w_l = 0$ and $\sum_{l=-m_1}^{m_2} w_l^2 = 1$ for some fixed $m_1, m_2 \geq 0$. Note that the matrix V is not a covariance matrix and that we have simply used the convenience of the matrix

notation in order to write concisely $d \times d$ (coordinate-wise) equations. In the analysis that follows, we have used $m_1 = 1$, $m_2 = 0$ and $w_1 = w_2 = 1/\sqrt{2}$.

If the coordinates of \mathbf{S} are twice differentiable with continuous second derivatives and $E|\underline{\varepsilon}_1|^{5+\delta} < \infty$ for some $\delta > 0$ then, as $n \rightarrow \infty$

(i)

$$|E \hat{\sigma}_{ij}(t) - \sigma_{ij}(t)| \leq c(h^2 + n^{-1}),$$

as $h \rightarrow 0$, $nh \rightarrow \infty$, for some unspecified positive constant c , uniformly on any compact in $(0,1)$.

(ii) as $h \rightarrow 0$, $nh \rightarrow \infty$, $\hat{\sigma}_{ij}(t) - \sigma_{ij}(t)$ is approximately $N\left(\sigma_{ij}''(t)h^2B, \frac{v_{ij}(t)}{nh}U\right)$,

(iii) if the bandwidth \tilde{h} is chosen as $\tilde{h} \sim (\log n)/n^{1/5+\delta'}$, where $0 < \delta' < \delta$, then as $n \rightarrow \infty$

$$\hat{v}_{ij}(t) \rightarrow v_{ij}(t),$$

almost surely, uniformly on any compact in $(0,1)$. Moreover

$$(3.11) \quad \left[\hat{\sigma}_{ij}(t) - z_{\alpha/2} \sqrt{\frac{\hat{v}_{ij}(t)U}{nh}}, \hat{\sigma}_{ij}(t) + z_{\alpha/2} \sqrt{\frac{\hat{v}_{ij}(t)U}{nh}} \right]$$

are approximate $(100 - \alpha)\%$ point-wise confidence intervals for $\sigma_{ij}(t)$, where $z_{\alpha/2}$ are the $(100 - \alpha/2)\%$ normal quantile.

In the analysis of the multivariate return time series in Section 5, 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 the kernel function makes little difference; see the monographs by Müller [24] and Wand and Jones [37].

As we have already emphasized, the non-stationary paradigm under discussion can be used both for understanding the nature of past changes in the dynamics of multivariate data as well as for short horizon forecasting. The methodological difference between the use of the paradigm for data description and that for forecasting consists in the type of

kernel used in estimation of the regression function. A symmetric kernel will be used when interested in describing the dynamic of the changes in the historical sample while an asymmetric one, giving weights only to the past and current observations will be applied in the forecasting exercises. See Sections 5 and 6 for detailed applications of the paradigm in the two set-ups.

3.3. Bandwidth selection. The equations (3.7) and (3.8) yield the asymptotic integrated square error (MISE) of $\hat{\mu}_{\text{He}}(t)$, the estimator of μ in (3.1) given by (3.2) :

$$MISE = h_{\mu}^4 B^2 \int \mu''(u)^2 du + \frac{\int \sigma^2(u) du}{nh_{\mu}} U.$$

Minimizing the MISE with respect to the bandwidth h_{μ} yields the globally optimal bandwidth

$$(3.12) \quad h_{\mu}^{(g)} = \left(\frac{\int \sigma^2(u) du U}{4nB^2 \int \mu''(u)^2 du} \right)^{1/5}.$$

The choice of smoothing parameter or bandwidth is crucial when applying non-parametric regression estimators, such as kernel estimators. For this reason we applied a set of different methods of bandwidth selection. Cross-validation is a method based on minimizing residual mean squared error criteria frequently used to infer the optimal smoothing parameter. Another method builds on estimating the asymptotically optimal global bandwidth (3.12) from the data. Since estimators for the residual variance and for an asymptotic expression for the bias (3.7) are plugged into the asymptotic formula (3.12), such selection rules are called ‘plug-in’ estimators. The functional that quantifies bias is approximated by the integrated squared second derivative of the regression function. This functional is determined by an iterative procedure introduced in Gasser *et al.* [11] based on a kernel estimator $\hat{\mu}''(t; h_{\mu}'')$ for the derivative. Such an estimator has the form (3.2) with the kernel K tailored to estimate second derivatives (see Gasser *et al.* [12]; for our application we used the optimal (2,4) kernel).

4. A HEAVY-TAILED MODEL FOR THE DISTRIBUTION OF THE INNOVATIONS

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

$$(4.13) \quad \hat{\underline{\varepsilon}}_k := \hat{\mathbf{S}}^{-1}(t_k) \hat{\mathbf{r}}_k, \quad k = 1, 2, \dots, n$$

with $\hat{\mathbf{S}}(t)$, the square root of the estimate $\hat{\Sigma}(t)$ of $\mathbf{S}(t)\mathbf{S}'(t)$ in (3.9). 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 5.3 for evidence supporting this claim), the use of the empirical cdf¹² 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 d independent coordinates, it is sufficient to specify the distributions of $\hat{\varepsilon}_i$, $i = 1, \dots, d$. A flexible and parcimonious 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 [8]. The density of this distribution is

$$(4.14) \quad f(x; m, c) = \frac{2\Gamma(m)}{c\Gamma(m-1/2)\pi^{1/2}} \left(1 + \left(\frac{x}{c}\right)^2\right)^{-m}, \quad x > 0.$$

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 [29].

¹²Using the empirical cdf is tantamount to assuming that the worse case scenarios cannot be any worse than what we have in the sample. Using extreme value techniques for modeling the tails of the innovations allows to extrapolate outside the range of the observed data producing events that are more extreme than the limited history available and that are in line with the distributional features of the observed sample.

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

$$(4.15) \quad f^{VII}(x; m_-, c_-, m_+, c_+) = \frac{1}{2} \left(f_-(x; m_-, c_-) 1_{(-\infty, 0)}(x) + f_+(x; m_+, c_+) 1_{[0, \infty)}(x) \right).$$

We refer to the distribution with density (4.15) (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 four 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 $\hat{\mathbf{S}}(t)$ fully specify the distribution of the time series of returns in the model (2.1).

5. UNDERSTANDING THE DYNAMICS OF 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 tri-variate setup. Note that similar modeling results have been achieved with higher dimensional vectors of returns.

The goal of the discussion in this section is to provide a picture of the changes in the dynamic of the multivariate vector of returns and to check the quality of the non-parametric

paradigm applied in the set-up of modeling. The next section will consider the performance of the paradigm in the forecasting set-up.

Figures 5.1 and 5.2 display the SACF-SCCF of the data and that of the absolute values of the data.

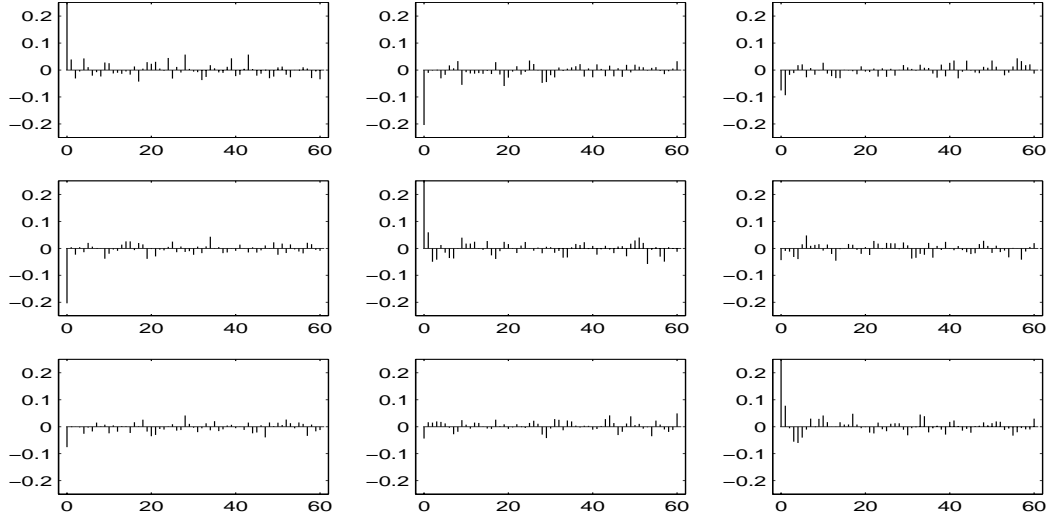


Figure 5.1. *SACF-SCCF of the data (EU returns, the first coordinate, FTSE returns, the second coordinate, the 10 year T-bond returns, the third coordinate respectively). On the diagonal the SACF of the 3 series. Off the diagonal the SCCF of pairs. Since the dependency structure in the data is unknown, no confidence intervals for the correlations are displayed.*

The SACF/SCCF of the returns (Figure 5.1) show extremely small auto- or cross-correlations at lags greater than 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 5.2 show larger correlations in the absolute values.

Note that an SACF/SCCF that displays positive correlations at large lags (like that in Figure 5.2) is not evidence of dependent data. Independent and *non-stationary* observations with a time-varying unconditional variance can produce SACF/SCCF like the ones in Figure 5.2. Positive correlations at large lags could be a sign of non-stationarities in

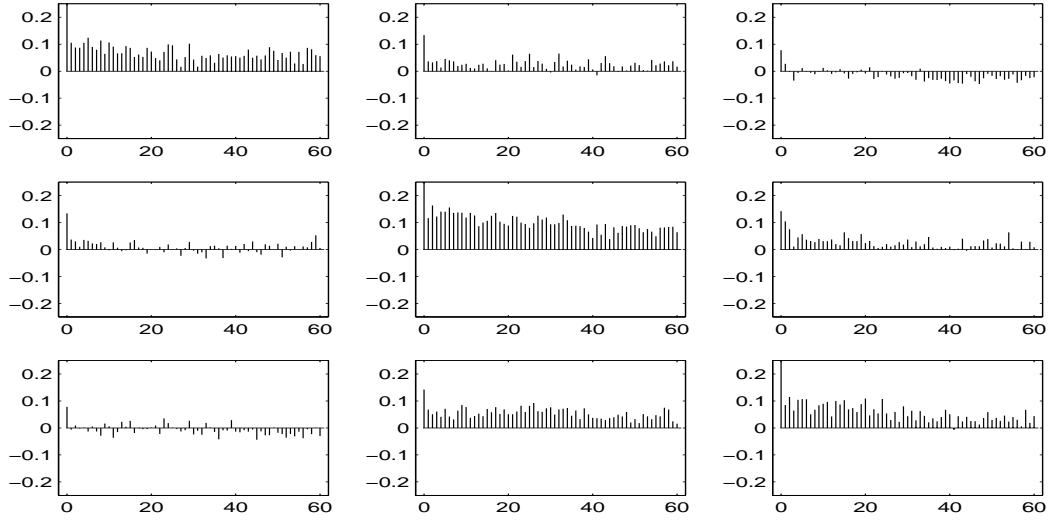


Figure 5.2. *SACF-SCCF of the absolute values of the data (EU returns, the first coordinate, FTSE returns, the second coordinate, the 10 year T-bond returns, the third coordinate respectively). On the diagonal the SACF of the 3 series. Off the diagonal the SCCF of pairs. Since the dependency structure in the data is unknown, no confidence intervals for the correlations are displayed.*

the second moment structure of the time series as well as a proof of stationary, non-linear, long-range dependence; see Mikosch and Stărică [23]. As emphasized earlier, our working paradigm is consistent with the non-stationary interpretation of the SACF/SCCF.

5.1. The evolution of the unconditional covariance structure. We estimated the optimal bandwidth in the set-up of the model (3.1) with $y_k = |\mathbf{r}_k \mathbf{r}_k'|$, $k = 1, 2, \dots, n$, using cross-validation and the method of Gasser *et al.* [11]¹³. Figure 5.3 displays the cross-validation graph. Based on this graph, the choice for the bandwidth is $h_\mu^{(c)} \in [0.005, 0.008]$ with a minimum at 0.006. The procedure of Gasser *et al.* [11] produced $h_\mu^{(g)} = 0.0076$. This is the bandwidth that we use in defining $\hat{\Sigma}(t)$.

The graphs in Figure 5.4 display two estimates of the time-varying standard deviations (sd's) of the three time series. Those in Figure 5.5 show two estimates of the time-dependent

¹³Using $y_k = |\mathbf{r}_k|$ yields qualitatively equal results.

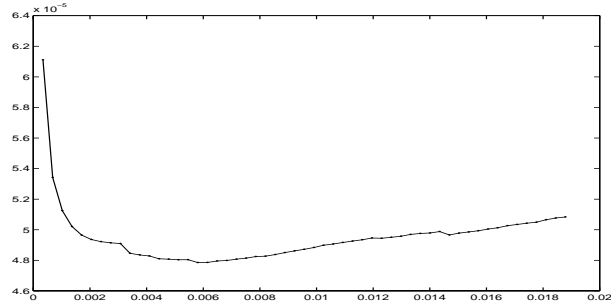


Figure 5.3. *The cross validation graph for the choice of the bandwidth h_μ for $\hat{\mu}$ in (3.2) and $y_k = |\mathbf{r}_k \mathbf{r}'_k|$. The bandwidths $h_\mu^{(c)}$ that minimize the cross validation function belongs to the interval $[0.005, 0.008]$.*

correlation between the three pairs of univariate time series (in the top graph, EU and FTSE, in the middle FTSE and T-bill, in the lower one, EU and T-bill). In all the pictures, the solid line is the estimate obtained using $\hat{\Sigma}$, defined in (3.9) with bandwidth $h = 0.0076$. The dotted line is the estimate obtained using the estimator

$$(5.1) \quad \hat{\Sigma}_1(t) := \sum_{k=1}^n \tilde{W}_{k,n}(t) \mathbf{r}_k \mathbf{r}'_k,$$

where the weights $\tilde{W}_{k,n}$ are defined as in (3.3) with the symmetric kernel K replaced by $\tilde{K}(u) = K(u)1_{u \leq 0}$. The bandwidth used in (5.1) was $h = 0.007$ ¹⁴. Note that $\hat{\Sigma}_1(t)$, estimate of $\mathbf{S}(t)\mathbf{S}'(t)$, uses only the information available at day t . This estimator will be used to produce the forecasting results presented in Section 6¹⁵.

The 95% confidence intervals given by (3.11) are also plotted. Note that the estimated volatilities and correlations that use only the past information belong almost always to the 95% confidence intervals. Hence using only past information seems to yield a rather precise estimates.

¹⁴For the choice of this value, see Section 6.

¹⁵The boundary modification proposed in Rice [30] has been used to take care of the boundary effect.

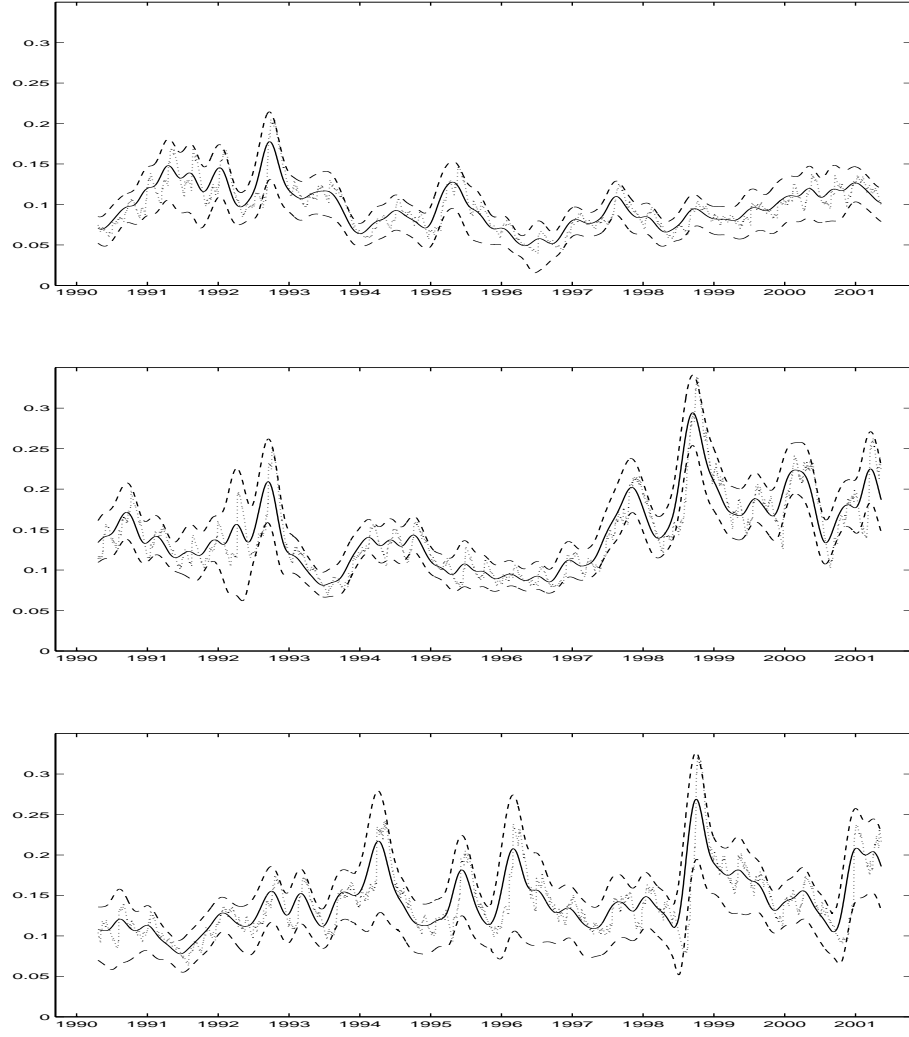


Figure 5.4. Local estimates ($\hat{\Sigma}(t)$, $h = 0.0076$, solid line, $\hat{\Sigma}_1(t)$, $h = 0.007$, dotted line) of the (annualized) standard deviation (sd) of the data: EU, (Top), FTSE (Middle) and the 10-year T-bond (Bottom). The annualized sd is obtained by multiplying the daily sd by a factor of $\sqrt{250}$. The 95% confidence intervals given by (3.11) are also displayed.

The graphs in Figure 5.4 and 5.5 show rather large variations in the estimated standard deviations as well as in the estimated correlation structure of the series. They individuate the existence of periods with unconditional volatilities and unconditional correlations

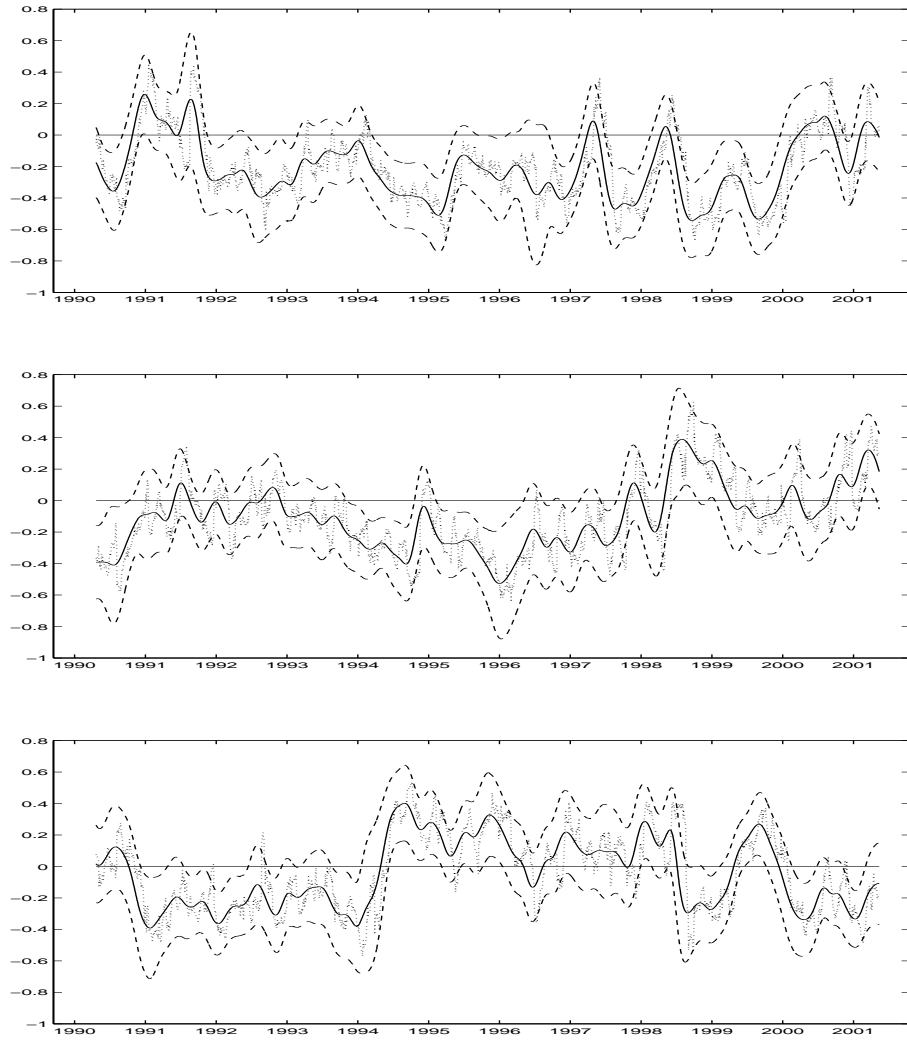


Figure 5.5. Local estimates ($\hat{\Sigma}(t)$, $h = 0.0076$, solid line, $\hat{\Sigma}_1(t)$, $h = 0.0076$, dotted line) of the correlations between the data: EU and FTSE, (Top), FTSE and 10-year T-bond (Middle) and 10-year T-bond and EU (Bottom). The 95% confidence intervals given by (3.11) are also displayed.

that are statistically significant different. 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. The two figures support the assumption of time-varying unconditional covariance structure that is the basis of our non-stationary paradigm.

5.2. The dependence structure of the standardized innovations. In this section we analyze the dependency structure of the estimated standardized innovations $\hat{\underline{\varepsilon}}_t$ defined in (4.13). A battery of three tests is used to achieve this goal. In the sequel we are ignoring the fact that the innovations come from a kernel regression and we treat them as if they were directly observed. In doing this we neglect the possible effect of the estimation error on the asymptotic properties of the statistics we present. As a consequence, the p -values of the tests should be interpreted more as upper limits than as precise values.

With this caveat in mind, 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 cumulative distribution functions 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 (5.3) 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

$$(5.2) \quad H_0 : F_i^{(1)} \text{ and } F_i^{(2)} \text{ are identical.}$$

Table 5.6 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 5.6 supports the hypothesis of stationarity of the coordinates of the sequence of estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$.

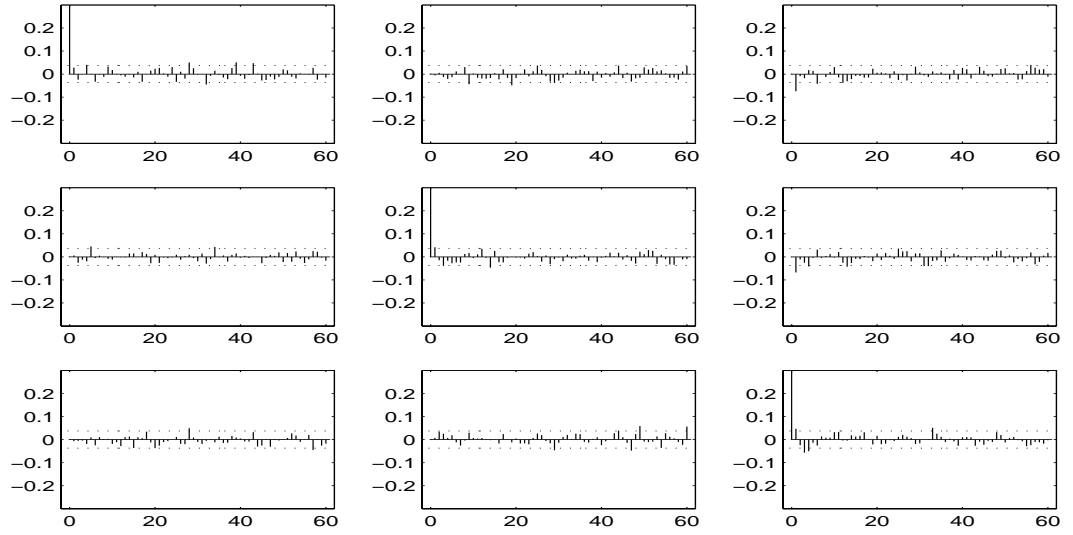


Figure 5.7. *SACF/SCCF of the estimated standardized innovations $(\hat{\underline{\varepsilon}}_t)$. The covariance structure was estimated using $\hat{\underline{\Sigma}}$.*

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 5.7 and 5.8) to test the hypothesis

$$(5.3) \quad H_0 : \hat{\underline{\varepsilon}}_t \text{ are iid vectors with independent coordinates.}$$

	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 5.6. *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.*

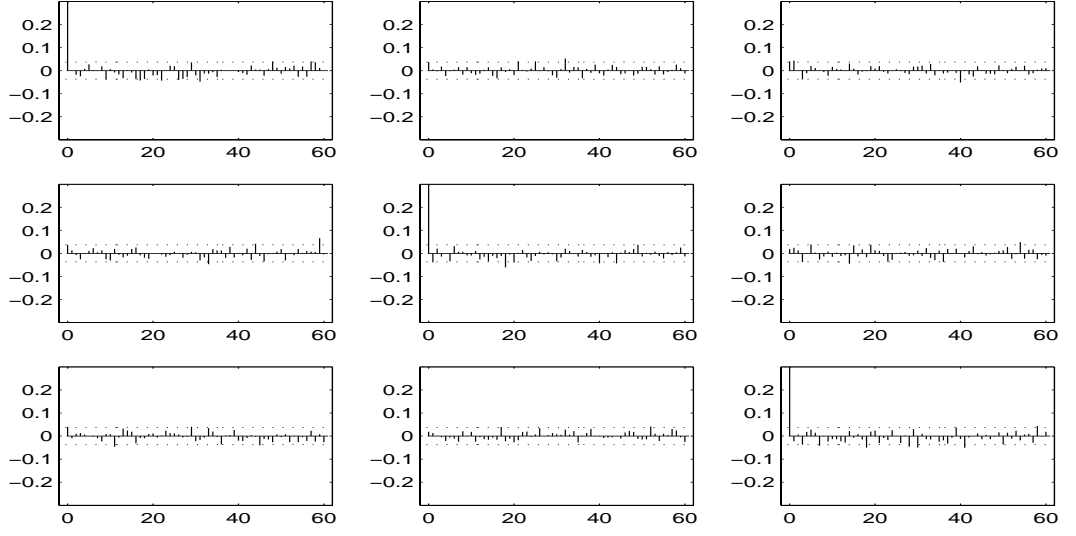


Figure 5.8. *SACF/SCCF of the absolute values of the estimated standardized innovations ($|\hat{\varepsilon}_t|$). The covariance structure was estimated using $\hat{\Sigma}$.*

The confidence intervals in Figures 5.7 and 5.8 correspond to the null hypothesis (5.3). 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 5.2. They support the choice of modeling the standardized innovations as a sequence of iid vectors with independent coordinates.

The visual test of the hypothesis (5.3) is complemented by a Ljung-Box test for the first 25 lags. Table 5.9 gives the p -values for the estimated standardized innovations ($\hat{\varepsilon}_t$) (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 (5.3) at 5% significance levels.

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.

	1	2	3		1	2	3
1	0.15	0.17	0.03		0.20	0.89	0.12
2	0.81	0.12	0.11		0.62	0.21	0.16
3	0.70	0.88	0.07		0.25	0.50	0.22

Table 5.9. *The p -values for the Ljung-Box test at lag 25 of the estimated standardized innovations $(\hat{\varepsilon}_t)$ (left) and their absolute values $(|\hat{\varepsilon}_t|)$ (right). The row and column numbers represent the coordinates.*

Kendall's τ takes values between -1 and 1 (independent variables have $\tau=0$) 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 [18].)

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. The null hypothesis is

$$(5.4) \quad H_0 : \quad \text{the random variables } \hat{\varepsilon}_i \text{ and } \hat{\varepsilon}_j \text{ are independent.}$$

The resulting p -values are given in Table 5.10. For all pairs the hypothesis of independence (5.4) is not rejected at usual statistical levels of significance..

At this point, we conclude that the battery of test described above do not reject the hypothesis that the estimated standardized innovations $(\hat{\varepsilon}_t)$ is a stationary sequence of iid vectors with independent coordinates.

5.3. The multivariate distribution of the standardized innovations. In this section we concentrate on modeling the marginal distribution of the estimated standardized innovations ($\hat{\varepsilon}_t$). 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 5.11 displays the standard normal plots of three coordinate series of the estimated standardized innovations ($\hat{\varepsilon}_t$). 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.

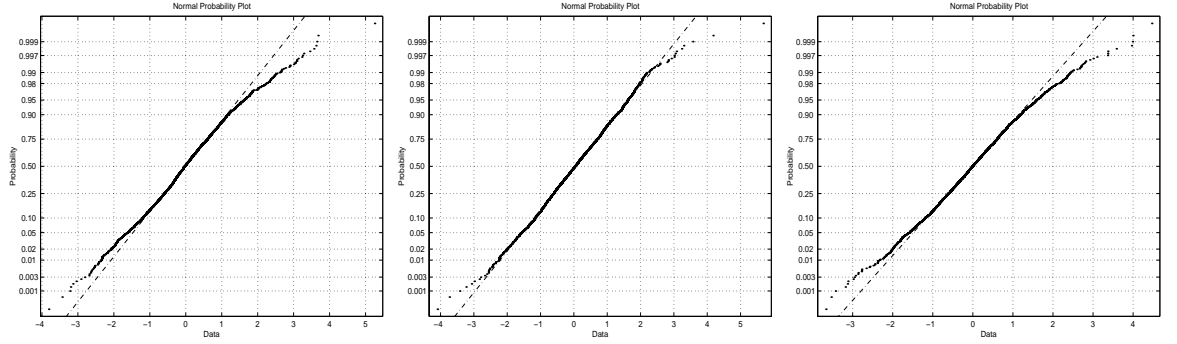


Figure 5.11. Normal probability plots of the three series of coordinates the of estimated standardized innovations ($\hat{\varepsilon}_t$).

The impression given by Figure 5.11 is confirmed by the p -values of the Kolmogorov-Smirnov (K-S) and Andersen-Darling (A-D) tests (for details on these tests see [32]) applied to the coordinate series ($\hat{\varepsilon}_i$), $i = 1, 2, 3$ reported in the left half of Table 5.12. The null

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

hypothesis is

$$(5.5) \quad H_0 : F_i \text{ is the standard normal distribution.}$$

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
$\hat{\varepsilon}_3$	0.001	0.003	$\hat{\varepsilon}_3$	0.10	0.25

Table 5.12. *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 ($\hat{\varepsilon}_t$).*

Figure 5.11 and the values on the left side of Table 5.12 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 ($\hat{\varepsilon}_t$) as asymmetric Pearson type VII heavy tailed distributions. Table 5.13 contains the estimated parameters obtained by fitting an asymmetric Pearson VII distribution (4.15) to the three coordinates of the estimated standardized innovations ($\hat{\varepsilon}_t$).

The estimated parameters in Table 5.13 confirm the results of the previous tail analysis: the first and the third coordinates have heavier tails than the second, with the right tail being heavier than the left one. To test the hypothesis

$$(5.6) \quad H_0 : \text{Var}(\hat{\varepsilon}_t) = I_d,$$

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

two estimates of the variances of the coordinates of the estimated standardized innovations ($\hat{\varepsilon}_t$) together with the corresponding standard deviations are produced. 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 (4.15). Since the variance of any coordinate is a function of the parameters reported in Table 5.13, 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 5.14. The right half of the same table reports the sample covariance together with the corresponding standard deviation. According to the values in Table 5.14 the hypothesis that $Var(\hat{\varepsilon}_t) = I_d$ 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 5.14. *The estimated variances of the coordinates of the estimated standardized innovations ($\hat{\varepsilon}_t$). 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 ($\hat{\varepsilon}_t$) using the estimated asymmetric Pearson VII distributions are displayed in Figure 5.15. A good fit of the asymmetric Pearson VII distributions should translate in linear graphs close to the first diagonal. The null hypothesis is

(5.7) $H_0 : F_i$ is the asymmetric Pearson VII distribution with parameters given in Table 5.12.

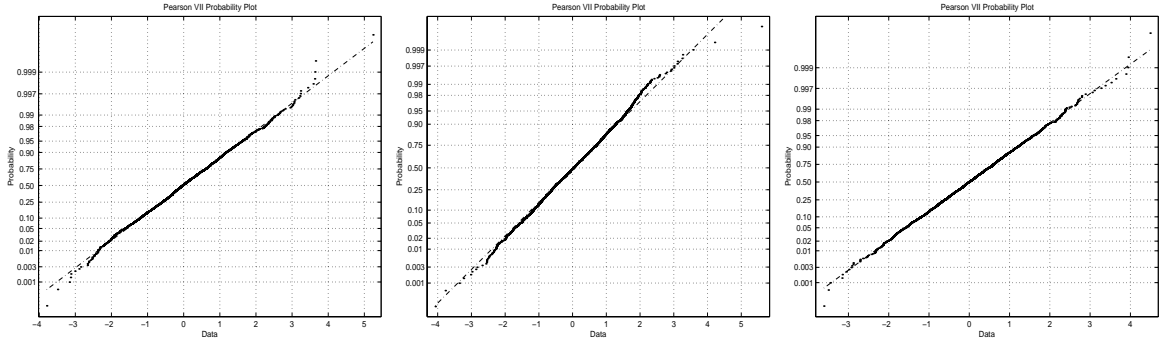


Figure 5.15. *The asymmetric Pearson VII probability plots of the three coordinate series of the estimated standardized innovations ($\hat{\varepsilon}_{i,t}$), $i = 1, 2, 3$.*

The straight plots in Figure 5.15 are a confirmation of the good fit of the asymmetric Pearson VII distribution.

The hypothesis (5.7) is formally tested using the Kolmogorov-Smirnov and Andersen-Darling tests. The p -values of these tests are reported on the right in Table 5.12. The hypothesis is not rejected at usual levels of significance.

The plots in Figure 5.15 and the results in Table 5.12 provide evidence that the parametric family described by (4.15) is indeed an appropriate model for the estimated standardized innovations ($\hat{\varepsilon}_t$).

This concludes the evaluation of the goodness of fit of the model (2.1). The statistical analysis seems to show that the model provides an overall good description of the multivariate data set considered. We now direct our attention towards evaluating the forecasting performance of the non-stationary paradigm.

6. FORECASTING MULTIVARIATE RETURNS

In this section we discuss aspects related to forecasting the multivariate returns using the non-stationary paradigm described in Section 2. We emphasize that we are interested in forecasting the whole distribution of the vector of future returns and not only the second moment structure.

We begin by specifying the m -day ahead forecasting methodology. Then we check the quality of our 1-day multivariate distributional forecasts. We end the section with a comparison (in the univariate framework) between the forecasting behavior of the industry standard Riskmetrics and that of our methodology on randomly generated portfolios containing the three instruments EU, the FTSE, and the US T-bond at one-day, ten-day and twenty-day horizons.

6.1. The m -day ahead forecasting methodology. Given $\hat{\Sigma}_1(\cdot)$, an estimate of the unconditional covariance matrix $\Sigma(\cdot) = \mathbf{S}(\cdot)\mathbf{S}'(\cdot)$ based only on past information, denote by $\hat{F}_{i,t}^{VII}$, $i = 1, 2, 3$, the asymmetric Pearson type VII distributions (4.15) with parameters estimated on the coordinates of the series $(\hat{\mathbf{S}}_1^{-1}(1) \mathbf{r}_1, \hat{\mathbf{S}}_1^{-1}(2) \mathbf{r}_2, \dots, \hat{\mathbf{S}}_1^{-1}(t) \mathbf{r}_t)$, where $\hat{\mathbf{S}}_1(\cdot)$ is the square root of $\hat{\Sigma}_1(\cdot)$.

Based on the model (2.1), the distributional forecast at time t of the m -day ahead return $\mathbf{r}_{t+1,m} := \mathbf{r}_{t+1} + \dots + \mathbf{r}_{t+m}$ is given by

$$\begin{aligned} \mathbf{r}_{t,m}^{REG} &= \sum_{l=1}^m \hat{\mathbf{S}}_1(t) \underline{\epsilon}_{lt}, \\ (6.1) \quad \underline{\epsilon}_{lt}, \quad l &= 1, \dots, m, \text{ are iid } d\text{-dimensional random vectors,} \\ \epsilon_{i,lt} &\text{ are mutually independent with distributions } \hat{F}_{i,t}^{VII}, \quad (i = 1, 2, 3). \end{aligned}$$

In other words, since the covariance matrix evolves slowly through time, to produce the m -day ahead forecast, the next m multivariate returns are assumed iid with a covariance matrix and parameters of the distribution of the standardized innovations estimated on recent past data.

For our forecasting exercise we use the one-sided-kernel estimate of the unconditional covariance matrix $\mathbf{S}(\cdot)\mathbf{S}'(\cdot)$ defined in (5.1). While the theoretical discussion in Section 2 focused on symmetric kernels, similar results are available for estimators of the type (5.1) (see Gijbels, Pope and Wand [13] for the homoscedastic case). In particular, the bias and the variance of these estimators are also given by (3.7) and (3.8). Moreover, for forecasting, cross-validation can be safely employed as a method of bandwidth selection even when the errors are serially correlated ¹⁶.

6.2. One-day ahead multivariate density forecast evaluation. Evaluating the multivariate distributional forecast (see Diebold *et al.* [7]) is particularly simple in the case of the model (2.1), due to the assumption of independence of the sequence (\mathbf{r}_t) . Verifying that the distribution of $\mathbf{r}_{t,1}^{REG}$ defined by (6.1) coincides with that of $\mathbf{r}_{t+1,1} = \mathbf{r}_{t+1}$ is equivalent to checking that the m -dimensional vectors (\mathbf{z}_t)

$$(6.2) \quad z_{i,t} := F_{i,t}^{VII}(v_{i,t}), \quad i = 1, 2, 3, \quad \text{where} \quad \mathbf{v}_t = \hat{\mathbf{S}}_1^{-1}(t)\mathbf{r}_t,$$

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.

For an informed decision on the bandwidth to be used in the estimation of the unconditional covariance matrix (5.1), the cross-validation was run (using only the first 1000

¹⁶Cross-validation mistakes the smoothness of the series caused by positive correlation for low variability, yielding bandwidth choices usually smaller than the optimal one. While this can be disastrous in mean estimation, it is the correct type of behavior in the forecasting context since averaging over a small number of past observations is more likely to be close to the next value in the series when there are positive correlations.

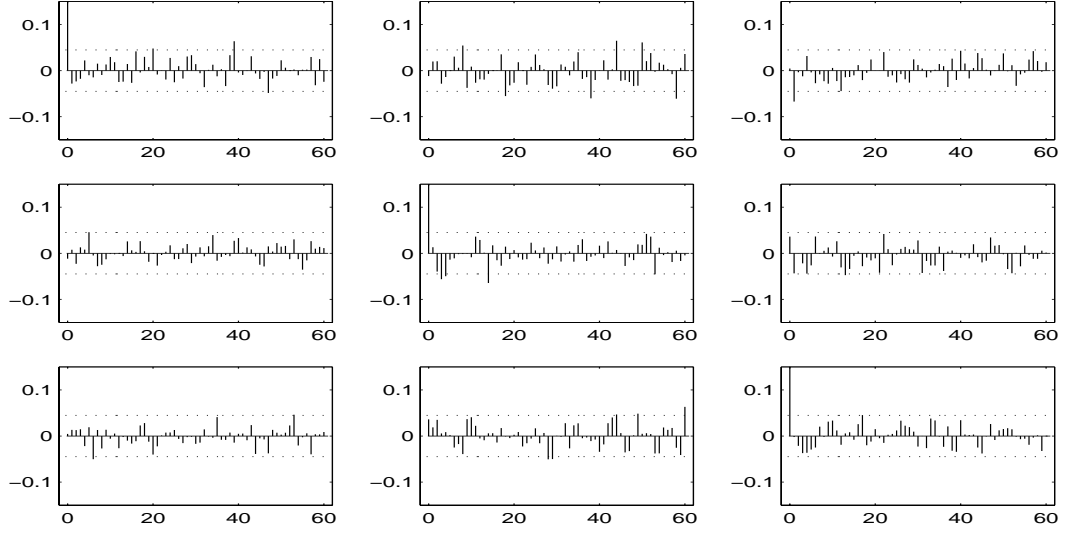


Figure 6.1. *SACF of the sequence $(\mathbf{z}_t - \bar{\mathbf{z}})$.*

observations) in the set-up of the model (3.1) with $y_k = |\mathbf{r}_k \mathbf{r}_k'|$, $k = 1, 2, \dots, n$ both for K , the symmetric Gaussian kernel and for the asymmetric \tilde{K} , $\tilde{K}(u) = K(u)1_{u \leq 0}$. The results are displayed in Figure 6.2: on the left, the graph for the Gaussian kernel K , on the right, the one for \tilde{K} . The cross-validation optimal bandwidth seem to belong to the interval $[0.0025, 0.008]$ for the Gaussian kernel and to the interval $[0.004, 0.007]$ for \tilde{K} . The empirical relationship between the two intervals of optimal bandwidth is in accordance with the equivalent kernel theory according to which the optimal bandwidths corresponding to the two kernels are related by: $h_{(\tilde{K})} = 2^{1/5} h_{(K)}$. Although in the forecasting exercise a fixed band-width, $h_\mu = 0.007$ was used, an adaptative choice is also available. A time-depending bandwidth can be obtained by running the cross-validation on the sample up to the moment when the forecast is made.

A battery of tests similar to the one in Section 5 is employed to verify the hypotheses 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 5. The same caveat on the impact of the uncertainty with respect to the estimated covariance matrix as in Section 5 applies. Figures 6.1 and 6.3

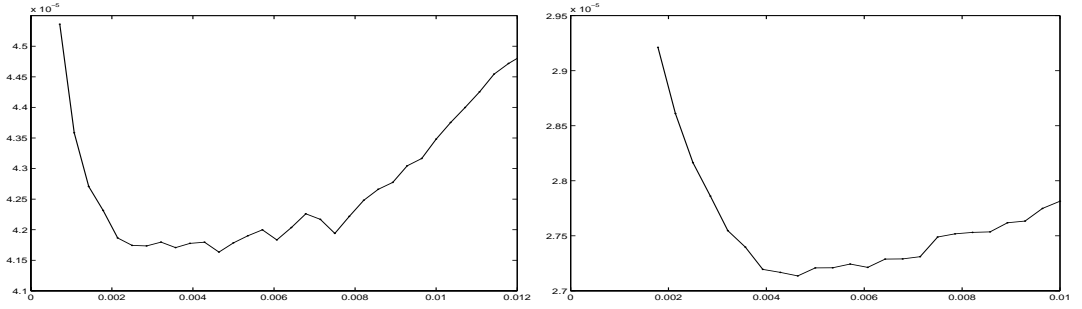


Figure 6.2. The cross validation graph for the choice of the bandwidth h_μ for $\hat{\mu}$ in (3.2) and $y_k = |\mathbf{r}_k \mathbf{r}_k'|$ using only the first 1000 observations for the Gaussian kernel (Left: $h_\mu^{(c)} \in [0.0025, 0.008]$) and the kernel \tilde{K} (Right: $h_\mu^{(c)} \in [0.004, 0.007]$).

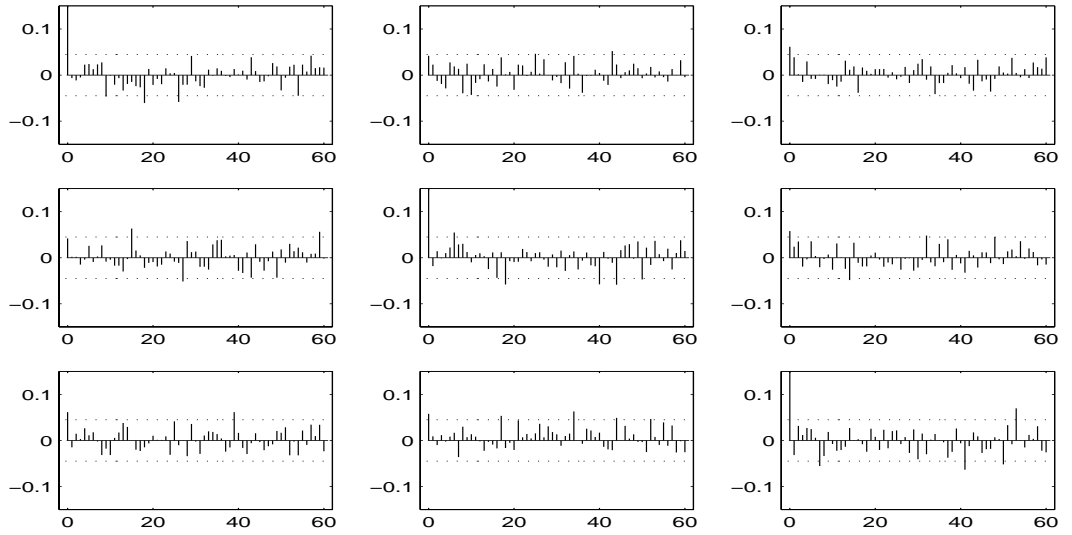


Figure 6.3. SACF of the absolute values of the sequence $(\mathbf{z}_t - \bar{\mathbf{z}})$.

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 6.4). 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 6.4. 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 6.5. 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 6.5. 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 6.6 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 6.7 support the conclusion that the marginal distributions of the three sequences $(z_{i,t})$, $i = 1, 2, 3$ are uniform $(0,1)$.

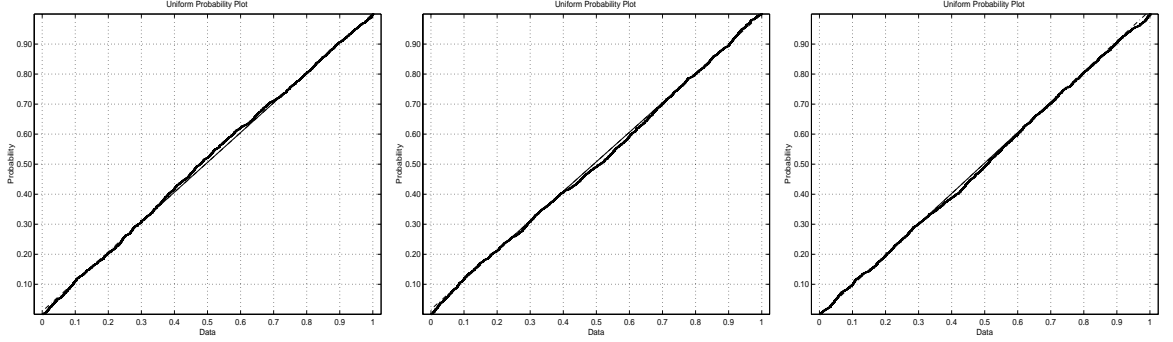


Figure 6.6. The uniform probability plot of the three series of coordinates z_i , $i = 1, 2, 3$, $t = 1, \dots, 1926$.

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

Table 6.7. p -values for the Andersen-Darling and Kolmogorov-Smirnov tests of uniformity applied to the coordinates of the sequence (\mathbf{z}_t) .

6.3. 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 and the approach described in Section 6.1 for forecasting horizons of one, ten and twenty days. 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 distributional forecasts for the next day vector of returns. For RiskMetrics, the distributional forecast is

$$(6.3) \quad \mathbf{r}_{t,m}^{RM} \stackrel{d}{=} N(0, m\hat{\Sigma}_t^2),$$

where

$$(6.4) \quad \hat{\Sigma}_t^2 := \sum_{i=0}^{l-1} \lambda^{t-i} \mathbf{r}_{t-i} \mathbf{r}_{t-i}' / \sum_{i=0}^{l-1} \lambda^{t-i},$$

is the exponential moving average estimate of the conditional covariance matrix Σ_t^2 . The parameters used were $\lambda = 0.94$ and $l = 120$ for one-day ahead forecasts and $\lambda = 0.97$ and $l = 200$ for ten- and twenty-day ahead forecasts (as stipulated in the RiskMetrics documents [31]). For the regression-type model (2.1), the m -day forecast $\mathbf{r}_{t,m}^{REG}$ is given by (6.1).

Note that our comparison focuses on the most common specification of the distributional forecast of RiskMetrics, i.e. that where the future returns are jointly normal. We chose this specification due to the fact that it is widely used in practice. Comparisons with other specifications (normal mixture models, GED models) are currently under investigation and the results will be reported elsewhere.

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

As explained in Diebold *et al.* [6], evaluating the correct distributional forecast $F_{mt,m}$ at the realized portfolio returns $r_{mt+1,m}^{(\mathbf{w})}$, $t = 1, \dots, [n/m] - 1$ yields an iid sequence $(F_{mt,m}(r_{mt+1,m}^{(\mathbf{w})}))$ of uniform (0,1) random variables. Hence the quality of a distributional forecast $G_{mt,m}$ can then be assessed by testing the hypothesis

$$(6.5) \quad H_0: (G_{mt,m}(r_{mt+1,m}^{(\mathbf{w})})) \text{ is an iid sequence with uniform (0,1) marginal distribution.}$$

In the sequel we test hypothesis (6.5) for $G_{t,m} = F_{t,m}^{RM}$ and $G_{t,m} = F_{t,m}^{REG}$. More concretely the sequences $(F_{mt,m}^{RM}(r_{mt+1,m}^{(\mathbf{w})}))$ and $(F_{mt,m}^{REG}(r_{mt+1,m}^{(\mathbf{w})}))$, $t = 1, \dots, [n/m] - 1$ 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.

For every horizon ($m = 1, m = 10, m = 20$) 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 6.2, 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 ¹⁷ ($F_{mt,m}^{RM}(r_{mt+1,m}^{(\mathbf{w})})$) and ($F_{mt,m}^{REG}(r_{mt+1,m}^{(\mathbf{w})})$), $t = 1, \dots, [n/m] - 1$, 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.

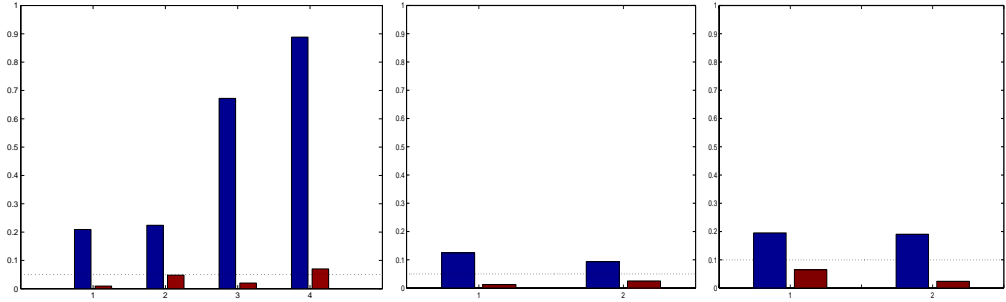


Figure 6.8. *The percentage of the p -values for the K-S (1), A-D (2), the L-B at lag 10 (3) and the variance test (4) that are smaller than 5% (Left and Center) and 10% (Right). For a given test, the first bar concerns the RiskMetrics methodology while the second one refers to the forecasting methodology described in Section 6.1. Left: One-day ahead, Center: Ten-day ahead, Right: Twenty-day ahead.*

The results of these simulations are summarized in Figure 6.8 where the percentage of p -values smaller than 5% for $m = 1$, $m = 10$ and than 10% for $m = 20$ is reported ¹⁸. For a given test, the first bar concerns RiskMetrics while the second one refers to the forecasting methodology described in Section 6.1. It is interesting to notice that, for one-day ahead forecasting, for almost 90% of the portfolios, plain-vanilla RiskMetrics fails (at the 5% level) the variance test. This should be compared to the 94% acceptance rate

¹⁷The sequence of the m -days ahead forecast would exhibit an intertemporal $m - 1$ dependence. To keep the independence between observations, necessary for the statistical tests, we restricted the sample to the sub-sequence of m -days apart forecasts.

¹⁸For $m = 20$, a higher percentage of 10 has been used due to the small number of observations in the sequences ($G_{mt,m}(r_{mt+1,m}^{(\mathbf{w})})$), $t = 1, \dots, [n/m] - 1$.

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 fails at least one of the four tests in 94% of the cases compared to only 9% for our methodology.

For $m = 10$ and $m = 20$, the empirical percentage of p -values for the last two tests were, for both methods, below the fixed theoretical level of 5% for ten-day forecasts and 10% for the twenty-day forecasts¹⁹. For the variance test, this is not surprising, since, due to averaging, for ten- and twenty-day returns, the multivariate normality assumption of RiskMetrics is more adequate than for daily returns. However, the normality tests show once again the superiority of our methodology over the plain-vanilla RiskMetrics²⁰. As mentioned before, comparisons involving other RiskMetrics specifications of the conditional distribution are under study and the results will be reported elsewhere.

7. RISKMETRICS VS. NON-PARAMETRIC REGRESSION

We conclude with a few remarks on the relationship between our approach and RiskMetrics. Univariately, the probabilistic model that forms the basis of RiskMetrics forecasting methodology outlined in (6.3) and (6.4) is the following conditional, multiplicative process

$$(7.6) \quad r_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1),$$

(see page 73 of [31]) where

$$(7.7) \quad \sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2,$$

according to Section B.2.1 of the Appendix B of [31]. 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.

¹⁹For this reason, they are not reported in the graphs in the center and right of Figure 6.8.

²⁰This forecasting methodology has been thoroughly investigated in the univariate case in a companion paper by Drees and Stărică [8]. There the authors show by the example of the S&P 500 time series of returns that this apparently structureless forecasting methodology outperforms conventional GARCH-type models both over one day and over time horizons of up to forty days.

From a probabilistic point of view, the model (7.6) and (7.7) is faulty. Results by Kesten [19] and Nelson [26] imply that a time series evolving according to the dynamics (7.6) and (7.7) will tend to 0 almost surely.

The claimed close relationship between the RiskMetrics methodology and GARCH-type models, prompted by the deceiving formal analogy between the GARCH(1,1) specification

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

and (7.7) and emphasized by the comparisons in Section 5.2.3 of ([31]), is hence misleading. Instead, the RiskMetrics approach can be motivated by the non-stationary model (2.1).

Note that the forecast (6.4) is just a kernel smoother of the type (5.1) with an one-sided exponential kernel $K_{exp}(x) = a^x 1_{[-\infty, 0]}(x)$, $a = \lambda^m$ and $h = 1/m$. Our experience shows that replacing the normal kernel with the exponential leads to results very similar to the ones reported in Section 6. 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 5 and 6 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.

Besides providing a solid statistical framework, the set-up of the non-stationary paradigm introduced in Section 2 allows for a optimal choice of the bandwidth, motivated by results from the statistical theory of curve estimation. By contrast, the choice of the parameters λ and l is empirical.

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 [31]. 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 are 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.

8. CONCLUSIONS

In this paper a simple multivariate non-stationary paradigm for modeling and forecasting the distribution of returns on financial instruments is discussed.

Unlike most of the multivariate econometric models for financial returns, our approach supposes the volatility to be exogenous. The vectors of returns are assumed to be independent and to have a changing unconditional covariance structure. The methodological frame is that of non-parametric regression with fixed equidistant design points where the regression function is the evolving unconditional covariance. The vectors of standardized innovations have independent coordinates and asymmetric heavy tails and are modeled parametrically. The use of the non-stationary paradigm is exemplified on a tri-variate sample of risk factors consisting of a foreign exchange rate Euro/Dollar (EU), an index, FTSE 100 index, and an interest rate, the 10 year US T-bond. The paradigm provides both a good description of the changes in the dynamic of the three risk factors and good multivariate distributional forecasts.

We believe that the careful parametric modeling of the extremal behavior of the standardized innovations makes our approach amenable for precise VaR calculations. Evaluating its behavior in these settings is, however, subject of further research.

REFERENCES

- [1] BARONE-ADESI, G., GIANNPOULOS, K., VOSPER, L. (1999) VaR without correlations for portfolio of derivative securities. *Journal of Futures Markets*, **19**, 583–602.
- [2] CAMPBELL, J., LO A., AND MACKINLAY, A. (1996) *The Econometrics of Financial Markets*. Princeton University Press.
- [3] CAI, J. (1994) A Markov model of unconditional variance in ARCH. *J. Business and Economic Statist.* **12**, 309–316.
- [4] DIEBOLD, F.X. (1986) Modeling the persistence of the conditional variances: a comment. *Econometric Reviews* **5**, 51–56.
- [5] DIEBOLD, F.X. AND INOUE, A. (2001) Long memory and regime switching. *J. Econometrics* **105**, 131–159.
- [6] DIEBOLD, F.X., GUNTHER, T. AND TAY, A. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, **39**, 863–883.
- [7] 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.

- [8] DREES, H. AND STĂRICĂ, C. (2002) A simple non-stationary model for stock returns. Available at <http://www.math.chalmers.se/~starica/>
- [9] FAN, J., JIANG, J., ZHANG C. AND ZHOU Z. (2002) Time-dependent Diffusion Models for Term Structure Dynamics and the Stock Price Volatility, *Statistica Sinica*, forthcoming Available at <http://www.stat.unc.edu/faculty/fan>
- [10] FAN, J. AND YAO, Q. (1998). Efficient Estimation of Conditional Variance Functions in Stochastic Regression. *Biometrika*, **85**, 645–660.
- [11] GASSER, T., KNEIP, A., KÖHLER, W. (1991) A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.*, **86**, 643–652.
- [12] GASSER, T., MÜLLER, H.-G., MAMMITZSCH, V. (1985) Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B*, **47**, 238–252.
- [13] GIJBELS, I., POPE, A. AND WAND, M. P. (1999) Understanding exponential smoothing via kernel regression. *J. R. Stat. Soc. Ser. B*, **61**, 39–50.
- [14] GRANGER, C.W.J., HYUNG, N. (1999) Occasional structural breaks and long-memory. Discussion paper 99-14, University of California, San Diego.
- [15] GRANGER, C., W., TERÄSVIRTA, T. (1999) A simple non-linear time series model with misleading linear properties. *Economics Letters* **62**, 161–165.
- [16] HIDALGO, J. AND ROBINSON, P.M. (1996) Testing for structural change in a long-memory environment. *J. Econometrics* **70**, 159 – 174.
- [17] HSU, D.A., MILLER, R., AND WICHERN, D. (1974) On the stable Paretian behavior of stock-market prices. *J. of American Statistical Association* **69**, 108–113.
- [18] KENDALL, M. AND STUART, A. (1979). The advanced theory of statistics. Charles Griffin, London.
- [19] KESTEN, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.
- [20] LOBATO, I. N. AND SAVIN, N. E (1998). Real and Spurious Long-Memory Properties of Stock-Market Data *Journal of Business & Economic Statistics*, 261–268.
- [21] 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.
- [22] 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.
- [23] MIKOSCH, T. AND STĂRICĂ C. (2004) Non-stationarities in financial time series, the long range dependence and the IGARCH effects, with Mikosch, T., *Review of Economics and Statistics*, **86**, 378–390.
- [24] MÜLLER, H., G. (1988) Nonparametric Regression Analysis of Longitudinal Data. Springer, Berlin.
- [25] MÜLLER, H.-G. AND STADTMÜLLER, U. (1987) Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610–625.
- [26] NELSON, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.
- [27] OFFICER, R. (1971) A time series examination of the market factor of the New York Stock Exchange. Ph.D. Dissertation. University of Chicago.
- [28] RODRÍGUEZ-POO, J. M. AND LINTON, O. (2001). Nonparametric factor analysis of residual time series. *Test*, **10**, 161–182.
- [29] NAGAHARA, Y., AND KITAGAWA, G. (1999). A Non-Gaussian Stochastic Volatility Model. *The Journal of Computational Finance*, **2**, 33–47.
- [30] RICE, J. (1984). Boundary modification for kernel regression. *Comm. Statist. A—Theory Methods* **13**, 893–900.
- [31] RiskMetricsTM. Technical Document. Available at <http://www.riskmetrics.com/techdoc.html>.
- [32] SHORACK, G. R. AND WELLNER J. A. (1986) Empirical processes with applications to statistics. John Wiley & Sons, New York.

- [33] SIMONATO, J., G. (1992) Estimation of GARCH processes in the presence of structural change. *Economic Letters* **40**, 155–158.
- [34] STĂRICĂ C. (2003). Is Garch(1,1) as good a model as the accolades of the Noble prize would imply?, available at <http://www.math.chalmers.se/~starica/>
- [35] STĂRICĂ C. AND GRANGER, CLIVE (2004) Non-stationarities in stock returns. To appear in *Review of Economics and Statistics*.
- [36] STOCK, J. AND WATSON, M. (1996) Evidence on structural instability in macroeconomic time series relations. *J. Business and Economic Statist.*, 1–10.
- [37] WAND, M.P. AND JONES, M.C. (1995) *Kernel Smoothing*. Chapman and Hall, London.