VILNIUS UNIVERSITY

Milda Pranckevičiūtė

High frequency data aggregation and Value-at-Risk

Doctoral dissertation

Physical sciences, Mathematics (01P)

Vilnius, 2011

The scientific work was carried out during 2006-2010 at Vilnius University

Scientific supervisor:

Prof. Dr. Habil. Alfredas Račkauskas (Vilnius University, Physical Sciences, Mathematics – 01P)

VILNIAUS UNIVERSITETAS

Milda Pranckevičiūtė

AUKŠTO DAŽNIO DUOMENŲ AGREGAVIMAS IR VERTĖS POKYČIO RIZIKA

Daktaro disertacija

Fiziniai mokslai, matematika (01P)

Vilnius, 2011

Disertacija rengta 2006-2010 metais Vilniaus universitete

Mokslinis vadovas:

Prof. habil. dr. Alfredas Račkauskas (Vilniaus universitetas, fiziniai mokslai, matematika – 01P)

Contents

Introduction

Risk management has become one of the most important tasks for financial institutions in recent years. The global financial crisis drew even more attention to the issues of risk measurement. An accurate estimation of risk exposure is highly important to financial institutions since the appropriate risk quantification is the basis for managing possible future losses and keep adequate capital. Financial institutions hold a risky portfolio consisting of financial assets, such as equity, bonds, foreign exchange, commodities or derivative securities. They face market risk arising due to unknown future price changes in their portfolio financial assets. Value-at-Risk (VaR) has been the most popular methodology to quantify market risk since 1996, when the Bank for International Settlements adopted an amendment to the Capital Accord allowing the use of internal models to estimate risk and to calculate capital requirements. VaR is a statistical model defined as the maximum future loss due to likely changes in the value of financial assets portfolio during a certain period with a certain probability. The estimate of risk obtained by the VaR model can be applied both to regulatory requirements in the calculation of capital adequacy and management of portfolio exposure risk.

The increasing volume of data in financial markets and a fast development of information technologies influenced the accessibility of high frequency data. Such data sets consist of the so-called "ticks" containing information about the financial market activity (price, volume, trader, etc.) and the time moment this information was recorded. "Tick-by-tick" data began to be collected in the early eighties. Soon the first empirical studies appeared whilst analyzing high frequency data behavior and stylized facts (see, for example, Goodhart and Figliuoli (1991), Zhou (1993)). Later, Engle (2000) introduced the definition of ultra high frequency data trying to emphasize that such data sets contain a full record of transactions and their associated characteristics, and it is not possible to access any more information. The main features of tick-by-tick data series are a huge number of observations, a random time interval between two subsequent events as well as a random number of daily observations. The analysis of high frequency data is complex since econometric theory is specified for regularly spaced data. There are mainly two possible ways to deal with high frequency observations. The first one is a "tick-by-tick" analysis. Special models are developed to treat randomly spaced data. Extensive information about handling high frequency data is summarized in Dacorogna *et al.* (2001). The other way is data regularization, where "tick-by-tick" observations are aggregated to obtain regularly spaced data series. In this thesis, the aggregation of high frequency data is considered.

Aims and problems. The main topic of the thesis is the data aggregation problem in risk measurement. We consider the Value-at-Risk model, as a tool to estimate the market risk. The following objectives are formulated to analyze data aggregation problem in VaR models:

- Define an aggregated VaR model and illustrate the VaR estimator dependence on the choice of the data aggregation method.
- Construct a functional GARCH model with univariate volatility and analyze its properties.
- Introduce a functional GARCH model in the Hilbert space and analyze its properties.
- Present the data aggregation problem from the view of stylized facts of high frequency returns.

Methods. The methods of advanced probability theory, statistics and functional analysis are applied.

Novelty. The new approach of using high frequency aggregated data to estimate VaR as a daily measure of risk is presented in the thesis. In relation, two new functional GARCH type models are introduced to model volatility of functional risk factors: a $\rho-\mathrm{GARCH}(1,1)$ model with volatility dependent on some features of functional returns and a Hilbert space valued $\mathrm{GARCH}(1,1)$ model with univariate volatility.

Maintaining statements.

- The aggregated Value-at-Risk model was defined and model estimator dependence on data aggregation was analyzed, taking high frequency foreign exchange rates.
- A functional $\rho-\mathrm{GARCH}(1,1)$ model, depending on some features of functional data, was constructed. The existence of a stationary solution and the consistency of maximum likelihood estimators of model parameters

were proved. Several examples with the known aggregated returns density function were given.

- The Hilbert space-valued GARCH$(1,1)$ with univariate volatility model was introduced. The existence of a stationary solution, the consistency and the asymptotic normality of quasi-maximum likelihood estimators of model parameters were proved; the asymptotic properties of residuals were analyzed.

- The dependence of the Hurst exponent, as a long memory parameter, on data aggregation was researched taking absolute returns of foreign exchange rates.

Main results. Let $\{(\tau_j, y_j)\}_{j=1}^{N}$ be an irregular time series, where τ_j and y_j indicate respectively the time and the value of the j'th observation. Fix a time interval between two observations at $\delta > 0$, and let $\tau_t^* = t\delta$, $t = 1, \ldots, N^*$. Using an appropriate aggregation scheme g one defines the regular time series

$$y_t^* = y_t^*(g) = g(\{(\tau_j, y_j), \tau_j \in (\tau_{t-1}^*, \tau_t^*]\}), \quad t = 1, \ldots, N^*.$$

This basically implies that the aggregated observation value y_t^* is constructed using information available from the moment τ_{t-1}^* to the moment τ_t^*. Note that the dimension of the aggregation g in the definition is not fixed; therefore both finite and infinite dimensional aggregation schemes can be used. For example, Brownlees and Gallo (2006) suggested several univariate aggregation rules, such as taking the first, the last, the maximum, the minimum or the sum of the values y_j in the interval $(\tau_{t-1}^*, \tau_t^*]$. Additionally, the methods based on the interpolation at τ_t^* of the previous and next observation in data series can be chosen (see, e.g., [26]). In this case, when the aggregation produces univariate time series, the standard econometric theory can be applied. Furthermore, one might construct functional observations from high frequency data. Ramsay and Silverman (1997) introduced several techniques for converting raw data into a functional form, such as basis functions methods, smoothing by local weighing, and the roughness penalty approach. The direct constructions of functional data can be used as well. For example, the consecutive maximal values of high frequency observations produce the non-decreasing functions,

$$y_t^*(s) = \max\{y_j | \tau_j \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}, \quad s \in [0,1].$$

Aggregated functional observations can be analyzed applying functional data models.

It is agreed that high frequency data might improve the quality of risk model estimates. However, there are only a few studies about the VaR estimation using high frequency data. One part of papers (see, e.g., [13],[43]) concerning high frequency data VaR analyze regularized data and apply standard VaR models. The other part of publications (see, e.g., [22], [31]) develop special VaR models for tick-by-tick observations. High frequency data VaR models, usually based on 5 to 30 minutes returns, are applied to intraday risk management purposes. The estimates of VaR obtained by such models enable us to manage the exposure of risky positions or the whole portfolio of risky assets during the day. To our knowledge, high frequency data VaR models are not applied to measure market risk for capital adequacy calculation purposes. The reason of this is the Capital Accord, adopted by the Bank for International Settlements (2006), where the requirements to the VaR model are stipulated, indicating that *"Value-at-Risk must be computed on a daily basis"*. In general, daily data are obtained either taking the closing or the last price of the day (equity markets) or fixing the price at a certain moment or period of the day (foreign exchange markets). For example, in the Bloomberg[1] system the daily foreign exchange rates are fixed at the end of the day, while the Bank of Lithuania sets official foreign exchange rates for the next day according to foreign exchange rates observed at around 10 a.m. local time. Obviously, both daily foreign exchange rates - one taken from the Bloomberg system and the other one provided by the Bank of Lithuania - are different. Furthermore, daily data can be obtained taking the maximum, the minimum, the average financial asset price during the day, or even applying more complex data aggregation methods. However, the official requirements to VaR models do not include any statement about aggregation.

Aggregated VaR. In the thesis, a new concept of aggregated VaR was introduced following the classical VaR definition, given in [34]. Consider aggregated financial assets prices $p_t(g)$, where the aggregation rule g is from the class $\mathcal{G}$. Denote by $\tau_t := t\delta$ the sequence of regular data series time moments and suppose f is a mapping function. Having taken logarithms of financial assets aggregated prices, the portfolio loss can be written as

$$L_{t+1}(g) = -[f\{\tau_{t+1}, \mathbf{Z}_t(g) + \mathbf{X}_{t+1}(g)\} - f\{\tau_t, \mathbf{Z}_t(g)\}],$$

[1]Bloomberg is a major global provider of 24-hour financial news and information including real-time and historic price data, financials data, trading news and analyst coverage, as well as general news and sports. Its services, which span their own platform, television, radio and magazines, offer professionals analytic tools.(*www.investopedia.com*)

here $\mathbf{X}_{t+1}(g) := \mathbf{Z}_{t+1}(g) - \mathbf{Z}_t(g)$ are changes of risk factors with the risk factors expressed as $\mathbf{Z}_t(g) = (Z_{t,1}(g), ..., Z_{t,d}(g)) = (\ln p_{t,1}(g), ..., \ln p_{t,d}(g))$. Denote the portfolio of the financial assets loss distribution function as $F_{L(g)}$, $g \in \mathcal{G}$. Assume that the confidence level $\alpha \in (0,1)$. The portfolio aggregated VaR at a fixed confidence level α is given by the smallest number l such that the probability that the loss $L(g)$, $g \in \mathcal{G}$ exceeds l is not larger than $(1-\alpha)$ over the time horizon $\delta \geq 0$:

$$VaR_\alpha(g) = inf\{l \in \mathbb{R} : P(L(g) > l) \leq 1 - \alpha\} = inf\{l \in \mathbb{R} : F_{L(g)}(l) \geq \alpha\}.$$

According to this definition, VaR not only depends on the confidence level α and the holding period δ, but also on the aggregation rule g from a given class $\mathcal{G}$.

The empirical study was performed taking high frequency foreign exchange rates to illustrate the market risk estimator's dependence on data aggregation. Foreign currencies - the US dollar (USD) versus the euro (EUR), the British pound (GBP) and the Japanese yen (JPY) - were taken to calculate VaR. The following aggregation rules were chosen for analysis:

- pointwise aggregation

$$p_t^{DAILY}(s) = \{p_i | \tau_i = \max\{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}\}, \quad s \in [0,1],$$

- maximum value aggregation

$$p_t^{MAX}(s) = \max\{p_i | \tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}, \quad s \in [0,1],$$

- minimum value aggregation

$$p_t^{MIN}(s) = \min\{p_i | \tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}, \quad s \in [0,1],$$

- average value aggregation

$$p_t^{AVE}(s) = \frac{1}{m_t(s)} \sum_{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]} p_i, \quad s \in [0,1],$$

  where

$$m_t(s) = \#\{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}.$$

Here the class of aggregation rules $\mathcal{G}$ is taken as an interval $[0,1]$ and the number s denotes an aggregation rule. For example, taking the maximal or average values for each time moment $s \in [0,1]$, we obtain the corresponding aggregation rule.

There is a large variety of different methodologies to calculate VaR (see, e.g., [38]). One of the most popular methods to estimate VaR, *historical simulation* methodology, can be easily generalized for aggregated data. Historical simulation aggregated VaR can be written as

$$VaR_{\alpha}(s) = q_{\alpha}(F_{L(s)}), \quad s \in [0, 1],$$

where $F_{L(s)}$ denotes the empirical distribution of the foreign exchange position loss and the symbol q_{α} denotes the α quantile of the empirical distribution.

Applying the VaR model and taking high frequency foreign exchange rates and daily foreign exchange rates, the market risk was estimated (the maximum possible loss over one day due to rate fluctuations). The empirical study has showed that the possible loss of a financial institution depends on the chosen aggregation rule.The presented analysis not only shows the difference of the risk estimate, depending on the chosen aggregation scheme, but also clearly indicates that the estimate of risk, calculated by using daily foreign exchange rates represents only a small part of the view what happens during the whole day.

According to the official requirements of the Bank for International Settlements, daily historical data are sufficient to estimate the market risk. However, such calculations are based only on a very small part of information available during the day. Therefore the market risk of a financial institution is measured inaccurately. The use of aggregated data in VaR models would account for the whole information observable in the markets during the day and let us estimate the risk more accurately.

Functional GARCH. In the second chapter of the thesis, the functional $\rho-$ GARCH$(1, 1)$ model has been developed. Consider a functional time series $(X_t, t \in \mathbb{Z})$, where for each t, $X_t = (X_t(g), g \in \mathcal{G})$ is a random function defined on the set $\mathcal{G}$. We assume all random elements to be defined on a fixed probability space $(\Omega, \mathcal{F}, P)$. We also assume, that for each t, $X_t \in \mathbb{E} \subset \mathbb{R}^{\mathcal{G}}$, is an $\mathbb{E}$-valued random element, where $\mathbb{E}$ is a separable topological vector space endowed with its Borel σ-field. Let $\rho : \mathbb{E} \to \mathbb{R}$ be a measurable semi-norm.

The process $(X_t, t \in \mathbb{Z})$ is defined as a functional $\rho-$ GARCH$(1, 1)$ model if, for each $g \in \mathcal{G}$ and $t \in \mathbb{Z}$, it satisfies

$$X_t(g) = \sigma_t \varepsilon_t(g),$$
$$\sigma_t^2 = \omega + \alpha\rho^2(X_{t-1}) + \beta\sigma_{t-1}^2,$$

where $(\varepsilon_t(g), g \in \mathcal{G}), t \in \mathbb{Z}$ are independent identically distributed random functions.

The strong stationarity (see Theorem 2.1) and the 2nd order stationarity (see Theorem 2.2) of the functional $\rho - \mathrm{GARCH}(1,1)$ model were proved following [40].

To estimate model parameters, the maximum likelihood method was chosen. The maximum likelihood estimator of the true parameters vector $\nu_0 = (\theta_0, \lambda_0)^T$, $\theta_0 = (\omega_0, \alpha_0, \beta_0)^T$ is obtained by maximizing the likelihood function:

$$\widehat{\nu}_n = \operatorname*{argmax}_{\nu \in K \times M} \widehat{L}_n(\nu),$$

and the notation is provided in section 2.4. The consistency (see Theorem 2.3) of the maximum likelihood estimator $\widehat{\nu}_n$ of $\rho - \mathrm{GARCH}(1,1)$ model parameters was proved following Theorem 6.1.4. in [71].

At the end of the chapter, several examples were given where the density function of random elements $\rho(\varepsilon_t)$ was known. The consistency of maximum likelihood parameter estimators was proved checking the conditions of Theorem 2.3.

Example 0.1. Assume that the class $\mathcal{G} = [0,1]$, i.e., each $s \in [0,1]$, corresponds to a certain aggregation rule. Consider the $C[0,1]$ valued time series $(X_t, t \in \mathbb{Z})$ with functional elements obtained applying the aggregation rule and expressed as

$$\begin{aligned} X_t(s) &= \sigma_t \varepsilon_t(s), \\ \sigma_t^2 &= \omega + \alpha \Big(\max_{s \in [0,1]} X_{t-1}(s) \Big)^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

where $(\varepsilon_t(s), s \in [0,1]), t \in \mathbb{Z}$ are independent identically distributed Wiener processes.

This example corresponds to the functional $\rho - \mathrm{GARCH}(1,1)$ model, where $\mathbb{E} = C[0,1]$ and the semi-norm $\rho(f) = \max_{0 \le s \le 1} f(s)$. Since the density function of $\rho(\varepsilon_t) = \max_{s \in [0,1]} \varepsilon_t(s)$ is well known, we can express the maximum likelihood function and show that the conditions of Theorem 2.3 are satisfied. Therefore, the maximum likelihood estimator of model parameters is consistent.

Example 0.2. Consider the $C[0,1]$-valued time series $(X_t, t \in \mathbb{Z})$, where

$$\begin{aligned} X_t(s) &= \sigma_t \varepsilon_t(s), \\ \sigma_t^2 &= \omega + \alpha \min_{s \in [0,1]} X_{t-1}^2(s) + \beta \sigma_{t-1}^2, \end{aligned}$$

where $(\varepsilon_t(s), s \in [0,1]), t \in \mathbb{Z}$ are independent identically distributed Wiener processes.

So this example corresponds to the functional $\rho-$GARCH$(1,1)$ model, where $\mathbb{E} = C[0,1]$ and the semi-norm $\rho(f) = \min_{0\leq s\leq 1} f(s)$. The consistency of the maximum likelihood estimator of model parameters is proved similarly as in the first example.

Example 0.3. In this example, we take the case, where the class $\mathcal{G} = \{g_1, \ldots, g_d\}$ consists of d aggregation rules and consider the R^d-valued time series $(X_t, t \in \mathbb{Z})$:

$$X_{jt} = \sigma_t \varepsilon_{jt}, \quad j = 1, \ldots, d,$$
$$\sigma_t^2 = \omega + \alpha \rho^2(X_{t-1}) + \beta \sigma_{t-1}^2,$$

where $(\varepsilon_{jt}, j = 1, \ldots .d), t \in \mathbb{Z}$ are independent identically distributed Gaussian random vectors with zero mean and the covariance matrix $\Lambda = (\lambda_{ij}, i, j = 1, \ldots, d)$ and ρ is a semi-norm on R^d. As an example, we consider the semi-norm $\rho(x) = \left(\sum_{i=1}^d x_i^2\right)^{\frac{1}{2}}$ and assume that $\lambda_{ii} = 1, i = 1, \ldots d$. Then $\rho^2(\varepsilon_{jt}) = \sum_{i=1}^d \varepsilon_{jt}^2$, $t \in \mathbb{Z}$ has χ^2-distribution. Therefore we can express the maximum likelihood function and prove the consistency of the model parameter estimator.

Example 0.4. In this example, we take a point-wise aggregation. Consider the $C[0,1]$ valued time series $(X_t, t \in \mathbb{Z})$, where

$$X_t(s) = \sigma_t \varepsilon_t(s),$$
$$\sigma_t^2 = \omega + \alpha (X_{t-1}(1) - X_{t-1}(0))^2 + \beta \sigma_{t-1}^2,$$

where $(\varepsilon_t(s), s \in [0,1]), t \in \mathbb{Z}$ are independent identically distributed Gaussian processes.

The example corresponds to the functional $\rho-$ GARCH$(1,1)$ model, where $\mathbb{E} = C[0,1]$ and the semi-norm $\rho(f) = |f(1) - f(0)|$. This case has also a practical explanation, when the returns of financial asset prices are analyzed. Assume, e.g., that the returns of a share, traded on the stock exchange, are taken. From this model equations we can see that the volatility depends on the difference between the returns taken at the beginning and at the end of the day, i.e., the opening price and the closing price returns.

The density function of $[\varepsilon_t(1) - \varepsilon_t(0)]$ is well known, so we can derive the maximum likelihood function to estimate the model parameters and to verify that the conditions of Theorem 2.3 are satisfied.

GARCH model in the Hilbert space. Let $\mathbb{H}$ be a real separable Hilbert space of infinite or finite dimensions with the inner product $\langle\cdot,\cdot\rangle$ and the corresponding norm $||\cdot||$, $||x||^2 = \langle x, x\rangle$, $x \in \mathbb{H}$. Let $(X_t, t \in \mathbb{Z})$ be an $\mathbb{H}$-valued random process, $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$.

We say that (X_t) is GARCH$(1,1)$ with univariate volatility (uvGARCH$(1,1)$), if

$$X_t = \sigma_t \varepsilon_t, \quad (\varepsilon_t) \sim \text{iid } (0, Q_\varepsilon),$$
$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \langle X_{t-1}, z\rangle^2, \quad t \in \mathbb{Z},$$

where $\omega > 0, \beta \geq 0$ and $z \in \mathbb{H}$ are parameters of interest. We also assume that $(\varepsilon_t, t \in \mathbb{Z})$ are independent identically distributed $\mathbb{H}$-valued random elements with zero mean and covariance Q_ε. It is clear, that (X_t) when projected in the direction z, namely, the time series $(\langle z, X_t\rangle)$, follows the classical GARCH$(1,1)$ model. However, the direction z is unknown.

The strong stationarity (see Theorem 3.1) and the 2nd order stationarity (see Theorem 3.3) of the Hilbert space-valued uvGARCH$(1,1)$ model have been proved using the results in [40] and [71].

The quasi-maximum likelihood approach was chosen to estimate model parameters. The quasi-maximum likelihood estimator $\widehat{\theta}_n$ is by obtained maximizing the likelihood function:

$$\widehat{\theta}_n = \text{argmax}_{\theta \in K} \widehat{L}_n(\theta),$$

and the notation is provided in Section 3.3. The consistency (see Theorem 3.4) of the quasi-maximum likelihood estimator $\widehat{\theta}_n$ was proved according Theorem 5.1.7 in [71]. The asymptotic normality (see Theorem 3.5) of the maximum likelihood estimator $\widehat{\theta}_n$ was proved following Theorem 7.2 in [40] and Theorem 5.6.1 in [71].

At the end of the chapter, the analysis of asymptotic properties of model residuals was made (see Theorem 3.6).

Aggregation and Stylized Facts. In the final chapter of the thesis, the problem of aggregation when analyzing the statistical properties of high frequency data is presented (for details of stylized facts see [23], [26] and [40]). Common statictical properties of financial assets prices have been studied for many years and such facts as *almost no autocorrelation of returns*, *non-stationarity of prices series* and *long memory* are well known. The analysis of low and high frequency returns volatility has revealed several statistical properties: *volatility clustering*, *seasonality*, and the *leverage effect*. The research of the institutional frameworks, news and other exogenous impacts also drew a lot of attention. Later on, high frequency data sets stylized facts, such as *the discreteness of quoted bid-ask spread*, *the short-term triangular arbitrage* and *scaling laws*, were summarized. When working with the aggregated high frequency data an important question arrises, whether the aggregation changes the basic stylized facts and how. To our knowledge, such an analysis has not been made as yet.

Long memory, as one of the basic stylized facts, was chosen for an empirical study. According to [56] the covariance stationary process, y_t is said to exhibit a long memory if the following condition is satisfied

$$\sum_{k=n}^{-n} |\rho_k| \to \infty, \, n \to \infty,$$

where ρ_k is the autocorrelation function at lag k. Applying the R/S statistic [49], the Hurst exponent was estimated and if its values fall into the interval $0,5 < H < 1$, it is said that the process exhibits a long-range dependence. The same aggregation rules as in Chapter 1 were taken: pointwise, maximum value, minimum value, and average value aggregation. The empirical study performed for foreign currencies (USD, EUR, GBP, JPY) has confirmed the widely known stylized fact that absolute returns have a long memory. However, it has also showed that the Hurst exponent depends on the applied aggregation rule and fluctuates all the day. This fact should be considered when modeling functional returns and applying long memory models.

Publications and presentations.

The main results are published in the following articles:

1. Pranckevičiūtė M. and Račkauskas A. Hilbert space valued GARCH with univariate volatility. *Vilnius University Faculty of Mathematics and Informatics Preprint 2011-06*, (2011).

2. Pranckevičiūtė M. and Račkauskas A. GARCH models depending on data aggregation. *Vilnius University Faculty of Mathematics and Informatics Preprint 2011-05*, (2011).

3. Pranckevičiūtė M. High Frequency Data Aggregation in Value-at-Risk Models: is Daily Data Enough? *FindEcon Monograph Series: Advances in Financial Market Analysis Number 10*, *accepted*.

4. Pranckevičiūtė M. High Frequency Data Aggregation in Historical Value-at-Risk Models. *Pinigų studijos*, 2 (2010) 42–52.

5. Pranckevičiūtė M. Long Memory in High Frequency FX Rates: Hurst Exponents Dependence on Data Aggregation. *Liet. Mat. Rink. 51*, (spec. nr.) (2010), 357–361.

Several presentations at conferences were given on the topics of the thesis:

1. Pranckevičiūtė M. *High Frequency Data Aggregation in Value-at-Risk Models.* 10th International Vilnius Conference on Probability and Mathematical Statistics, Vilnius University, 28 June - 02 July 2010, Vilnius, Lithuania.

2. Pranckevičiūtė M. *Long Memory in High Frequency FX Rates: Hurst Exponents Dependence on Data Aggregation.* LI Conference of the Lithuanian Mathematical Society, Šiauliai University, 17-18 June 2010, Šiauliai, Lithuania.

3. Pranckevičiūtė M. *Ultra High Frequency Data Aggregation in Value-at-Risk Models.* 9th Annual Conference, Forecasting Financial Markets and Economic Decision-Making (FindEcon), University of Lodz, 13-15 May 2010, Lodz, Poland.

4. Pranckevičiūtė M. *Functional Data Analysis of Foreign Exchange Rates.* L Conference of the Lithuanian Mathematical Society, Vilnius University Institute of Mathematics and Informatics, 18-19 June 2009, Vilnius, Lithuania.

Structure of the thesis.

The thesis consists of an introduction, four chapters, general conclusions, two appendices and the bibliography.

- Chapter 1 is designated to define the aggregated Value-at-Risk model. The standard VaR concept is explicitly introduced and the most common VaR method - historical simulation is presented. A new definition of the aggregated VaR is given, and the empirical study on the foreign exchange rate (USD versus EUR, GBP, JPY) position VaR estimators' dependence on the data aggregation functions (pointwise, maximum value, minimum value and average value) is provided.

- At the beginning of Chapter 2, pointwise GARCH models are considered in the context of risk measurement purposes. Next, a functional $\rho-\text{GARCH}(1,1)$ model is defined and analyzed. Finally, some examples of the $\rho-\text{GARCH}(1,1)$ model taking the known density function of aggregated observations are given.

- In Chapter 3, the general Hilbert space valued time series is presented and the $\text{GARCH}(1,1)$ model with univariate volatility is investigated. The estimation of this model is considered and the asymptotic properties of quasi-maximum likelihood parameter estimators are provided.

- In Chapter 4, stylized facts of high frequency data are presented. The long memory of absolute daily foreign exchange returns is analyzed, using classical R/S statistic and estimating the Hurst exponent. The empirical study of the dependence of the Hurst index intraday value on the data aggregation rule for the USD versus EUR, GBP and JPY currencies is provided.

Chapter 1

Aggregated Value-at-Risk model

VaR methodology is based on modeling the risk factors of financial assets. Logarithmic prices are usually taken for risk factors and the returns distribution is analyzed. When the portfolio consists of highly liquid financial instruments, "tick-by-tick" data sets are available for risk analysis. Despite the common agreement that high frequency data might improve the quality of risk model estimates, there are only a few studies about VaR estimation using high frequency data. One part of papers concerning high frequency data VaR analyze regularized data and apply standard VaR models. For instance, Beltratti and Morana (1999) used half hour German mark versus the US dollar exchange rate to estimate daily and high frequency data VaR. Giot (2005) took 15 and 30 minute returns to estimate intraday VaR applying parametric (Normal, Normal GARCH, Student GARCH, RiskMetrics and high-frequency duration models) and non-parametric (empirical quantile, extreme distributions models) to three stocks traded on the New York Stock Exchange. Other authors developed special VaR models based on "tick-by-tick" data. Dione, Duchesne and Pacurar (2006) introduced a study of intraday VaR estimates calculated using the ultra-high frequency GARCH model as an extension of the framework proposed by Engle (2000). Colletaz, Hurlin, and Topkavi (2007) suggested combining Autoregressive Conditional Duration (ACD) models and a non parametric quantile estimation to model irregularly spaced intraday VaR with a stochastic forecast horizon. Most of such studies estimate VaR in short time horizons for the portfolio of risky assets management purposes. However, there is no research about using high frequency data to estimate daily VaR which would be applied to the capital adequacy calculation.

Despite the continued research of high frequency VaR models, in practice, daily VaR models are still wideused. This is also due to the fact that even the latest version of capital standards published by the Basel Committee of Bank for International Settlements in 2006 allows using daily data for risk measurement without any notice about higher frequency data. An extensive survey about standard VaR methods is provided by Engle and Manganelli (2001). They classify VaR methodologies into three categories: parametric, nonparametric and semi-parametric models. Parametric models are based on making an explicit assumption on the distribution function of risk factors and mainly deal with returns parametrization and volatility modeling. The most common approaches to estimate variance are RiskMetrics methodology [50], [58] and ARCH family models introduced by Engle (1982) and Bollerslev (1986). Nonparametric models are based on the assumption that future behavior of risk factors will be similar as in the past and the past behavior of risk factors is analyzed. The model of historical simulation is relatively simple to estimate and is widely applied in practice. Furthermore, there is a number of historical simulation method improvements. For example Boudoukh, Richardson and Whitelaw (1998) introduced a hybrid approach as a combination of RiskMetrics and historical simulation methodologies where the exponentially declining weights were applied to past returns. Hull and White (1998) proposed a volatility adjusted version of the classical historical simulation approach. The extensions of nonparametric models can also be referred as semiparametric models. In addition, there is a large group of semiparametric models using the extreme value theory to measure VaR. Applications of extreme value theory are proposed in Embrechts, Kluppelberg and Mikosch (1997) and Danielson and de Vries (2000). The models based on quasi-maximum likelihood GARCH are introduced by Diebold, Schuermann and Stroughair (1999) and McNeil and Frey (2000). The Conditional Autoregressive VaR model is proposed by Engle and Manganelli (1999) with the extended version that incorporates the extreme value theory, presented in [38]. Finally, a separate group of VaR models based on the Monte-Carlo simulation can be distinguished (see, for example, Jorion (2007)).

1.1 Standard Value-at-Risk

A standard VaR method was developed and is still widely used taking daily financial asset prices for risk estimation. Despite the huge variety of methodologies, VaR is an easy concept to interpret; therefore it is often applied to measure the potential loss of a decrease in a financial assets portfolio market value. A VaR model estimates the maximum loss due to changes in financial asset prices during the assets holding period with a chosen confidence level. Below we provide the common concept of VaR, following McNeil, Frey and Embrechts (2005), as the basis to define an aggregated high frequency data VaR model.

1.1.1 Loss distribution

Let $(\Omega, \mathcal{F}, P)$ denote the probability space of all random variables. Consider a portfolio of risky assets such as stocks, bonds, foreign exchange or commodities and denote by $V(\tau)$ the value of this portfolio at time τ. By assumption, a random variable $V(\tau)$ is observable at time τ. The portfolio loss over the period $[\tau, \tau+\delta]$ for a fixed time horizon δ is defined by $L_{[\tau,\tau+\delta]} := -(V(\tau+\delta) - V(\tau))$. The distribution of $L_{[\tau,\tau+\delta]}$ is called a *loss distribution*.

Assume that, for any generic d-dimensional stochastic process $\mathbf{U}(\tau)$, the discrete-parameter time series $(\mathbf{U}_t)_{t\in\mathbb{N}}$ is defined by setting $\mathbf{U}_t := \mathbf{U}(\tau_t)$, where $\tau_t := t\delta$. Then the portfolio loss can be expressed as

$$L_{t+1} := L_{[\tau_t,\tau_{t+1}]} = -(V_{t+1} - V_t). \tag{1.1}$$

It should be noted that an index t can vary indicating minutely, hourly, or daily data depending on the time horizon δ. For example, taking δ equal to one day, L_{t+1} represents the loss between days t and $t+1$, while V_t and V_{t+1} are portfolio values.

Risk factors

The risky assets portfolio value is usually modeled as the function of time and risk factors. Consider a d-dimensional random vector of risk factors $\mathbf{Z}_t = (Z_{t,1}, ..., Z_{t,d})^T$ and a measurable function $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. Then the portfolio value is

$$V_t := f\{\tau_t, \mathbf{Z}_t\}. \tag{1.2}$$

It is assumed that the random vector of risk factors $\mathbf{Z}_t$ is observable at time t. Depending on the portfolio at hand, the risk factors and function f are chosen. Usually the logarithmic prices of financial assets are taken as risk factors. For example, consider a fixed portfolio of stocks or foreign exchange and denote by w_i the number of financial asset i in the portfolio at time t. Denote the price process of the financial asset by $(p_{t,i})_{t\in\mathbb{N}}$ and take risk factors as logarithmic prices $Z_{t,i} := \ln p_{t,i}$, $1 \le i \le d$. Then the portfolio value can be expressed as $V_t = \sum_{i=1}^d w_i \exp(Z_{t,i})$. However, there can be a lot of other choices of risk factors. Not only the price, but also the features of risky assets, such as volume, price volatility and other derived characteristics, can be taken as risk factors. Furthermore, market news announcements, macroeconomic indicators can also be selected as parameters to model the distribution of financial assets. The function f is called a *mapping function* and the representation of the portfolio value in (1.2) is regarded as *risk mapping.* Risk factors and their changes are the main object of interest in risk modeling.

Suppose that logarithmic prices of financial assets are taken for risk factors. Then portfolio loss (1.1) can be expressed as

$$L_{t+1} = -[f\{\tau_{t+1}, \mathbf{Z}_t + \mathbf{X}_{t+1}\} - f\{\tau_t, \mathbf{Z}_t\}]. \tag{1.3}$$

Here $\mathbf{Z}_t = (\ln p_{t,1}, ..., \ln p_{t,d})^T$ is the vector of logarithmic prices of financial assets at time t and $\mathbf{X}_{t+1} := \mathbf{Z}_{t+1} - \mathbf{Z}_t$ is the change in the risk factor value. Taking the example of stocks or foreign exchange portfolio considered above, the portfolio loss can be written as $L_{t+1} = -(V_{t+1} - V_t) = -\sum_{i=1}^{d} w_i p_{t,i}(\exp(X_{t+1,i}) - 1)$. Equation (1.3) shows that the portfolio loss distribution is determined by the distribution of risk factor changes $\mathbf{X}_{t+1}$, as the distribution of risk factors $\mathbf{Z}_t$ is known at time t.

Conditional loss distribution

In this subsection, a distinction between conditional and unconditional loss distributions is considered in short. Both the conditional and unconditional distributions are relevant in the risk management. The unconditional loss distribution is usually used in the credit risk management and insurance, where losses are measured over long time horizons. Since we take a particular interest in measuring the market risk over a relatively short time period, the conditional loss distribution will be considered in the rest part of the thesis. Denote by $\mathcal{F}_t$ the σ-algebra, generated by risk factor changes $(X_s)_{s \leq t}$, and denote by $F_{[X_{t+1}|\mathcal{F}_t]}$ the conditional distribution of X_{t+1} given the information $\mathcal{F}_t$. Before giving a definition of the conditional loss distribution, the notation of a *loss operator* that maps changes of risk factors into losses, is introduced. Recall a portfolio loss as in (1.3) and define the loss operator as

$$l_{[t]}(x) := -[f\{\tau_{t+1}, Z_t + x\} - f\{\tau_t, Z_t\}], \quad x \in \mathbb{R}^d.$$

Note that $L_{t+1} = l_{[t]}(X_{t+1})$.

The *conditional loss distribution* $F_{[L_{t+1}|\mathcal{F}_t]}$ can be defined as the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{[\mathbf{X}_{t+1}|\mathcal{F}_t]}$ and written as

$$F_{[L_{t+1}|\mathcal{F}_t]}(a) = P(l_{[t]}(X_{t+1}) \leq a|\mathcal{F}_t) = P(L_{t+1} \leq a|\mathcal{F}_t), \quad a \in \mathbb{R}. \tag{1.4}$$

The conditional loss distribution gives the conditional distribution of the next period loss L_{t+1}, given all the current information $\mathcal{F}_t$.

1.1.2 Value-at-Risk

In this section, a general definition of VaR is introduced. Consider a portfolio of risky assets with the conditional loss L distribution function $F_L(l)$. With reference to the definition of the standard VaR we assume the confidence level $\alpha \in (0,1)$. The portfolio VaR at the fixed confidence level α is given by the smallest number l such that the probability that loss L exceeds l is no larger than $(1-\alpha)$ over the time horizon $\delta \geq 0$:

$$VaR_\alpha = inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}. \qquad (1.5)$$

According to this definition, VaR depends on the confidence level α and the holding period δ that is chosen while setting the discrete time series, as defined in Section 1.1.1. Under regulatory requirements (BCBS 2006) to VaR parameters, the confidence level α is 99% and the holding period equals 10 days. However, for everyday risk management purposes the confidence level is usually taken 95% and a 1 day time horizon is used. Additionally, the VaR value depends on the choice of a loss distribution model. Therefore there is a great variety of different methodologies to estimate VaR. Two approaches are usually chosen to model the portfolio or a single asset loss distribution. According to the first one, the assumption that future behavior of risk factors will be the same as in the past is made. Then the past behavior of risk factors is analyzed. The other approach is based on making an explicit assumption on the distribution of risk factors. These methods are briefly presented in the next two sections.

Historical Value-at-Risk

One of the most popular methods to estimate VaR is the *historical simulation* methodology. The method, based on the concept of rolling windows, is easily implemented. Suppose that the data set of n historical prices of the financial asset such as equity, bond or foreign exchange is available. Assume that the financial asset price returns are identically distributed with the same empirical distribution function over the time horizon analyzed. The historical VaR model can be written as:

$$VaR_\alpha^{hist} = q_\alpha(F_L), \qquad (1.6)$$

where F_L denotes the empirical distribution of the financial asset position loss L, and the symbol q_α denotes the α quantile of the empirical distribution. In practice, a daily estimate of the historical VaR model is obtained by calculating all the possible changes in the current price over the analyzed time horizon and choosing a sufficiently large price change at the required confidence level. To estimate VaR for the following day, the whole window is moved forward by one day and the entire procedure for calculating all the possible changes in the price is repeated.

Parametric Value-at-Risk

Another widely used VaR methodology is *parametric* VaR. The main feature of the parametric VaR methodology is the assumption that the risk factors distribution determines the distribution of losses.

Assume that the loss distribution function F_L is normal with mean a_L and variance σ_L^2. The parametric VaR measure can be expressed as

$$VaR_\alpha^{par} = a_L + \sigma_L q_\alpha\{\Phi\}, \tag{1.7}$$

where $q_\alpha\{\Phi\}$ denotes the α quantile of the standard normal distribution Φ.

When using the parametric VaR methodology in practice, mean a_L and variance σ_L^2 are not known and have to be estimated. Usually, the mean of returns is very close to zero and is not regarded in the VaR estimation. Besides, Kim, Malz and Mina (1999) have showed that mean forecasts for a shorter than three months period do not produce accurate estimates. In the market risk the forecasts are usually taken up to 10 days. In this case, the forecast of future returns is determined by the volatility parameter σ estimate. One of the most common methods to estimate variance is the RiskMetrics approach (see, e.g., [50]), where the variance can be calculated using *Exponentially Weighted Moving Average (EWMA)*. Another broadly used approach to estimate risk factor volatility is the GARCH model introduced by Bollerslev (1986).

The presented standard VaR definition, based on the distribution of loss, can be straightforwardly generalized when the aggregated high frequency risk factors data are considered. A new concept of aggregated VaR is introduced in the next section.

1.2 Aggregated Value-at-Risk

Consider "tick-by-tick" series of financial asset prices, where the jth observation consists of two variables - moment τ_j and value p_j. The time series of N such observations can be written as $\{(\tau_j, p_j)\}_{j=1}^N$. To construct regular time series, the time interval between two observations δ is fixed and a new time scale is obtained by taking $\tau_t^* = t\delta$, $t = 1, ..., N^*$. Suppose that the aggregation rule g is from a class $\mathcal{G}$ (for a more general approach to aggregation see Kvedaras and Račkauskas (2010)). Data series of regular financial asset prices are defined as

$$p_t(g) := g(\{(\tau_j, p_j), \tau_j \in (\tau^*_{t-1}, \tau^*_t]\}), \quad t = 1, ..., N^*. \tag{1.8}$$

There are several possible ways of choosing the aggregation rule. For example, daily data are obtained either taking the closing or the last price of the day (equity markets) or fixing the price at a certain moment or period of the day (foreign exchange markets). In such a case, all the remaining information on price behavior during the day is not taken into account. Therefore we would like to try taking such data aggregation that captures more information about the financial asset price than the fixed price at a certain moment. Having aggregated the prices, using different rules, it is important to analyze the VaR estimator's dependence on data aggregation. In this chapter, we use pointwise, maximum value, minimum value and average value data aggregation rules. There are a lot of possibilities for other choices of aggregation rules; however, our purpose is not to analyze all the possible aggregation methods, but to illustrate the dependence of the VaR estimate on the selected aggregation rules.

Suppose a "tick-by-tick" series of financial asset prices is regularized, using the aggregation rule $g \in \mathcal{G}$, as defined in equation (1.8). Having taken logarithms of financial asset aggregated prices (1.8), the portfolio loss in (1.3) can be written as

$$L_{t+1}(g) = -[f\{\tau_{t+1}, \mathbf{Z}_t(g) + \mathbf{X}_{t+1}(g)\} - f\{\tau_t, \mathbf{Z}_t(g)\}], \tag{1.9}$$

where $\mathbf{X}_{t+1}(g) := \mathbf{Z}_{t+1}(g) - \mathbf{Z}_t(g)$ are changes of risk factors with $\mathbf{Z}_t(g) = (Z_{t,1}(g), ..., Z_{t,d}(g))^T = (\ln p_{t,1}(g), ..., \ln p_{t,d}(g))^T$. Assume that the conditional distribution function the portfolio of financial assets loss is $F_{L(g)}(l)$, $g \in \mathcal{G}$.

Definition 1.1 (Aggregated Value-at-Risk)**.** Assume the confidence level $\alpha \in (0, 1)$. The portfolio VaR at a fixed confidence level to be α is given by the smallest number l such that the probability that the loss $L(g)$, $g \in \mathcal{G}$ exceeds l is no larger than $(1 - \alpha)$ over the time horizon $\delta \geq 0$:

$$VaR_\alpha(g) = inf\{l \in \mathbb{R} : P(L(g) > l) \leq 1 - \alpha\} = inf\{l \in \mathbb{R} : F_{L(g)}(l) \geq \alpha\}. \tag{1.10}$$

According to this definition, VaR depends not only on the confidence level α and the holding period δ, but also on the aggregation rule g from the given class $\mathcal{G}$. This is the general definition of aggregated VaR, where neither the aggregation rule nor the class of these rules is specified. For example, we can define the aggregation rule taking the aggregated price as a point equal to the closing price of the day. In this case, the aggregated VaR corresponds to the standard definition of the daily VaR and its estimate. However, applying more complex aggregation rules, VaR values will be different. In the next section, we present a numerical example of the aggregated VaR estimate using the historical simulation methodology.

1.3 Numerical example

Foreign currencies (USD, EUR, GBP, JPY) were taken to calculate VaR and to illustrate the market risk estimator's dependence on the data aggregation rule. These currencies were chosen as frequently traded on the foreign exchange market and, therefore, large data sets are available for research. Foreign exchange rates were taken every minute over one-year period, i.e., the final data set for each currency consisted of 1440 minutely rates for 252 working days of the year. The average price of the bid and ask rates was used. By the assumption a position of 1 million of USD was held. The possible loss due to the USD exchange rate decrease in respect of the local currency (EUR, GBP or JPY) was considered over one day, i.e., due to foreign exchange rate fluctuations a smaller equivalent value of 1 million USD position was estimated. The confidence level was taken 95 percent.

According to the portfolio loss definition (1.9), risk factors and the risk mapping function should be specified. As usual, logarithmic foreign exchange rates are taken as risk factors. Consider any foreign currency and denote by w the amount of this currency equivalent in respect of the fixed local currency. Suppose $p_t(g)$, $g \in \mathcal{G}$ is a regular data series of foreign exchange rates. Accordingly, risk factors $Z_t(g)$, $g \in \mathcal{G}$ are logarithmic foreign exchange rates and risk factor changes $X_t(g)$, $g \in \mathcal{G}$ are logarithmic foreign exchange returns. The portfolio of a single currency loss in equation (1.9), after mapping the risk can be given by $L_{t+1}(g) := -w\, p_t(g)(\exp(X_{t+1}(g)) - 1)$. Using the Taylor series of an exponent, the linearized loss $\tilde{L}_{t+1}(g)$ can be expressed as

$$\tilde{L}_{t+1}(g) := -w\, p_t(g) X_{t+1}(g). \tag{1.11}$$

In practice, the linearized loss is usually taken to estimate VaR. To specify the aggregation rules that were applied to aggregate foreign exchange rates, consider "tick-by-tick" observations $\{(\tau_j, p_j)\}_{j=1}^{N}$, where p_j indicates the financial asset price (it could also be volume, number of news announcements or some other feature) recorded at time τ_j. We assume that each $s \in [0, 1]$ corresponds to the aggregation scheme $g \in \mathcal{G}$. So we identify class $\mathcal{G}$ with the interval $[0, 1]$ and the aggregation rule g with s.

The following aggregation schemes were chosen for analysis:

- pointwise aggregation

$$p_t^{DAILY}(s) = \{p_i | \tau_i = \max\{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}\}, \quad s \in [0,1],$$

- maximum value aggregation

$$p_t^{MAX}(s) = \max\{p_i | \tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}, \quad s \in [0,1],$$

- minimum value aggregation

$$p_t^{MIN}(s) = \min\{p_i | \tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}, \quad s \in [0,1],$$

- average value aggregation

$$p_t^{AVE}(s) = \frac{1}{m_t(s)} \sum_{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]} p_i, \quad s \in [0,1],$$

  where

$$m_t(s) = \#\{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}.$$

The parameter $s \in [0,1]$ is assumed to be continuous. In practice, however, risk factors are recorded at discrete time moments and therefore this parameter can be chosen depending on the available data frequency - one second, one minute, five minutes, etc. In the thesis, a parameter s equal to one minute, was taken.

Application of the aggregation rules defined above results in different aggregated values during a day. Using the pointwise (POINT) data aggregation, the aggregated value is obtained taking the corresponding tick observation at every moment of the day (see Figure 1.3). The maximum (MAX) value aggregation at a certain moment of the day is defined as the maximum tick observation up to that moment. On the contrary, under the minimum (MIN) value aggregation, the aggregated value is obtained taking the minimum tick observation up to the corresponding moment of the day. The average (AVG) value aggregation means that at every moment of the day, the aggregated value is calculated as an arithmetic average of all the tick observations up to that moment. Figure 1.1 illustrates the EUR/USD (i.e. the price of 1 EUR in USD) aggregated exchange rate during a day and using all the aggregation rules listed above.

The historical simulation methodology was chosen to estimate the aggregated VaR of the foreign currency position. Assume that the foreign exchange position

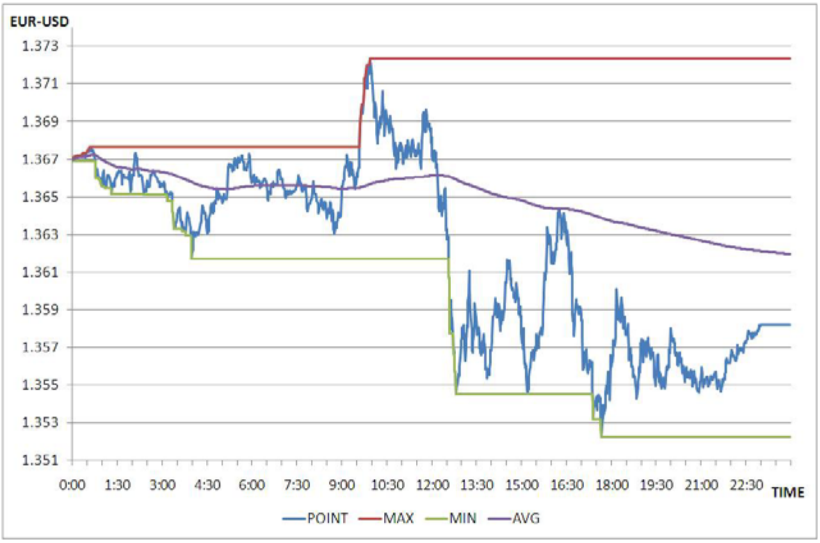

Figure 1.1: EUR/USD aggregated exchange rate

loss is defined as in equation (1.11). Then the aggregated VaR can be written as

$$VaR_\alpha(s) = q_\alpha(F_{L(s)}), \quad s \in [0,1],$$

where $F_{L(s)}$ denotes the empirical distribution of the foreign exchange position loss and the symbol q_α denotes the α quantile of the empirical distribution.

The VaR estimate is the empirical α quantile calculated as

$$\widehat{VaR}_\alpha(s) = Y^*_{[\alpha n]+1}(s), \quad s \in [0,1],$$

where $Y^*_{[\alpha n]+1}(s)$, is the $[\alpha n]+1$ member of ordered time series for every $s \in [0,1]$ consisting of

$$Y_j(s) = -w\,p_n(s)X_j(s), \quad j = 1,...,n, \quad Y^*_1(s) \le Y^*_2(s) \le ... \le Y^*_n(s).$$

The results of the aggregated VaR estimates for the exchange rates of EUR/USD, GBP/USD and JPY/USD are presented in Figures 1.2, 1.3, and 1.4. In addition, the horizontal line in each picture corresponds to the daily VaR estimate calculated using the daily foreign exchange rates provided by Bloomberg. The daily foreign exchange rates in the Bloomberg system are fixed as that of the end of the day. All the pictures show that the pointwise value aggregation yields the most volatile VaR. Another common characteristic is that the average value aggregation gives a very low VaR estimate at the end of the day as this aggregation method results in smoother prices and, consequently, lower returns. However, there are some considerable differences in the presented charts for all the currencies analyzed.

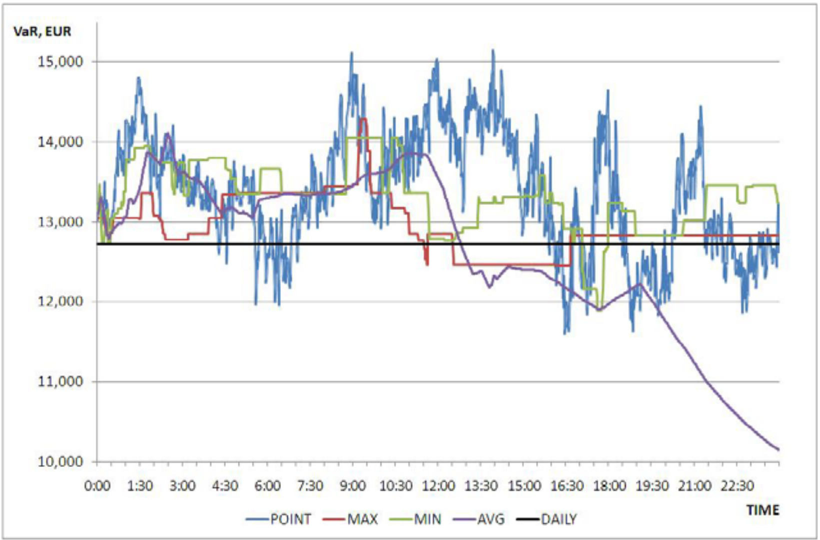

Figure 1.2: Aggregated VaR of the USD position in respect of the EUR. currency

Figure 1.2 shows VaR estimates of the EUR/USD foreign exchange position. The maximum VaR value, amounting to 15,2 th.EUR, is obtained by pointwise aggregation at 13:56. The pointwise aggregation produces the largest VaR estimates between 9 a.m. and approximately 3 p.m. when there is the most active trade in the European financial market as well as the start of trade in America. Consequently, larger exchange rate fluctuations result in larger returns. At the end of the day, the EUR/USD exchange rate does not change considerably and VaR values are respectively low. Therefore the daily foreign exchange rate fixed at the end of the day results in rather a low daily VaR estimate amounting to 12,7 th.EUR or 16 per cent less than the maximum VaR value in the pointwise aggregation. Even the maximum and minimum value aggregation rules give larger VaR estimates most of the time than daily VaR. It should be noted that the minimum value aggregation gives a higher VaR estimate almost all the day compared to the maximum value aggregation, since decreases in the EUR/USD foreign exchange rate were larger than increases during the analyzed period. The average value aggregation resulted in the largest deviation between the smallest VaR value of 10,1 th.EUR and the largest VaR value of 14,1 th.EUR, i.e., a difference of 39 per cent. Nevertheless, the average value aggregation rule is not very suitable to estimate the risk for such a long aggregation interval as one day. The chart shows that in the second part of the day new observations do not make any impact and the average value aggregation gives smoother and smoother exchange rates with minor fluctuations, and results in a low risk estimate.

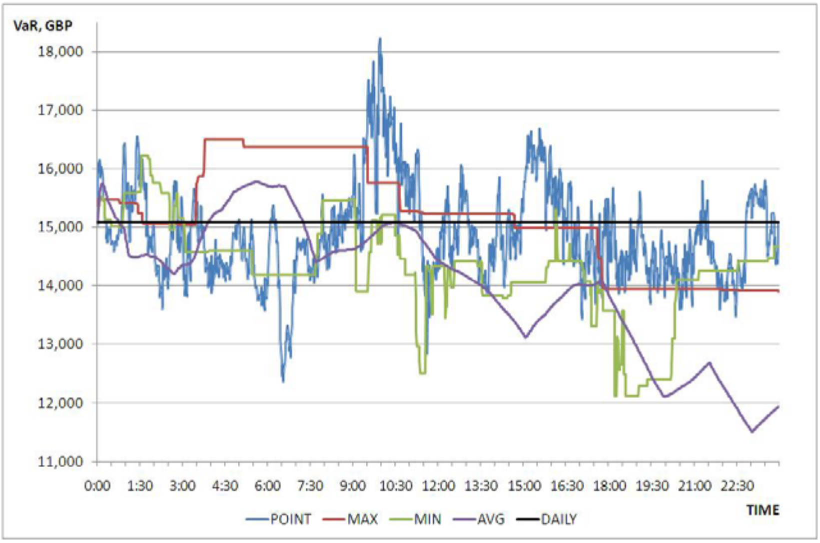

Figure 1.3: Aggregated VaR of the USD position in respect of the GBP. currency

Figure 1.3 shows VaR estimates of the GBP/USD foreign exchange position. Overall, this picture is quite similar to the EUR/USD graph but still there are some differences. There are two prominent peaks in the pointwise aggregation case. The first one is between 10 a.m. and 11 a.m., when the business day in Europe starts and the market is opened, and the second one around 4 p.m., when the market is going to be closed in Europe. The British pounds are mostly traded in Europe; therefore this chart represents active trading hours in Europe when the highest risk estimates are obtained. The largest VaR value amounting to 18,2 th. GBP is obtained by the pointwise aggregation at 9:57. Similarly as in the EUR/USD case, the daily VaR estimate is smaller in some intervals compared to VaR estimates obtained using data aggregation. The daily VaR 15,1 th. GBP in total is 17 per cent smaller than the maximum VaR value of the pointwise aggregation. The pointwise aggregation also resulted in the largest deviation between the smallest VaR value of 12,4 th.GBP and the largest VaR value of 18,2 th.GBP, which makes even 47 per cent difference. Unlike in the EUR/USD case, the maximum value aggregation most of the day gives larger VaR estimates than the minimum value aggregation. Increases of the foreign exchange rate at this time were sharper and resulted in larger fluctuations than decreases. VaR estimates of the average value aggregation rule tend to be lower in the second part of the day and the same comments suit as in the EUR/USD case.

Figure 1.4 shows VaR estimates of the JPY/USD foreign exchange position. This picture is rather different from EUR/USD and GBP/USD graphs. This is mainly due to time differences between Europe, America, and Asia. The Japanese

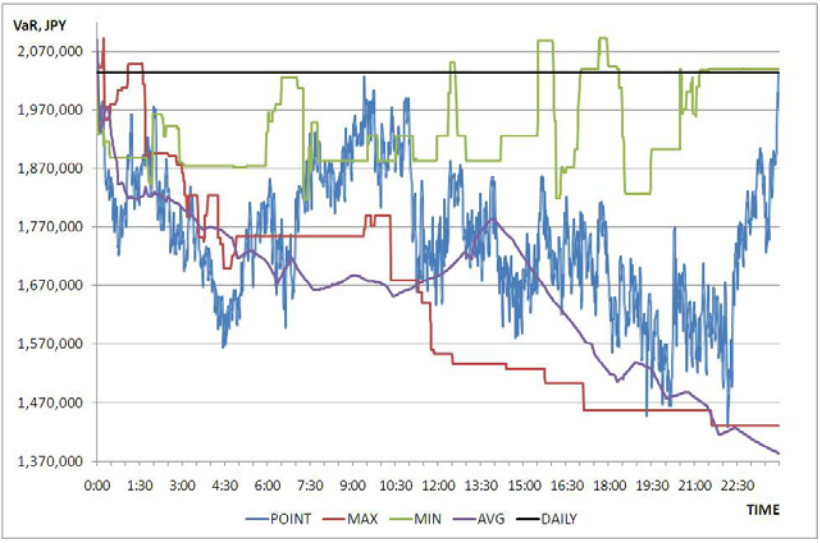

Figure 1.4: Aggregated VaR of the USD position in respect of the JPY. currency

yen currency trade is more active between America and Asia in the second part of the day and in the late hours (if London time is considered). Therefore at the end of the day, fluctuations of the JPY/USD foreign exchange rate are larger and returns series have sharper increases and decreases. The pointwise aggregation VaR estimates have two peaks. One is between 9 a.m. and 11 a.m., which is related to the start of trading in Europe, and the other one is between 11. p.m. and 2 a.m. when the trading is active in Asia. Unlike it was in the case of EUR/USD and GBP/USD foreign exchange rates; the maximum VaR value 2.092,6 th. JPY in total is obtained by the minimum value aggregation at 17:42. The daily VaR value amounting to 2.034,2 th.JPY is very close to the maximum VaR. This is due to time differences, as the largest foreign exchange rate fluctuations are also observed at the end of the day. It should be noted that, during most part of the day, there is a very large difference between the VaR estimates, using the maximum and the minimum value aggregation rules. The main reason is that the foreign exchange rate decreases were more sharper and larger than increases, and the foreign exchange rate with the aggregated minimum value was more volatile than under the maximum value aggregation. The VaR result, when using the average aggregation scheme, is quite similar as that of other currencies with a very large difference of 51 per cent between the minimum (1.383,5 th.JPY) and the maximum (2.089,8 th.JPY) VaR estimates. The JPY/USD currency example shows that not only the aggregation rule, but also the aggregation period is very important and makes a difference when estimating the risk.

1.4 Conclusions

The analysis presented in this chapter shows not only the difference of the risk estimate depending on the chosen aggregation rule, but also clearly indicates that the estimate of risk, calculated using daily foreign exchange rates, represents only a small part of the view what happens during the whole day. According to the official requirements of the Bank for International Settlements, daily historical data are sufficient to estimate the market risk. However, financial markets operate at a very high frequency nowadays when millions of trades are performed during one minute. Then the question is whether it is enough to measure risk and to calculate the capital requirement on a daily basis. Perhaps instead of taking daily data financial institutions should estimate risk with regard to all the information available during the day. This question is very important not only for financial institutions, but also for the supervisory authorities, since the main task for the supervisory authorities is to ensure that financial institutions would accurately estimate the risk assumed and hold sufficient amounts of capital to be able to meet their obligations.

The empirical study performed on the VaR value dependence on the choice of the four analyzed risk factor aggregation methods (pointwise, maximum value, minimum value and average value) illustrates how much the estimates may vary. Looking at the presented dynamics of the aggregated VaR of the USD positions in respect of EUR, GBP and JPY and different results of risk estimates thereby, the question can be posed which aggregation rule should be taken to obtain the best estimate of risk. However, the answer to this problem is still open and can only be solved after a thorough theoretical research on the risk estimate, based on aggregated risk factors. We provided an example of the aggregated VaR estimate, using the historical simulation methodology; however, to calculate the aggregated VaR estimate using the parametric methodology, one needs a special model to estimate volatility. Therefore, in the next chapter, the functional GARCH model and its asymptotic properties are introduced.

Chapter 2

Functional $\rho - \mathrm{GARCH}(1,1)$ model

The uncertainty, usually measured by volatility, is the central issue in financial analysis. The financial asset return volatility estimation is the main object of interest in risk measurement, asset pricing and portfolio allocation. In financial economics, volatility is often defined as an instantaneous standard deviation of a random Wiener-driven component in the continuous-time diffusion model. Estimation of financial assets, such as stocks, foreign exchange or interest rates, volatility is based on returns of their logarithmic prices, i.e., $r_j = \log p_j - \log_{j-1}$, $j = 1, 2, \ldots$, is the return of the jth price p_j of a financial asset. There is a huge number of models proposed for financial asset returns and volatility (see, e.g., [4] and [5] for surveys).

It is observed that the conditional volatility of financial returns series changes over time, therefore GARCH family models are one the most popular tools in the financial risk management. In 1982, Engle introduced an ARCH(p) model, assuming that returns can be written as

$$r_j = \sigma_j \varepsilon_j,$$

where σ_j is volatility and satisfies

$$\sigma_j^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k r_{j-k}^2,$$

with non-negative parameters α_i, $i = 0, 1, \ldots, p$, and (ε_j) is a white noise process with variance 1. However, empirical works have shown that the ARCH(p) model fits real data only for large p. In 1986, Bollerslev proposed a GARCH(p, q) model, where volatility is given by

$$\sigma_j^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k r_{j-k}^2 + \sum_{k=1}^{q} \beta_k \sigma_{j-k}^2.$$

There are many generalizations of the GARCH model to represent other stylized facts of financial asset returns. The basic GARCH model assumes that both positive and negative shocks of the same absolute size have the same impact on the future conditional variance. However, the asymmetry, usually referred as a leverage effect, is often observed in financial returns series, i.e., a fall of the price tends to cause a higher volatility increase compared to the same size growth of the price. There are three most frequently used GARCH models to describe this type of asymmetry: the Threshold GARCH (see, e.g., [73]), the Asymmetric GARCH (see, e.g., [39]), and the Exponential GARCH (see, e.g., [62]). Furthermore, to account for long memory, observed in absolute or squared returns, the Fractionally Integrated GARCH model was developed (see, e.g., [10]). However, GARCH(1,1) is the model, most often applied in practice, despite the advantages of that discussed above and numerous other generalizations of ARCH models. For example, Hansen and Lunde (2005) compared various types of ARCH models (AGARCH, EGARCH, TGARCH and many others) in their ability to describe conditional variance. The analysis has shown no evidence that other more sophisticated models outperformed the GARCH(1,1) model.

Stochastic volatility models can be distinguished as another broad class of volatility models (see, e.g., [2], [41] and [68] for surveys). The volatility dynamics expression of models of this type includes an unobserved shock to the return variance. The variance process becomes inherently latent, i.e., even with all the available past information and knowledge of the data generating process, the exact value of the current volatility state is unknown. This feature implies that the volatility process is not measurable with respect to observable information. Therefore, data filtering and smoothing techniques together with simulation procedures (simulated moments, Markov Chain Monte Carlo, etc.) are used in the estimation and forecasting of stochastic volatility. In contrast, the conditional variance is assumed to be observable, given past information, in GARCH models and usually the maximum likelihood method is applied. Despite these differences, stochastic volatility and GARCH type models are closely related. In practice, the class of GARCH models is more often chosen for volatility forecasting due to the easier parameter estimation.

One more group of volatility models is a realized volatility or sometimes referred as the historical volatility, since it measures what happened in fact in the past. The notion of realized volatility represents a model-free approach to a consistent estimation of the quadratic return variation under general assumptions, such as arbitrage-free financial markets. High-frequency returns are usually taken for the assessment of a lower frequency return volatility. For more details on the realized volatility measurement and forecasting refer to [6] and [24].

All the volatility models discussed above were univariate. Generalizations of these models to applications in a higher dimension, when several financial assets volatility is measured together, can be attributed to the class of multivariate volatility (see, e.g., [7], [12] and [47] for reviews). However, due to some difficulties in parameter estimation and in setting the sufficient conditions to ensure that the covariance matrix forecasts remain positive definite for all forecasting horizons, multivariate volatility models are rarely applied in practice.

In this chapter, we present a completely different - functional generalization of the basic GARCH$(1,1)$ model. We consider functional returns, subject to a certain aggregation rule g, varying in the class $\mathcal{G}$. The functional $\rho-$ GARCH$(1,1)$ is introduced and its properties are analyzed.

2.1 Point-wise GARCH

Suppose for each aggregation rule g from a given class $\mathcal{G}$ we have a time series $(X_t(g), t \in \mathbb{Z})$. Consider a classical GARCH(p, q) model:

$$X_t(g) = \sigma_t(g)\varepsilon_t(g), \tag{2.1}$$

$$\sigma_t^2(g) = \omega(g) + \alpha(g)\sum_{i=1}^{q} X_{t-i}^2(g) + \beta(g)\sum_{i=1}^{p}\sigma_{t-i}^2(g), \tag{2.2}$$

where for each $g \in \mathcal{G}$, $(\varepsilon_t(g), t \in \mathbb{Z})$ are random variables and $(\omega(g), \alpha(g), \beta(g))$ is the vector of parameters.

If for each g the random variables $(\varepsilon_t(g), t \in \mathbb{Z})$ are i.i.d., the classical GARCH theory applies (see, e.g., [40]) pointwise with respect to $g \in \mathcal{G}$. Then, with respect to a data set, the best aggregation rule can be found, for example, by minimizing a distance between the data set observations $X_t(g), t = 1, \dots, n$ and estimates $\hat{X}_t(g)$:

$$g_0 = \operatorname*{argmin}_{g \in \mathcal{G}} d(\hat{X}_t(g), X_t(g)). \tag{2.3}$$

The problem of choosing an aggregation rule appears, for instance, in the risk measurement. To illustrate this problem, consider Value-at-Risk (VaR), a statistical model, defined as the maximum future loss due to possible changes in the value of financial asset portfolio during a certain period with a certain probability. Assume a position of foreign exchange, where the potential loss arises due to foreign exchange rate fluctuations, is held. In practice, a daily market risk is calculated using official foreign exchange rates (the Bank of Lithuania fixes daily foreign exchange rates at around 10 a.m. for the next working day) or daily rates taken from financial data systems (in Bloomberg, daily foreign exchange rates are taken as that of the end of the day). As already discussed in the previous chapter, there are a lot of other possibilities to choose daily rates, if the aggregation is applied. The presented numerical example has clearly shown the VaR estimate dependence on the aggregation when the high frequency foreign exchange rates are aggregated.

Recall the definition of the parametric VaR model, based on an explicit assumption about the distribution function of risk factors and it mainly deals with the parametrization of returns and volatility modeling. Assume that the distribution function F_L of loss L is normal with zero mean and variance σ_L^2. The parametric VaR can be expressed as

$$VaR_{\alpha}^{par} = \sigma_L q_{\alpha}\{\Phi\}, \tag{2.4}$$

where $q_{\alpha}\{\Phi\}$ denotes the α quantile of the standard normal distribution Φ. Suppose that the GARCH$(1,1)$ model is taken to estimate the risk factor volatility. The variance equation can be written as

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{2.5}$$

Assume that a set of "tick-by-tick" observations $\{(\tau_j, p_j)\}_{j=1}^N$ consists of foreign currency exchange rates p_j recorded at times $\tau_j, j = 1, \ldots, N$. Consider the case where $\tau_t^* = t\delta$ with δ corresponding to one day and let

$$p_t(s) = \{p_j | \tau_i = \max\{\tau_j \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}\}$$

be the last price in the interval up to s of the tth day. To illustrate the variability of the VaR value during the day, time s was fixed every 15 minutes and a parametric VaR using (2.4) was calculated for each s, taking the volatility estimates obtained from the GARCH$(1,1)$ model as in (2.5). The assumption of holding 1 million USD position was made. The confidence level was chosen 95% and the holding period was taken one day. Figure 2.1 shows the EUR/USD position of the VaR estimate variability during the day. If, for instance, the foreign exchange rate is fixed at the end of the day, the estimate of the VaR model will be near to the minimum during the day; meanwhile, taking the foreign exchange rate at around 10 a.m. gives the maximum daily VaR estimate. What is the best choice for s? One of the ways to select an aggregation rule is to choose the one, where the aggregated data fit the model best. Taking the same example as considered above, we use the Euclidean distance and find the point s_0 that has the minimum

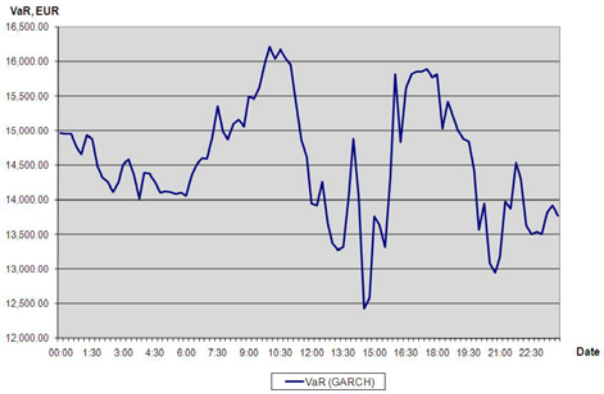

Figure 2.1: Value-at-Risk of EUR/USD using GARCH$(1,1)$

distance between the estimate obtained by GARCH$(1,1)$ and the aggregated data:

$$s_0 = \arg \min_{s\in[0,1]} \sum_{t=1}^{n} (\hat{X}_t(s) - X_t(s))^2.$$

Calculations performed by taking the foreign exchange data show that the point, where the aggregated data fits the model best, is at 23:45. However, there might be different criteria of choosing the aggregation rule. For instance, from the point of view of supervisory authorities of financial institutions or clients, the daily rate should be chosen so that it had the maximum estimate of risk. This conservative approach to measure risk ensures that financial institutions hold sufficient amounts of capital to cover possible losses. In that case, the point of 10:00 would be the best. Nevertheless, the problem of both examples is that the situation changes every day. Next day estimating VaR the picture should look similar to that presented above, however the point that yields the maximum risk estimate or the minimum distance might be different. The natural solution to this problem would be to fit the model, where volatility does not vary during the day, but it depends on a certain information known at that day, similarly as analyzed in Alizadeth, Brandt and Diebold (2002). This approach is presented in the next section.

Another way to overcome the problem of aggregation for VaR is to take $\sigma_t^2 = T(\sigma_t^2(g), g \in \mathcal{G})$, where T is a functional. The conservative approach to measure risk would be choosing a functional, that yields the maximum volatility estimate, i.e. $\sigma_t^2 = \sup_{g\in\mathcal{G}} \sigma_t^2(g)$. From the point-wise model we have the following estimate of $\sigma_t^2(g)$:

$$\widehat{\sigma}_t^2(g) = \frac{\widehat{\omega}(g)}{1-\widehat{\beta}(g)} + \widehat{\alpha}(g)\sum_{j=1}^{t} X_{t-1}^2(g), \quad g \in \mathcal{G}.$$

Hence, one could take $\widehat{\sigma}_t^2 = \sup_{g\in\mathcal{G}} \widehat{\sigma}_t^2(g)$.

2.2 Model

Let $(\Omega, \mathcal{F}, P)$ denote the probability space of all random elements. Consider a functional time series $(X_t, t \in \mathbb{Z})$, where for each t, $X_t = (X_t(g), g \in \mathcal{G})$ is a random function defined on a set $\mathcal{G}$. Assume, that for each t, $X_t \in \mathbb{E} \subset \mathbb{R}^{\mathcal{G}}$, is an $\mathbb{E}$-valued random element, where $\mathbb{E}$ is a separable topological vector space endowed with its Borel σ-field. Let $\rho : \mathbb{E} \to \mathbb{R}$ be a measurable semi-norm.

Definition 2.1. The process $(X_t, t \in \mathbb{Z})$ is a functional $\rho -$ GARCH$(1,1)$, if for each $g \in \mathcal{G}$ and $t \in \mathbb{Z}$ it satisfies,

$$X_t(g) = \sigma_t \varepsilon_t(g), \tag{2.6}$$

$$\sigma_t^2 = \omega + \alpha\rho^2(X_{t-1}) + \beta\sigma_{t-1}^2, \tag{2.7}$$

where $(\varepsilon_t(g), g \in \mathcal{G}), t \in \mathbb{Z}$ are independent identically distributed random functions.

In the next two sections we summarize some properties of the functional $\rho -$ GARCH$(1,1)$ process.

2.3 Stationarity

Consider the functional $\rho -$ GARCH$(1,1)$ process $X_t, t \in \mathbb{Z}$, defined by (2.6), (2.7), where $\alpha \geq 0$, $\beta \geq 0$, $\omega > 0$.

Theorem 2.1. *If inovations (ε_t) are iid, and*

$$-\infty \leq \gamma := E\log\{\alpha\rho^2(\varepsilon_0) + \beta\} < 0, \tag{2.8}$$

then the series

$$h_t := \omega + \omega \sum_{n=1}^{\infty} \prod_{j=1}^{n} \left(\alpha\rho^2(\varepsilon_{t-j}) + \beta\right) \tag{2.9}$$

converges a.s. and the process (X_t) defined as

$$X_t(g) = h_t^{1/2}\varepsilon_t(g), \quad g \in \mathcal{G}, \tag{2.10}$$

is the unique strictly stationary solution of the model (2.6), (2.7).

Remark 2.1. The proof goes along the lines of the proof of strict stationarity for the classical GARCH$(1,1)$ model (see, e.g., Theorem 2.1 in [40]).

It is just needed to notice that $\rho^2(X_t) = \sigma_t^2 \rho^2(\varepsilon_t)$ and therefore (σ_t^2) satisfies

$$\sigma_t^2 = \omega + (\alpha\rho^2(\varepsilon_{t-1}) + \beta)\sigma_{t-1}^2.$$

This difference equation has the solution $\sigma_t^2 = h_t$.

Proof. By iterating equation (2.7) and using $\rho^2(X_t) = \sigma_t^2 \rho^2(\varepsilon_t)$, we obtain

$$\begin{aligned}
\sigma_t^2 &= \omega + (\alpha\rho^2(\varepsilon_{t-1}) + \beta)\sigma_{t-1}^2 \\
&= \omega + \omega(\alpha\rho^2(\varepsilon_{t-1})^2 + \beta) + (\alpha\rho^2(\varepsilon_{t-1}) + \beta)(\alpha\rho^2(\varepsilon_{t-2}) + \beta)\sigma_{t-2}^2 \\
&= \omega\Big\{1 + \sum_{n=1}^{N}(\alpha\rho^2(\varepsilon_{t-1})^2 + \beta)...(\alpha\rho^2(\varepsilon_{t-n}) + \beta)\Big\} \\
&+ (\alpha\rho^2(\varepsilon_{t-1}) + \beta)...(\alpha\rho^2(\varepsilon_{t-(N+1)}) + \beta)\sigma_{t-(N+1)}^2 \\
&:= h_t(N) + (\alpha\rho^2(\varepsilon_{t-1}) + \beta)...(\alpha\rho^2(\varepsilon_{t-(N+1)}) + \beta)\sigma_{t-(N+1)}^2,
\end{aligned}$$

Since the process $(h_t(N))$ consists of positive terms, the limit of the process $h_t = \lim_{N\to\infty} h_t(N)$ exists in the interval $[0, +\infty]$. Furthermore, we have that

$$\begin{aligned}
&(\alpha\rho^2(\varepsilon_{t-1}) + \beta)h_{t-1}(N-1) \\
=&(\alpha\rho^2(\varepsilon_{t-1}) + \beta)\omega\Big\{1 + \sum_{n=1}^{N-1}(\alpha\rho^2(\varepsilon_{t-2})^2 + \beta)\dots(\alpha\rho^2(\varepsilon_{t-1-n}) + \beta)\Big\} \\
=&\omega\Big\{(\alpha\rho^2(\varepsilon_{t-1}) + \beta) + \sum_{n=1}^{N-1}(\alpha\rho^2(\varepsilon_{t-1}) + \beta)(\alpha\rho^2(\varepsilon_{t-2})^2 + \beta)\dots(\alpha\rho^2(\varepsilon_{t-1-n}) + \beta)\Big\} \\
=&\omega\sum_{n=1}^{N}(\alpha\rho^2(\varepsilon_{t-1}) + \beta)\dots(\alpha\rho^2(\varepsilon_{t-n}) + \beta) = h_t(N) - \omega,
\end{aligned}$$

and, as $N \to \infty$, the relation $h_t(N) = \omega + (\alpha\rho^2(\varepsilon_{t-1}) + \beta)h_{t-1}(N-1)$ becomes

$$h_t = \omega + (\alpha\rho^2(\varepsilon_{t-1}) + \beta)h_{t-1}. \tag{2.11}$$

Assume that $\gamma < 0$. Using the Cauchy theorem for the series of positive terms [1], we obtain

$$\Big[(\alpha\rho^2(\varepsilon_{t-1}) + \beta)...(\alpha\rho^2(\varepsilon_{t-n}) + \beta)\Big]^{1/n} = \exp\left[\frac{1}{n}\sum_{j=1}^{n}\log(\alpha\rho^2(\varepsilon_{t-j}) + \beta)\right] \to e^{\gamma} \quad a.s., \tag{2.12}$$

as $n \to \infty$, since by applying the strong law of large numbers to the series

[1] If $(\sum a_n)$ is the series consisting of positive numbers and $\lambda = \overline{\lim} a_n^{1/n}$, then (i) if $\lambda < 1$, the series $(\sum a_n)$ converges, (ii) if $\lambda > 1$, the series $(\sum a_n)$ diverges.

$(\log(\alpha\rho^2(\varepsilon_t)+\beta))$ and condition (2.8), we obtain

$$\frac{1}{n}\sum_{j=1}^{n}\log(\alpha\rho^2(\varepsilon_{t-j})+\beta)\to E\log(\alpha\rho^2(\varepsilon_0)+\beta)=\gamma,\quad a.s.\quad \text{as}\quad n\to\infty.$$

Hence, we have shown that the process, defined in (2.11), converges *a.s.* Consequently, the process $X_t(g)$ defined as

$$X_t(g)=h_t^{1/2}\varepsilon_t(g)=\left\{\omega+\omega\sum_{n=1}^{\infty}(\alpha\rho^2(\varepsilon_{t-1})+\beta)...(\alpha\rho^2(\varepsilon_{t-n})+\beta)\right\}^{1/2}\varepsilon_t(g),\tag{2.13}$$

is strictly stationary and it fits the model (2.6), (2.7). Furthermore, this process is ergodic by applying Theorem A.2. in [40] (see, Theorem A.1 in Appendix 1).

To show the uniqueness of the solution, we conversely assume that $\tilde{X}_t(g)=\tilde{h}_t^{1/2}\varepsilon_t(g)$, $g\in\mathcal{G}$ is another strictly stationary solution. Suppose $\mathbb{P}(h_t\neq\tilde{h}_t)>0$ for a certain t. Iterating (2.11), we obtain

$$\begin{aligned}h_t&=\omega+(\alpha\rho^2(\varepsilon_{t-1})+\beta)h_{t-1}\\&=\omega+\omega(\alpha\rho^2(\varepsilon_{t-1})+\beta)+(\alpha\rho^2(\varepsilon_{t-1})+\beta)(\alpha\rho^2(\varepsilon_{t-2})+\beta)h_{t-2}=\dots\\&=\omega+\omega\sum_{j=1}^{n-1}(\alpha\rho^2(\varepsilon_{t-1})+\beta)\dots(\alpha\rho^2(\varepsilon_{t-j})+\beta)\\&+(\alpha\rho^2(\varepsilon_{t-1})+\beta)\dots(\alpha\rho^2(\varepsilon_{t-n})+\beta)h_{t-n},\end{aligned}$$

and analogously

$$\begin{aligned}\tilde{h}_t=\omega+\omega\sum_{j=1}^{n-1}&(\alpha\rho^2(\varepsilon_{t-1})+\beta)\dots(\alpha\rho^2(\varepsilon_{t-j})+\beta)\\&+(\alpha\rho^2(\varepsilon_{t-1})+\beta)\dots(\alpha\rho^2(\varepsilon_{t-n)})+\beta)\tilde{h}_{t-n}.\end{aligned}$$

Then we have

$$|h_t-\tilde{h}_t|=(\alpha\rho^2(\varepsilon_{t-1})+\beta)...(\alpha\rho^2(\varepsilon_{t-n})+\beta)|h_{t-n}-\tilde{h}_{t-n}|.\tag{2.14}$$

Since the series h_t converges *a.s.*, as $n\to\infty$, $(\alpha\rho^2(\varepsilon_{t-1})+\beta)...(\alpha\rho^2(\varepsilon_{t-n})+\beta)\to 0$ with probability 1. Therefore, $\mathbb{P}(|h_{t-n}-\tilde{h}_{t-n}|\to\infty)>0$ implying that $h_{t-n}\to\infty$ or $\tilde{h}_{t-n}\to\infty$ with a positive probability. This is impossible, since both processes (h_t) and $(\tilde{h}_t)$ are stationary. Hence, we conclude that $h_t=\tilde{h}_t$ for all $t>0$ *a.s.*

□

Definition 2.2. The process $(X_t, t\in\mathbb{Z})$ of $\mathbb{E}$ valued random elements is 2nd order stationary if for all $t,h\in\mathbb{Z}$,

- $EX_t^2(g) < \infty$, $g \in \mathcal{G}$,
- $EX_t(g) = \mu(g)$, $g \in \mathcal{G}$,
- $Cov(X_t(g), X_{t+h}(f)) := E(X_t(g) - \mu(g))(X_{t+h}(f) - \mu(f)) = \Gamma_h(g, f), \quad g, f \in \mathcal{G}$.

Theorem 2.2. *Let $(\varepsilon_t(g), t \in \mathbb{Z})$ be iid square integrable random elements, i.e., $E\varepsilon_t^2(g) < \infty$ for all $g \in \mathcal{G}$. If $\alpha E\rho^2(\varepsilon_0) + \beta < 1$, the process $(X_t(g), t \in \mathbb{Z})$ defined by (2.10) is the unique 2nd order stationary solution of the model (2.6), (2.7).*

Remark 2.2. The proof runs along the lines of the proof of 2nd order stationarity for the classical GARCH$(1,1)$ model (see e.g., Theorem 2.2 in [40]).

It is just needed to notice that the strict stationarity condition (2.8) is satisfied, when $\alpha E\rho^2(\varepsilon_0) + \beta < 1$ and, therefore, to conclude the proof, one needs to show that the solution (2.10) has a finite variance.

Proof. Applying Jensen's inequality $E[\varphi(Z)] \leq \varphi(E[Z])$ to a concave function $\varphi(Z) = \log(Z)$ and using $\alpha E\rho^2(\varepsilon_0) + \beta < 1$ we obtain:

$$E\log\{\alpha\rho^2(\varepsilon_t) + \beta\} \leq \log E\{\alpha\rho^2(\varepsilon_0) + \beta\} < 0.$$

Hence the condition (2.8) is satisfied, therefore a strictly stationary solution of equations (2.6), (2.7) exists and is given by (2.10). Next, it suffices to show that the process defined by (2.10) has a finite variance.

Set $\psi(g, f) = E\varepsilon_0(g)\varepsilon_0(f)$, $f, g \in \mathcal{G}$. Since h_t is $\mathcal{F}_{t-1}$ measurable and $\varepsilon_t(g)$ is independent of $\mathcal{F}_{t-1}$, by a monotone convergence theorem for all $g \in \mathcal{G}$ we derive

$$\begin{aligned}
EX_t^2(g) &= E[h_t\varepsilon_t^2(g)] = E[E(h_t\varepsilon_t^2(g)|\mathcal{F}_{t-1}] = Eh_tE\varepsilon_t^2(g) \\
&= \psi(g,g)\omega\Big\{1 + \sum_{n=1}^{\infty} E\Big(\prod_{j=1}^{n}(\alpha\rho^2(\varepsilon_{t-j}) + \beta)\Big)\Big\} \\
&= \psi(g,g)\omega\Big\{1 + \sum_{n=1}^{\infty}(\alpha E\rho^2(\varepsilon_0) + \beta)^n\Big\} \\
&= \frac{\psi(g,g)\,\omega}{1 - (\alpha E\rho^2(\varepsilon_0) + \beta)}.
\end{aligned}$$

To prove the uniqueness, we use exactly the same arguments as in the proof of Theorem 2.1. Assume that $\tilde{X}_t(g) = \tilde{h}_t^{1/2}\varepsilon_t(g)$ is the other second order stationary solution of (2.6), (2.7). Suppose $\mathbb{P}(h_t \neq \tilde{h}_t) > 0$ for a certain t. Then

$$\begin{aligned}
E|h_t - \tilde{h}_t| &= E\Big\{\prod_{j=1}^{n}(\alpha\rho^2(\varepsilon_{t-j}) + \beta)\Big\}E|h_{t-n} - \tilde{h}_{t-n}| \\
&= (\alpha E\rho^2(\varepsilon_0) + \beta)^n E|h_{t-n} - \tilde{h}_{t-n}|.
\end{aligned}$$

Since $E|h_{t-n} - \tilde{h}_{t-n}| \leq E|h_{t-n}| + E|\tilde{h}_{t-n}|$, that is finite and due to stationarity independent of n and $(\alpha E\rho^2(\varepsilon_0)+\beta)^n \to 0$, as $n \to \infty$, we find that $E|h_t - \tilde{h}_t| = 0$ and therefore for each t, $h_t = \tilde{h}_t$ *a.s.* □

2.4 Estimation

Consider the vector $\theta = (\omega, \alpha, \beta)^T$ of parameters to be estimated. Unknown distributional parameters of ε_0 can be estimated from the residuals $(\hat{\varepsilon}_t)$ of the model (2.6), (2.7). It follows from (2.6) that

$$\rho(X_t) = \sigma_t \rho(\varepsilon_t). \tag{2.15}$$

Hence, $\theta = (\omega, \alpha, \beta)^T$ can be estimated similarly as in the case of the univariate GARCH(1, 1) model.

Denote the true value of parameters by $\theta_0 = (\omega_0, \alpha_0, \beta_0)^T$. Suppose that the random variable $\rho(\varepsilon_0)$ has a density p_λ from the parametric class of Lebesque densities on $\mathbb{R}$, $\{p_\lambda | \lambda \in M\}$, $M \subset \mathbb{R}^d$. Then, the conditional density of $\rho(X_t)$, given $\mathcal{F}_{t-1}$ where $\mathcal{F}_t = \sigma(X_s, s \leq t)$ is

$$p_{X,\lambda}(x|\mathcal{F}_{t-1}) = \frac{1}{\sigma_t} p_\lambda(x/\sigma_t).$$

To construct the likelihood function, unobserved σ_t^2 is replaced by $\tilde{\sigma}_t^2$, where $\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta)$, $\theta = (\omega, \alpha, \beta)^T$, satisfies

$$\tilde{\sigma}_t^2 = \omega + \alpha\rho^2(X_{t-1}) + \beta\tilde{\sigma}_{t-1}^2, \quad t = 1, \ldots, n,$$

with the initial values $\tilde{\sigma}_0^2 = 0$ and $X_0 = 0$. We easily find that

$$\tilde{\sigma}_t^2(\theta) = \omega \sum_{j=0}^{t-1} \beta^j + \alpha \sum_{j=1}^{t-1} \beta^{j-1} \rho^2(X_{t-j}), \quad t = 1, \ldots, n.$$

It should be noted that one can show that the estimator does not depend on the choice of initial values.

Assume that $\theta \in K$, $\lambda \in M$ and $K \subset \mathbb{R}^3$ and $M \subset \mathbb{R}^d$ are compact sets and denote the vector of parameters by $\nu = (\theta, \lambda)^T$.

Let $C(K \times M)$ be the Banach space of continuous functions $f : K \times M \to \mathbb{R}$ endowed with the uniform distance

$$d_{K\times M}(f, g) = \sup_{x \in K\times M} |f(x) - g(x)|.$$

Set $||f||_{K\times M} := d_{K\times M}(f,0)$ for $f \in C(K\times M)$.

Now we consider the likelihood function

$$\widehat{L}_n(\nu) = \widehat{L}_n(\nu; X_1, X_2, \ldots, X_n) = \sum_{t=1}^{n} \widehat{\ell}_t(\nu, X_t, X_{t-1}, \ldots, X_1), \tag{2.16}$$

$$\widehat{\ell}_t(\nu, X_t, X_{t-1}, \ldots, X_1) = \log[\tilde{\sigma}_t^{-1} p_\lambda(\rho(X_t)/\tilde{\sigma}_t)], \quad t = 1, ..., n. \tag{2.17}$$

The maximum likelihood estimator of the true parameters vector $\nu_0 = (\theta_0, \lambda_0)^T$ is obtained by maximizing the likelihood function:

$$\widehat{\nu}_n = \underset{\nu \in K\times M}{\operatorname{argmax}}\, \widehat{L}_n(\nu). \tag{2.18}$$

It is noteworthy, that $(\tilde{\sigma}_t^2)$ is nonstationary in general, therefore the ergodic theorem cannot be applied while proving the consistency of the estimator. To establish the limit properties of the maximum likelihood estimator, we consider a stationary approximation (h_t) to $(\tilde{\sigma}_t^2)$ (for details regarding this approach see [71]). Following this idea, for $\theta = (\omega, \beta, \alpha)^T$ in a compact set $K \subset \mathbb{R}^3$, define

$$h_t(\theta) = \frac{\omega}{1-\beta} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} \rho^2(X_{t-j}).$$

It is essential that this approximation satisfies

$$\sup_{\theta \in K} |h_t(\theta) - \tilde{\sigma}_t^2(\theta)| \stackrel{a.s.}{\to} 0,$$

exponentially fast[2] as $t \to \infty$ and $(h_t(\theta)) = (\sigma_t^2)$ a.s. if and only if $\theta = \theta_0$.

Now, replacing the sequence $(\tilde{\sigma}_t^2)$ by its stationary approximation (h_t) in (2.16),(2.17), we define the maximum likelihood function

$$L_n(\nu) = \sum_{t=1}^{n} \ell_t(\nu),$$

where

$$\ell_t(\nu) = \log \frac{1}{(h_t(\theta))^{1/2}} p_\lambda \left(\frac{\rho(X_t)}{(h_t(\theta))^{1/2}} \right).$$

Set

$$\nu_n = \operatorname{argmax}_{\nu \in K\times M} L_n(\nu).$$

Before establishing the strong consistency of the estimator, we list several regularity assumptions for the class of densities $\mathcal{D} = \{p_\lambda | \lambda \in M\}$, where $M \subset \mathbb{R}^d$ is a

[2] A sequence $(\xi_t)_{t\in T}$ of random elements with values in a normed vector space $(B, ||\cdot||)$ is said to converge to zero exponentially fast almost surely as $t \to \infty$, if there exists $\gamma > 1$ with $\gamma^t ||\xi_t|| \stackrel{a.s.}{\to} 0$.

compact set.

M.1 $p_\lambda(x) > 0$ for all $\lambda \in M$ and $x \in \mathbb{R}$.

M.2 The map $\mathbb{R} \times M \to (0, \infty) : (x, \lambda) \mapsto p_\lambda(x)$ is continuous.

M.3 From $p_\lambda = p_{\lambda'}$ it follows that $\lambda = \lambda'$.

Now we can establish the conditions for the consistency of the maximum likelihood estimator $\widehat{\nu}_n$ following Theorem 6.1.4 in [71].

Theorem 2.3. *Consider the model* (2.15), (2.7) *and let the following conditions hold:*

(i) $E \log\{\alpha\rho^2(\varepsilon_0) + \beta\} < 0$ *and* $E\rho^2(X_0) < \infty$.

(ii) $K \subset (0, \infty) \times [0, \infty) \times [0, 1)$ *and* $M \subset \mathbb{R}^d$ *are compact subsets that contain true parameters* $\theta_0 \in K$, $\lambda_0 \in M$ *and* $(\alpha_0, \beta_0) \neq 0$, $\omega_0 \geq \omega_1 > 0$.

(iii) The class of densities $\mathcal{D} = \{p_\lambda | \lambda \in M\}$ *is such that the conditions M.1-M.3 hold.*

(iv) $\int_0^\infty |\log p_{\lambda_0}(x)| p_{\lambda_0}(x) dx < \infty$.

(v) $\frac{1}{n} \sup_{\nu \in K \times M} |\widehat{L}_n(\nu) - L_n(\nu)| \xrightarrow[n\to\infty]{\text{a.s.}} 0$.

Then the maximum likelihood estimator $\hat{\nu}_n$ *is strongly consistent:*

$$(\widehat{\theta}, \widehat{\lambda}) \xrightarrow[n\to\infty]{\text{a.s.}} (\theta_0, \lambda_0).$$

Remark 2.3. We prove following Theorems 5.3.1 and 6.1.1 from Straumann [71] and has a standard structure. First, it is shown that $\hat{L}_n/n \overset{a.s.}{\to} L$ in $\mathbb{C}(K \times M)$ as $n \to \infty$, where

$$L(\nu) = E\ell_0(\nu) = -\frac{1}{2} E(\log h_0(\theta)) + E\left[\log p_\lambda\left(\frac{\rho(X_0)}{(h_0(\theta))^{1/2}}\right)\right], \quad \nu \in K \times M.$$

Next we have to prove that $L(\nu)$ is uniquely maximized at $\nu = \nu_0$. In the end, using the standard arguments we can show that a strong consistency follows from the almost sure convergence of $\hat{L}_n/n$ towards L together with the fact that the limit L is uniquely maximized.

Proof. First we will show that $L_n/n \overset{a.s.}{\to} L$ in $\mathbb{C}(K \times M)$. The sequence

$$\ell_t = -(1/2) \log h_t + \log p_\lambda\Big(\frac{\rho(X_t)}{h_t^{1/2}}\Big)$$

consists of random elements with the values in $C(K \times M)$ as the density function p_λ is continuous according to assumption M.2. Besides, (ℓ_t) is of the form $(\ell_t) = (f(X_t, X_{t-1}, \dots))$, where f is a measurable function and the process $(X_t, t \in \mathbb{Z})$ is strictly stationary and ergodic, since condition (2.8) of Theorem 2.1 is satisfied by the assumption (i). Hence, from Theorem A.2. in [40] it follows that the sequence (ℓ_t) is stationary and ergodic. To apply Theorem 2.2.1 in [71] (see, Theorem A.2 in Appendix 1) to the sequence (ℓ_t) we have to verify if $E \sup_{\nu \in K \times M} |\ell_0(\nu)| < \infty$. Recall that

$$h_0(\theta) = \frac{\omega}{1-\beta} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} \rho^2(X_{-j}),$$

and define the compact $K = [\underline{\omega}, \bar{\omega}] \times [\underline{\alpha}, \bar{\alpha}] \times [\underline{\beta}, \bar{\beta}]$, $\underline{\omega} > 0$, $\bar{\beta} < 1$. Since by assumption (i) $E\rho^2(X_0) < \infty$, we can show that

$$E \sup_{\theta \in K} |h_0(\theta)| = E\Big(\frac{\bar{\omega}}{1-\bar{\beta}} + \bar{\alpha} \sum_{j=1}^{\infty} \bar{\beta}^{j-1} \rho^2(X_{-j}) \Big) = \frac{\bar{\omega}}{1-\bar{\beta}} + \frac{\bar{\alpha}}{1-\bar{\beta}} E\rho^2(X_0) < \infty. \tag{2.19}$$

Therefore by Jensen's inequality it follows that

$$E \sup_{\theta \in K} |\log h_0(\theta)| \leq \log E \sup_{\theta \in K} h_0(\theta) < \infty. \tag{2.20}$$

Since ν_0 is the maximizer of the likelihood function and using the assumption (iv), we obtain:

$$\begin{aligned} E \sup_{\nu \in K \times M} \left| \log p_\lambda \Big(\frac{\rho(X_0)}{h_0^{1/2}(\theta)} \Big) \right| &= E \left| \log p_{\lambda_0} \Big(\frac{\rho(X_0)}{h_0^{1/2}(\theta_0)} \Big) \right| = E |\log p_{\lambda_0}(\rho(\varepsilon_0))| \\ &= \int_0^{\infty} |\log p_{\lambda_0}(x)| p_{\lambda_0}(x) dx < \infty. \end{aligned} \tag{2.21}$$

From (2.20) and (2.21) we can conclude that

$$\begin{aligned} E \sup_{\nu \in K \times M} |\ell_0| &= E \sup_{\nu \in K \times M} \left| \frac{1}{2} \log h_0(\theta) + \log p_\lambda \Big(\frac{\rho(X_0)}{h_0^{1/2}(\theta)} \Big) \right| \leq \\ &\leq \frac{1}{2} E \sup_{\theta \in K} |\log h_0(\theta)| + E \sup_{\nu \in K \times M} \left| \log p_\lambda \Big(\frac{\rho(X_0)}{h_0^{1/2}(\theta)} \Big) \right| < \infty. \end{aligned}$$

Thus, from Theorem 2.2.1 in [71] it follows that $L_n/n \overset{a.s.}{\to} L$ in $\mathbb{C}(K \times M)$. Since by assumption (v) $\frac{1}{n} \sup_{\nu \in K \times M} |\widehat{L}_n(\nu) - L_n(\nu)| \xrightarrow[n \to \infty]{\text{a.s.}} 0$, we can conclude that $\hat{L}_n/n \overset{a.s.}{\to} L$.

Next, we will show that the function $L(\nu)$ is uniquely maximized at $\nu = \nu_0$. We have established $E \sup_{\nu \in K \times M} |\ell_0(\nu)| < \infty$, therefore it follows that the function

$L(\nu_0) < \infty$. Set

$$f_t(\nu) = \frac{1}{(h_t(\theta))^{1/2}} p_\lambda\Big(\frac{\rho(X_t)}{(h_t(\theta))^{1/2}}\Big).$$

Since $\log x \leq x - 1$, $\forall x > 0$ with equality if and only if $x = 1$, we can write that

$$L(\nu) - L(\nu_0) = E\Big(\log \frac{f_0(\nu)}{f_0(\nu_0)}\Big) \leq E\Big(\frac{f_0(\nu)}{f_0(\nu_0)}\Big) - 1 \tag{2.22}$$

with equality if and only if $f_0(\nu) = f_0(\nu_0)$ a.s.

Denote $r(\theta) = \sigma_0/\sqrt{h_0(\theta)}$. Recall that $\rho(X_0) = \sigma_0\rho(\varepsilon_0) = \sqrt{h_0(\theta_0)}\rho(\varepsilon_0)$, then we have

$$f_0(\nu) = \frac{1}{\sqrt{h_0(\theta)}} p_\lambda\left(\frac{\rho(X_0)}{\sqrt{h_0(\theta)}}\right) = \frac{1}{\sqrt{h_0(\theta)}} p_\lambda\left(\frac{\sigma_0\rho(\varepsilon_0)}{\sqrt{h_0(\theta)}}\right) = \frac{1}{\sqrt{h_0(\theta)}} p_\lambda(r(\theta)\rho(\varepsilon_0)),$$

$$f_0(\nu_0) = \frac{1}{\sqrt{h_0(\theta_0)}} p_{\lambda_0}\left(\frac{\rho(X_0)}{\sqrt{h_0(\theta_0)}}\right) = \frac{1}{\sigma_0} p_{\lambda_0}\Big(\frac{\sigma_0\rho(\varepsilon_0)}{\sigma_0}\Big) = \frac{1}{\sigma_0} p_{\lambda_0}(\rho(\varepsilon_0)),$$

and since $\rho(\varepsilon_0)$ is independent of $r(\theta)$, it follows that

$$\begin{aligned} E\Big(\frac{f_0(\nu)}{f_0(\nu_0)}\Big) &= E\Big(E\Big[\frac{r(\theta)p_\lambda(r(\theta)\rho(\varepsilon_0))}{p_{\lambda_0}(\rho(\varepsilon_0))}\Big| r(\theta)\Big]\Big) = E\Big(\int \frac{r(\theta)p_\lambda(r(\theta)x)}{p_{\lambda_0}(x)} p_{\lambda_0}(x)dx\Big) \\ &= E\Big(\int p_\lambda(r(\theta)x)d(r(\theta)x)\Big) = E(1) = 1. \end{aligned}$$

Hence and from inequality (2.22) we can conclude that $L(\nu) \leq L(\nu_0)$ with equality if and only if $f_0(\nu) = f_0(\nu_0)$ a.s.

Observe that $f_0(\nu) = f_0(\nu_0)$ a.s. is equivalent to

$$p_{\lambda_0}(\rho(\varepsilon_0)) = r(\theta)p_\lambda(r(\theta)\rho(\varepsilon_0)) \quad \text{a.s.} \tag{2.23}$$

We will show that (2.23) implies $\theta = \theta_0$ and $\lambda = \lambda_0$. Suppose by contrast that $\theta \neq \theta_0$. Then, since $\sigma_0^2 = h_0(\theta)$ a.s. if and only if $\theta = \theta_0$, we have $P[r(\theta) \neq 1] > 0$. So, by Lemma 6.2.1. from [71][3] we derive

$$P[p_{\lambda_0}(\rho(\varepsilon_0)) \neq r(\theta)p_\lambda(r(\theta)\rho(\varepsilon_0))|r(\theta)] > 0 \quad \text{on} \quad \{r(\theta) \neq 1\}, \tag{2.24}$$

and therefore

$$P[p_{\lambda_0}(\rho(\varepsilon_0)) \neq r(\theta)p_\lambda(r(\theta)\rho(\varepsilon_0))] = E[P[p_{\lambda_0}(\rho(\varepsilon_0)) \neq r(\theta)p_\lambda(r(\theta)\rho(\varepsilon_0))|r(\theta)]] > 0, \tag{2.25}$$

which contradicts (2.23). So we can conclude that $\theta = \theta_0$. From assumption

[3]Let $a > 0$ be a constant. Then for all $\lambda \in M$: $a \neq 1 \Rightarrow P[p_{\lambda_0}(\rho(\varepsilon_0)) \neq a p_\lambda(a\rho(\varepsilon_0))] > 0$.

M.3 it follows that $p_{\lambda_0}(\rho(\varepsilon_0)) = p_\lambda(\rho(\varepsilon_0))$ with probability 1 implies that $\lambda = \lambda_0$. Altogether, we have shown that $L(\nu) \le L(\nu_0)$ with equality if and only if $\nu = \nu_0$.

Finally, it remains to show that, almost sure uniform convergence of $\hat{L}/n$ towards L together with the fact that the limit L has a unique maximum, implies a strong consistency. Suppose $\delta > 0$ is arbitrary and $P(\limsup_{n\to\infty} |\widehat{\nu}_n - \nu_0| > \delta) > 0$. Define the set $K' \times M' = K \times M \cap \{\nu : |\nu - \nu_0| \ge \delta\}$. Since the set $K' \times M'$ is compact and $\widehat{L}/n \stackrel{a.s.}{\to} L$ in $C(K \times M)$ there is an event $W \subset \{\limsup_{n\to\infty} |\widehat{\nu}_n - \nu_0| > \delta\}$ with a positive probability and being such that for every $w \in W$, one can find a convergent subsequence $(\widehat{\nu}_{n_k}) \subset K' \times M'$ with $\lim \widehat{\nu}_{n_k} = \nu$ and $\widehat{L}_{n_k}/n_k \to L$ in $C(K \times M)$, where (n_k) and $\nu \in K' \times M'$ depend on the realization w. On the other hand, by the definition of the maximum likelihood estimation $L(\nu) = \lim \widehat{L}_{n_k}(\widehat{\nu}_{n_k})/n_k \ge \lim \widehat{L}_{n_k}(\widehat{\nu}_0)/n_k = L(\nu_0)$ on W. Since $W \ne \emptyset$, there exists at least one point $\nu \in K' \times M'$ with $L(\nu) \ge L(\nu_0)$ and it is a contradiction, since L is uniquely maximized at $\nu = \nu_0$. Note that $\delta > 0$ was chosen arbitrarily, therefore we can conclude that $\widehat{\nu}_n \stackrel{a.s.}{\to} \nu_0$ as $n \to \infty$. □

2.5 Some examples

In this section, we consider a case of the model (2.6), (2.7) where the density function of $\rho(\varepsilon_0)$ is known.

Example 2.1. Assume that $\mathcal{G} = [0,1]$, i.e., each $s \in [0,1]$ corresponds to a certain aggregation rule. Consider the $C[0,1]$-valued time series $(X_t, t \in \mathbb{Z})$ expressed as

$$X_t(s) = \sigma_t \varepsilon_t(s), \quad s \in [0,1],$$
$$\sigma_t^2 = \omega + \alpha\Big(\max_{s\in[0,1]} X_{t-1}(s)\Big)^2 + \beta\sigma_{t-1}^2,$$

where $(\varepsilon_t(s), s \in [0,1]), t \in \mathbb{Z}$ are independent identically distributed Wiener processes.

This example corresponds to the model (2.6), (2.7), where $\mathbb{E} = C[0,1]$ and the semi-norm $\rho(f) = \max_{0\le s\le 1} f(s)$.

The density of $\rho(\varepsilon_t) = \max_{s\in[0,1]} \varepsilon_t(s)$ is well known (see, e.g., [17]) and can be expressed as

$$f_{\rho(\varepsilon)}(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right).$$

To show almost sure convergence of the maximum likelihood estimator we verify the assumptions for the density function of Theorem 2.3.

To fulfil the assumption (i), define the parameter space as follows

$$\omega \in [\omega_1, \omega_2], \quad \omega_1 > 0, \quad \omega_2 < \infty$$

$$\alpha \in [\alpha_1, \alpha_2] \quad \alpha_1 > 0, \quad \alpha_2 < 1$$

$$\beta \in [\beta_1, \beta_2], \quad \beta_1 > 0, \quad \beta_2 < 1$$

To check the assumption (iv) we have

$$E|\log p(\rho(\varepsilon_t)| = E\left|\log\sqrt{\frac{2}{\pi}} - \frac{1}{2}\log\sigma_t^2 - \frac{1}{2}\frac{\rho^2(X_t)}{\sigma_t^2}\right| \le C + E\log\sigma_0^2 + E\rho^2(\varepsilon_0) < \infty,$$

where C is a constant.

For the assumption (v) first observe

$$\sup_{\theta\in K}|\tilde{\sigma}_t^2 - \sigma_t^2| = \sup_{\theta\in K}|\beta^n[\alpha(\tilde{\rho}^2(X_0) - \rho^2(X_0)) + (\tilde{\sigma}_0^2 - \sigma_0^2)]| \le \beta_2^n K, \qquad (2.26)$$

where K>0 is a constant and $\tilde{\rho}^2(X_0)$, $\tilde{\sigma}_0^2$ are fixed initial values. Now we estimate

$$\begin{aligned}
\frac{1}{n}\sup_{\theta\in K}|\widehat{L}_n - L_n| &= \frac{1}{n}\sup_{\theta\in K}\left|\sum_{t=1}^{n}(\log\tilde{\sigma}_t^2 - \log\sigma_t^2) + \frac{1}{2}\sum_{t=1}^{n}\rho^2(X_t)\left(\frac{1}{\tilde{\sigma}_t^2} - \frac{1}{\sigma_t^2}\right)\right| \\
&\le \frac{1}{n}\sup_{\theta\in K}\left|\sum_{t=1}^{n}\left(\frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2}\right) + \frac{1}{2}\sum_{t=1}^{n}\rho^2(X_t)\left(\frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\tilde{\sigma}_t^2\sigma_t^2}\right)\right| \\
&\le \frac{1}{\omega_1 n}\sum_{t=1}^{n}\sup_{\theta\in K}|\tilde{\sigma}_t^2 - \sigma_t^2| + \frac{1}{2\omega_1^2 n}\sum_{t=1}^{n}\rho^2(X_t)\sup_{\theta\in K}|\tilde{\sigma}_t^2 - \sigma_t^2| \\
&\le \frac{c}{n}\sum_{t=1}^{n}(1 + \rho^2(X_t))\sup_{\theta\in K}|\tilde{\sigma}_t^2 - \sigma_t^2|,
\end{aligned}$$

where $c = \max(\frac{1}{\omega_1}, \frac{1}{2\omega_1^2}) > 0$ denotes a constant. To obtain the first inequality, we have used $\log x \le x - 1$, if $x > 0$. Together with (2.26) we get that

$$\frac{1}{n}\sup_{\theta\in K}|\widehat{L}_n - L_n| \le \frac{cK}{n}\sum_{t=1}^{n}\beta_2^t(1 + \rho^2(X_t)) \to 0 \quad \text{a.s.}, \quad n \to \infty$$

since $E\log^+(1 + \rho^2(X_t)) < \infty$ by Lemma 2.5.3 of Straumann [71]; and using Proposition 2.5.1 of Straumann [71] (see, Proposition A.3 in Appendix 1), one can conclude that $\sum_{t=1}^{\infty}\beta_2^t(1 + \rho^2(X_t)) < \infty$ a.s.

In this example, we have made an assumption that the innovations are Wiener processes. However, the distribution of the maximum random variable is known only for several stochastic processes, such as the Brownian motion (as in example 2.1), the Brownian bridge, the Brownian motion with a linear drift and several cases of the stationary Gaussian processes with a specified covariance function (see [8] for the list of processes and the corresponding references). Even if the distribution of the maximum process of innovations is unknown, one can use the quasi-maximum likelihood approach for parameter estimation, i.e., an assumption about Gaussian innovations can be made and the Gaussian maximum likelihood function is then analyzed.

Example 2.2. Consider the $C[0,1]$-valued time series $(X_t, t \in \mathbb{Z})$, where

$$\begin{aligned} X_t(s) &= \sigma_t \varepsilon_t(s), \quad s \in [0,1], \\ \sigma_t^2 &= \omega + \alpha \min_{s \in [0,1]} X_{t-1}^2(s) + \beta \sigma_{t-1}^2, \end{aligned}$$

and $(\varepsilon_t(s), s \in [0,1]), t \in \mathbb{Z}$ are independent identically distributed Wiener processes.

Thus this example corresponds to the model (2.6), (2.7), where $\mathbb{E} = C[0,1]$ and the semi-norm $\rho(f) = \min_{0 \le s \le 1} f(s)$. It is known that the random processes $\Big(\max_{s \in [0,1]} \varepsilon_t(s)\Big)$ and $\Big(-\min_{s \in [0,1]} \varepsilon_t(s)\Big)$ have the same distributions, hence the conditions of the maximum likelihood estimator consistency can be verified similarly as in example 2.1.

Example 2.3. Here, we take the case, where the class $\mathcal{G} = \{g_1, \ldots, g_d\}$ consists of d aggregation rules.

Consider the R^d-valued time series $(X_t, t \in \mathbb{Z})$:

$$\begin{aligned} X_{jt} &= \sigma_t \varepsilon_{jt}, \quad j = 1, \ldots, d, \\ \sigma_t^2 &= \omega + \alpha \rho^2(X_{t-1}) + \beta \sigma_{t-1}^2, \end{aligned}$$

where $(\varepsilon_{jt}, j = 1, \ldots . d), t \in \mathbb{Z}$ are independent identically distributed Gaussian random vectors with zero mean and the covariance matrix $\Lambda = (\lambda_{ij}, i, j = 1, \ldots, d)$ and ρ is a semi-norm on R^d.

As an example, we consider the following semi-norm

$$\rho(x) = \Big(\sum_{i=1}^{d} x_i^2\Big)^{\frac{1}{2}},$$

and assume that $\lambda_{ii} = 1, i = 1, \ldots d$. Then $\rho^2(\varepsilon_{jt}) = \sum_{i=1}^{d} \varepsilon_{jt}^2$, $t \in \mathbb{Z}$ has a χ^2-distribution with the known density function:

$$f_{\chi^2}(x) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} x^{\frac{d}{2}-1} \exp\Big(-\frac{x}{2}\Big).$$

The conditional distribution function of $\rho(X_{jt}) = \sigma_t\rho(\varepsilon_{jt})$ can be written as

$$P(\rho(X_{jt}) \le x|\mathcal{F}_{t-1}) = P(\sigma_t\rho(\varepsilon_{jt}) \le x|\mathcal{F}_{t-1}) = P\left(\left(\sum_{j=1}^{d}\varepsilon_{jt}^2\right)^{\frac{1}{2}} \le \frac{x}{\sigma_t}\middle|\mathcal{F}_{t-1}\right)$$
$$= P\left(\left(\sum_{j=1}^{d}\varepsilon_{jt}^2\right) \le \frac{x^2}{\sigma_t^2}\middle|\mathcal{F}_{t-1}\right),$$

and the conditional density function is as follows

$$f_{\rho(X)}(x) = \frac{2x}{\sigma_t^2}f_{\chi^2}\left(\frac{x^2}{\sigma_t^2}\right) = \frac{1}{2^{\frac{d}{2}-1}\Gamma(\frac{d}{2})}\frac{x^{d-1}}{\sigma_t^d}\exp\left(-\frac{x^2}{2\sigma_t^2}\right).$$

Using the same considerations as in Example 2.1, we can verify the validity of assumptions in Theorem 2.3. For the assumption (iv) we can write

$$\begin{aligned}
&E|\log f_X(\rho(X_{jt}))| \\
=&E\left|\frac{d-1}{2}\log\frac{\sum_{j=1}^{d}X_{jt}^2}{\sigma_t^2} - \log 2^{\frac{d}{2}-1} - \log\Gamma(\frac{d}{2}) - \frac{1}{2}\log\sigma_t^2 - \frac{1}{2\sigma_t^2}\sum_{j=1}^{d}X_{jt}^2\right| \\
\le&c_1 + c_2E\log\sum_{j=1}^{d}\varepsilon_{j0}^2 + E\log\sigma_0^2 + E\sum_{j=1}^{d}\varepsilon_{j0}^2 < \infty,
\end{aligned}$$

where c_1 and c_2 are positive constants. Following the same arguments as in Example 2.1, for the assumption (v) we have:

$$\frac{1}{n}\sup_{\theta\in K}|\widehat{L}_n - L_n| =$$
$$\frac{1}{n}\sup_{\theta\in K}\left|\frac{d}{2}\sum_{t=1}^{n}(\log\sigma_t^2 - \log\tilde{\sigma}_t^2) + \frac{1}{2}\sum_{t=1}^{n}\sum_{j=1}^{d}X_{jt}^2\left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right)\right| \to 0 \quad \text{a.s.},$$

as $n \to \infty$.

Example 2.4. We use here the point-wise aggregation. Consider the $C[0,1]$-valued time series $(X_t, t \in \mathbb{Z})$, where

$$X_t(s) = \sigma_t\varepsilon_t(s), \quad s \in [0,1], \tag{2.27}$$
$$\sigma_t^2 = \omega + \alpha(X_{t-1}(1) - X_{t-1}(0))^2 + \beta\sigma_{t-1}^2, \tag{2.28}$$

and $(\varepsilon_t(s), s \in [0,1]), t \in \mathbb{Z}$ are independent identically distributed Gaussian processes.

So this example corresponds to the model (2.6), (2.7), where $\mathbb{E} = C[0,1]$ and the semi-norm $\rho(f) = |f(1) - f(0)|$. This case also has a practical explanation, when the returns of a financial asset prices are analyzed. Assume, for example, that the returns of a share, traded on the stock exchange, are taken. From the model defined by equations (2.27), (2.28), we can see that the volatility depends on the difference between the returns taken at the beginning and at the end of the day, i.e., the opening and closing price returns.

The density function of $[\varepsilon_t(1) - \varepsilon_t(0)]$ is well known and, using the density function of the bivariate Normal distribution with zero mean and the covariance matrix $\begin{pmatrix} 1 & \xi \\ \xi & 1 \end{pmatrix}$, can be written as

$$f_{[\varepsilon_t(1)-\varepsilon_t(0)]} = \frac{1}{2\sqrt{\pi(1-\xi)}} \exp\left\{ -\frac{x^2}{4(1-\xi)} \right\}.$$

The distribution of $|\varepsilon_t(1) - \varepsilon_t(0)|$ is symmetric and the conditional density function of $\rho(X_t(s)) = \sigma_t \rho(\varepsilon_t(s)) = \sigma_t |\varepsilon_t(1) - \varepsilon_t(0)|$ can be written as

$$f_{\rho(X)}(x) = \frac{1}{\sigma_t \sqrt{\pi(1-\xi)}} \exp\left\{ -\frac{x^2}{4\sigma_t^2(1-\xi)} \right\}.$$

Since the density function is the same, except for constants, as the density function analyzed in Example 2.1, the assumptions of Theorem 2.3 can be verified similarly as in Example 2.1.

2.6 Conclusions

In the thesis, the data aggregation problem in the risk measurement is considered. The empirical analysis shows that risk estimates depend on aggregation. Consequently, there is a question, how to choose an aggregation rule. One of the ways to select an aggregation rule is to take the one, where the aggregated data fit the model best. Another way, with a conservative approach to the risk measurement, could be choosing such a rule, which gives the largest estimate of risk. However, the difficulty of this question also consists in the fact that the aggregation scheme, selected under an appropriate rule, might change every day. With regard to this observation, we have developed a functional GARCH-type model. The main idea of this model is the definition of volatility, which is stable during the day, but depends on some features of high frequency returns. We have established the conditions for the existence of a stationary solution and for the consistency of the maximum likelihood estimator in this model. Finally, several practical examples of the model were presented.

Chapter 3

uvGARCH(1, 1) model in a Hilbert space

In this chapter, following Kvedaras and Račkauskas (2010) we consider a functional aggregation of high frequency data producing a regular functional time series with values in a certain Hilbert space. There are several methods to construct functional observations from high frequency data. One can use the methodology introduced by Ramsay and Silverman (1997). They presented several techniques for converting raw data into a functional form, such as basis functions methods, smoothing by local weighing and the roughness penalty approach. Moreover, the direct construction of functional data can be chosen as well. For example, consecutive maximal values of high frequency observations produce nondecreasing functions

$$y_t^*(s) = \max\{y_j | \tau_j \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*], \quad s \in [0,1]$$

whereas consecutive averages result in

$$y_t^*(s) = m_t^{-1}(s) \sum_{\tau_j \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]} y_j,$$
$$m_t(s) = \#\{\tau_j \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}, \quad s \in [0,1].$$

Here $(y_t^*(s), j = 1, \ldots, N^*)$, $s \in [0,1]$ are functional observations, aggregated from high frequency irregularly spaced time series $\{(\tau_j, y_j)\}_{j=1}^N$ with a fixed time interval $\delta > 0$ and $\tau_t^* = t\delta$, $t = 1, \ldots, N^*$.

With any of the mentioned regularization of high frequency observations one receives the functional time series $(y_t^*(s), s \in [0,1])$. One of the classical Hilbert spaces $\mathbb{H}$, e.g., $\mathbb{H} = L_2(0,1)$ the space of square Lebesgue integrable functions, is usually considered as a path space. For detailed information about statistical modeling of functional data we refer to [18] and [66]. We are interested in extending the real-valued conditional heteroscedastic models to a functional framework. In this chapter, the general Hilbert space-valued time series is considered and the GARCH$(1,1)$ model with univariate volatility is introduced and its properties are investigated.

3.1 Model

Let $\mathbb{H}$ be a real separable Hilbert space of infinite or finite dimension with the inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $|| \cdot ||$, $||x||^2 = \langle x, x \rangle$, $x \in \mathbb{H}$. Classical Hilbert spaces include the Lebesgue space $L_2(0,1)$ of measurable square integrable functions $x : [0,1] \to \mathbb{R}$ endowed with the inner product $\langle x, y \rangle = \int_0^1 x(t)y(t)dt$. Another important framework for functional data analysis is the Hilbert space $L_{2,1}(0,1)$ of differentiable functions $x : (0,1) \to \mathbb{R}$ such that $\int_0^1 (x'(t))^2 dt < \infty$ with the inner product $\langle x, y \rangle = \int_0^1 x(t)y(t)dt + \int_0^1 x'(t)y'(t)dt$.

The space of bounded linear operators $u : \mathbb{H} \to \mathbb{H}$ is denoted by $L(\mathbb{H})$. We consider $L(\mathbb{H})$ as the Banach space with the usual uniform norm $||u|| = \sup_{||x|| \le 1} ||ux||$. For $x, y \in \mathbb{H}$, we denote as $x \otimes y$ the linear operator on $\mathbb{H}$,

$$x \otimes y(z) = \langle x, z \rangle y, \quad z \in \mathbb{H}.$$

The covariance operator of an $\mathbb{H}$-valued random element X is $Q_X = E(X \otimes X)$, $Q_X x = E\langle X, x \rangle X$, $x \in \mathbb{H}$. It exists and is bounded linear operator whenever $E\langle X, z \rangle^2 < \infty$ for any $z \in \mathbb{H}$. If $E||X||^2 < \infty$, then the operator Q_X is nuclear.

Definition 3.1. Let $(X_t, t \in \mathbb{Z})$ be an $\mathbb{H}$-valued random process,$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$. We say that (X_t) is GARCH$(1,1)$ with univariate volatility (in short uvGARCH$(1,1)$) if

$$X_t = \sigma_t \varepsilon_t, \quad (\varepsilon_t) \sim \text{iid } (0, Q_\varepsilon), \tag{3.1}$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \langle X_{t-1}, z \rangle^2, \quad t \in \mathbb{Z}, \tag{3.2}$$

where $\omega > 0, \beta \ge 0$ and $z \in \mathbb{H}$ are parameters of interest.

We also assume that $(\varepsilon_t, t \in \mathbb{Z})$ are independent identically distributed $\mathbb{H}$-valued random elements with zero mean and covariance Q_ε. It is clear that (X_t) when projected in the direction z, namely, the time series $(\langle X_t, z \rangle, t \in \mathbb{Z})$, follows the classical GARCH$(1,1)$ model. However, the direction z is unknown.

3.2 Stationarity

In this section, the conditions for the existence of the strong and the 2nd order stationary solution of the model (3.1), (3.2) are stipulated.

Theorem 3.1. *If*

$$-\infty \leq \gamma := E \log\{\langle \varepsilon_0, z\rangle^2 + \beta\} < 0, \tag{3.3}$$

then the series

$$h_t := \omega + \omega \sum_{n=1}^{\infty} \prod_{j=1}^{n} \Big(\langle \varepsilon_{t-j}, z\rangle^2 + \beta \Big) \tag{3.4}$$

converges a.s. and the process $(X_t, t \in \mathbb{Z})$ *defined as*

$$X_t = h_t^{1/2} \varepsilon_t, \quad t \in \mathbb{Z}, \tag{3.5}$$

is the unique strictly stationary solution of the model (3.1), (3.2).

Theorem 3.1 can be proved following the lines of the strong stationarity Theorem 2.1 proof, presented in the previous chapter. We will use here the method based on the techniques of stochastic recurrence equations (SREs), introduced by Straumann (2005).

More general Hilbert space valued GARCH models with univariate volatility are obtained replacing (3.2) by

$$\sigma_t^2 = g_\theta(X_{t-1}, \sigma_{t-1}^2), \quad t \in \mathbb{Z}, \tag{3.6}$$

where the volatility process (σ_t) is a nonnegative real-valued process and (ε_t) is a sequence of iid $\mathbb{H}$-valued random elements with zero mean and the known covariance Q_ε. Nonnegative functions $\{g_\theta, \theta \in \Theta\}$ are defined on $\mathbb{H} \times [0, \infty)$ and it is necessary that σ_t were $\mathcal{F}_{t-1} = \sigma(X_s, s \leq t-1)$-measurable for $t \in \mathbb{Z}$.

The solution of stationarity for such models is obtained via stochastic recurrence equations (for details of this approach see [71] and references therein):

$$s_{t+1} = \psi_t(s_t), \quad t \in \mathbb{Z}, \tag{3.7}$$

on $[0, \infty)$, where

$$\psi_t(s) = g_\theta(s^{1/2}\varepsilon_t, s), \quad s \in [0, \infty).$$

Assume that (ψ_t) is the stochastic process with values in a complete separable metric space $\mathbb{E}$ endowed with the σ-algebra $\mathcal{E}$. It is clear that, if (s_t) is a solution of (3.7) and s_t is $\mathcal{F}_{t-1}$-measurable, then the sequence $(s_t^{1/2}\varepsilon_t, s^{1/2})$ is stationary and fits (3.1), (3.6).

For an integer $m \geq 0$ set $\psi_t^{(m)} = \psi_t \circ \cdots \circ \psi_{t-m}$. For a function f set

$$\Lambda(f) = \sup_{s \neq t} \frac{|f(s) - f(t)|}{|s - t|}.$$

The following result is taken from [71].

Proposition 3.2. *Let the functional process* (ψ_t) *be stationary and ergodic. Fix an arbitrary* $s_0^2 \in [0, \infty)$ *and suppose that the following conditions hold:*

(*a*) $E(\log^+ |\psi_0(s_0^2)|) < \infty$*;*

(*b*) $E[\log^+ \Lambda(\psi_0)] < \infty$ *and for some integer* $r \geq 1$ *it holds that* $E[\log \Lambda(\psi_0^{(r)})] < 0$.

Then the stochastic recurrence equations (3.7) *admits a unique stationary ergodic solution* (σ_t^2) *such that* σ_t^2 *is* $\mathcal{F}_{t-1}$*-measurable for every* $t \in \mathbb{Z}$*. Moreover,*

$$\sigma_t^2 = \lim_{m \to \infty} \psi_{t-1} \circ \cdots \circ \psi_{t-m}(s_0^2), \quad t \in \mathbb{Z}, \tag{3.8}$$

and the limit does not depend on s_0^2.

For the model (3.1, 3.2), we have $g_\theta(x, s) = \omega + \beta s + \langle x, z \rangle^2$ and

$$\psi_t(s) = g_\theta(s^{1/2}\varepsilon_t, s) = \omega + (\beta + \langle \varepsilon_t, z \rangle^2)s, \quad s \in [0, \infty), \ t \in \mathbb{Z}.$$

The corresponding Lipshitz exponents are

$$\Lambda(\psi_0) = \beta + \langle \varepsilon_0, z \rangle^2 \quad \text{and} \quad \Lambda(\psi_0^{(r)}) = (\beta + \langle \varepsilon_{-1}, z \rangle^2) \cdots (\beta + \langle \varepsilon_{-r}, z \rangle^2), \quad r \geq 1.$$

Theorem 3.1 follows immediately from Proposition 3.2. Indeed, since $s_0^2 \in [0, +\infty)$ is taken arbitrary, for the condition S.1. we can write

$$E(\log^+ |\psi_0(0)|) = E(\log^+ \omega) < \infty.$$

Note that $\log \Lambda(\psi_0^{(r)}) = \sum_{i=1}^r \log(\beta + \langle \varepsilon_{-i}, z \rangle^2)$, and for this reason the condition S.2. is equivalent to

$$E(\log(\beta + \langle \varepsilon_0, z \rangle^2) < 0.$$

Since relation (3.8) is valid for arbitrary initial values, we obtain

$$\begin{aligned}
\sigma_t^2 &= \lim_{m \to \infty} \psi_{t-1} \circ \cdots \circ \psi_{t-m}(0) = \lim_{m \to \infty} \omega \Big(1 + \sum_{k=1}^m \prod_{i=1}^k (\beta + \langle \varepsilon_{t-i}, z \rangle^2)\Big) \\
&= \omega \Big(1 + \sum_{k=1}^\infty \prod_{i=1}^k (\beta + \langle \varepsilon_{t-i}, z \rangle^2)\Big), \quad a.s.
\end{aligned}$$

Definition 3.2. The $\mathbb{H}$-valued process $(X_t, t \in \mathbb{Z})$ is second-order stationary if for all $t, h \in \mathbb{Z}$ it holds that

(i) $E||X_t||^2 < \infty$,

(ii) $EX_t = \mu$,

(iii) $E\langle X_t - \mu, x\rangle\langle X_{t+h} - \mu, y\rangle = \Gamma_h(x, y)$, $x, y \in \mathbb{H}$.

Theorem 3.3. *Suppose that $(\varepsilon_t, t \in \mathbb{Z})$ are iid square integrable random elements, i.e., is $E||\varepsilon_0||^2 < \infty$. If $E\langle\varepsilon_0, z\rangle^2+\beta < 1$, the process $(X_t, t \in \mathbb{Z})$ defined by (3.5) is the unique second order stationary solution of equations* (3.1), (3.2). *In addition, for $p \geq 2$, if $E||\varepsilon_0||^p < \infty$ and $E(\langle\varepsilon_0, z\rangle^2 + \beta)^{p/2} < 1$, solution (3.5) has a finite p'th order moment.*

Remark 3.1. For the proof of Theorem 3.3 one has to verify, that the conditions yield the existence of a strong stationary solution, which has a finite second moment indeed. Therefore, the proof is analogous to that of the 2nd order stationarity of the model $\rho -$ GARCH(1, 1), obtained by Theorem 2.2.

Proof. By Jensen's inequality for a concave function we get

$$E \log\{\langle\varepsilon_{t-j}, z\rangle^2 + \beta\} \leq \log E\{\langle\varepsilon_0, z\rangle^2 + \beta\} < 0.$$

Hence, condition (3.3) is satisfied, therefore a strictly stationary solution of equations (3.1), (3.2) exists and is given by (3.5). Thus, it is enough to show that the process defined by (3.5) has a finite variance. Set $q_0 = E\langle\varepsilon_0, \varepsilon_0\rangle$. By a monotone convergence theorem we derive:

$$\begin{aligned} E\|X_t\|^2 = q_0 E h_t &= q_0\omega \left\{1 + \sum_{n=1}^{\infty} E \prod_{j=1}^{n} (\langle\varepsilon_{t-j}, z\rangle^2 + \beta)\right\} \\ &= q_0\omega \left\{1 + \sum_{n=1}^{\infty} \left(E\langle\varepsilon_0, z\rangle^2 + \beta\right)^n\right\} = \frac{q_0\,\omega}{1 - (E\langle\varepsilon_0, z\rangle^2 + \beta)}. \end{aligned}$$

To prove the uniqueness, assume that $\tilde{X}_t = (\omega + \omega\tilde{h}_t)^{1/2}\varepsilon_t$ is another second-order stationary solution of (3.1), (3.2). Suppose $\mathbb{P}(h_t \neq \tilde{h}_t) > 0$ for a certain t. Then

$$\begin{aligned} E|h_t - \tilde{h}_t| &= E\left\{\prod_{j=1}^{n} (\langle\varepsilon_{t-j}, z\rangle^2 + \beta)\right\} E|h_{t-(n+1)} - \tilde{h}_{t-(n+1)}| \\ &= (E\langle\varepsilon_0, z\rangle^2 + \beta)^n E|h_{t-(n+1)} - \tilde{h}_{t-(n+1)}|. \end{aligned}$$

Since $E|h_{t-(n+1)} - \tilde{h}_{t-(n+1)}| \leq E|h_{t-(n+1)}| + E|\tilde{h}_{t-(n+1)}|$ is finite and due to stationarity independent of n; and $(\alpha E\langle\varepsilon_0, z\rangle^2 + \beta)^n$ approaches 0, as $n \to \infty$, we obtain that $E|h_t - \tilde{h}_t| = 0$ and therefore for every t, $h_t = \tilde{h}_t$ *a.s.*

To prove that the process $(X_t, t \in \mathbb{Z})$ has a finite p'th moment, we have to show that

$$E||X_t||^p = E||h_t^{1/2}\varepsilon_t||^p = Eh_t^{p/2}E||\varepsilon_t||^p < \infty. \tag{3.9}$$

By the assumption of the theorem $E||\varepsilon_t||^p < \infty$. It remains to prove that $Eh_t^{p/2} < \infty$. Recalling the definition of h_t

$$h_t = \omega\Big(1 + \sum_{n=1}^{\infty}\prod_{j=1}^{n}(\langle\varepsilon_{t-j}, z\rangle^2 + \beta)\Big)$$

it suffices to check that

$$E\Big(\sum_{n=1}^{\infty}\prod_{j=1}^{n}(\langle\varepsilon_{t-j}, z\rangle^2 + \beta)\Big)^{p/2} < \infty. \tag{3.10}$$

Since $p/2 \geq 1$, by the Minkovski inequality [1] and taking into account the independence and identical distribution of the sequence (ε_t), (3.10) reduces to

$$\begin{aligned}\left[E\Big(\sum_{n=1}^{\infty}\prod_{j=1}^{n}(\langle\varepsilon_{t-j}, z\rangle^2 + \beta)\Big)^{p/2}\right]^{2/p} &\leq \sum_{n=1}^{\infty}\Big(E\prod_{j=1}^{n}(\langle\varepsilon_{t-j}, z\rangle^2 + \beta)^{p/2}\Big)^{2/p} \\ &= \sum_{n=1}^{\infty}\Big(E(\langle\varepsilon_0, z\rangle^2 + \beta)^{p/2}\Big)^{2n/p} < \infty. \end{aligned} \tag{3.11}$$

The convergence is guaranteed by the assumption of the theorem $E(\langle\varepsilon_0, z\rangle^2 + \beta)^{p/2} < 1$. Thus, we have showed that (3.9) is valid; the proof is complete. □

3.3 Estimation

In this section, we investigate a quasi-maximum likelihood estimator of the model (3.1), (3.2). We assume throughout that $E||\varepsilon_0||^2 < \infty$. In this case, the covariance operator Q_ε is nuclear. Let $(\phi_j) \subset \mathbb{H}$ be the orthonormal basis for $\mathbb{H}$ consisting of eigenfunctions of the operator Q_ε. Let the corresponding eigenvalues be $\mu_1^2 \geq \mu_2^2 \geq \cdots$. We shall consider only the case where the eigenfunctions (ϕ_j) and eigenvalues (μ_j^2) are known.

Assume that $X_1, X_2, \ldots, X_n$ is a sample of the unique stationary ergodic solution to equations (3.1), (3.2) with the true parameter $\theta_0 = (\omega_0, \beta_0, z_0) \in \mathbb{R}^2 \times \mathbb{H}$, where $z_0 = \lambda_1^o\phi_1 + \cdots + \lambda_d^o\phi_d$, with known $d \geq 1$. In this case,

[1] $\Big\|\sum_{i=1}^{\infty} a_i\Big\| \leq \sum_{i=1}^{\infty}||a_i||$, where $||a|| = \Big(E|a|^{p/2}\Big)^{2/p}$.

$\theta_0 = (\omega_0, \beta_0, \lambda_1^o, \ldots, \lambda_d^o) \in [0, \infty)^2 \times \mathbb{R}^d$. Consider the set $K \subset \mathbb{R}^{2+d}$:

$$K = [\underline{\omega}, \overline{\omega}] \times [\underline{\beta}, \overline{\beta}] \times [-\alpha, \alpha]^d, \tag{3.12}$$

where $\underline{\omega} > 0, \overline{\omega} < \infty, \alpha > 0, \underline{\beta} \geq 0$ and $\overline{\beta} < 1$. Suppose that the true parameter θ_0 belongs to the set K.

Denote an Euclidean space $\mathbb{R}^{2+d}$ norm by $\|\cdot\|_d$, which for $x = (x_1, \ldots, x_{d+2}) \in \mathbb{R}^{2+d}$ is defined as

$$\|x\|_d = \max_{1 \leq i \leq d+2} |x_i|.$$

Recall that $C(K)$ is the Banach space of continuous functions $f : K \to \mathbb{R}$ endowed with the uniform norm

$$||f||_K = \sup_{x \in K} |f(x)|.$$

Set $\theta = (\omega, \beta, \ \lambda_1, \ldots, \lambda_d) \in K$ and $z = \lambda_1 \phi_1 + \cdots + \lambda_d \phi_d$,

$$h_t(\theta) = \frac{\omega}{1 - \beta} + \sum_{j=1}^{\infty} \beta^{j-1} \langle X_{t-j}, z \rangle^2. \tag{3.13}$$

This is the solution to the equation

$$h_t(\theta) = \omega + \beta h_{t-1}(\theta) + \langle X_{t-1}, z \rangle^2, \quad \theta = (\omega, \beta, \lambda_1, \ldots, \lambda_d) \in K,$$

in the space $C(K)$ irrelevant to the initial value $h_0 \in C(K)$. We also see that $h_t(\theta_0) = \sigma_t^2$. Next, let us define an approximation to (h_t):

$$\widehat{h}_t(\theta) = \frac{\omega}{1 - \beta} + \sum_{j=1}^{t} \beta^{j-1} \langle X_{t-j}, z \rangle^2, \quad \theta = (\omega, \beta, \lambda_1, \ldots, \lambda_d) \in K. \tag{3.14}$$

The approximation satisfies

$$||\widehat{h}_t - h_t||_K \xrightarrow{\text{a.s.}} 0 \quad \text{exponentially fast as} \quad t \to \infty.$$

Indeed, for any $\theta \in K$,

$$\begin{aligned} |\widehat{h}_t(\theta) - h_t(\theta)| &= \sum_{j=t+1}^{\infty} \beta^{j-1} \langle X_{t-j}, z \rangle^2 \\ &\leq C_0 \sum_{j=t+1}^{\infty} \overline{\beta}^{j-1} ||X_{t-j}||^2, \end{aligned}$$

where $C_0 > 0$ is a constant. Evidently $\sum_{j=t+1}^{\infty} d^{j-1} ||X_{t-j}||^2 \xrightarrow{\text{a.s.}} 0$, as $t \to \infty$, for any $0 < d < 1$. Taking ℓ such that $\overline{\beta}_0 < \ell < 1$ we see that $\ell^t |\widehat{h}_t(\theta) - h_t(\theta)| \leq$

$\sum_{j=t+1}^{\infty} d^{j-1}||X_{t-j}||^2 \xrightarrow{\text{a.s.}} 0$, where $d = \overline{\beta}_0/\ell$.

If (ε_t) were Gaussian, then the conditional density function of $(\langle X_t, \phi_k\rangle, k = 1, \ldots, d)$, given (σ_t), is $\prod_{j=1}^{d}(\sqrt{2\pi}\mu_j\sigma_t)^{-1}\exp\{-x_k^2/(2\mu_j^2\sigma_t^2)\}$. However, (σ_t) are not known. Since the function $(\widehat{h}_t)$ serves as an estimate of (σ_t^2) with the parameters that compose the vector θ_0, we consider the following quasi-maximum likelihood function

$$\widehat{L}_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{d}\Big(\frac{\langle X_t, \phi_i\rangle^2}{\mu_i^2\widehat{h}_t(\theta)} + \log(\widehat{h}_t(\theta))\Big).$$

Now choose a measurable $\widehat{\theta}_n$ so that

$$\widehat{\theta}_n = \text{argmax}_{\theta\in K}\widehat{L}_n(\theta).$$

Let us note that $\Big(\sum_{i=1}^{d}\langle X_t, \phi_i\rangle^2/(\mu_i^2\widehat{h}_t) + \log\widehat{h}_t\Big)$ is neither stationary nor ergodic. Therefore, in order to investigate limit properties of the estimator $\widehat{\theta}_n$ we replace this process by a stationary and ergodic one $\Big(\sum_{i=1}^{d}\langle X_t, \psi_i\rangle^2/(\mu_i^2 h_t) + \log h_t\Big)$. Define

$$L_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{d}\Big(\frac{\langle X_t, \phi_i\rangle^2}{\mu_i^2 h_t(\theta)} + \log(h_t(\theta))\Big)$$

and set

$$\theta_n = \text{argmax}_{\theta\in K}L_n(\theta).$$

3.3.1 Consistency

Theorem 3.4. *Assume that* $E\langle\varepsilon_0, z\rangle^2 + \beta < 1$, *for all* $\theta = (\omega, \beta, \lambda_1, \ldots, \lambda_d) \in K$, *where* $z = \lambda_1\phi_1 + \cdots + \lambda_d\phi_d \in \mathbb{H}$ *and the set* K *is defined by* (3.12). *Then*

$$\widehat{\theta}_n \quad \xrightarrow[n\to\infty]{\text{a.s.}} \quad \theta_0.$$

Proof. We follow the steps of the proof of Theorem 5.3.1 in [71]. Clearly (L_n) is a sequence of random elements of the space $C(K)$ and, moreover,

$$L_n = \sum_{t=1}^{n}\ell_t,$$

where

$$\ell_t = -\frac{1}{2}\sum_{i=1}^{d}\Big(\frac{\langle X_t, \phi_i\rangle^2}{\mu_i^2 h_t} + \log h_t\Big), \quad t = 1, \ldots, n. \tag{3.15}$$

Since $E\langle\varepsilon_t,\phi_i\rangle^2=\langle Q_\varepsilon\phi_i,\phi_i\rangle=\mu_i^2$, we deduce

$$\begin{aligned}L(\theta):&=En^{-1}L_n(\theta)=E\Big[-\frac{1}{2n}\sum_{t=1}^{n}\sum_{i=1}^{d}\Big(\frac{\langle X_t,\phi_i\rangle^2}{\mu_i^2h_t(\theta)}+\log(h_t(\theta))\Big)\Big]\\&=-\frac{d}{2}E\Big(\frac{\sigma_0^2}{h_0(\theta)}+\log h_0(\theta)\Big).\end{aligned}$$

The random elements $\ell_t,t\geq 1$ are of the form $\ell_t=f(X_t,X_{t-1},\dots)$, for each $t\geq 1$, where the function f is measurable. Hence, the $C(K)$-valued process (ℓ_t) is stationary and ergodic. Now we verify that $E||\ell_t||_K<\infty$. Indeed, we have

$$||\ell_t||_K\leq\frac{1}{2}\sum_{i=1}^{d}\Big[\frac{\langle X_t,\phi_i\rangle^2}{\mu_i^2}||1/h_t||_K+||\log h_t||_K\Big]$$

Since $h_t(\theta)\geq\underline{\omega}$ and $E\langle X_t,\phi_i\rangle^2=E\sigma_t^2E\langle\varepsilon_t,\phi_i\rangle^2=\mu_i^2E\sigma_t^2=\mu_i^2E\sigma_0^2$, we deduce

$$E||\ell_t||_K\leq\frac{d}{2\underline{\omega}}E\sigma_0^2+\frac{d}{2}E||\log h_0||_K.$$

Next we write

$$\begin{aligned}E\sigma_0^2&=E\Big\{\omega_0+\omega_0\sum_{n=1}^{\infty}\prod_{j=1}^{n}\Big(\langle\varepsilon_{-j},z_0\rangle^2+\beta_0\Big)\Big\}=\omega_0+\omega_0\sum_{n=1}^{\infty}(E\langle\varepsilon_{-j},z_0\rangle^2+\beta_0)^n\\&=\omega_0+\frac{\omega_0}{1-(E\langle\varepsilon_0,z_0\rangle^2+\beta_0)}<\infty,\end{aligned}$$

by the assumption of the theorem, where $z_0=\lambda_1^o\phi_1+\dots+\lambda_d^o\phi_d$. In view that

$$\begin{aligned}\sup_{\theta\in K}h_0(\theta)&\leq\frac{\overline{\omega}}{1-\overline{\beta}}+\sum_{j=1}^{\infty}\overline{\beta}^{j-1}\max_{\lambda_1,\dots,\lambda_d\in[-\alpha,\alpha]^d}\Big(\sum_{i=1}^{d}\lambda_i\langle X_{-j},\phi_i\rangle\Big)^2\\&=\frac{\overline{\omega}}{1-\overline{\beta}}+\alpha^2\sum_{j=1}^{\infty}\overline{\beta}^{j-1}\Big(\sum_{i=1}^{d}\langle X_{-j},\phi_i\rangle\Big)^2,\end{aligned}$$

and using $E\langle X_t,\phi_i\rangle^2\leq\mu_1^2E\sigma_0^2<\infty$, we obtain

$$\begin{aligned}E\sup_{\theta\in K}h_0(\theta)&\leq\frac{\bar{\omega}}{1-\bar{\beta}}+\alpha^2\sum_{j=1}^{\infty}\bar{\beta}^{j-1}E\Big(\sum_{i=1}^{d}\langle X_{-j},\phi_i\rangle\Big)^2\\&\leq\frac{\bar{\omega}}{1-\bar{\beta}}+d\alpha^2\mu_1^2E\sigma_0^2\sum_{j=1}^{\infty}\bar{\beta}^{j-1}<\infty.\end{aligned}$$

Now it follows $E\|\log h_0\|_K<\infty$ by applying Jensen's inequality. Hence $E||\ell_t||_K<\infty$ and by the law of large numbers (see, e.g., Theorem 2.2.1 in [71]), we conclude that

$$n^{-1}L_n\xrightarrow[n\to\infty]{\text{a.s.}}L,\quad\text{in}\quad C(K).\tag{3.16}$$

Next, we demonstrate that $n^{-1}||\widehat{L}_n - L_n||_K \xrightarrow[n\to\infty]{\text{a.s.}} 0$. Since $h_t(\theta) \geq \underline{\omega}$ and $\widehat{h}_t(\theta) > \underline{\omega}$ for each $\theta \in K$, we have

$$||(\widehat{h}_t)^{-1} - (h_t)^{-1}||_K = \sup_{\theta\in K} \frac{|\widehat{h}_t(\theta) - h_t(\theta)|}{\widehat{h}_t(\theta)h_t(\theta)} \leq \underline{\omega}^{-2}||\widehat{h}_t - h_t||_K.$$

The application of the mean value theorem leads to

$$||\log \widehat{h}_t - \log h_t||_K \leq c_0^{-1}||\widehat{h}_t - h_t||_K.$$

Thus, there exists a constant $c > 0$ such that

$$||\widehat{L}_n - L_n||_K \leq c\sum_{t=1}^{n}\sum_{k=1}^{d}\Big(1 + \frac{\langle X_t, \phi_k\rangle^2}{\mu_k^2}\Big)||\widehat{h}_t - h_t||_K \tag{3.17}$$

Since $||\widehat{h}_t - h_t||_K \xrightarrow{\text{a.s.}} 0$ exponentially fast, as $t \to \infty$, and

$$E\sum_{k=1}^{d}\Big(1 + \frac{\langle X_t, \phi_k\rangle^2}{\mu_k^2}\Big) = d + E\sigma_0^2\sum_{k=1}^{d}\frac{\langle \varepsilon_0, \phi_k\rangle^2}{\mu_k^2} \leq d(1 + E\sigma_0^2) < \infty,$$

by Proposition 2.5.1 in [71], we deduce that $\sum_{t=1}^{\infty}\sum_{k=1}^{d}(1+\langle X_t, \phi_k\rangle^2/\mu_k^2)||\widehat{h}_t - h_t|| < \infty, \quad a.s.$ Therefore $||\widehat{L}_n - L_n||_K/n \to 0, \quad a.s.$ and (3.16) yields

$$\widehat{L}_n/n \to L, \quad a.s. \quad \text{in} \quad C(K). \tag{3.18}$$

Further we prove the uniqueness of the maximum of L. We have to check that $L(\theta) < L(\theta_0)$ for each $\theta \in K, \theta \neq \theta_0$. It is equivalent that the function $(2/d)L(\theta) + \log \sigma_0^2$, $\theta \in K$ is uniquely maximized as $\theta = \theta_0$. We define

$$U(\theta) := (2/d)L(\theta) + \log \sigma_0^2 = \log \frac{\sigma_0^2}{h_0(\theta)} - E\frac{\sigma_0^2}{h_0(\theta)}.$$

Since $U(\theta_0) = -1$ and $\log(x) - x \leq -1$ for all $x > 0$ with equality only if $x = 1$, we obtain $U(\theta) \leq -1 = U(\theta_0)$ with equality, only if $\theta = \theta_0$, because $\sigma_0^2/h_0(\theta) = 1$ if and only if $\theta = \theta_0$. This shows that L is uniquely maximized as $\theta = \theta_0$.

Now, suppose that $(\widehat{\theta}_n)$ does not converge a.s. to θ_0. Then, for an $\varepsilon > 0$, we write $P(\limsup_{n\to\infty} ||\widehat{\theta}_n - \theta_0||_d > \varepsilon) > 0$. First, note that the set $K' = K \cap \{\theta : ||\theta - \theta_0||_d \geq \varepsilon\}$ is compact and $K' \subset K$. Since $\widehat{L}/n \to L$ a.s. in $C(K)$, there is an event $D \subset \{\limsup_{n\to\infty} ||\widehat{\theta}_n - \theta_0||_d > \varepsilon\} \neq \emptyset$ such that $\lim_n ||n^{-1}L_n(\omega) - L||_K = 0$ for each $\omega \in D$ and the sequence $(\widehat{\theta}_n(\omega)) \subset K'$ has an convergent subsequence, say, $(\widehat{\theta}_{n'}) \subset K'$. Let $\theta' = \lim_{n'\to\infty} \widehat{\theta}_{n'}(\omega)$. Then

$$L(\theta') = \lim \widehat{L}_{n'}(\widehat{\theta}_{n'})/n' \geq \lim \widehat{L}_{n'}(\theta_0)/n' = L(\theta_0).$$

Thus, there exists at least one point $\theta \in K', \theta \neq \theta_0$ with $L(\theta) \geq L(\theta_0)$. This is a contradiction. The proof is complete. □

3.3.2 Asymptotic normality

Denote

$$A_\varepsilon = E\Big[\sum_{i=1}^{d}\Big(1 - \frac{\langle \varepsilon_0, \phi_i\rangle^2}{\mu_i^2}\Big)\Big]^2 = \sum_{i,j=1}^{d}\Big(\frac{E\langle \varepsilon_0, \phi_i\rangle^2\langle \varepsilon_0, \phi_j\rangle^2}{\mu_i^2\mu_j^2} - 1\Big)$$

and the matrix

$$J = E\sigma_1^{-4}h_1'(\theta_0)[h_1'(\theta_0)]^T.$$

As usual, $N(0,\Sigma)$ denotes a normal random vector with zero mean and the covariance matrix Σ.

Theorem 3.5. *Assume that the following conditions are satisfied:*

(i) $\theta_0 \in K^0$, *where* K^0 *is the interior of* K, *which is defined by (3.12);*

(ii) For $0 < \delta < 1$, $E||\varepsilon_0||^{4+\delta} < \infty$ *and* $E(\langle \varepsilon_0, z\rangle^2 + \beta)^2 < 1$ *for all* $\theta = (\omega, \beta, \lambda_1, \ldots, \lambda_d) \in K$, *where* $z = \lambda_1\phi_1 + \cdots + \lambda_d\phi_d$;

(iii) The matrix J *is invertible.*

Then,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \frac{A_\varepsilon}{d^2}J^{-1}).$$

Proof. We follow the steps of the proofs of Theorem 7.2. in [40] and of Theorem 5.6.1. in [71]. The first two derivatives of the function $\ell_t(\theta), \theta \in K$ are

$$\begin{aligned}\ell_t'(\theta) &= (\partial\ell_t(\theta)/\partial\theta_j,\ j = 1, \ldots, d+2)\\ &= -\frac{1}{2}\frac{h_t'(\theta)}{h_t(\theta)}\sum_{i=1}^{d}\left(1 - \frac{\langle X_t, \phi_i\rangle^2}{\mu_i^2 h_t(\theta)}\right),\end{aligned} \tag{3.19}$$

and

$$\begin{aligned}l_t''(\theta) &= (\partial^2\ell_t(\theta)/\partial\theta_j\partial\theta_i,\ i,j = 1, \ldots, d+2)\\ &= \frac{(h_t'(\theta))^T h_t'(\theta)}{2h_t^2(\theta)}\sum_{i=1}^{d}\left(1 - \frac{2\langle X_t, \phi_i\rangle^2}{\mu_i^2 h_t(\theta)}\right) - \frac{h_t''(\theta)}{2h_t(\theta)}\sum_{i=1}^{d}\left(1 - \frac{\langle X_t, \phi_i\rangle^2}{\mu_i^2 h_t(\theta)}\right).\end{aligned} \tag{3.20}$$

The assumptions of this theorem imply that $\theta_n \xrightarrow{\text{a.s.}} \theta_0$ (see the proof of Theorem

3.4). For large enough n the following Taylor expansion is valid

$$L'_n(\theta_n) = L'_n(\theta_0) + L''_n(\tilde{\theta}_n)(\theta_n - \theta_0), \tag{3.21}$$

where $||\tilde{\theta}_n - \theta_0|| < ||\theta_n - \theta_0||$. Since θ_n is the maximizer of L_n, one has that $L'_n(\theta_n) = 0$. Therefore, (3.21) is equivalent to

$$L''_n(\widetilde{\theta}_n)(\theta_n - \theta_0) = -L'_n(\theta_0). \tag{3.22}$$

Next, the following will be proved:

(A1) $n^{-1/2} L'_n(\theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \frac{1}{4} A_\varepsilon J)$;

(A2) $n^{-1} L''_n(\theta_0) \xrightarrow[n\to\infty]{\mathrm{P}} -\frac{d}{2} J$;

(A3) $n^{-1} L''_n(\widetilde{\theta}_n) \xrightarrow[n\to\infty]{\mathrm{P}} -\frac{d}{2} J$;

(A4) $\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow[n\to\infty]{\mathrm{P}} 0$.

Then, from (3.22), (A2), (A3) and Slutsky's theorem we deduce $\sqrt{n}(\theta_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \frac{A_\varepsilon}{d^2} J^{-1})$ and (A4) completes the proof of the theorem.

First we prove (A1). Recall that $h_t(\theta_0) = \sigma_t^2$ and $X_t = \sigma_t \varepsilon_t$. Hence, by (3.19) we have

$$L'_n(\theta_0) = \sum_{t=1}^{n} l'_t(\theta_0) = \frac{1}{2} \sum_{t=1}^{n} \frac{h'_t(\theta_0)}{\sigma_t^2} \sum_{i=1}^{d} \left(\frac{\langle \varepsilon_t, \phi_i \rangle^2}{\mu_i^2} - 1 \right)$$

Let $\mathcal{F}_t = \sigma(\varepsilon_s, s \le t), t \ge 1$. The random vector $h'_t(\theta_0)/\sigma_t^2$ is $\mathcal{F}_{t-1}$ measurable and $\mathcal{F}_{t-1}$ is independent of ε_t and $E\langle \varepsilon_t, \phi_i \rangle^2 = \mu_i^2$. Consequently, $(l'_t(\theta_0))_{t\in\mathbb{N}} \subset \mathbb{R}^{d+2}$ is a stationary ergodic martingale difference sequence with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$. If the following two conditions are satisfied

(C1) $n^{-1} \sum_{t=1}^{n} E\Big(\ell'_t(\theta_0)[\ell'_t(\theta_0)]^T | \mathcal{F}_{t-1} \Big) \xrightarrow[n\to\infty]{\mathrm{P}} Q$,

(C2) for a $\delta \in (0,1)$ $\lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{t=1}^{n} E||\ell'_t(\theta_0)||_d^{2+\delta} = 0$,

then we obtain $n^{-1/2} \sum_{t=1}^{n} \ell'_t(\theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, Q)$ (it follows from a slight generalization of Theorem 3.1. in [65]).

First we establish the finiteness of the moment $E||\ell'_t(\theta_0)||_d^{2+\delta}$. Since $E||\varepsilon_t||^{4+\delta} < \infty$, it suffices to verify that

$$E||h'_t(\theta_0)||_d^{2+\delta} \sigma_t^{-2(2+\delta)} < \infty. \tag{3.23}$$

As far as $\sigma_t \ge \underline{\omega}$, (3.23) reduces to $E||h'_t(\theta_0)||_d^{2+\delta} < \infty$.

Recall that

$$h_t(\theta) = \frac{\omega}{1-\beta} + \sum_{j=1}^{\infty} \beta^{j-1}(\lambda_1\langle X_{t-j}, \phi_1\rangle + ... + \lambda_d\langle X_{t-j}, \phi_d\rangle)^2$$
$$= \frac{\omega}{1-\beta} + \sum_{j=1}^{\infty} \beta^{j-1}\langle X_{t-j}, z\rangle^2,$$

where $\theta = (\omega, \beta, \lambda_1, \dots, \lambda_d)$, $z = \lambda_1\phi_1 + \dots + \lambda_d\phi_d$, and calculate the derivatives

$$\frac{\partial h_t(\theta)}{\partial \omega} = \frac{1}{1-\beta}, \quad \frac{\partial h_t(\theta)}{\partial \beta} = \frac{\omega}{(1-\beta)^2} + \sum_{j=1}^{\infty}(j-1)\beta^{j-2}\langle X_{t-j}, z\rangle^2,$$
$$\frac{\partial h_t(\theta)}{\partial \lambda_k} = 2\sum_{j=1}^{\infty} \beta^{j-1}\langle X_{t-j}, z\rangle\langle X_{t-j}, \phi_k\rangle, \quad k = 1, ..., d.$$

Now we find

$$E||\ell_t'(\theta)||_d^{2+\delta} \le \frac{1}{(1-\bar{\beta})^{2+\delta}} + \frac{\bar{\omega}^{2+\delta}}{(1-\bar{\beta})^{4+2\delta}} + E\Big(\sum_{j=1}^{\infty}(j-1)\bar{\beta}^{j-2}\langle X_{t-j}, \alpha\rangle^2\Big)^{2+\delta}$$
$$+ 4\sum_{k=1}^{d} E\Big(\sum_{j=1}^{\infty} \bar{\beta}^{j-1}\langle X_{t-j}, \alpha\rangle\langle X_{t-j}.\phi_k\rangle\Big)^{2+\delta}.$$

Hence, $E||\ell_t(\theta)||_d^{2+\delta} < \infty$ follows from $E||X_t||^{4+\delta} = E||X_0||^{4+\delta} < \infty$ (this in turn follows from the conditions of the theorem and Theorem 3.3). So we get

$$n^{-1}\sum_{t=1}^{n} E\Big(\ell_t'(\theta_0)[\ell_t'(\theta_0)]^T|\mathcal{F}_{t-1}\Big) = n^{-1}\frac{1}{4}\sum_{t=1}^{n}\frac{h_t'(\theta_0)[h_t'(\theta_0)]^T}{\sigma_t^4}E\Big[\sum_{i=1}^{d}\Big(\frac{\langle\varepsilon_t, \phi_i\rangle^2}{\mu_i^2} - 1\Big)\Big]^2$$
$$= \frac{A_\varepsilon}{4}n^{-1}\sum_{t=1}^{n}\frac{h_t'(\theta_0)[h_t'(\theta_0)]^T}{\sigma_t^4} \xrightarrow[n\to\infty]{\text{P}} \frac{A_\varepsilon}{4}J$$

by the law of large numbers (see, e.g., Theorem 2.2.1 in [71]).

Now we prove (A2). It is easy to see that $L_n''(\theta_0) = \sum_{t=1}^{n}\ell_t''(\theta_0)$ and $(\ell_t''(\theta_0))$ is a stationary and ergodic process. Hence, the law of large numbers applies, if we establish the finiteness of $E||\ell_0''(\theta_0)||_{d^2}$, where $\|\cdot\|_{d^2}$ denotes the norm of the $(d+2)\times(d+2)$ matrix, defined as

$$\|A\|_{d^2} = \max_{i,j=1,\dots,d+2}|a_{ij}|, \quad \text{for} \quad A = (a_{ij}).$$

We write

$$
\begin{aligned}
&E\|\ell_t''(\theta_0)\|_{d^2}\\
\le&\frac{1}{2}E\Big[\frac{||h_t'(\theta_0)[h_t'(\theta_0)]^T||_{d^2}}{h_t^2(\theta_0)}\sum_{i=1}^{d}\Big|\frac{2\langle X_t,\phi_i\rangle^2}{\mu_i^2 h_t(\theta_0)}-1\Big|+\frac{||h_t''(\theta_0)||_{d^2}}{h_t(\theta_0)}\sum_{i=1}^{d}\Big|1-\frac{\langle X_t,\phi_i\rangle^2}{\mu_i^2 h_t(\theta_0)}\Big|\Big]\\
=&\frac{1}{2}E\Big[\frac{||h_t'(\theta_0)[h_t'(\theta_0)]^T||_{d^2}}{\sigma_t^4}\sum_{i=1}^{d}\Big|\frac{2\langle \varepsilon_t,\phi_i\rangle^2}{\mu_i^2}-1\Big|+\frac{||h_t''(\theta_0)||_{d^2}}{\sigma_t^2}\sum_{i=1}^{d}\Big|1-\frac{\langle \varepsilon_t,\phi_i\rangle^2}{\mu_i^2}\Big|\Big]
\end{aligned}
$$

and we see that $E||\ell_t''(\theta_0)||_{d^2}<\infty$ follows from

$$E\frac{||h_t'(\theta_0)[h_t'(\theta_0)]^T||_{d^2}}{\sigma_t^4}<\infty \tag{3.24}$$

and

$$E\frac{||h_t''(\theta_0)||_{d^2}}{\sigma_t^2}<\infty. \tag{3.25}$$

As (3.24) follows from (3.23), it remains to check (3.25). We calculate the second derivative $h_t''(\theta)$:

$$
\begin{aligned}
\frac{\partial^2 h_t(\theta)}{\partial\omega^2}&=\frac{\partial^2 h_t(\theta)}{\partial\omega\partial\lambda_i}=0,\quad \frac{\partial^2 h_t(\theta)}{\partial\omega\partial\beta}=\frac{1}{(1-\beta)^2},\\
\frac{\partial^2 h_t(\theta)}{\partial\beta^2}&=\frac{2\omega}{(1-\beta)^3}+\sum_{j=1}^{\infty}(j-1)(j-2)\beta^{j-3}\langle X_{t-j},z\rangle^2\\
\frac{\partial^2 h_t(\theta)}{\partial\beta\partial\lambda_k}&=2\sum_{j=1}^{\infty}(j-1)\beta^{j-2}\langle X_{t-j},z\rangle\langle X_{t-j},\phi_k\rangle,\\
\frac{\partial^2 h_t(\theta)}{\partial\lambda_k\partial\lambda_l}&=2\sum_{j=1}^{\infty}\beta^{j-1}\langle X_{t-j},\phi_k\rangle\langle X_{t-j},\phi_l\rangle,
\end{aligned}
$$

where $k,l=1,\dots,d$. Since $\sigma^2>\underline{\omega}$, we easily see that (3.25) follows from $E||X_t||^2<\infty$. Hence, by the law of large numbers

$$\frac{1}{n}L_n''(\theta_0)\xrightarrow[n\to\infty]{\mathrm{P}}E\ell_0''(\theta_0).$$

It is easy to see that

$$E\ell_0''(\theta_0)=-\frac{d}{2}E\sigma_0^{-4}(h_0'(\theta_0))^T h_0'(\theta_0)=-\frac{d}{2}J.$$

(A3) follows from continuity of the function $\theta\to L''(\theta)$ on K which is obvious in definition (3.20).

Finally we prove (A4). First we have to show that

$$\frac{1}{\sqrt{n}}\|\hat{L}_n' - L_n'\|_K \xrightarrow[n\to\infty]{\mathrm{P}} 0, \quad n\to\infty. \tag{3.26}$$

Denote $Y_t^2 = \sum_{i=1}^d \frac{\langle X_t,\phi_i\rangle^2}{\mu_i^2}$ and take the following derivatives

$$\hat{l}_t' = -\frac{1}{2}\Big(d - \frac{Y_t^2}{\hat{h}_t}\Big)\frac{\hat{h}_t'}{\hat{h}_t}, \quad l_t' = -\frac{1}{2}\Big(d - \frac{Y_t^2}{h_t}\Big)\frac{h_t'}{h_t}.$$

Applying the mean value theorem to the function[2] $f(u,v) = \frac{u}{v}(d - \frac{Y_t^2}{v})$, we obtain the following expression

$$\begin{aligned}
&\|\hat{l}_t' - l_t'\|_K\\
=&\frac{1}{2}\Big\|\frac{\hat{h}_t'}{\hat{h}_t}\Big(d-\frac{Y_t^2}{\hat{h}_t}\Big) - \frac{h_t'}{h_t}\Big(d-\frac{Y_t^2}{h_t}\Big)\Big\|_K\\
=&\frac{1}{2}\Big\|\frac{1}{\bar{h}_t}\Big(d-\frac{Y_t^2}{\bar{h}_t}\Big)(\hat{h}_t'-h_t') - \frac{\bar{h}_t'}{\bar{h}_t^2}\Big(d-\frac{Y_t^2}{\bar{h}_t}\Big)(\hat{h}_t-h_t) + \frac{\bar{h}_t'}{\bar{h}_t}\Big(d+\frac{Y_t^2}{\bar{h}_t^2}\Big)(\hat{h}_t-h_t)\Big\|_K\\
\le&c(d+Y_t^2)\|\hat{h}_t'-h_t'\|_K + c(d+Y_t^2)\|\bar{h}_t'\|_K\|\hat{h}_t-h_t\|_K\\
\le&c(d+Y_t^2)\{\|\hat{h}_t'-h_t'\|_K + \|\hat{h}_t'-h_t'\|_K\|\hat{h}_t-h_t\|_K + \|h_t'\|_K\|\hat{h}_t-h_t\|_K\},
\end{aligned}$$

where c is a constant and $|\bar{h}_t'| \le |\hat{h}_t' - h_t'| + |h_t'|$.

Recall that $\hat{h}_t = \frac{\omega}{1-\beta} + \sum_{j=1}^t \beta^{j-1}\langle X_{t-j}, z\rangle^2$ and calculate the derivatives

$$\begin{aligned}
&\frac{\partial\hat{h}_t}{\partial\omega} = \frac{1}{1-\beta}, \quad \frac{\partial\hat{h}_t}{\partial\beta} = \frac{\omega}{(1-\beta)^2} + \sum_{j=1}^t (j-1)\beta^{j-2}\langle X_{t-j}, z\rangle^2,\\
&\frac{\partial\hat{h}_t}{\partial\lambda_k} = 2\sum_{j=1}^t \beta^{j-1}\langle X_{t-j}, z\rangle\langle X_{t-j}, \phi_k\rangle, \quad k = 1, ..., d.
\end{aligned}$$

Since

$$\frac{\partial h_t}{\partial\omega} = \frac{\partial\hat{h}_t}{\partial\omega},$$

[2]For every function $f:\mathbb{R}^2\to\mathbb{R}$ with continuous partial derivatives f_u and f_v and for all distinct pairs (u_1,v_1) and (u_2,v_2) in $\mathbb{R}^2$, there exists an intermediate point (u^*,v^*) on the line segment, joining the points (u_1,v_1) and (u_2,v_2), such that $f(u_2,v_2)-f(u_1,v_1) = f_u(u^*,v^*)(u_2-u_1) + f_v(u^*,v^*)(v_2-v_1)$. (see, e.g., [67])

and

$$\sup_{\theta\in K}\Big|\frac{\partial h_t(\theta)}{\partial \beta}-\frac{\partial \hat{h}_t(\theta)}{\partial \beta}\Big| = \sup_{\theta\in K}\Big|\sum_{j=t+1}^{\infty}(j-1)\beta^{j-2}\langle X_{t-j},z\rangle^2\Big|$$
$$\leq c_1\sum_{j=t+1}^{\infty}(j-1)\bar{\beta}^{j-2}\|X_{t-j}\|^2,$$
$$\sup_{\theta\in K}\Big|\frac{\partial h_t(\theta)}{\partial \lambda_k}-\frac{\partial \hat{h}_t(\theta)}{\partial \lambda_k}\Big| = \sup_{\theta\in K}\Big|\sum_{j=t+1}^{\infty}2\beta^{j-1}\langle X_{t-j},z\rangle\langle X_{t-j},\phi_k\rangle\Big|$$
$$\leq 2c_2\sum_{j=t+1}^{\infty}\bar{\beta}^{j-1}\|X_{t-j}\|^2,$$

where c_1, c_2 are finite constants, we conclude that $\|\hat{h}'_t - h'_t\|_K \xrightarrow{\text{a.s.}} 0$ exponentially fast as $t\to\infty$, the same as shown earlier for $\|\hat{h}_t - h_t\|_K$. Furthermore, $E\|h'_t\|_K < \infty$ and

$$EY_t = E\Big\{\sum_{i=1}^{d}\frac{\langle X_t,\phi_i\rangle^2}{\mu_i^2}\Big\} \leq dEh_0 < \infty.$$

We can apply Proposition 2.5.1 and Lemma 2.5.3 in [71] to conclude that

$$\|\hat{L}'_n - L'_n\|_K \leq \sum_{t=1}^{\infty}\|\hat{l}'_t - l'_t\|_K < \infty, \quad \text{a.s.},$$

which completes the proof of (3.26).

From the mean value theorem

$$L'_n(\hat{\theta}_n) - L'_n(\theta_n) = L''_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_n), \tag{3.27}$$

where $\bar{\theta}_n$ lies on the line segment that connects $\hat{\theta}_n$ and θ_n, which is completely contained in the interior of K. Since $L'_n(\hat{\theta}_n) = L'_n(\theta_n) = 0$, equation (3.27) is equivalent to

$$\frac{1}{\sqrt{n}}(L'_n(\hat{\theta}_n) - L'_n(\theta_n)) = \frac{1}{n}L''_n(\bar{\theta}_n)\sqrt{n}(\hat{\theta}_n - \theta_n), \tag{3.28}$$

and both sides of this equation tend to 0, as $n\to\infty$. Using (A3) and applying Theorem 2.2.1 in [71] to L''_n/n together with $\bar{\theta}_n \xrightarrow{a.s.} \theta_0$, we conclude that $L''_n(\bar{\theta}_n)/n \xrightarrow{a.s.} -\frac{d}{2}J$. Since by assumption iii) matrix J is invertible, we can deduce that $\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{a.s.} 0$. This completes the proof. □

3.4 Analysis of residuals

In this section, we investigate the asymptotic properties of residuals $(\widehat{\varepsilon}_t)$ of the model (3.1), (3.2). Recalling

$$\sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{j=1}^{\infty} \beta^{j-1} \langle X_{t-j}, z \rangle^2 \tag{3.29}$$

we define

$$\widehat{\sigma}_t^2 = \frac{\widehat{\omega}}{1-\widehat{\beta}} + \sum_{j=1}^{m} \widehat{\beta}^{j-1} \langle X_{t-j}, \widehat{z} \rangle^2, \tag{3.30}$$

where $1 < m < n$ and set

$$\widehat{\varepsilon}_t = \widehat{\sigma}_t^{-1} X_t, \quad t = 1, \dots, n.$$

Theorem 3.6. *Assume that* $E\langle \varepsilon_0, z \rangle^2 + \beta < 1$, *for all* $\theta = (\omega, \beta, \lambda_1, \dots, \lambda_d) \in K$, *where* $z = \lambda_1 \phi_1 + \cdots + \lambda_d \phi_d \in \mathbb{H}$, *and the set* K *is defined by* (3.12). *Then, for any* $x, y \in \mathbb{H}$, *we have*

$$n^{-1} \sum_{k=1}^{n} \langle \widehat{\varepsilon}_k, x \rangle \langle \widehat{\varepsilon}_k, y \rangle \xrightarrow[n\to\infty]{\mathrm{P}} \langle Q_\varepsilon x, y \rangle.$$

Proof. We write

$$\sum_{k=1}^{n} \widehat{\varepsilon}_k \otimes \widehat{\varepsilon}_k = \sum_{k=1}^{n} \varepsilon_k \otimes \varepsilon_k + V_n,$$

where

$$V_n = \sum_{k=1}^{n} \left(\frac{\sigma_k^2}{\widehat{\sigma}_k^2} - 1 \right) \varepsilon_k \otimes \varepsilon_k.$$

Since

$$n^{-1} \sum_{k=1}^{n} \langle \varepsilon_k, x \rangle \langle \varepsilon_k, y \rangle \xrightarrow[n\to\infty]{\mathrm{P}} \langle Q_\varepsilon x, y \rangle$$

for any $x, y \in \mathbb{H}$ by the law of large numbers, we have to prove

$$n^{-1} \langle V_n x, y \rangle \xrightarrow[n\to\infty]{\mathrm{P}} 0. \tag{3.31}$$

To this end we decompose

$$
\begin{aligned}
\sigma_j^2 - \widehat{\sigma}_j^2 &= \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \beta^{i-1} \langle X_{j-i}, z\rangle^2 - \frac{\widehat{\omega}}{1-\widehat{\beta}} - \sum_{i=1}^{m} \widehat{\beta}^{i-1} \langle X_{j-i}, \widehat{z}\,\rangle^2 \\
&= \frac{\omega}{1-\beta} - \frac{\widehat{\omega}}{1-\widehat{\beta}} + \sum_{i=m+1}^{\infty} \beta^{i-1} \langle X_{j-i}, z\rangle^2 + \sum_{i=1}^{m} (\beta^{i-1} - \widehat{\beta}^{i-1}) \langle X_{j-i}, z\rangle^2 \\
&+ \sum_{i=1}^{m} \widehat{\beta}^{i-1} [\langle X_{j-i}, z\rangle^2 - \langle X_{j-i}, \widehat{z}\rangle^2] \\
&= \tau_{1j} + \tau_{2j} + \tau_{3j} + \tau_{4j}, \quad j = 1, \ldots, n.
\end{aligned}
$$

We get accordingly

$$
|\langle V_n x, y\rangle| \leq \sum_{v=1}^{4} V_n^{(v)},
$$

where

$$
V_n^{(v)} = \sum_{j=1}^{n} \frac{\tau_{vj}}{\widehat{\sigma}_j^2} \langle \varepsilon_j x, y\rangle^2, \quad v = 1, \ldots, 4.
$$

On the set $|\omega - \widehat{\omega}| + |\beta - \widehat{\beta}| \leq \tau_0$ we have $\widehat{\sigma}_k^2 \geq (\omega_0 - \tau_0)/(1 - \beta_1 + \tau_0) := c_0^{-1}$. Then, we have to check

$$
n^{-1} \sum_{j=1}^{n} \tau_{vj} ||\varepsilon_j||^2 \xrightarrow[n\to\infty]{\mathrm{P}} 0, \quad v = 1, \ldots, 4. \tag{3.32}
$$

Assumptions of this theorem imply that the conditions of Theorem 3.4 are valid, therefore the result in (3.32) follows from the consistency of the model parameter estimates. □

Chapter 4

Stylized facts and aggregation

While examining the specific behavior of financial asset prices, market analysts usually relates that to the political and economic events or announcements. For example, Cutler, Poterba and Summers (1989) have found that the news proxies can explain about one-third of stock returns volatility. Then, it might seem that the behavior of different financial asset (stock, foreign exchange, commodity, etc.) prices is influenced by specific market events and have different statistical properties. However, the statistical analysis shows that data series of financial asset prices have common features, the so-called basic stylized facts. Therefore, the analysis of basic stylized facts is an important issue, when modeling financial asset prices and their return series. The requirements to models include not only the desirable statistical properties, for example, stationarity, parameter estimate consistency and asymptotic normality as studied in the previous two chapters, but also the capability to reproduce stylized facts in the data series observed.

The statistical properties of financial asset prices have been studied for many years and are well known. The appearance of high frequency data opened new fields for search of the stylized facts that characterize such data sets. The statistical properties are analyzed for both regular and "tick-by-tick" data series. Moreover, regularly spaced data series can be constructed from high frequency data by applying data aggregation. It is known that even a linear aggregation that reduces the frequency of data series, changes the stylized facts, for example, the Gaussianity of financial asset returns, i.e., high frequency returns are distributed non-normally, but with an increasing interval between prices, the distribution of returns approaches to normal. However, it is not known if the same frequency prices, constructed applying different and more complex data aggregation, really changes the basic statistical properties, and if so, then how. The analysis of this problem is important when developing models for functional returns of series of aggregated financial asset prices.

In this chapter, a long memory as one of basic stylized facts is considered. Dependence of the long memory parameter on aggregation is presented via a numerical example for foreign exchange rates. The chapter is organized as follows. At first, a short review of the basic stylized facts on financial asset prices and return data series is provided. Next, the concept of long memory and a review of papers on the long memory analysis, using foreign exchange rates, is presented. Then, the related concepts of an R/S statistic and a Hurst exponent are introduced. Finally, a numerical example of absolute returns Hurst exponent estimates of foreign currencies (USD versus EUR, GBP, JYP), using data aggregation, is presented.

4.1 Basic stylized facts

In this section a short review of basic stylized facts is provided following [23], [26] and [40]. Gathering of common statistical facts started with the research of returns of low frequency financial asset prices. Among the earliest observations about the behavior of financial data series there was *almost no autocorrelation of returns*. The series of returns of financial asset prices usually has a very weak autocorrelation and is almost a white noise process. However, the series of high frequency returns might have significant autocorrelations due to market microstructure effects. For example, Goodhart and Figliuoli (1991) were the first who demonstrated the existence of the negative first-order autocorrelation of the highest frequency returns (up to 4 minutes). One of the explanations for this observation is that market traders have different opinions about the effect of news on the direction of price movements, which is a contradiction to the hypothesis of a homogenous market and results in the negative correlation of returns. Furthermore, it has been observed that the series of squared returns or absolute returns is often strongly autocorrelated. Another feature observed in financial data is *non-stationarity of price series*. In general, the trajectory of prices series is close to a random walk process without the constant term. Differently, the series of returns is usually the second order stationary process. Although the assumption on the Gaussian distribution of returns is usually made when modeling financial data series, it is obvious that the returns *distribution is non-normal*. Empirical studies about the distribution of returns reveal several common properties. The most important observation is that the distribution of returns is fat-tailed. The means of returns are close to zero. The absolute values of skewness are significantly smaller than 1 and the empirical distribution is almost symmetric. The empirically determined kurtosis exceeds the value of 0, which is the theoretical value of Gaussian distribution. Furthermore, the distribution of returns has a finite variance and the third moment, while the fourth moment usually diverges. The temporal aggregation tends to diminish the effects of non-normality and it is noticed that the distribution of weekly data approaches to normal. One more important statistical property of financial data series is *long memory*. The estimate of correlation between absolute returns is slowly decreasing with an increasing interval between returns. The same applies to the correlation between volatility estimates. More details about the long memory will be provided in the next section.

The analysis of low and high frequency return volatility has revealed several statistical properties. First, the so-called *volatility clustering*, characterized by the observation that large price changes are usually followed by large ones and small price changes are followed by small ones. High volatility events tend to cluster in time, since a positive autocorrelation of different volatility measures might be observed for several days. In addition, returns have generally a non-constant conditional variance (conditional heteroscedasticity). Second, the *seasonality* property is revealed by volatility variation, when intraday and intraweek returns are analyzed. It has been observed that volatility tends to increase after weekends or festivals, when the markets do not function. In addition, the seasonality effect is also present in the intra-day data series. The daily patterns may be explained by the behavior of active periods of the three main markets (European, Asian and American) that partially overlap. Finally, the *leverage effect*, described by an observation that positive and negative price changes have an asymmetric impact on volatility. The fall of the price tends to cause a higher volatility increase as compared to the same size of the price growth. This can be explained by the fact that reactions of market participants are more sensitive to negative information than to the positive one.

One more group of statistical properties of financial asset prices can be found analyzing institutional frameworks and exogenous impacts. Currency systems serves as an example of an institutional framework, where foreign exchange rates are kept within fixed bounds. For instance, Lithuanian litas is pegged to the Euro at a fixed rate of LTL 3,4528 = EUR 1 and Latvian lats floats within 1 percent of the central rate of LAT 0,702804 = EUR 1. Similarly it is evident that official interventions of central banks have a positive impact on the market. The interventions by central banks may be direct via official announcements or indirect via unannounced interventions. As an example of an official intervention can be decisions on the key interest rates for the euro area (the rates on the main refinancing operations, the deposit facility and the marginal lending facility) made by European Central Bank[1]. Goodhard and Hesse (1993) have presented the study, which shows that official interventions have positive effects in the long run. Moreover, the impact of news is broadly analyzed. Major economic news announcements, general political and economic news, economic forecasts and even discussions among traders are understood as news. Since it is difficult to quantify news, the analysis of news impact is quite complex. Different investigations have been performed on a large variety of news and showed of mixed effects.

A typical stylized fact observed in high frequency data sets is *scaling laws*. It has been noticed that mean absolute returns and mean squared returns can be expressed as functions of their time intervals. One of the most important scaling laws, also applied in the risk management, is the observation that the size of the average absolute value of returns is scale-invariant in the time interval of its occurrence. The analysis of such properties is important since there is no agreement of the common frequency to fix and analyze data. Therefore, the search for a high frequency financial data scaling laws is still carried out rather extensively. In the recent paper, Dupuis, Glattfelder and Olsen (2011) have announced the discovery of 12 new empirical scaling laws for high frequency foreign exchange data. They believe that with an extended collection of stylized facts, the space of possible theoretical explanations of market mechanisms becomes more constrained. Among other statistical properties observed especially for the "tick-by-tick" financial data is *the discreteness of quoted bid-ask spread*, i.e., the difference between the ask price and bid price has discrete values, such as 5, 7, 10 basis points. *The short-term triangular arbitrage* is also the fact noticed in high frequency data sets. The opportunity of arbitrage appears since a short time is needed for smaller currency traders to adjust to the price movements of the leading currencies, such as EUR or USD. However, the transactional costs are often higher than the profit from the difference of currency prices.

4.2 Long memory in foreign exchange returns

One of the most important stylized facts is a long memory, usually observed in absolute or squared returns. According to McLeod and Hippel (1978) the covariance stationary process y_t is said to exhibit the long memory if the following condition is satisfied

$$\sum_{k=n}^{-n} |\rho_k| \rightarrow \infty,\ n \rightarrow \infty,$$

[1] http://www.ecb.int/mopo/decisions/html/index.en.html

where ρ_k is the autocorrelation function at lag k. For example, fractionally integrated processes are long memory processes, where the process y_t is said to be integrated of order d or I(d), if

$$(1-L)^d y_t = u_t,$$

where L is the lag operator, $-0,5 < d < 0,5$ and u_t is a stationary and ergodic process with a bounded and positively valued spectrum at all frequencies (see, e.g. [9]). For $0 < d < 0,5$ the process y_t is a long memory, the autocorrelations are positive and decaying at a hyperbolic rate. For $-0,5 < d < 0$ the process is a short memory as the sum of absolute values of autocorrelations tends to a constant. Additionally, an important class can be distinguished when u_t is I(0) and covariance stationary process.

There has been a great interest in the analysis of the long memory property of foreign exchange returns in recent years. Researchers that examine a long-range dependence in foreign exchange returns have come to different conclusions. One part of papers gaive evidence that foreign exchange returns have a long memory, however, the other part rejected the hypothesis on the long memory in foreign exchange returns. One of the earliest studies on a long memory in foreign exchange returns was performed by Booth, Kaen and Koveos in 1982. They applied the R/S statistic to find out that the spot rate of the US dollar versus the British pound, the French frank and the German mark exhibit a long-term dependence. Cheung (1993) found some evidence of long memory in weekly exchange rate returns series of the British pound, the German mark, the Swiss frank, the French frank and the Japanese yen currencies, using the Geweke and Porter-Hudak (GPH) test. Bhar (1994) tested for a long-term memory in the US dollar versus the Japanese yen exchange rate changes, using Lo's methodology and found no evidence of a long-term memory. The same methodology, applied in the estimation of daily volatility of the exchange rate, has shown the presence of a long-term memory. Tschernig (1995) found some evidence for the weak long memory in the changes of the U.S. dollar versus the German mark and the Swiss frank spot rates. However, there was no evidence of long memory in the German mark versus the Swiss frank spot rate changes. Moody and Wu (1995) made a rescaled range and Hurst exponent analysis on "tick-by-tick" interbank foreign exchange rates to find out that they are mean-reverting. Later on in the research of 1996, Moody and Wu improved Lo's R/S statistic and have concluded that the German mark versus the US dollar exchange rate series is mildly trending on time scales from 10 to 100 ticks. Nath and Reddy (2002) showed that the R/S analysis, applied to the US dollar versus the Indian rupee exchange rate, had indications of a long-term memory, but with noise. Oh, Um and Kim (2006) studied a long-term memory in diverse stock market indices and foreign exchange rates, using the Detrended Fluctuation Analysis (DFA). In all daily and high-frequency market data explored no significant long-term memory property was detected in the return series, while a strong long-term memory property was found in the volatility time series. Soofi, Wang, Zhang (2006) applied plug-in and Whittle methods to test the long memory property of 12 Asian currencies versus the US dollar exchange rates. The plug-in method results have shown that, with the exception of the Chinese renminbi, all foreign exchange returns series may have a long memory. On the other hand, the results, based on Whittle method, have indicated that only the Japanese yen and the Malaysian ringgit may have long memory.

Researchers looking for a long memory in foreign exchange returns volatility usually analyze absolute or squared foreign exchange returns. The results of such an analysis are mostly the same, proving that foreign exchange returns volatility is a long memory process. Dacorogna *et al.* (1993) examined the squared returns for intra-daily exchange rate data and defined a the slow decay of autocorrelations. Ding and Granger (1996) analyzed speculative returns series from different markets. They have found that absolute returns and their power transformations have long, positive autocorrelations. The exchange rate of the US dollar versus the German mark was an exception with the strongest property when taking the power of 1/4, while usually this property is the strongest one for absolute returns. Andersen and Bollerslev (1997) have demonstrated that the volatility of the German mark versus the US dollar exchange rate five-minute returns exhibits long-run dependence. Ohanissian, Russel and Tsay (2003) applied a specific test, based on the GPH estimates of the aggregated series, to the actual intra-daily squared returns of the US dollar versus the German mark and the Japanese yen exchange rates. They have concluded that the volatility of these series is a true long memory process.

4.3 R/S statistic and the Hurst exponent

There are several methods for a long-term dependence analysis. For example, the Rescaled Range (R/S) analysis and the Hurst exponent can be used to test a long memory in high frequency foreign exchange data. The concept of rescaled range was first introduced by Hurst (1951), where he investigated the data of the river Nile level. Mandelbrot (1972) has developed and popularized the classical R/S statistic methodology. Later on, there was a number of improvements of the classical R/S statistic (see, e.g., [42], [54], [60]).

The R/S statistic is a range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. The R/S statistic for the returns $(X_t(g))$, depending on the aggregation rule $g \in \mathcal{G}$, can be defined as

$$[R/S](N,g) := \frac{1}{M}\sum_{t_0=1}^{M}\frac{R(N,t_0,g)}{S(N,t_0,g)},$$

where

$$R(N,t_0,g) = \max_{1\le\tau\le N}\sum_{t=t_0+1}^{t_0+\tau}[X_t(g)-\bar{X}(N,t_0,g)] - \min_{1\le\tau\le N}\sum_{t=t_0+1}^{t_0+\tau}[X_t(g)-\bar{X}(N,t_0,g)],$$

$$\bar{X}(N,t_0,g) = \frac{1}{N}\sum_{t=t_0+1}^{t_0+N}X_t(g)$$

and

$$S(N,t_0,g) = \left\{\frac{1}{N}\sum_{t=t_0+1}^{t_0+N}[X_t(g)-\bar{X}(N,t_0,g)]^2\right\}^{\frac{1}{2}}.$$

Assuming that the scaling law exists, the Hurst exponent H can be estimated using the following expression:

$$[R/S](N,g) \approx cN^{H(g)},$$

where c is a constant. There are three cases of the Hurst exponent values that describe a different behavior of time series:

- $H = 0,5$ is a random walk;
- $0,5 < H < 1$ is a persistent or trend reinforcing behavior;
- $0 < H < 0,5$ is an anti-persistent or mean-reverting behavior.

In the case of $0,5 < H < 1$, the time series is characterized by a long memory process.

4.4 Numerical example

Just like in Chapter 1 foreign currencies (USD, EUR, GBP, JPY) were chosen for the empirical research of the Hurst exponent dependence on data aggregation. High frequency foreign exchange rates were available at every minute of the day over one year period; therefore, the final data set for each currency consisted of 1440 minutely rates for 252 working days of the year. In general, we consider an irregular time series of foreign exchange rates $\{(\tau_j, p_j)\}_{j=i}^{N}$, where τ_j and p_j

indicate, respectively, the time and the value of the j'th observation. Then we fix a time interval between two observations at $\delta > 0$ and let $\tau_t^* = t\delta, t = 1, \ldots, N^*$. In this case, δ is chosen equal to one minute. Again, we consider a class of aggregation rules with $g \in \mathcal{G}$, where $g : [0,1] \to \mathbb{R}$ and the interval $[0,1]$ represents one day. Aggregation rules were chosen in the same way as in Chapter 1. Let us recall them. For each $s \in [0,1]$ we take:

- *pointwise aggregation*

$$p_t^{DAILY}(s) = \{p_i | \tau_i = \max\{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}\}$$

- *maximum price aggregation*

$$p_t^{MAX}(s) = \max\{p_i | \tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\},$$

- *minimum price aggregation*

$$p_t^{MIN}(s) = \min\{p_i | \tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*], \}$$

- *average price aggregation*

$$p_t^{AVE}(s) = \frac{1}{m_t(s)} \sum_{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]} p_i,$$

$$m_t(s) = \#\{\tau_i \in (\tau_{t-1}^*, (1-s)\tau_{t-1}^* + s\tau_t^*]\}.$$

The Hurst exponents (H) were estimated taking absolute returns of the aggregated prices, expressed as:

$$X_t(s) = \left| \frac{\log p_t(s)}{\log p_{t-1}(s)} \right|, \quad s \in [0,1].$$

The results of Hurst index estimates that depend on the aggregation rule of the analyzed currencies are provided in Figures 4.1, 4.2, and 4.3.

The estimates of Hurst exponents of all the currencies vary from 0,63 and 0,81 and confirm the stylized fact of a long memory in foreign exchange absolute daily returns. The pictures show that Hurst indices vary depending on the aggregation rule and also change during the day with s varying in the interval $[0,1]$. The largest fluctuations of the Hurst index values estimated are observed when the pointwise aggregation scheme is used since by applying other aggregation rules we get smoother functions of aggregated prices and returns. The most stable intraday Hurst exponent is obtained when the average aggregation scheme is chosen. At

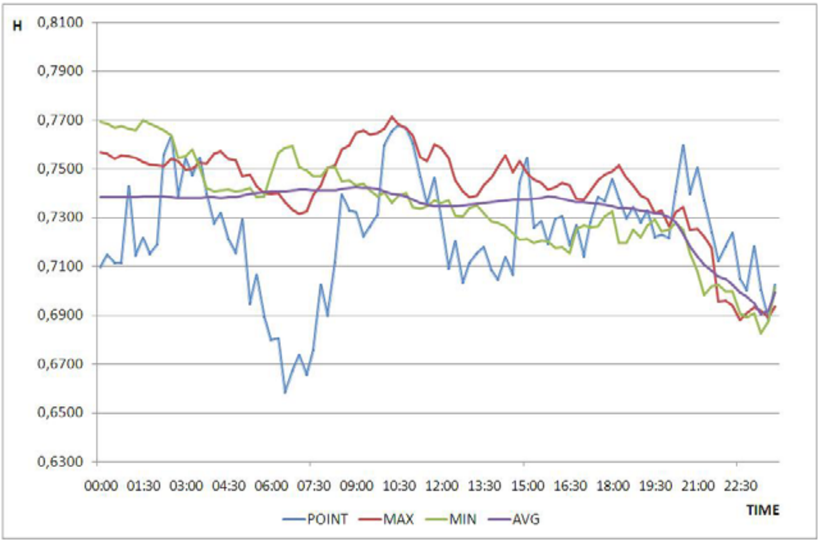

Figure 4.1: EUR/USD Hurst exponents dependence on data aggregation

the end of the day all the Hurst indices of currencies, estimated by using various aggregation rules, converge to the common values.

Nevertheless, different patterns of the Hurst index dependence on the exchange rate aggregation are observed for each currency analyzed. In the case of EUR/USD currency, the largest differences between various aggregation rules appear in the first part of the day. The maximum Hurst index value, amounting to 0,771, is obtained at 10:15 by the maximum value aggregation and the smallest Hurst index value of 0,658 is reached at 06:30 by the pointwise aggregation. In the second part of the day the Hurst indices decrease. On the contrary, the Hurst indices of GBP/USD currency increase in the second part of the day and exhibit a slight decrease at the very end of the day. The maximum and the minimum Hurst index values in the case of GBP/USD exchange rate are given by the pointwise aggregation and total, respectively 0,807 at 19:45 and 0,679 at 07:15. The Hurst exponents of JPY/USD currency have the largest variability compared to EUR/USD and GBP/USD currencies, except for the average aggregation rule, where the JPY/USD Hurst index is the most stable one. The maximum Hurst index value of 0,808 is obtained by the pointwise aggregation at 11:45 and the minimum Hurst index value, amounting to 0,637, appears at the very beginning of the day at 00:00 by the maximum value aggregation. More values of the Hurst index estimates are provided in the tables of Appendix 2. In conclusion, the analysis made shows that Hurst exponents depend on the choice of data aggregation. Therefore this feature should be regarded when modeling the aggregated price returns and volatility.

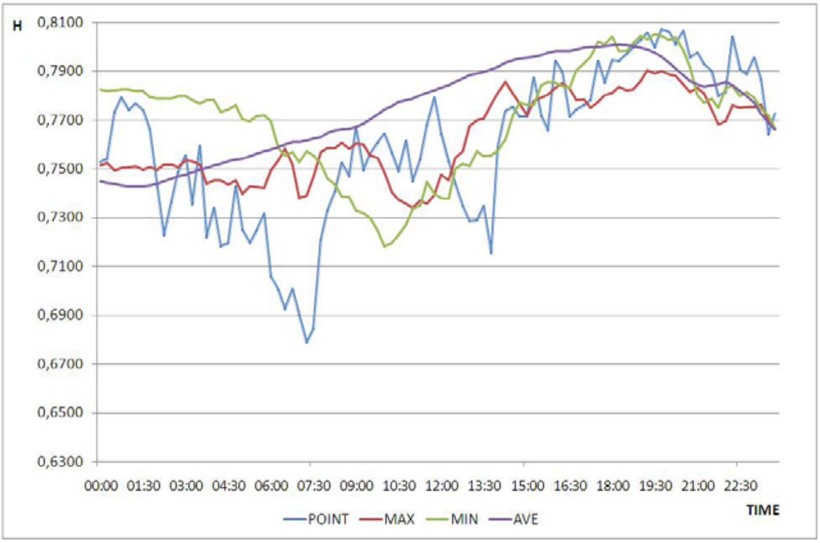

Figure 4.2: GBP/USD Hurst exponents dependence on data aggregation

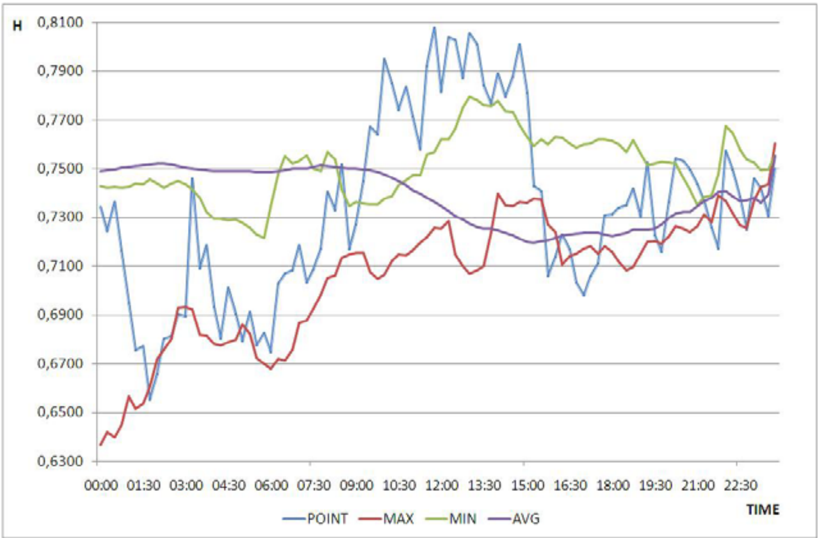

Figure 4.3: JPY/USD Hurst exponents dependence on data aggregation

4.5 Conclusions

In this chapter, we consider the basic stylized facts of high frequency data and aggregation. The analysis and representation of empirical statistical properties is very important when modeling series of financial asset prices. Having applied the aggregation to obtain functional data, we have to verify whether the aggregation affects well-known stylized facts. A long memory as one of the most important statistical properties, observed in absolute or squared returns, was chosen to illustrate the impact of aggregation on the Hurst index, as a parameter of long memory, which was calculated applying the R/S statistic. The empirical analysis made of the Hurst index dependence on the choice of the four risk factor aggregation methods analyzed (pointwise, maximum value, minimum value, and average value) illustrates the variability of the foreign exchange rates EUR/USD, GBP/USD and JPY/USD. We have illustrated that the data aggregation did not influence the general statistical property of a long memory in absolute returns series, however, the parameter of the long memory fluctuated, depending on the aggregation as well as on the time moment it was measured.

General Conclusions

In the thesis, data aggregation in risk measurement models was considered. The tasks set in the introduction were accomplished and can be summarized as follows:

- The aggregated Value-at-Risk model has been defined and dependence of model estimates on the data aggregation has been analyzed using high frequency foreign exchange rates. The empirical study has showed that the estimates of currency risk considerably depended on the applied aggregation rule and varied during the day.

- The functional $\rho - \text{GARCH}(1,1)$ model that depends on some features of functional data has been constructed. The existence of a stationary solution and the consistency of maximum likelihood estimators of model parameters were proved. Several examples with the known aggregated returns density function were given with a reference to applications in the analysis of financial assets risk.

- The Hilbert space-valued $\text{GARCH}(1,1)$ model with univariate volatility was introduced. The existence of a stationary solution, the consistency and the asymptotic normality of quasi-maximum likelihood estimators of the model parameters have been proved; the asymptotic properties of residuals have been analyzed.

- The dependence of the Hurst exponent, as a long memory parameter, on data aggregation have been investigated, using absolute returns of foreign exchange rates. The analysis has confirmed the well known stylized fact that absolute returns exhibit a long memory; however the Hurst index fluctuated, depending on the aggregation as well as on the time moment it was measured.

The thesis posed a problem of applying aggregated high frequency observations when calculating risk. The analysis made has clearly shown that the values of estimates varied, depending on the aggregation rule and on the time moment the risk was measured. This observation means that the choice of how daily data are fixed matters a lot. Up till now the daily data are understood as one value attributed to the day, usually the closing or last price of the day. However, there are much more different ways to construct daily observations when aggregation is considered. Thus, it is very important to set up certain aggregation rules to be sure that all market participants measure risk in the same way. The field is quite new, therefore, in the thesis, only the problem of aggregation is illustrated and some steps in constructing new functional data models are made. An extensive research is still required to set up a theoretical background, since the theory of functional data modeling is little developed.

Appendix 1

Theorem A.1 (Theorem A.2. [40]). *If $(Z_t)_{t\in\mathbb{Z}}$ is a strictly stationary ergodic process and $(Y_t)_{t\in\mathbb{Z}}$ is the process defined as*

$$Y_t = f(\dots, Z_{t-1}, Z_t, Z_{t+1}, \dots),$$

where f is a measurable function of $\mathbb{R}^\infty$ on $\mathbb{R}$, then the process $(Y_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

Theorem A.2 (Theorem 2.2.1. [71]). *Let (v_t) or $(v_t)_{t\in\mathbb{N}}$, respectively, be a stationary ergodic sequence of random elements with values in $\mathbb{C}(K, \mathbb{R}^{d'})$. Then the uniform strong law of large numbers is implied by $E\|v_0\|_K < \infty$.*

Proposition A.3 (Proposition 2.5.1. [71]). *Let $(\xi_t)_{t\in T}$ be a sequence of real random variables with $\xi_t \stackrel{e.a.s.}{\to} 0$ and $(v_t)_{t\in T}$ be a sequence of identically distributed random elements with values in a separable Banach space $(B, \|\cdot\|)$. If $E(\log^+\|v_0\|) < \infty$, then $\sum_{t=0}^{\infty} \xi_t v_t$ converges a.s., and one has $\xi_n \sum_{t=0}^{n} v_t \stackrel{e.a.s.}{\to} 0$ and $\xi_n v_n \stackrel{e.a.s.}{\to} 0$, as $n \to \infty$.*

Appendix 2

Table A.1: Hurst Exponents of EUR/USD absolute daily aggregated returns

$TIME$	$POINT$	MAX	MIN	AVE
00 : 00	0,7097	0,7569	0,7691	0,7382
01 : 00	0,7428	0,7551	0,7664	0,7384
02 : 00	0,7188	0,7513	0,7672	0,7388
03 : 00	0,7542	0,7495	0,7550	0,7379
04 : 00	0,7275	0,7560	0,7409	0,7383
05 : 00	0,7292	0,7469	0,7410	0,7389
06 : 00	0,6798	0,7398	0,7474	0,7406
07 : 00	0,6738	0,7313	0,7505	0,7413
08 : 00	0,6897	0,7505	0,7501	0,7410
09 : 00	0,7322	0,7647	0,7431	0,7424
10 : 00	0,7594	0,7665	0,7403	0,7407
11 : 00	0,7603	0,7636	0,7343	0,7374
12 : 00	0,7296	0,7587	0,7358	0,7345
13 : 00	0,7113	0,7383	0,7339	0,7354
14 : 00	0,7044	0,7506	0,7277	0,7366
15 : 00	0,7544	0,7482	0,7213	0,7375
16 : 00	0,7293	0,7424	0,7175	0,7382
17 : 00	0,7139	0,7374	0,7266	0,7363
18 : 00	0,7456	0,7489	0,7324	0,7347
19 : 00	0,7276	0,7392	0,7218	0,7330
20 : 00	0,7216	0,7263	0,7249	0,7302
21 : 00	0,7502	0,7253	0,7080	0,7140
22 : 00	0,7182	0,6960	0,6996	0,7049
23 : 00	0,7180	0,6933	0,6907	0,6950

Table A.2: Hurst Exponents of GBP/USD absolute daily aggregated returns

TIME	*POINT*	*MAX*	*MIN*	*AVE*
00 : 00	0,7526	0,7515	0,7825	0,7448
01 : 00	0,7738	0,7507	0,7826	0,7429
02 : 00	0,7424	0,7497	0,7788	0,7438
03 : 00	0,7554	0,7532	0,7798	0,7477
04 : 00	0,7339	0,7450	0,7780	0,7513
05 : 00	0,7251	0,7395	0,7703	0,7541
06 : 00	0,7058	0,7493	0,7694	0,7579
07 : 00	0,6899	0,7381	0,7526	0,7611
08 : 00	0,7329	0,7584	0,7460	0,7645
09 : 00	0,7666	0,7603	0,7328	0,7672
10 : 00	0,7642	0,7487	0,7182	0,7407
11 : 00	0,7448	0,7339	0,7335	0,7788
12 : 00	0,7644	0,7477	0,7381	0,7832
13 : 00	0,7283	0,7675	0,7510	0,7882
14 : 00	0,7589	0,7810	0,7575	0,7918
15 : 00	0,7710	0,7717	0,7762	0,7954
16 : 00	0,7940	0,7830	0,7853	0,7982
17 : 00	0,7758	0,7782	0,7929	0,7994
18 : 00	0,7944	0,7812	0,8040	0,8007
19 : 00	0,8028	0,7860	0,8044	0,7996
20 : 00	0,8061	0,7887	0,8026	0,7935
21 : 00	0,7974	0,7830	0,7801	0,7844
22 : 00	0,7813	0,7694	0,7832	0,7857
23 : 00	0,7954	0,7752	0,7790	0,7772

Table A.3: Hurst Exponents of JPY/USD absolute daily aggregated returns

$TIME$	$POINT$	MAX	MIN	AVE
00 : 00	0,7342	0,6366	0,7427	0,7488
01 : 00	0,6948	0,6565	0,7426	0,7507
02 : 00	0,6659	0,6719	0,7436	0,7520
03 : 00	0,6894	0,6931	0,7434	0,7504
04 : 00	0,6932	0,6782	0,7296	0,7491
05 : 00	0,6793	0,6859	0,7276	0,7490
06 : 00	0,6747	0,6677	0,7345	0,7486
07 : 00	0,7185	0,6868	0,7531	0,7499
08 : 00	0,7405	0,7050	0,7567	0,7511
09 : 00	0,7269	0,7155	0,7363	0,7501
10 : 00	0,7946	0,7067	0,7376	0,7476
11 : 00	0,7711	0,7163	0,7473	0,7411
12 : 00	0,7815	0,7252	0,7621	0,7345
13 : 00	0,8053	0,7068	0,7793	0,7275
14 : 00	0,7889	0,7398	0,7778	0,7246
15 : 00	0,7812	0,7359	0,7631	0,7197
16 : 00	0,7137	0,7239	0,7629	0,7214
17 : 00	0,6979	0,7172	0,7600	0,7235
18 : 00	0,7310	0,7158	0,7613	0,7223
19 : 00	0,7300	0,7152	0,7562	0,7251
20 : 00	0,7362	0,7222	0,7523	0,7299
21 : 00	0,7433	0,7265	0,7349	0,7345
22 : 00	0,7571	0,7367	0,7673	0,7407
23 : 00	0,7457	0,7370	0,7523	0,7379

Bibliography

[1] Alizadeth S., Brandt M.W., Diebold F.X. (2002) Range-Based Estimation of Stochastic Volatility Models. *The Journal of Finance* **57**, 1047–1091.

[2] Andersen T.G., Benzoni L. (2009) *Stochastic Volatility*, Federal Reserve Bank of Chicago, Working Paper No. 2009-04.

[3] Andersen T.G., Bollerslev T. (1997) Heterogenous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns. *The Journal of Finance* **52**, 975–1005.

[4] Andersen T.G., Bollerslev T., Christoffersen P.F., Diebold F.X. (2005) *Volatility Forecasting*, PIER Working Paper No. 05-011.

[5] Andersen T.G., Bollerslev T., Diebold F.X. (2004) *Parametric and Nonparametric Volatility Measurement.* In: L.P. Hansen and Y. Ait-Sahalia (eds.) "Handbook of Financial Econometrics", Amsterdam: North-Holland.

[6] Andersen T.G., Bollerslev T., Diebold F.X. (2003) Modeling and Forecasting Realized Volatility. *Econometrica* **71**, 529–626.

[7] Asai M., McAleer M., Yu J. (2006) Multivariate Stochastic Volatility: A Review. *Econometric Reviews* **25**, 145–175.

[8] Azaïs J.M., Wschebor M. (2002) The Distribution of the Maximum of a Gaussian Process: Rice Method Revisited, *In and out of equilibrium: probability with a physical flavour*, Progress in Probability, Birkhaüser, 321–348.

[9] Baillie R.T. (1996), Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**, 5–59.

[10] Baillie R.T., Bollerslev T., Mikkelsen H.O. (1996), Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **74**, 3–30.

[11] Basel Committee of Banking Supervision (2006) International Convergence of Capital Measurement and Capital Standards. *Available from www.bis.org.*

[12] Bauwens L., Laurent S., Rombouts J.V.K. (2006), Multivariate GARCH models: A Survey. *Journal of Applied Econometrics* **21**, 79–109.

[13] Beltratti A., Morana C. (1999) Computing value at risk with high frequency data. *Journal of Empirical Finance* **6**, 431–455.

[14] Bhar R. (1994) Testing for long-term memory in yen/dollar exchange rate. *Asia-Pacific Financial Markets* **1**, 101–109.

[15] Bollerslev T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

[16] Booth G.G., Kaen F.R., Koveos P.E. (1982) R/S analysis of foreign exchange rates under two international monetary regimes. *Journal of Monetary Economics* **10**, 407–415.

[17] Borodin A.N., Salminen P. (1996) *Handbook of Brownian Motion*, Berlin: Birkhäuser Verlag.

[18] Bosq D. (2000) *Linear Processes in Function Spaces*, Lecture Notes in Statistics, **149**, Springer.

[19] Boudoukh J., Richardson M., Whitelaw R.F. (1998) The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk. *Risk* **11**, 64–67.

[20] Brownless C., Gallo G. (2006) Financial econometric analysis at ultra-high frequency: data handling concerns. *Computational Statistics & Data Analysis* **51**, 2232–2245.

[21] Cheung Y.W. (1993) Long Memory in Foreign-Exchange Rates. *Journal of Business and Economic Statistics* **11**, 93–101.

[22] Colletaz G., Hurlin C., Topkavi S. (2007) *Irregularly Spaced Intraday Value at Risk Models (ISIVaR) Forecasting and Predictive Abilities*, Working paper, Univercity of Orléans.

[23] Cont R. (2001) Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance* **1**, 223–236.

[24] Corsi F., Dacorogna M.M., Müller U.A., Zumbach G. (2001) Consistent High-precision Volatility from High-frequency Data. *Economic Notes by Banca Monte dei Pasci di Siena SpA* **30**, 183–204.

[25] Cutler D.M., Poterba J.M., Summers L.H. (1989) What moves stock prices? *Journal of Portfolio Management* **15**, 4–12.

[26] Dacorogna M.M., Gençay R., Müller U.A., Olsen R.B., Pictet O.V. (2001) *An introduction to high frequency finance*, London: Academic Press.

[27] Dacorogna M.M., Müller U.A., Nagler R.J., Olsen R.B., Pictet O.V. (1993) A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* **12**, 413–438.

[28] Danielson J., de Vries C.G. (2000) Value-at-Risk and Extreme Returns. *Annales d'Économie et de Statistique* **60**.

[29] Diebold F.X., Schuermann T., Stroughair J.D. (1999) *Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management.* In "Advances in Computational Finance", Kluwer Academic Publishers, Amsterdam.

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

[31] Dione G., Duchesne P., Pacurar M. (2006) *Intraday Value at Risk (IVaR) Using Tick-by-Tick Data with Application to the Toronto Stock Exchange*, Working paper, HEC Montreal.

[32] Dupuis A., Glattfelder J.B., Olsen, R.B. (2011) Patterns in high-frequency FX data: Discovery of 12 empirical scaling laws. *Quantitative Finance* **11(4)**, 599–614.

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

[34] Embrechts P., Frey R., McNeil A.J. (2005) *Basic Concepts in Risk Management* In: P.Embrechts, R. Frey and A.J. McNeil (eds.) "Quantitative Risk Management", Princeton, NJ: Princeton University Press.

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

[36] Engle R.F. (2000), The Econometrics of Ultra-High Frequency Data. *Econometrica* **68**, 1–22.

[37] Engle R.F., Manganelli S. (1999) *CAViaR: Conditional Value at Risk by Quantile Regressiom*, NBER Working Paper 7341.

[38] Engle R.F., Manganelli S. (2001), *Value at Risk Models in Finance*, European Central Bank Working Paper No.**75**.

[39] Engle R.F., Ng V.K. (1993), Measuring and Testing the Impact of News on Volatility. *Journal of Finance* **48**, 1749–1778.

[40] Francq Ch., Zakoian J.M. (2009) *Modéles GARCH: Structure, inférence statistique et applications financières*, Collection "Économie et statistiques avancées", Economica.

[41] Ghysels E., Harvey A.C., Renault E. (1996) *Stochastic Volatility.* In: G.S. Maddala and C.R. Rao (eds.) "Handbook of Statistics" **14**, Amsterdam: North-Holland.

[42] Giraitis L., Kokoszka P., Leipus, R. (2001) Testing for Long Memory in the Presence of a General Trend. *Journal of Applied Probability* **38**, 1033–1054.

[43] Giot P. (2005) Market risk models for intraday data. *The European Journal of Finance* **11**, 309–324.

[44] Goodhart C.A.E., Figliuoli L. (1991) Every minute counts in financial markets. *Journal of International Money and Finance* **10**, 23–52.

[45] Goodhart C.A.E., Hesse T. (1993) Central bank FX intervention assessed in continuous time. *Journal of International Money and Finance* **12**, 368–389.

[46] Hansen P.R., Lunde A. (2005) A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* **20**, 873–889.

[47] Harvey A.C., Ruiz E., Shephard N. (1994) Multivariate stochastic variance models. *Review of Economic Studies* **61**, 247Ű-264.

[48] Hull J., White A. (1998) Incorporating volatility updating into the historical simulation method for value at risk. *Journal of Risk* **1**, 5–19.

[49] Hurst H.E. (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116**, 770–799.

[50] J.P. Morgan/Reuters (1996) *RiskMetrics Technical Document*, 4th ed., available from www.riskmetrics.com.

[51] Jorion P. (2007) *Monte Carlo Methods*, Chapter 12, In "Value at Risk: The New Benchmark for Managing Financial Risk", 3rd Edition.

[52] Kim J., Malz A.M., Mina J. (1999) *LongRun Technical Document*, available from www.riskmetrics.com.

[53] Kvedaras V., Račkauskas A. (2010) Regression Models with Variables of Different Frequencies: The Case of a Fixed Frequency Ratio. *Oxford Bulletin of Economics and Statistics* **72**, 600–620 .

[54] Lo A.W. (1991) Long-term memory in stock market prices. *Econometrica* **59**, 1279–1313.

[55] Mandelbrot B.B. (1972) Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis. *Annals of Economic and Social Measurement* **1**, 259–290.

[56] McLeod A.I., Hipel K.W. (1978) Preservation of the Rescaled Adjusted Range 1: A Reassessment of the Hurst Phenomenon. *Water Resources Research* **14**, 491–508.

[57] McNeil A.J., Frey R. (2000) Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. *Journal of Empirical Finance* **7**, 271–300.

[58] Mina J., Xiao J.Y. (2001) *Return to RiskMetrics: The Evolution of a Standard.*, available from www.riskmetrics.com.

[59] Moody J., Wu L. (1995) Price Behavior and Hurst Exponents of tick-by-tick Interbank Foreign Exchange Rates. *In Proceedings of the IEEE/IAFE 1995 Computational Intelligence for Financial Engineering*, IEEE Service Center, 26-Ű30.

[60] Moody J., Wu L. (1996) *Improved Estimates for the Rescaled Range and Hurst Exponents.* In: Refenes, A., Abu-Mostafa, Y., Moody, J. and Weigend, A. (eds.) "Neural Networks in Financial Engineering", Proceedings of the Third International Conference, 537–553, London, World Scientific.

[61] Nath G.C., Reddy Y.V. (2002) Long Memory in Rupee-Dollar Exchange Rate - An Empirical Study. *Capital Market Conference 2002, Indian Institute of Capital Markets.*

[62] Nelson D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **59**, 347–370.

[63] Oh G.J., Um C.J., Kim S. (2006) Long-term Memory and Volatility Clustering in Daily and High-frequency Price Changes, arXiv:physics/0601174.

[64] Ohanissian A., Russel J.R., Tsay R.S. (2003) *Using Temporal Aggregation to Distinguish between True and Spurious Long Memory*, Working paper, GSB - University of Chicago.

[65] Račkauskas A. (1995) On the conditional covariance condition in the martingale CLT. *Lithuanian Mathematical Journal* **35**, 93–103.

[66] Ramsay J.O., Silverman B.W. (1997) *Functional Data Analysis*, New York: Springer.

[67] Riedel T., Sahoo P.K. (1998) *Mean value theorems and functional equations*, Singapore: World Scientific Publishing Co. Pte. Ltd.

[68] Shephard, N. (2004) *Stochastic Volatility: Selected Readings*, Oxford, UK: Oxford University Press.

[69] So M.K.P., Kwok S.W.Y. (2006) A multivariate long memory stochastic volatility model. *Physica A: Statistical Mechanics and its Applications* **362**, 450–464.

[70] Soofi A.S., Wang S., Zhang Y. (2006) Testing for Long Memory in the Asian Foreign Exchange Rates. *Journal of System Science and Complexity* **19**, 182–190.

[71] Straumann D. (2005) *Estimation of Conditionally Heteroscedastic Time Series Models*, Lecture Notes in Statistics, Springer.

[72] Tschernig R. (1995) Long Memory in Foreign Exchange Rates Revisited. *Journal of International Financial Markets, Institutions and Money* **5**, 53–78.

[73] Zakoïan J.M. (1994) Threshold Heteroskedastic Models, *Journal of Economic Dynamics and Control* **18**, 931–955.

[74] Zhou B. (1993) High frequency data and volatility in foreign exchange rates, *Journal of International Money and Finance* **10**, 23–52.