Value at Risk Using Stochastic Volatility Models

Mattias Bengtsson and Viktor Olsbo

10th September 2003

Abstract This master's thesis deals with Value at Risk (VaR). Estimations are done in several different ways, using parametric and non-parametric volatility models. Underlying distributions that are used are the Generalized Hyperbolic distribution, various special cases of it, and the Generalized Pareto distribution. The models are fitted to three different data sets, namely, DAX, Olsen USD/DM and Siemens stock prices. For one of the data sets, we also look at how the mean-variance mixing relations between Non-Gaussian Ornstein-Uhlenbeck (OU) process volatility and Brownian motion perform in a VaR setting. We show that the Generalized Hyperbolic distribution, used together with the Nadaraya-Watson or the Variance Window volatility model, may very well be used when calculating Value at Risk.

Keywords: Devolatization; FOEL; GARCH-AR; Generalized Hyperbolic distribution; Generalized Pareto distribution; Integrated Volatility; Lévy process; LPFA; Mixing relations; Nadaraya-Watson estimator; Non-Gaussian OU process; Pearson VII distribution; Variance Window.

Acknowledgements We are greatly indebted to Patrik Albin for being the fantastic inspirator that he is, as well as for his comments and suggestions. We would also like to thank Erik Brodin, Johan Tykesson and Cătălin Stărică for fruitful discussions.

On a personal level, Mattias would like to thank Annah, Malkolm, Barbro, Daniel, Linda, Robert, Essie and Willy, Marianne Johansson and Marianne Persson. Of course, thanks also goes to Viktor for fruitful and inspiring collaboration.

On a personal level, Viktor would like to thank his family Maja & Leon, Maj-Britt, Lennart, Oskar, Catherine, Ida, Åke, Ann-Marie, Lisa, Inga, his in-laws Christer, Berit, Nanna, Johan and of course his friends. A special thanks goes to Mattias for inspiring and fruitful collaboration.

Contents

Chapter 1

Introduction

One of the main objectives of risk management, is to protect an institution against unacceptable losses. Further, as JP Morgan and Reuters, [40], put it:

> *Setting limits in terms of risk, helps business managers to allocate risk to those areas which they feel offer the most potential, or in which their firms' expertise is greatest. This motivates managers of multiple risk activities to favor risk reducing diversification strategies.*

Firms (e.g. banks, financial institutions etc.) are faced with many risks in their daily business activities. Some of these are (see [22]):

- *Business Risks*; the risks that are specific to a certain industry.
- *Market Risks*; the risks of losses arising from adverse movements in market prices (e.g. equity prices) or market rates (e.g. interest rates and exchange rates).
- *Credit Risks*; the risks of loss arising from the failure of a counterpart to honour a promised payment.
- *Liquidity Risks*; the risks arising from the cost of inconvenience of unwinding a position.
- *Operational Risks*; the risks arising from failure of internal systems or the people who operate in them.
- *Legal Risks*; the risks arising from the prospect that contracts may not be enforced.

Managers must therefore manage their firms' exposures to these various risks. They must decide on which risks they want to bear, assess the risks they currently bear, and alter their exposures accordingly, so that they bear the risks they want.

There are many examples of the effect of large losses incurred to companies or (and) institutions with inadequate risk management. Some of these are displayed in the table below, taken from [30] and [34].

Company	Country	Year	Loss (Billion)
Pension Funds	Sweden	2001-02	SEK 452
Banks	Sweden	1991-94	$15.0
Orange County, (Public Fund)	California, USA	1994	$1.64
Showa Shell Sekiyu	Japan	1993	$1.58
Kashima Oil	Japan	1994	$1.45
Metallgesellschaft	Germany	1994	$1.34
Barings Bank	UK	1995	$1.33

Table 1.1: *Examples of great losses due to improper risk management.*

We will focus on measuring market risk, which is the type of risk that is most easily quantifiable.

As stated above, market risk is the risk of loss resulting from a change in the value of tradable assets. For example, if a person holds a position of 10000 Ericsson B stocks, this person is exposed to the risk that the value of the stock decreases. Market risk thus arises from the changes in prices of assets and liabilities. Market risk is also called *Price Risk*.

Remark It should be noted that, we will measure market risk from a mathematical point of view. That is, we will not take into great account managerial perspectives (i.e. how proposed models would work from a managers point of view). The main focus is; finding models that work well in the backtesting procedures, which we work with, not caring about transaction fees and courtages etc.

Chapter 2

Risk Measures

Although we are often interested in the price of financial assets, the performance of securities is measured in terms of price changes. Financial theory assumes that returns (not prices or price changes) are the compensation for risk. That is, it is assumed that the higher the expected return, the higher the risk.

Modern portfolio theory is also based on returns, using a mean-variance approach, see [25]. Here mean represents an asset's expected return, and variance (or standard deviation) their risk, ([43]).

There are two standard definitions of returns, *percentage*, R_t^p, and *logarithmic*, *log-returns*, (or *geometric*), R_t^l. They are defined as follows:

Definition 1 Let P_t denote, the price at time t. Then percentage returns are defined as

$$R_t^p = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1,$$

and log-returns are defined as

$$R_t^l = \log(\frac{P_t}{P_{t-1}}).$$

$\square$

Percentage returns, which arguably is the most intuitive return concept, fulfill the principle of limited liabilities (i.e. you can not lose more than you have invested). On the other hand, R_t^l may take values on the whole real line. (For percentage returns to have this property, it would imply negative prices.)

Remark Although percentage returns only take values in $[-1, \infty)$, they are commonly modelled as random variables taking values on the whole real line (e.g. Gaussian).

Standard results from interest rate theory, applied on returns, lead us to the following relationship between *continuously compounded returns*, R_t^c, and 1-period percentage returns, see [43]

$$R_t^c = \log\left(1 + R_t^p\right) = \log\left(\frac{P_t}{P_{t-1}}\right) = R_t^l.$$

Log-returns are therefore sometimes called continuously compounded returns.

Remark Since log-returns usually are small, we can use a first order Taylor expansion of R_t^l. This gives,

$$R_t^l = \log\left(1 + R_t^p\right) \approx R_t^p + o(R_t^p),$$

which means, that the difference between R_t^l and R_t^p usually is relatively small.

Further, if we assume the price process to be of exponential type (i.e. $P_t = \mathrm{e}^{X_t}$), log-returns will coincide with the increments of the process X_t. As we will see later, this feature is very convenient. Throughout this thesis we will use log-returns.

2.1 Classical Approaches to Market Risk Measurement

The most simple measure for the *risk* of an asset, A, is the standard deviation, σ_{A}. This idea dates back to the 50's, and was proposed by Markowitz, [38].

Further, if we follow the theory of CAPM (*Capital Asset Pricing Model*), we will run across another risk measure, β. It follows from CAPM that risk can be divided into *systematic* and *nonsystematic* risk. These are defined as follows, see [28]:

- **Nonsystematic Risk** The risk of price change due to the unique circumstances of a specific security, as opposed to the overall market. This risk can be virtually eliminated from a portfolio through diversification.
- **Systematic Risk** The risk which is common to an entire class of assets or liabilities. The value of investments may decline over a given time period, simply because of economic changes or other events that impact large portions of the market. Asset allocation and diversification can protect against un-diversifiable risk, because different portions of the market tend to under perform at different times.

The risk measure β of CAPM is defined as

$$\beta_A = \frac{\rho_{A,M}}{\sigma_M^2},$$

where $\rho_{A,M}$ is the covariance between the asset and the market portfolio, M, and σ_M^2 is the variance of the market portfolio.

There are two portfolio performance measures built on these two risk measures, the *Sharpe ratio*, *SR*, and the *Treynor measure*, *T*. These are sometimes used as risk measures as well.

The Sharpe ratio is a ratio of an assets expected excess return (i.e. return exceeding the return on a risk free investment), to its standard deviation. That is,

$$SR = \frac{\bar{R}_A - R_F}{\sigma_A},$$

where $\bar{R}_A$ is the *assets expected return* and R_F is the *return on a risk free asset.*

The Treynor measure is similar to the Sharpe ratio, except that σ_A is replaced by β_A. In other words, it is the ratio of excess return and non-diversifiable risk

$$T = \frac{\bar{R}_A - R_F}{\beta_A}.$$

The main problem when using the above mentioned risk measures, is that they are relative measures. Thus, they do not express the exposure of risk in monetary terms. A risk measure that expresses the exposure in monetary terms is *Value at Risk* (VaR).

2.2 Value at Risk

One question each portfolio owner/manager is faced with is,

How much can I lose during a normal trading period?

The obvious answer to that question, is of course, *everything*! Though correct, it is not a very constructive answer. One way of producing a more stimulating answer is the VaR methodology.

VaR is an attempt to provide a single number, summarizing the total market risk in a portfolio of financial assets. It has become widely used by corporate treasurers and fund managers, as well as by financial institutions.

By using VaR, we can produce statements like, see [27]

"We are certain that, with probability α, we will not lose more than $\mathrm{VaR}_{\alpha,D}$ in the next D days."

$\mathrm{VaR}_{\alpha,D}$ is the D-day α-level VaR. If we let $R_{t,D}$ be a random variable, whose outcome determines the value of next D-day return, the definition of VaR is as follows, see e.g. [39]:
Definition 2

$$\mathrm{VaR}_{\alpha,D} = -\inf\{x \in \mathbb{R} : \mathbb{P}(R_{t,D} \leq x) \geq \alpha\}$$

□

The minus sign is there just to present VaR as a positive value. Calculations are often performed for $D = 1$. To obtain the D-day VaR, the following approximation is often used

$$\mathrm{VaR}_{\alpha,D} \approx \sqrt{D} \times \mathrm{VaR}_{\alpha,1}$$

Note that this formula holds with equality, only when the returns are additive, i.i.d. Gaussian with zero mean. Throughout this thesis we will restrict ourselves to the 1-day VaR, denoted by VaR_α. The second choice to be made when calculating VaR, is the probability level. Most banks use $\alpha = 0.99 = 99\%$. This is mainly due to suggestions from the Basel Committee on Banking Supervision, see [11]. We will estimate VaR for three probability levels, namely, $95\%, 97.5\%$ and 99%.

There are several different methods for calculating VaR (or estimating it, to be precise). We now give a short explanation of the most popular three. See [12], [22] and [27] for more information on these methods.

2.2.1 Historical Simulation (HS)

This method uses historical data for predicting future VaR values. The first step is to identify the market variables (i.e. market factors), that affect the value of the asset. Examples of such variables are exchange rates, equity prices, interest rates and so on. After deciding on the length of the time horizon, that the prediction will be based upon, historical values of all market factors must be collected. If the time horizon is T days, we now have T different scenarios for the future value of the asset. These scenarios are ranked by value, and the VaR is then the $1-\alpha$ percentile, of this data set.

2.2.2 Monte Carlo Simulation (MCS)

Instead of using past values of market factors for prediction, this method builds on an assumed knowledge about the stochastic mechanism that generate the returns. Unkown parameters, for this mechanism, are then estimated using historical values. A set of possible future values of the asset are then generated, using stochastic simulation.

The main advantage of this method, is that it is possible to make the data set very large. On the other hand, it is a serious drawback, that we have to make distributional assumptions about the returns. In this case, as in many other, practitioners usually restrict themselves to the univariate or the multivariate Gaussian distribution.

2.2.3 Model Building Approach (MBA)

This approach is sometimes referred to as the *Variance-Covariance* approach. It comes in two versions, depending on what is assumed, about the distribution of the returns:

- **Non Parametric** This method uses a number of historical observations of returns (and can therefore also be included in HS). The VaR is calculated as the empirical $(1-\alpha)$-percentile of these observations. An advantage of this method is that it doesn't make any particular assumption about distribution of returns. The obvious drawback is that it is not possible to extrapolate beyond the range of the data at hand.

- **Parametric** The most widely used distribution is the Gaussian. This is mainly because of the computational simplifications this implies. Since it is (or at least should be) common knowledge that financial returns are not Gaussian, one must be highly aware that an assumption of this kind greatly increases the risk of underestimating the VaR.

Distributions with heavier tails than the Gaussian are closer to the truth. In this thesis we will use the MBA, and propose three alternative distributions, namely, the Generalized Pareto distribution, the Pearson type VII and two special special cases (the Hyperbolical and the Normal Inverse Gaussian), of the Generalized Hyperbolical distribution.

VaR is then calculated as

$$\mathrm{VaR}_\alpha \ \text{ s.t. } 1-\alpha = \int_{-\infty}^{-\mathrm{VaR}_\alpha} f_{R_t}(s)ds = F_{R_t}(-\mathrm{VaR}_\alpha),$$

where f_{R_t} and F_{R_t} are the assumed (continuous) density function and distribution function, respectively, for the returns. In other words,

$$\mathrm{VaR}_\alpha = -F_{R_t}^{-1}(1-\alpha),$$

and in most cases, the distribution function has to be inverted numerically.

It is well known that, though uncorrelated, returns exhibit dependence. We will model the returns in two ways, each incorprating dependence. One approach is to model the price process as an exponential semimartingale, and the stochastic volatility as a non-Gaussian OU process. The second approach, is to assume that returns have the representation

$$R_t = \mu + \sigma_t X_t,$$

where X_t is a Lévy process. The dependency structure is, in both cases, incorporated in the stochastic volatility, σ_t. Thus the devolatized innovations, X_t, are assumed independent. This means that, assuming $\sigma_t > 0$,

$$X_t = \frac{R_t - \mu}{\sigma_t},$$

form an i.i.d. sequence.

The stochastic volatility, will be modelled using three different approaches, namely, non-parametric (Nadaraya-Watson), Variance Window and GARCH-AR. These models will be benchmarked against a constant volatility.

Denoting the α-level VaR for R_t and X_t, $\mathrm{VaR}_\alpha^{R_t}$ and $\mathrm{VaR}_\alpha^{X_t}$, respectively, we work under the assumption

$$\mathrm{VaR}_\alpha^{R_t} = \mu + \sigma_t \mathrm{VaR}_\alpha^{X_t}.$$

2.2.4 VaR Limitations

VaR is not the final word in market risk management, but it is a step in the right direction. This is the reason for the widespread use of VaR in the financial industry.

Drawbacks of VaR can be divided into two main groups ([43]):

- **Limits intrinsic to the VaR concept**
  - VaR is only for traded assets or liabilities. That is, it does not work for deposits or loans which, of course, is of great concern to banks.
  - VaR does not incorporate credit risk. Credit risk is an active research field, and there is a need for an integrated risk management system.
  - VaR does not incorporate liquidity risk. This is one of VaR's major limitations. The problem can be very important for instruments traded in thin markets, where the buying or sales of relatively small quantities of a single instrument can cause large price variations.
  - VaR only measures risk for unusual, but normal, events. Major chocks to the financial system are not incorporated.

- **Limits due to the statistical methodology used to implement VaR** A poet, who lived in Rome in the nineteenth century, defined statistics as, that strange science which states: “If you have eaten two chickens, and I am starving, we have both eaten one chicken each.” The point made here, is that VaR is based on statistical assumptions and methodologies. If these assumptions are unrealistic, the inference will probably lead to the wrong conclusions.

  The obvious examples here are, the far too often used Gaussian distribution, and models that do not incorporate dependence. Models based on assumptions of independence, are bound to fail when applied to dependent data.

It is important to note, that the critique of the use of Gaussian distribution does not threaten the validity of VaR itself. This critique is directed against the practitioner, with the poor judgement of using the Gaussian distribution. The Gaussian assumption can invalidate some models used for calculating VaR, but not the utility of a measure of market risk, such as VaR.

The limitations in terms of credit and liquidity risk are of wider concern. It lies in the future to show if VaR will be the foundation of an integrated risk management system, incorporating market, credit and liquidity risk.

Chapter 3

Lévy processes

3.1 Introduction to Lévy processes

Definition 3 ([47]) An $\mathbb{R}$-valued stochastic process $\{X_t : t \geq 0\}$, is a Lévy process, if the following conditions are satisfied:

- For any choice of $n \geq 1$ and $0 \leq t_0 \leq t_1 \leq t_2 \leq ... \leq t_n$, the random variables

  $$X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$$

  are independent (i.e. the process has independent increments);

- $X_0 = 0$ a.s;

- The distribution of $X_{s+t} - X_s$ does not depend on s, that is, the process has stationary increments;

- For all $\epsilon > 0$, it holds that, $\lim_{t \to s} \mathbb{P}(|X_s - X_t| > \epsilon) = 0$ (i.e. stochastic continuity);

- X_t is right-continuous for $t \geq 0$, and has left limits for $t > 0$ a.s. This is also known as *càdlàg*[1].

$\square$

Definition 4 A Lévy process with non-negative increments is called a subordinator.

$\square$

Remark ([32]) The most important Lévy process is Brownian motion, which is a Lévy process with Gaussian distributed increments. When botanist R. Brown described the movement of a pollen particle in suspended in fluid in 1828, it was observed that it moved in an irregular, random fashion. A. Einstein argued, in 1905, that the movement is due to bombardment of the particle by the molecules of the fluid. He obtained the equations for Brownian motion. In 1900 L. Bachelier used the Brownian motion as a model for movement of stock prices, see [4]. The Mathematical foundation for Brownian motion as a stochastic process was laid by N. Wiener in 1923, see [52]. Brownian motion is frequently also called the Wiener process. However, in stochastic calculus, some authors give the concept of Brownian motion and the Wiener process different meaning, see e.g. [35].

[1] Càdlàg stands for *continú a droit avec limites a gauche.*

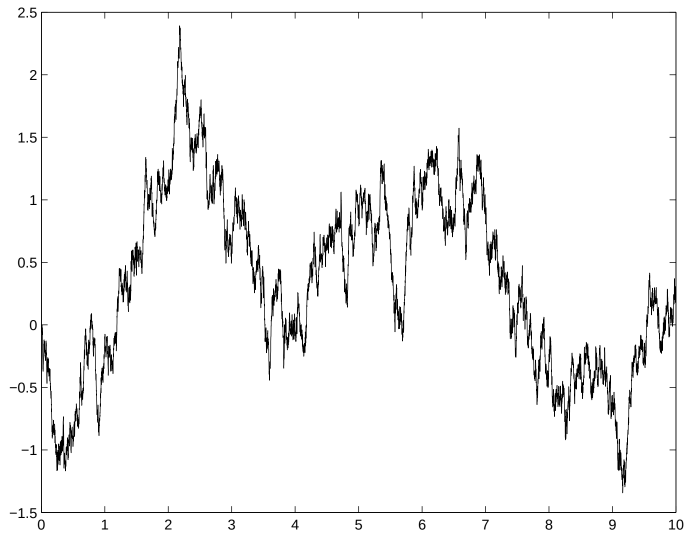


Figure 3.1: *Realization of a Brownian motion, with $\mu = 0$ and $\sigma = 1$.*

3.2 Lévy Processes Used in this Thesis

We work under the assumption that the price of an asset can be modelled by an exponential Lévy process. A special case is the Bachelier-Samuelson model, in which the price process S_t is taken to be exponential Brownian motion;

$$S_t = S_0 \mathrm{e}^{\mu t + \sigma B_t}, \ 0 \leq t \leq T,$$

where μ is the mean and σ is the (constant) volatility.

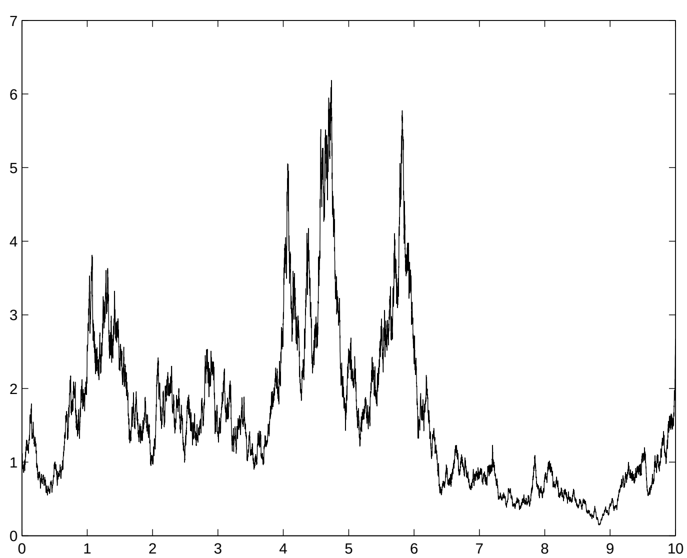


Figure 3.2: *Realization of an exponential Brownian motion, with $\mu = 0$ and $\sigma = 1$.*

This gives a model that is fairly easy to work with, but it does not really capture the behavior of a stock price, exchange rate or an index. This is due to the fact that the Bachelier-Samuelson model makes the log-returns,

$$X_t = \log S_t - \log S_{t-\Delta t},$$

where Δ typically is one day, of say a stock price at the time t, Gaussian distributed. The Gaussian distribution is disadvantageous in that it does not permit skewness, and in that the probability of extreme events is way too small.

A model that seems to better fit a price process, see [20], is the exponential Lévy motion

$$S_t = S_0 \mathrm{e}^{\mu t + \sigma_t L_t},\ 0 \leq t \leq T.$$

Here L_t is a more general Lévy process, than Brownian motion, and σ_t is stochastic volatility.

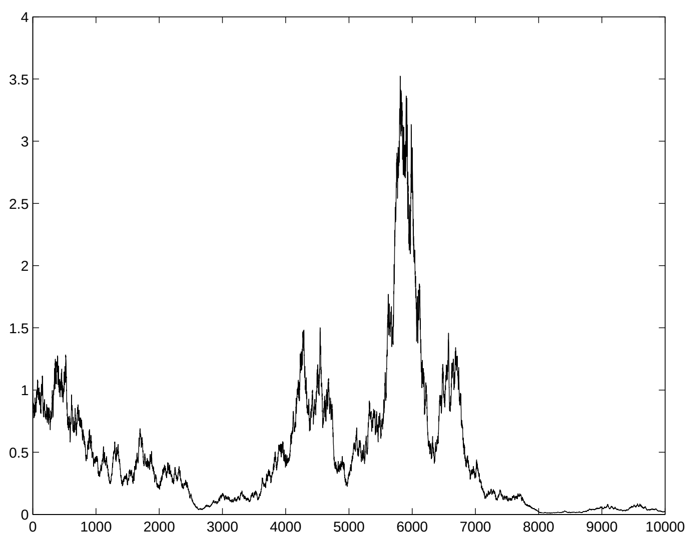


Figure 3.3: *Exponential Normal Inverse Gaussian Lévy process (see section 5 below), with* $\mu = 0$ *and* $\sigma_t \equiv 1$.

Chapter 4

Stochastic Volatility

Working under the assumption that the price process S_t of an asset follows an exponential Levy process, we have, as mentioned above, that

$$S_t = S_0 \mathrm{e}^{\mu t + \sigma_t L_t},\ 0 \leq t \leq T.$$

Here L_t is a Lévy process, and σ_t a non-negative stationary stochastic process (*volatility process*), independent of L_t. If we assume that σ_t moves slowly, compared to L_t, we have that

$$X_t = \log S_t - \log S_{t-1} = \mu + \sigma_t L_t - \sigma_{t-1} L_{t-1} \approx \mu + \sigma_t \Delta L_t = \mu + \sigma_t A_t{}^1$$

It is well known, that log-returns are not independent. Assumig that A_t is the increment process of a Lévy process, it is a random walk. This implies that the dependency stucture must be fully captured by the volatility process.

Definition 5 The *devolatized log-returns* of a price process, S_t, as above, are defined as

$$A_t = \frac{X_t - \mu}{\sigma_t}.$$

□

In order to fit a marginal distribution to A_t, it is necessary to estimate σ_t and μ. We estimate μ by the sample mean, (i.e $\hat{\mu} = \frac{1}{T}\sum_{j=1}^{T} X_{t-j}$).

We will, under these assumptions, consider three models and hence estimators, $\hat{\sigma}_t$, for the volatility process σ_t. Later on, we will also investigate a fourth volatility model (see section 4.4). We now give a presentation of the volatility models:

4.1 Variance Window

The most basic estimator for the volatility, is what we call the Variance Window. This is a very natural estimator for volatility, and it is often used to estimate the constant volatility

[1] The notation A_t for the increment process comes from the French word *augmentation*.

in the Bachelier-Samuelson model. The Variance Window is defined as the sample standard deviation of the log-returns, over the last n trading days. That is,

$$\sigma_t = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_{t-i}-\mu_n)^2},$$

where $\mu_n = \frac{1}{n}\sum_{j=1}^{n} X_{t-j}$.

In calculations, n differs between different data sets, and is optimized in each case to get the best devolatization. That is, optimized to get the least dependence in the A_t process. This is done by minimizing the *Ljung-Box test statistic*.

Definition 6 The *autocorrelation function* (acf) of a (weakly) stationary stochastic process, X_t, is defined as

$$\text{acf}(s) = \text{Corr}(X_{t+s}, X_t).$$

□

Definition 7 For a sample $x_1, x_2, ..., x_T$, the *sample autocorrelation function* (sacf), is defined as

$$\text{sacf}(i) = \frac{1}{Ts^2}\sum_{t=i+1}^{T}(x_t-\hat{\mu})(x_{t-i}-\hat{\mu}).$$

Here, s^2 is the sample variance, and $\hat{\mu}$ is the sample mean.

□

Theorem 1 Let r_i be the sacf, and T sample length. Then the Ljung-Box test statistic is defined by, see [41]

$$Q(k) = T(T-2)\sum_{i=1}^{k}\frac{r_i^2}{T-i}.$$

The statistic $Q(k)$ is asymptotically χ_k^2 distributed, when data are uncorrelated.

□

Remark The statistic $Q(k)$ is a modification of the *Portmanteau statistic* $Q^*(k) = T\sum_{i=1}^{k} r_i^2 \sim_{\text{approx}} \chi_k^2$, for which Box and Pearce [16] showed the asymptotic properties. Ljung and Box [15] later argued that the modified statistic has better asymptotic properties.

4.2 Time Series Model

A popular approach for modelling stochastic volatility, is to use autoregressive models. In 1982, Engle, see [24], introduced a model known as the *first order autoregressive heteroschedastic*, ARCH(1), process. In 1986, Bollerslev, see [14], extended it to the *Generalized* ARCH, (GARCH) process. This model is very popular, and is still widely used in the financial industry.

Definition 8 Consider the model, $X_t = \mu + \sigma_t \epsilon_t$, where X_t is the log-return process, σ_t is a stationary stochastic process, and ϵ_t is noise, that is independent of the σ_t process, and of X_s for $s < t$. We say that σ_t^2 is a GARCH(p, q) process, if

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where α_i, $\beta_j > 0$, $\forall\ i = 0, ..., p$ and $j = 1, ..., q$ are parameters. The condition for the existence of such a GARCH process is, see [14],

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

□

We will focus on the GARCH(1,1)-model, which simplifies the above expression to

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

Contrary to our beliefs, we will assume the noise, ϵ_t, to be N$(0, 1)$-distributed. That is, X_t is conditionally (on past observations) following a N$(0, \sigma_t^2)$-law, i.e. $X_t | \mathcal{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2)$, where $\mathcal{F}_{t-1}$ is the information up to time $t - 1$.

Remark This model could probably be improved by using other distributions for the noise. We work under the assumption of Gaussianity, since it is what is most frequently used in the financial industry, and we want to investigate how well this popular model works.

The process σ_t moves rapidly, and we would like to "cool it down" for better devolatization. The cooling down is done using an AR(1)-process (*Auto Regressive* process of order 1).

Definition 9 The AR(p)-process is defined by

$$X_t = \sum_{i=1}^{p} \rho_i X_{t-1} + \eta_t,$$

where η_t is noise, with variance σ^2, and independent of X_s for $s < t$.

□

The parameters, ρ_i, can be estimated by the following scheme, see [2]:

Yule-Walker: AR(p)-process First we estimate the covariance function, $r_X(\tau)$, for $\tau = 0, ..., p$. Then we estimate σ^2 and $\rho_1, ..., \rho_p$ by solving the following system of equations:

$$\begin{cases} r_X(j) - \rho_1 r_X(j-1) - ... - \rho_p r_X(j-p) = 0, & \text{for } j = 1, ..., p \\ r_X(0) - \rho_1 r_X(1) - ... - \rho_p r_X(p) = \sigma^2. \end{cases}$$

For an AR(1)-process, this simplifies to

$$\begin{cases} \hat{\rho}_1 = r_X(1)/r_X(0) \\ \hat{\sigma}^2 = [r_X(0)^2 - r_X(1)^2]/r_X(0). \end{cases}$$

Estimating GARCH parameters Estimation of α_0, α_1 and β, can be done, using the *maximum likelihood* (ML) method. Let $Y_1, ..., Y_n$ be de-meaned log-returns, $Y_t = X_t - \mu$, that is, we have that Y_t is conditionally Gaussian w.r.t $\sigma_1, ..., \sigma_{t-1}$, or equivalently, $Y_1, ..., Y_{t-1}$. Then we have have a joint density function, given by

$$\begin{aligned} p\left(Y_1, ..., Y_n\right) &= p\left(Y_n | Y_1, ..., Y_{n-1}\right) p\left(Y_{n-1} | Y_1, ..., Y_{n-2}\right) \times ... \times p\left(Y_2, | Y_1\right) p\left(Y_n\right) \\ &= p\left(Y_1\right) \prod_{t=2}^{n} p\left(Y_t | Y_1, ..., Y_{t-1}\right) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_{t-1}} \exp\left\{ -\frac{\sigma_t^2}{2\sigma_{t-1}^2} \right\}, \end{aligned}$$

where the conditioning on α_0, α_1 and β has been dropped for notational reasons. The above expression gives , the log-likelihood function, for observations $y_1, y_2, ..., y_n$,

$$l(\alpha_0, \alpha_1, \beta; y_1, y_2, ..., y_n) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ \log(2\pi\sigma_t^2) + \frac{y_t^2}{\sigma_t^2} \right\}.$$

We estimate the parameters, α_0, α_1 and β, by numerical maximization of $l(\alpha_0, \alpha_1, \beta; y_1, y_2, ..., y_n)$.

Finally, the GARCH(1,1)-AR(1)-process, is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \left(X_{t-1} - \rho_1 X_{t-2}\right)^2 + \beta\sigma_{t-1}^2.$$

4.3 Nonparametric Approach, Nadaraya-Watson

When trying to model volatility, it is crucial to determine how volatility evolves over time. In statistical language, we want to investigate the association of an explanatory variable, t (time), and a response variable (volatility or squared de-meaned log-returns) Y^2. Specifically, we want to express a functional relation between t and Y^2, of the form

$$\mathbb{E}(Y_t^2) = \sigma_t^2.$$

Here we assume that

$$Y_t^2 = \sigma_t^2 + \epsilon, \quad \mathbb{E}(\epsilon) = 0, \quad \mathbb{E}(\epsilon^2) = c.$$

In this nonparametric approach, the task is to estimate σ_t^2 without making any distribution assumptions on Y. This can be done by weighting and summing squared de-meaned log-returns,

$$\hat{\sigma}_t^2 = \sum_{i=1}^{n} w_i(t, h) Y_i^2,$$

in an appropriate manner. To do this, we need to determine the weights $w_i(t, h)$, where the number h remains to be defined.

In 1964, Nadaraya, [42] and Watson, [51], independently proposed the weights

$$w_i(t, h) = \frac{K\left((t - t_i)/h\right)}{\sum_{j=1}^{n} K\left((t - t_i)/h\right)},$$

for appropriate kernels $K(\cdot)$. In our case, we decided to use the following volatility estimator, see [49],

$$\hat{\sigma}_t^2 = \frac{\sum_{i=1}^{t-1} K_h(i - t) Y_i^2}{\sum_{i=1}^{t-1} K_h(i - t)}.$$

Here $Y_t = X_t - \hat{\mu}$, $\hat{\mu} = \frac{1}{t}\sum_{j=1}^{t} X_j$ and $K_h(\cdot) = h^{-1}K(\cdot/h)$, where $K(x) = \mathrm{e}^{-x^2}$.

The numbers h and n are known as the bandwidth and the window length, respectively. Thus the task at hand is to optimize h and n, to get the best devolatization of log-returns. This can be done in several different ways, and we will discuss a few of these next.

A rather primitive approach, is to run a loop over different values for h and n, and check the devolatization properties with a Ljung-Box statistic. This approach empirically supports that a larger n gives better devolitization. However, the differences are negligible for values exceeding 40, say, and working with large windows increases computer time.

From a statistical point of view, we would like to set h and n to values that, at the same time, minimize the bias and the variance of the volatility estimator, $\hat{\sigma}_t^2$. This can be done using cross-validation , see [26]. We first decide on a window length, n, and then go from there.

The cross-validation test statistic is given by

$$CV(h) = \sum_{i=1}^{n} \{Y_i^2 - \hat{\sigma}_{t,-i}(h)^2\}^2\,,$$

where $\hat{\sigma}_{t,-i}(\cdot)^2$ is the volatility estimate obtained by omitting the i:th data point. It can be shown that

$$CV(h) = \sum_{i=1}^{n} \{Y_i^2 - \hat{\sigma}_t(h)^2\}^2\, \Xi\,(W_{t_i}(h)),$$

where

$$\Xi(u) = (1-u)^{-2} \qquad \text{and} \qquad W_{t_i}(h) = \frac{K(0)}{\sum_{j=1}^{n} K(\frac{t_j - t_i}{h})}.$$

This means that cross-validation can be considered a residual sum of squares, and all that has to be done, in order to obtain a good value for h, is to minimize this sum of squares.

4.4 Integrated Volatility

Here we are working under a new assumption about the price process. Namely, we assume that the price process is an exponential semimartingale, see [5]. This means that

$$S_t = S_0 \mathrm{e}^{M_t}$$

for a semimartingale M_t.

Remark A semimartingale has a decomposition, $M_t = \alpha_t + m_t$, where α_t has locally bounded variation, and m_t is a local martingale, such that $\alpha_0 = m_0 = 0$. However, as we make quite specific choices of α_t and m_t, there is no need to make formal definitions of these concepts.

Here we assume that the semimartingale, M_t, is given by

$$M_t = \mu t + \beta \sigma_t^{*2} + B_{\sigma_t^{*2}},$$

where $B_{\sigma_t^{*2}}$ denotes Brownian motion, time changed by σ_t^{*2}.

If we build σ_t^{*2} as

$$\sigma_t^{*2} = \int_0^t \sigma_s^2 ds,$$

for a non-negative stochastic process σ_t^2, with the right properties, we can write this model as a *Stochastic Differential Equation* (SDE)

$$dM_t = (\mu + \beta\sigma_t^2)dt + \sigma_t dB_t.$$

The process σ_t^{*2} is called *integrated variance*. The increments

$$\xi_t = \int_{t-1}^t \sigma_s^2 ds,$$

of σ_t^{*2} are called *actual variance*, while σ_t^2 is called *spot variance*.

The solution to the above SDE is

$$M_t = \mu t + \beta \int_0^t \sigma_s^2 ds + \int_0^t \sigma_s dB_s.$$

Using the mentioned model, it follows that log-returns are given by,

$$X_t = \mu + \beta\xi_t + \xi_t^{1/2}\varepsilon_t,$$

where ε_t are i.i.d. $\mathrm{N}(0,1)$-distributed. If we, in the language of section 5 below, assume that $\xi_t \sim \mathrm{GIG}\,(\lambda, \delta, \gamma)$, then the mixing property implies $X_t \sim \mathrm{GH}\,(\lambda, \alpha, \beta, \mu, \delta)$. We can also allow dependency (or, rather, correlation) in ξ_t.

Using the semimartingale model, the devolatized log-returns are given by

$$\varepsilon_t = \frac{X_t - \mu - \beta\xi_t}{\xi_t^{1/2}},$$

and should be i.i.d. $\mathrm{N}(0,1)$. This means that we have to estimate the parameters, μ and β, as well as the increments of the integrated variance, ξ_t. Parameter estimations are done by fitting a Generalized Hyperbolic density to the log-returns (undevolatized). In order to estimate the integrated variance, we first need to introduce the notion of the *quadratic variation* of a stochastic process.

4.4.1 Quadratic Variation

Definition 10 ([32]) The quadratic variation of a process X_t, is defined as the following limit in probability

$$[X]_t = \lim \sum_{i=1}^n (X_{t_i^n} - X_{t_{i-1}^n})^2.$$

Here the limit is taken over partitions

$$0 = t_0^n < t_1^n < \ldots < t_n^n = t,$$

with $\max_{1 \le i \le n}(t_i^n - t_{i-1}^n) \to 0$.

□

Remark An alternative definition of quadratic variation, when X_t is a semimartingale, is, see e.g. [46]:

$$[X]_t = X_t^2 - 2\int_0^t X_{s-} dX_s.$$

Now, if we assume that σ_t^{*2} behaves nicely, see [5], it follows that

$$[M]_t = [\int_0^{\cdot} \sigma_s dB_s]_t = \int_0^t \sigma_s^2 ds = \sigma_t^{*2},$$

and further, trivially, the increments are

$$\Delta[M]_t = [M]_t - [M]_{t-1} = \xi_t.$$

This means that we can estimate ξ_t, using an estimate of the increment of the quadratic variation of the log-price process.

One natural estimator of $\Delta[M]_t$, would be to take log-returns on smaller and smaller intervals. This is done using *intra-day* observations, also called *tick data*, which we define as

$$X_{k,t} = M_{t-1+\frac{k}{N}} - M_{t-1+\frac{k-1}{N}}.$$

Now, it can be shown that, see e.g. [29],

$$[X_N]_t = \sum_{k=1}^{N} X_{k,t}^2$$

converges in probability, to ξ_t (i.e. actual variance), as $N \to \infty$. The sum is called the (*integrated*) *realized variance*. Here N is the number of observations during one day. Further, it can be shown that the convergence rate is $\sqrt{N}$.

What justifies (or spells the doom on) the use of the integrated volatility model, is the availability (or non-availability) of intra-day observations, of the asset in consideration. So to be able to get useful daily observations of the volatility, we need observations of the above mentioned kind. Figure 4.1 show plots of the realized variance (from USD/DM data). The realized variance is based on 144 intra-day log-returns (see section 6.1).

In order to make predictions we will now have to make assumptions about the process σ_t^2 (or in fact ξ_t). These assumptions will be based on the correlation structure and the marginal density.

For the marginal density we propose the *General Inverse Gaussian* (GIG) distribution. It is defined as follows:

Definition 11 A random variable is said to have an GIG$(\lambda, \delta, \gamma)$-*distribution* if its density function is given by

$$f_{\mathrm{GIG}}(x; \lambda, \delta, \gamma) = \frac{(\gamma/\delta)^{\lambda}}{2K_{\lambda}(\delta\gamma)}\, x^{\lambda-1} \exp\left\{ -\frac{1}{2}\,(\delta^2 x^{-1} + \gamma^2 x)\right\}\,,\; x > 0,$$

where $\lambda \in \mathbb{R}$ and $\delta, \gamma \in \mathbb{R}_+$ are constants.

Here, K_λ is the modified Bessel function of the third kind (see section 5.1).

□

Two special cases of the GIG distribution, which we use, are; the *Inverse Gaussian* (IG), and the *Gamma* (Γ) distribution. They are defined in the following way:

Definition 12 A random variable is said to have an IG(δ, γ)-*distribution*, if its density function is given by

$$f_{\mathrm{IG}}(x; \delta, \gamma) = \frac{\delta \mathrm{e}^{\delta\gamma}}{\sqrt{2\pi}}\, x^{-\frac{3}{2}} \exp\left\{ -\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}\,,\; x > 0,$$

where $\delta \in \mathbb{R}_+ \setminus \{\, 0\}$, and $\gamma \in \mathbb{R}_+$ are constants.

□

Definition 13 A random variable is said to have a $\Gamma(\nu, \alpha)$-*distribution* if its density function is given by

$$f_{\Gamma}(x; \nu, \alpha) = \frac{\alpha^{\nu}}{\Gamma(\nu)} x^{\nu-1} \exp\{-\alpha x\}\,,\; x > 0,$$

where $\Gamma(\nu) = \int_0^\infty z^{\nu-1} e^{-z} dz$, while $\alpha \in \mathbb{R}_+$, and $\nu \in \mathbb{R}_+ \setminus \{\, 0\}$ are constants.

□

For considerations in chapter 5, we also state,

Definition 14 A random variable is said to have an PH(δ, γ)-*distribution* if its density function is given by

$$f_{\mathrm{PH}}(x; \delta, \gamma) = \frac{(\gamma/\delta)}{2K_1(\delta\gamma)} \exp\left\{ -\frac{1}{2}\,(\delta^2 x^{-1} + \gamma^2 x)\right\},$$

where $\delta \in \mathbb{R}_+ \setminus \{\, 0\}$, and $\gamma \in \mathbb{R}_+$ are constants.

□

Remark The IG(δ, γ)-distribution and the $\Gamma(\nu, \alpha)$-distribution coincide with the GIG$(-\frac{1}{2}, \delta, \gamma)$-distribution, and the GIG$(\nu, 0, \gamma)$-distribution, respectively, where $\alpha = \gamma^2/2$.

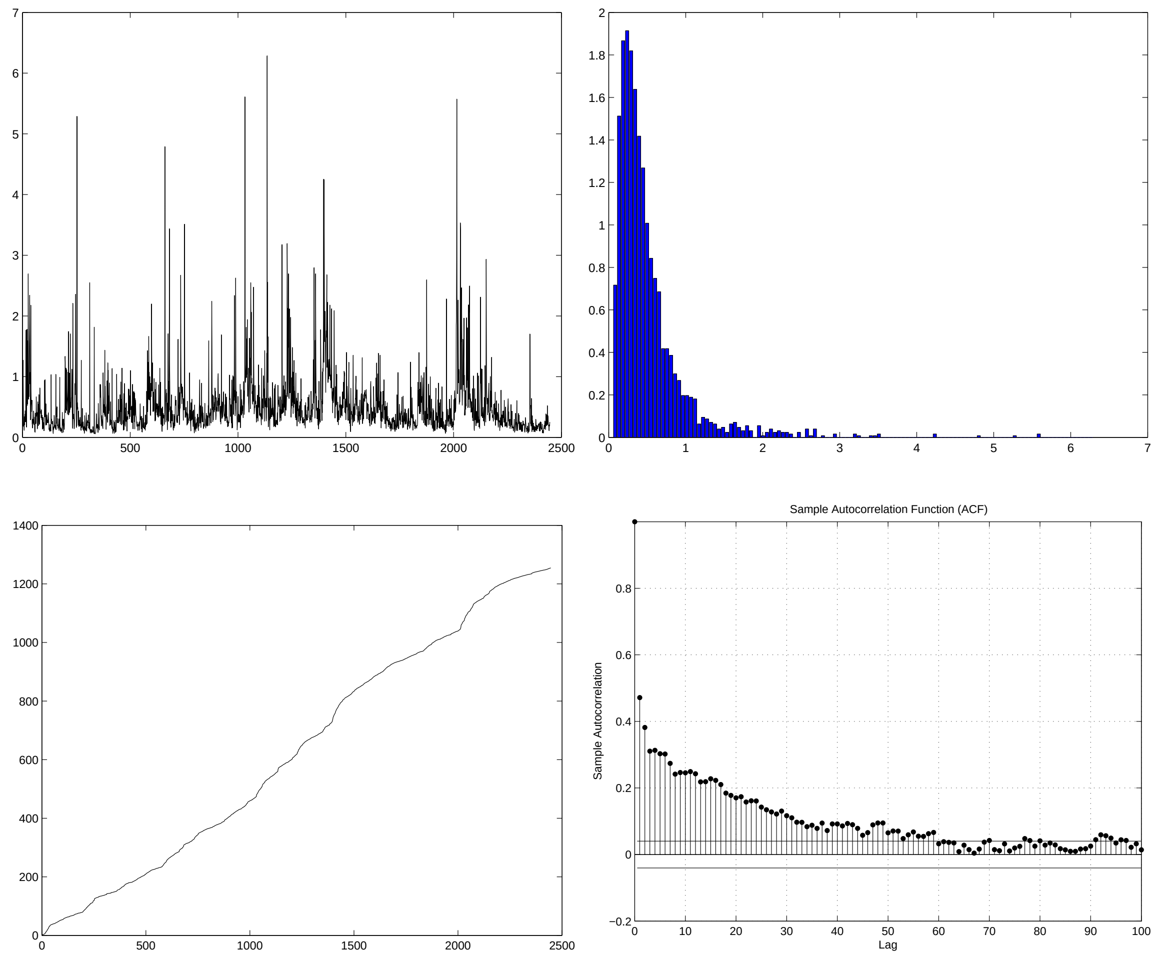


Figure 4.1: *Realized variance (top left), histogram of realized variance (top right), aggregated realized variance (i.e. the chronometer) (bottom left) and* sacf *of the realized variance (bottom right).* (USD/DM)

The parameters of the IG and Gamma distributions, are estimated using ML. As displayed, in Figure 4.2 below, an IG distribution seems to fit the USD/DM realized volatility data, much better than a Gamma. The parameter estimates are: $\delta = 0.8472$, $\gamma = 1.6518$, $\nu = 1.8711$ and $\alpha = 3.6481$.

When it comes to the correlation structure, it seems reasonable to use processes with acf:s of exponential damp-down type (i.e. $\mathrm{acf}_1(s) = \mathrm{e}^{-\alpha|s|}$).

The sacf in figure 4.1, indicates that one such exponential term is not enough, so we need to superpose two (or more) independent processes of this type. That is, we are looking for an acf of the following kind

$$\mathrm{acf}_k(s) = \sum_{i=1}^{k} m_i \mathrm{e}^{-\alpha_i|s|},$$

where $\sum_{i=1}^{k} m_i = 1$. We will denote this process P_k.

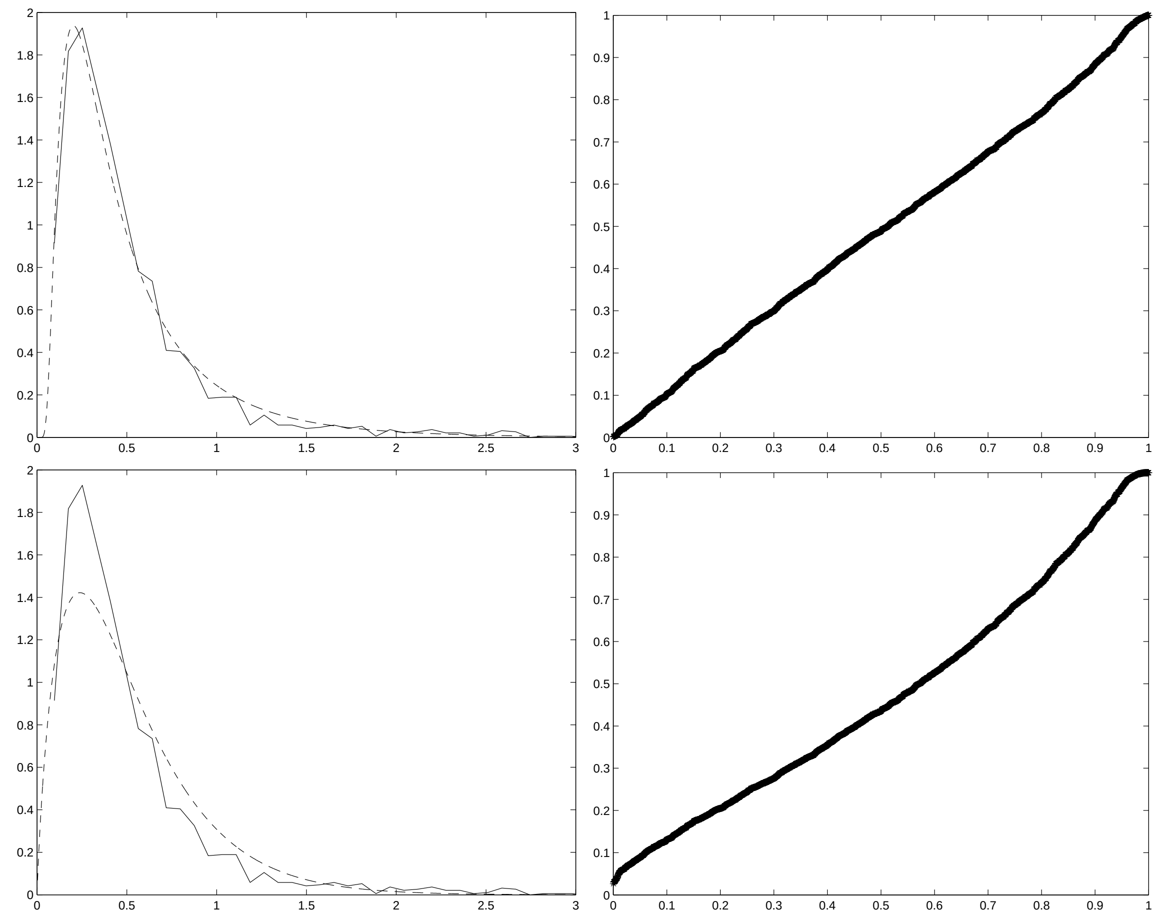


Figure 4.2: IG *(top) and* Γ *(bottom) densities estimated from USD/DM realized variance data (the jagged line represents the empirical density), together with the corresponding PP-plots.*

For only two processes, this simplifies to

$$\mathrm{acf}_2(s) = m\mathrm{e}^{-\alpha_1|s|} + (1-m)\mathrm{e}^{-\alpha_2|s|}.$$

The parameters $(m_i, \alpha_i,\ i = 1, 2, ..., k)$ are estimated by minimizing

$$\sum_{i=1}^{n} \left[\mathrm{sacf}(i) - \mathrm{acf}_k(i)\right]^2,$$

for n large enough.

The parameter estimates (for $k = 1, 2, 3$), are shown in table 4.1. The need for at least two processes is clear. Making the model more complex, by adding a third process, does not seem to enhance the fit to the correlation structure. However, one should be aware that this does not necessarily carry over to the dependency structures. In figure 4.3, the estimated acf using three processes, each with acf of exponential damp-down type, is almost indistinguishable from the acf with only two processes. It is clear though, that one is not enough.

Model	$\hat{m}_1$	$\hat{m}_2$	$\hat{m}_3$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
P_1	1.00			0.1694		
P_2	0.6499	0.3501		1.5192	0.0343	
P_3	0.0766	0.5734	0.3499	1.4174	1.5331	0.0343

Table 4.1: *Estimated parameters for autocorrelation functions of* P_k *processes,* $k = 1, 2, 3$*, (USD/DM volatility data).*

We will proceed under the assumption, that the increments of the realized volatility is the sum of two independent processes. Both having acf:s consisting of exponential damp-downs.

4.4.2 Ornstein-Uhlenbeck (OU) Processes

To get a process, with the above mentioned properties, we assume that ξ_t satisfies

$$d\xi_t = -\alpha\xi_t dt + dZ_{\alpha t}.$$

Here the process $Z_{\alpha t}$ is a Lévy process, with non-negative increments, the so called BDLP (*Background Driving Lévy Process*), which is a subordinator. The above SDE is a generalization, see e.g. [3], of the *Langevin equation*, see [36],

$$dU_t = -\beta U_t dt + \sigma dB_t, \quad t > 0,$$

Here B_t is Brownian motion, $\sigma \in \mathbb{R}$ and $\beta > 0$ are constants. This equation has solution

$$U_t = \mathrm{e}^{-\beta t} U_0 + \int_0^t \sigma \mathrm{e}^{-\beta(t-s)} dB_s,$$

which is known as the (Gaussian) Ornstein-Uhlenbeck (OU) process. This is the only stochastic process that is simultaneously Gaussian, Markov and stationary.

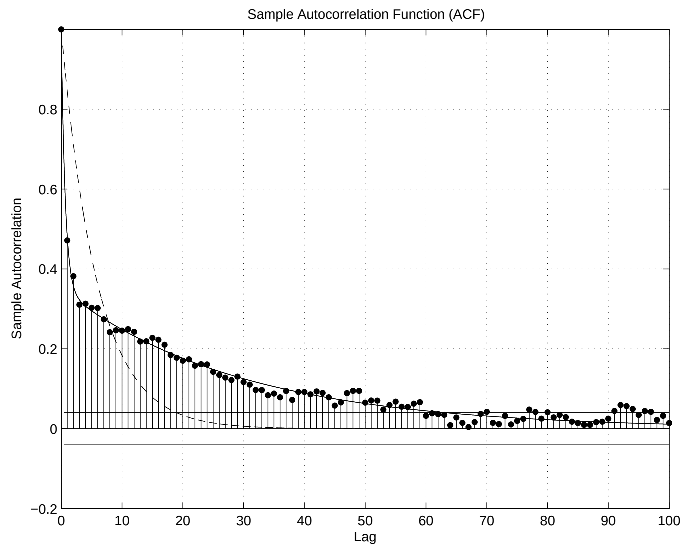


Figure 4.3: Sacf *for USD/DM volatility data and superpositioning of acfs from one (–), two (-) and three (-) independent processes.*

Theorem 2 For an Ornstein-Uhlenbeck process

$$\sigma_t^2 = \int_{-\infty}^{t} \mathrm{e}^{-\alpha(t-s)} dZ_s,$$

the correlation function is given by

$$\mathrm{acf}_1(s) = \mathrm{e}^{-\alpha|s|}.$$

□

The, above mentioned, Gaussian OU process is not applicable in our situation, since its driving process (i.e. the Brownian motion) can have negative increments, which implies that U_t can take negative values.

However, we can give the OU process a GIG marginal distribution, by choosing an appropriate BDLP. As mentioned we use OU processes with either a Γ-marginal or an IG-marginal.

Theorem 3 ([10]) The BDLP of an OU process with $\mathrm{IG}(\delta, \gamma)$ marginal distribution (IG-OU), is the compound Poisson process given by

$$Z_t = \gamma^{-2} \sum_{k=0}^{N_t} u_k^2 + Q_t,$$

where u_k are i.i.d. $\mathrm{N}(0,1)$ and independent of the Poisson process N_t. Further, N_t has rate $2/\delta\gamma$ and Q_1 is $\mathrm{IG}(\delta/2, \gamma)$-distributed.

□

Theorem 4 ([10]) The BDLP of an OU process with $\Gamma(\nu, \alpha)$ marginal distribution (Γ-OU), is the compound Poisson process given by

$$Z_t = \sum_{k=0}^{N_t} r_k,$$

where r_k are i.i.d. $\Gamma(1, \alpha)$, and N_t is a Poisson process with rate ν.

□

4.4.3 Superpositions of OU Processes

To capture the long range dependence in the realized variance process we can, as mentioned above, use a superposition of OU processes. That is, we assume that

$$\tau_t = \sum_{i=1}^{k} \tau_t^{(i)}.$$

Here, for each $i = 1, ..., k$, $\tau_t^{(i)}$, is an OU process, i.e. a solution to the SDE

$$d\tau_t^{(i)} = -\alpha_i \tau_t^{(i)} dt + dZ_{\alpha_i t}^{(i)},$$

where $Z_{\alpha_i t}^{(i)}$ are independent, not necessarily identically distributed, BDLP:s. We will denote τ_t by OU_k.

Both the IG and the Γ distributions are closed under convolution. More specifically, with obvious notation, $\sum_{i=1}^k \mathrm{IG}(\delta_i, \gamma) = \mathrm{IG}(\sum_{i=1}^k \delta_i, \gamma)$ and $\sum_{i=1}^k \Gamma(\delta_i, \gamma) = \Gamma(\sum_{i=1}^k \delta_i, \gamma)$. Therefore, fitting a superposition of IG-OU or Γ-OU processes, will generate a process with IG, or Gamma distributed univariate marginals, respectively. However, the resulting process will typically not be an OU process.

So, if we assume that

$$\tau_t^{(i)} \sim \mathrm{IG}(\delta_i, \gamma),$$

where $\sum_{i=1}^k \delta_i = \delta$, the resulting process, τ_t, will have $\mathrm{IG}(\delta, \gamma)$-distributed marginals. Since, see e.g. [5], $\mathrm{Var}(\tau_t^{(i)}) = \delta_i/\gamma^3$, the acf for this process is given by

$$\mathrm{acf}_k(s) = \mathrm{Corr}(\tau_{t+s}, \tau_t) = \frac{\sum_{i=1}^k \mathrm{Cov}(\tau_{t+s}^{(i)}, \tau_t^{(i)})}{\mathrm{Var}(\tau_t)} = \sum_{i=1}^{k} \frac{\delta_i}{\delta} \mathrm{e}^{-\alpha_i |s|}.$$

We will call τ_t an IG-OU_k process.

Remark For a Γ-OU process, there are similar results. If $\tau_t^{(i)} \sim \Gamma(\nu_i, \alpha)$, where $\sum_{i=1}^k \nu_i = \nu$, the resulting process will satisfy $\tau_t \sim \Gamma(\nu, \alpha)$. ($\Gamma$-$\mathrm{OU}_k$ processes).

Now, if we set each $\delta_i = m_i \delta$, where $\sum_{i=1}^k m_i = 1$, it follows trivially that

$$\mathrm{acf}_k(s) = \sum_{i=1}^{k} m_i \mathrm{e}^{-\alpha_i |s|},$$

and further, that $\tau_t \sim \mathrm{IG}(\delta, \gamma)$.

Using this result, and the results shown in table 4.1, we can assume that the USD/DM data (see section 6.1), is the sum of two IG-OU processes, having marginal densities $\mathrm{IG}(0.6499\delta, \gamma)$, and $\mathrm{IG}(0.3501\delta, \gamma)$, respectively. This gives an acf

$$\mathrm{acf}_2(s) = 0.6499\mathrm{e}^{-1.5192|s|} + 0.3501\mathrm{e}^{-0.0343|s|},$$

which, as seen in figure 4.3, seems to fit the correlation structure very well.

For an IG-OU process, $\tau_t^{(i)}$, the conditional distribution of $\tau_t^{(i)}$ given $\tau_0^{(i)}$ (i.e. $\tau_t^{(i)}|\tau_0^{(i)}$) has mean, see also [5],

$$\mathbb{E}[\tau_t^{(i)}|\tau_0^{(i)}] = \mathrm{e}^{-\alpha_i t}\tau_0^{(i)} + \delta_i\gamma\left(1 - \mathrm{e}^{-\alpha_i t}\right).$$

If we were able to observe today's value of both processes, $\tau_0^{(1)}$ and $\tau_0^{(2)}$, we could use this quantity to predict tomorrow's volatility.

It is easily seen that, for a sum, τ_t, of two OU processes, $\tau_t^{(1)}$ and $\tau_t^{(2)}$, it follows that

$$\eta_t = \tau_t - \mathrm{e}^{-\alpha_1}\tau_{t-1} - \mathrm{e}^{-\alpha_2}\left[\tau_{t-1} - \mathrm{e}^{-\alpha_1}\tau_{t-2}\right] \text{ , for } t \text{ even,}$$

is an i.i.d. sequence.

Unfortunately, when transforming our observations to η_t, in this manner, the BDS-test (see section 6.3) indicates that the i.i.d. property is not fulfilled. Hence the data does not seem to come from a superposition of two independent OU processes.

4.4.4 Integrated OU processes

Following O.E. Barndorff-Nielsen and N. Shepard, [6], we assume the spot variance, σ_t^2, to be an OU process.

So, we assume that σ_t^2 is a solution to

$$d\sigma_t^2 = -\alpha\sigma_t^2 dt + dZ_{\alpha t}.$$

Then it follows that, see [7]

$$\sigma_t^{*2} = \int_0^t \sigma_s^2 ds = \alpha^{-1}(1 - \mathrm{e}^{-\alpha t})\sigma_0^2 + \alpha^{-1}\int_0^t \{1 - \mathrm{e}^{-\alpha(t-s)}\} dZ_{\alpha s}.$$

Curiously, though σ_t^2 might have jumps, the integrated variance σ_t^{*2} has continuous sample paths.

Further, it can be shown that, see [5], the increments of the process σ_t^{*2}, ξ_t (i.e. actual variance), have an acf with exponential decay. That is,

$$\mathrm{Corr}(\xi_{t+s}, \xi_t) = \mathrm{e}^{-\alpha|s|},$$

which is precisely what we are looking for.

If we assume σ_t^2 to be an IG-OU process, it follows, see [5], that σ_t^{*2} also has IG marginal density. We can then, by the same arguments as before, superpose several integrated OU processes (intOU), $(\sigma_t^{*2})^{(i)}$, and get a resulting process having the right marginal density and acf

$$\mathrm{acf}_k(s) = \sum_{i=1}^{k} m_i \mathrm{e}^{-\alpha_i |s|}.$$

As in the OU case, the conditional expectation seems to be a natural predictor. For the intOU process, the conditional expectation is, see [6]

$$\mathbb{E}[\sigma_t^{*2} | \sigma_0^{*2}] = \alpha^{-1} \left(1 - \mathrm{e}^{-\alpha t}\right) \sigma_0 + \alpha^{-1} \delta \gamma \left(\alpha t - 1 + \mathrm{e}^{-\alpha t}\right),$$

where σ_0^{*2} denotes today's value. If we assume σ_t^{*2} to be a superposition of, say, two independent intOU processes, this predictor also requires that we can observe today's value for both of the underlying processes.

To solve the problems of prediction that we have run into, we will use today's value as an estimator for tomorrow's integrated volatility.

Chapter 5

Distributions

5.1 Generalized Hyperbolic Distribution

The GH distribution was introduced by Ole Barndorff-Nielsen, [8], in order to model the spread of windblown sands. We use it to model both log-returns (see section 4.4), and devolatized log-returns, of financial assets.

Definition 15 The density of a $GH(x;\lambda,\alpha,\beta,\mu,\delta)$-distributed random variable is given by

$$\begin{aligned} f_{\mathrm{GH}}\left(x;\lambda,\alpha,\beta,\mu,\delta\right) &= \frac{\left(\alpha^2-\beta^2\right)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\left(\delta^2+(x-\mu)^2\right)^{\left(\lambda-\frac{1}{2}\right)/2} \\ &\times \; K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\exp\left\{\beta\left(x-\mu\right)\right\}\,,\; x\in\mathbb{R}, \end{aligned}$$

where the K_λ is the modified Bessel function of the third kind, given by

$$K_\lambda(x)=\frac{1}{2}\int_0^\infty y^{\lambda-1}\mathrm{e}^{-\frac{1}{2}x\left(y^{-1}+y\right)}dy.$$

The domains of the parameters are

$$\begin{aligned} &\delta\geq 0,\quad |\beta|<\alpha \quad \text{if} \quad \lambda>0 \\ &\delta>0,\quad |\beta|<\alpha \quad \text{if} \quad \lambda=0 \\ &\delta>0,\quad |\beta|\leq\alpha \quad \text{if} \quad \lambda<0 \end{aligned}$$

□

The plots in figure 5.1 and figure 5.2, illustrate the fit of the GH distribution and the Gaussian distribution to the DAX data. There is no doubt that, at least when compared to the Gaussian distribution, the GH distribution displays a superior fit.

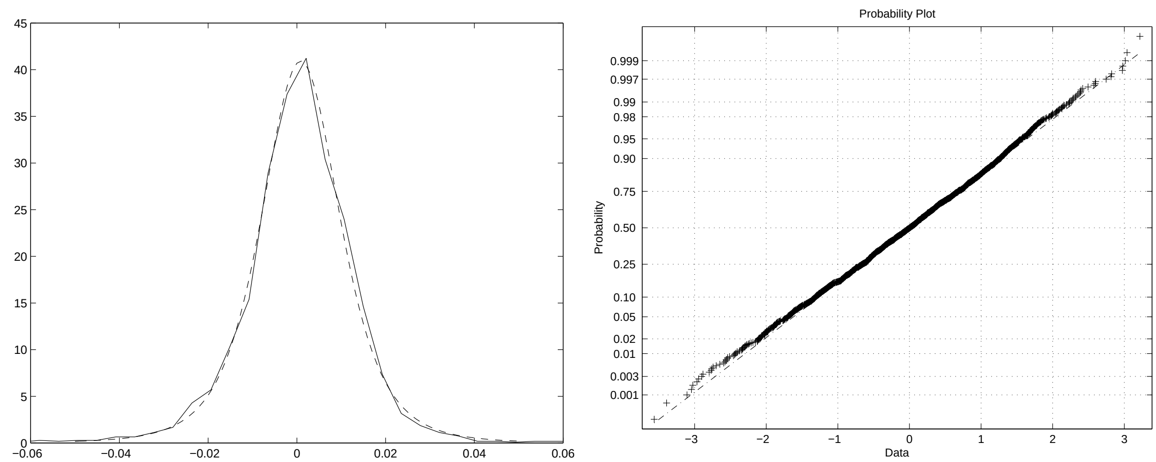


Figure 5.1: *Fit of* GH *distribution (–) to* DAX *data (-).*

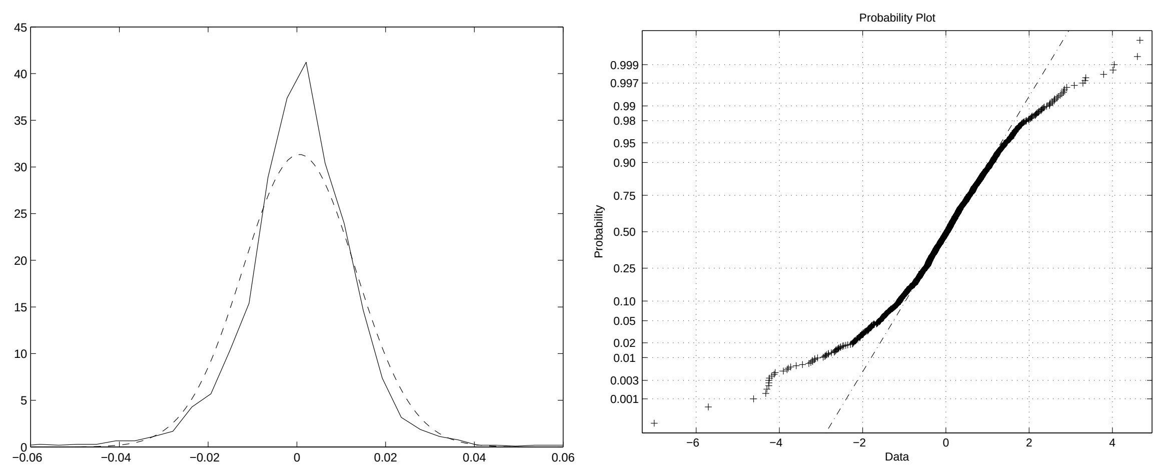


Figure 5.2: *Fit of Gaussian distribution (–) to* DAX *data (-).*

If unfamiliar with the Generalized Hyperbolic distribution, one might ask, why all the parameters, and what happens if we vary them? And what role does the Bessel function play?

The answer to; why all the parameters? is, of course, flexibility. Using the GH-model, we can get tails that are heavy enough to capture most extremal events. We can get skewness and illustrate the so called *leverage effect*, which is, when a stock falls the volatility seems to increase, which in turn gives a heavier left tail.

When it comes to the parameters, λ, is the index of the Bessel function, and affects the weight of the tails (scaling), μ is a location parameter, β a skewness parameter, and both α and δ are scale parameters. The plots in figure 5.3 below, demonstrate the effect of changing one parameter at a time, *ceteris paribus*[1].

[1] all else equal

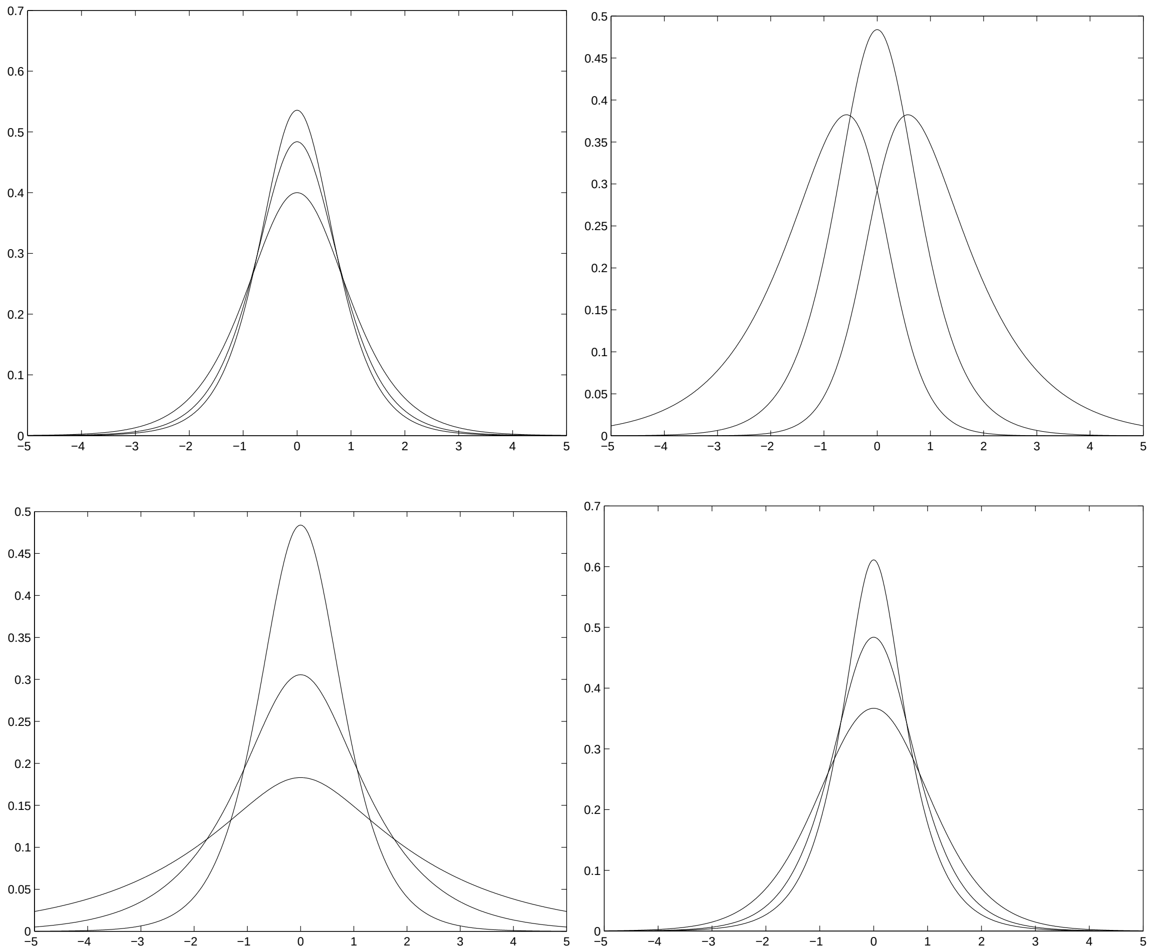


Figure 5.3: *Top left:* $\lambda = 0.5$ *(high),* $\lambda = 1$ *and* $\lambda = 2$ *(low). Top right:* $\beta = -1$ *(left),* $\beta = 0$ *and* $\beta = 1$ *(right). Bottom left:* $\alpha = 0.5$ *(high),* $\alpha = 1$ *and* $\alpha = 2$ *(low). Bottom right:* $\delta = 0.5$ *(high),* $\delta = 1$ *and* $\delta = 2$ *(low)*

A $\mathrm{GH}(\lambda, \alpha, \beta, \mu, \delta)$-distributed random variable has mean and variance, see [44],

$$\mathbb{E}[X] = \mu + \frac{\delta\beta K_{\lambda+1}(\zeta)}{\gamma K_\lambda(\zeta)}$$

$$\mathrm{Var}(X) = \frac{\delta K_{\lambda+1}(\zeta)}{\gamma K_\lambda(\zeta)} + \left(\frac{\beta\delta}{\gamma}\right)^2 \Big[\frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \frac{K_{\lambda+1}^2(\zeta)}{K_\lambda^2(\zeta)}\Big],$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $\zeta = \delta\gamma$. For the special cases Hyperbolic and NIG the mean simplifies to

$$\mathbb{E}[X] = \mu + \frac{\delta\beta}{\gamma},$$

and in the NIG case, it follows that

$$\mathrm{Var}(X) = \delta^2 \Big[\frac{1}{\zeta} + \left(\frac{\beta}{\gamma}\right)^2 \frac{1}{\zeta+1}\Big] \frac{\zeta}{\zeta+1}.$$

5.1.1 The Mixing Property

As we have seen, we need the fact that a GH distribution can be expressed as a mean-variance mixture, between a Generalized inverse Gaussian (GIG) distribution and a Gaussian distribution, in the sense that, see [44]

$$f_{\mathrm{GH}}(x;\lambda,\alpha,\beta,\mu,\delta)=\int_0^\infty f_{\mathrm{N}}\left(x;\mu+\beta t,t\right)f_{\mathrm{GIG}}\left(t;\lambda,\delta^2,\alpha^2-\beta^2\right)dt,$$

where f_{N} is the density function of a Gaussian random variable with mean $\mu+\beta t$ and variance t, and f_{GIG} is the density of a GIG distributed random variable.

Equivalently we can express the mixing property by, see [5],

$$X=\mu+\beta\sigma^2+\sigma\varepsilon,$$

where, for $\alpha=\sqrt{\beta^2+\gamma^2}$, $X\sim\mathrm{GH}(\lambda,\alpha,\beta,\mu,\delta)$, $\sigma^2\sim\mathrm{GIG}(\lambda,\delta,\gamma)$ and $\varepsilon\sim\mathrm{N}(0,1)$, with σ^2 and ε independent.

It can be shown that the mean-variance mixture of an $\mathrm{IG}(\delta,\gamma)$-distribution and a standard Gaussian gives the *Normal Inverse Gaussian*, $\mathrm{NIG}(\alpha,\beta,\mu,\delta)$-*distribution*, whereas the variance-mean mixture of an $\Gamma(\nu,\gamma^2/2)$-distribution and a standard Gaussian gives the *Normal Gamma*, $\mathrm{N}\Gamma(\nu,\gamma,\beta,\mu)$-*distribution*. Furthermore a mixture of a $\mathrm{PH}(\delta,\gamma)$-distribution and a standard Gaussian gives the *Hyperbolic*, $\mathrm{H}(\alpha,\beta,\mu,\delta)$-*distribution*, where $\alpha=\sqrt{\beta^2+\gamma^2}$.

Definition 16 The density function of the $\mathrm{NIG}(\alpha,\beta,\mu,\delta)$-distributed random variable is given by

$$f_{\mathrm{NIG}}(x;\alpha,\beta,\mu,\delta)=\frac{\alpha}{\pi}\exp\left\{\delta\sqrt{\alpha^2-\beta^2}-\beta\mu\right\}q\left(\frac{x-\mu}{\delta}\right)^{-1}K_1\left(\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right)\mathrm{e}^{\beta x},$$

where

$$q(x)=\sqrt{1+x^2}\quad\text{and}\quad\alpha=\sqrt{\beta^2+\gamma^2}.$$

□

Definition 17 The density function of the $\mathrm{N}\Gamma(\nu,\gamma,\beta,\mu)$-distributed random variable is given by

$$f_{\mathrm{N}\Gamma}(x;\nu,\gamma,\beta,\mu)=\frac{\gamma^{2\nu}\left(\gamma^2/2\right)^{1-2\nu}}{\sqrt{2\pi}\delta\Gamma\left(\nu\right)2^{\nu-1}}x^{\nu-\frac{1}{2}}K_{\nu-\frac{1}{2}}\left(\frac{\gamma^2}{2}|x-\mu|\right)\exp\left\{\beta\left(x-\mu\right)\right\}.$$

□

Definition 18 The density function of the $\mathrm{H}(\alpha,\beta,\mu,\delta)$-distributed random variable is given by

$$f_{\mathrm{H}}(x;\alpha,\beta,\mu,\delta)=\frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta\left(x-\mu\right)\right\}.$$

□

Remark It is readily shown that

$$\mathrm{NIG}(\alpha, \beta, \mu, \delta) = \mathrm{GH}(-\frac{1}{2}, \alpha, \beta, \mu, \delta),$$
$$\mathrm{N\Gamma}(\nu, \gamma, \beta, \mu) = \mathrm{GH}(\nu, \alpha, \beta, \mu, 0)$$

and

$$\mathrm{H}(\alpha, \beta, \mu, \delta) = \mathrm{GH}(1, \alpha, \beta, \mu, \delta).$$

Remark From a simulation point of view, the mixing properties of the various distributions are most interesting. If we consider a price process, for which the log-returns have GH marginal distribution, the mixing relations provide us with simple methods of simulating that process, given that we have simulated the OU volatility process. Simulation of an OU volatility process in turn, is done from the BDLP, so all we really have to do, to simulate a price process with GH marginals, is to generate the BDLP of a GIG-OU process. Below we see a simulation of an exponential NIG$(1.8244, -0.02, -0.0069, 0.9117)$-process, where the parameter values come from fitting a NIG distribution to USD/DM data.

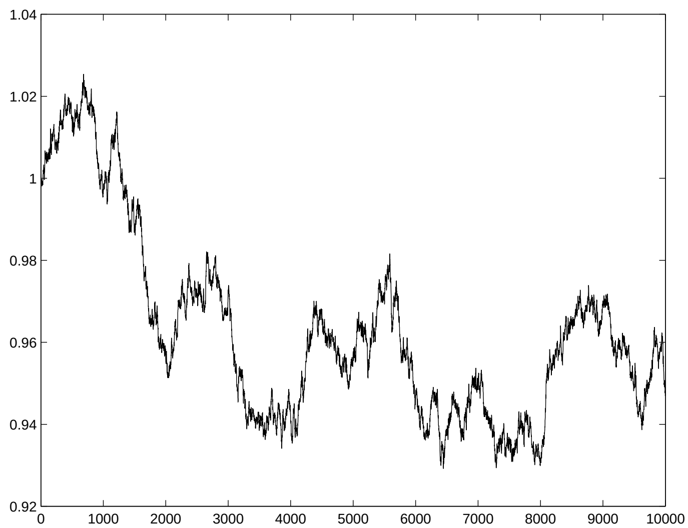


Figure 5.4: Realization of an exponential NIG process.

Fitting log-returns to the GH distribution, along with special cases of the GH distribution, is done by ML. Due to the complexity of the densities involved, this is done numerically.

Remark It should be noted that we use the GH distribution, along with special cases of the GH distribution, in two different contexts: The first is to fit distributions of non-devolatized log-returns, and the second is to fit distributions of devolatized log-returns.

5.2 Gaussian Distribution

Since it is the easiest to work with, and the one that is most often used in finance, the Gaussian is the distribution against which all others are compared. The Gaussian distribution is the cornerstone of the Bachelier-Samuelson model, and hence the price process model upon which the Black-Scholes option pricing theory is built, see [13].

Remark The Gaussian distribution can be obtained from the GH as

$$\mathrm{N}(\mu, \sigma^2) = \lim_{\gamma \to \infty} \mathrm{GH}(\lambda, \gamma, 0, \mu, \sigma^2\gamma).$$

5.3 Pearson VII Distribution

This distribution may not be well known, but it seems to fit financial log-returns quite well, see also [49]. It coincides with a mixture of two GH distributions. The Pearson VII distribution enables us to model the left and the right tail of the log-returns, each at a time. This can be used to model skewness, that often is present in financial data. The density function for the whole real line is given by

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

The positive and negative sides, of the density function, of a Pearson VII distributed random variable, are both given by

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

Here m is a shape parameter, and c a scale parameter. These parameters are estimated using numerical maximum likelihood.

Remark The density function $f(x; m, c)$, is the density function of a *Student-t* distributed random variable, with $\nu = 2m - 1$ degrees of freedom, multiplied by the scale parameter $c\nu^{-1/2}$. This gives a connection to the Gaussian distribution, since Student-t (with n degrees of freedom) converges weakly to the Gaussian distribution as $n \to \infty$.

One "ugly" feature of the Pearson VII density is that you are most likely to get a discontinuity at zero when you "glue" the positive and negative parts together.

Remark Either side of the Pearson VII distribution can be obtained from the GH as $\mathrm{P}_+^{VII}(x; m_+, c_+) = \mathrm{GH}(m_+ + \frac{1}{2}, 0, 0, 0, c_+)$ or $\mathrm{P}_-^{VII}(x; m_-, c_-) = \mathrm{GH}(m_- + \frac{1}{2}, 0, 0, 0, c_-)$. This indicates, by mixing properties , see [5], that if, say, the positive log-returns are distributed $\mathrm{P}_+^{VII}(x; m_+, c_+)$, as above, then the volatility of the positive returns have a Reciprocal Gamma $\mathrm{R}\Gamma(m_+ - \frac{1}{2}, \frac{c_+^2}{2})$ distribution, for which the density is given by

$$f_{\mathrm{R}\Gamma}(x; \nu, \alpha) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{-\nu-1} e^{-\frac{\alpha}{x}}, \ x > 0.$$

Below we see a fit of the Pearson VII density to the log-returns of Siemens stock price series.

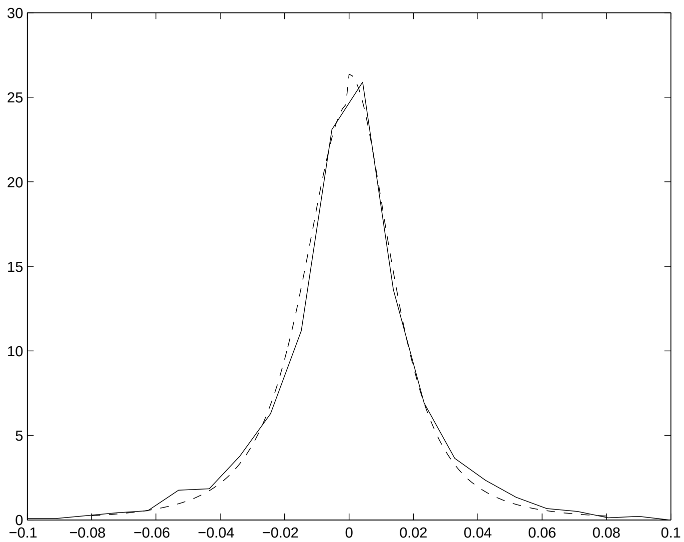


Figure 5.5: *Pearson* VII *density (–) fitted to* Siemens *data (-)*

5.4 Generalized Pareto Distribution

In this section we consider an extreme value model, that uses Generalized Pareto (GP) distribution. First we look at some preliminaries from extreme value theory. See e.g. [21] as a general reference, for the material on extremes that we present here.

For a sequence, $X_1, X_2, ...$, of independent and identically distributed random variables, with distribution function F, we are interested in the variables that exceed some prescribed level u. Expressed in terms of equations, we are dealing with the following

$$\mathbb{P}\left(X > u + y \,|\, X > u\right) = \frac{1 - F\left(u + y\right)}{1 - F\left(u\right)}, \quad y > 0.$$

That is, given the excess of a high threshold (level) u, what is the probability that this excess is larger than y? Had we known F, we would have been done. However, this is not the case in practice, and we need some theory to get around this obstacle.

Theorem 5 Let $X_1, X_2, ...$ be an i.i.d. sequence with common distribution F, and let

$$M_n = \max\{X_1, ..., X_n\}.$$

Let X an arbitrary term in the X_i sequence, and suppose that F is such that, for some sequences of constants $(a_n > 0)$ and (b_n), we have

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \quad \text{as} \quad n \to \infty,$$

(where G is non-degenerate), so that for n large

$$\mathbb{P}\left(M_n \leq z\right) \approx G(z).$$

Then we have

$$G(z) = \exp\left\{-\left(1+\xi\frac{z-\mu}{\sigma}\right)^{-1/\xi}\right\}$$

for some $\mu, \sigma > 0$ and $\xi \in \overline{\mathbb{R}}$.

□

We will, though, assume that $\xi \neq \pm\infty$. For large enough u, the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi},$$

defined on $\{y : y > 0 \text{ and } (1+\frac{\xi y}{\tilde{\sigma}}) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi\,(u-\mu) \qquad (*).$$

The family of distributions given by H , is known as the *Generalized Pareto family*, and seems to be well suited for our purposes, as will be seen below.

So, working with the GP model, we have

$$\mathbb{P}(X > x\,|X > u) = \left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}, \; x > u.$$

In other words, the distribution function is

$$F(x) = 1 - \zeta_u\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}, \; x > 0,$$

where $\zeta_u = P(X > u)$, ξ is a shape parameter, and σ is a scale parameter.

Remark Observations of the event $(X > u)$, are Binomial(n, p) distributed, with $p = \zeta_u$. The probability ζ_u, is thus estimated by $\hat{\zeta}_u = k/n$, where k is the number of excesses of u among n observations.

One nice feature of the GP distribution is the existence of a closed expression for the inverse,

$$x = F^{-1}(\alpha) = u + \frac{\sigma}{\xi}\left[\left(\frac{\zeta_u}{\alpha}\right)^{\xi} - 1\right].$$

This, as we have seen, comes in handy when calculating VaR.

The log-likelihood function is

$$l(\zeta_u, \xi, \sigma; x_i) = \log\zeta_u - \log\sigma - \left(\frac{1}{\xi}+1\right)\log\left(1+\xi\left(\frac{x_i-u}{\sigma}\right)\right).$$

However, before making ML estimations, a main issue is to choose an appropriate threshold. Of course, one wants as low a threshold as possible, to minimize the variance of parameter estimates. But choosing the threshold too low, might lead to bias problems, because of inclusion of values that are not really "extreme". The choice of threshold can be made in several ways, and we work along the lines of Coles' book , see [21]. It uses a so called mean residual life plot, which is a method based on the mean of a GP distributed random variable.

If a random variable, Y, has GP distribution, with parameters σ and ξ, then

$$E(Y) = \frac{\sigma}{1-\xi},$$

for $\xi < 1$. Otherwise, the mean is infinite. If we assume that the GP distribution is a valid model for excesses of a threshold σ_{u_0}, then we have that

$$E(X - u_0 \,|\, X > u_0) = \frac{\sigma_{u_0}}{1-\xi},$$

again given that $\xi < 1$. It follows from $(*)$ that

$$E(X - u \,|\, X > u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi\,(u - u_0)}{1-\xi},$$

which implies that $E(X - u \,|\, X > u)$ is a linear function in u. This expression, of course, also is the mean of the excesses of u, and can be approximated by the sample mean of the excesses of u. The above reasoning implies that these estimates depend linearly on u, at the levels at which the GP model is appropriate. So we look at the following:

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} \left(x_{(i)} - u\right) \right) : u < x_{\max} \right\},$$

where $x_{(1)} ... x_{(n_u)}$ are the ordered observations of excesses of the threshold u. This gives the mean residual life plot, and we choose our threshold at the lowest u at which the plot becomes linear. Hence, we obtain the largest number of observations, for which the model is valid.

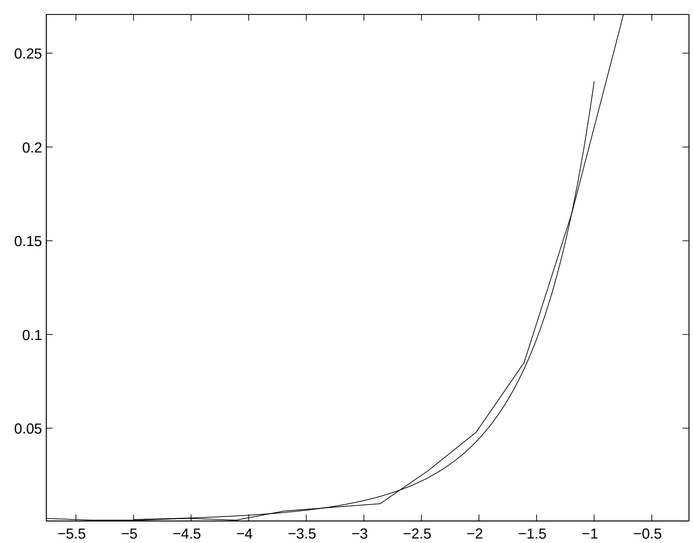


Figure 5.6: *Fit of a GP density (smooth line) to devolatized returns of a* DAX *data (jagged line). The devolatization is done using the Nadaraya-Watson estimator.*

Chapter 6

Test of Models

We have now come to the evaluation of our models. There are three main aspects that we wish to analyze.

- Are the devolatized log-returns independent?
- How well do the marginal densities fit empirical data?
- How good are the VaR estimates?

First we take a look at the data sets we are working with.

6.1 Data

We have used three data sets, for our price process S_t, $t \geq 0$:

- USD/DM; from 1 December 1986, and 2300 trading days forward. This data set also contains realized variances.
- DAX[1]; From 5 November 1991 to 15 January 1999.
- Siemens; From 29 Mars 1995 to 23 May 2002.

Concerning the availability of the used data, the DAX and Siemens data can be downloaded from e.g. Yahoo Finance. Also the USD/DM daily observations can be found here. The USD/DM observations were kindly made available to us by Erik Brodin (ECMI, Chalmers University of Technology).

The data sets are displayed in the following pages, and the plots are organized as follows: The top plot shows S_t/S_0, the middle plot shows log-returns (i.e $X_t = \log(S_t/S_{t-1})$) and the bottom one shows the devolatized log-returns (i.e. A_t). The realized variance data is displayed in figure 4.1, see section 4.4.

[1] der Deutsche Aktieindex

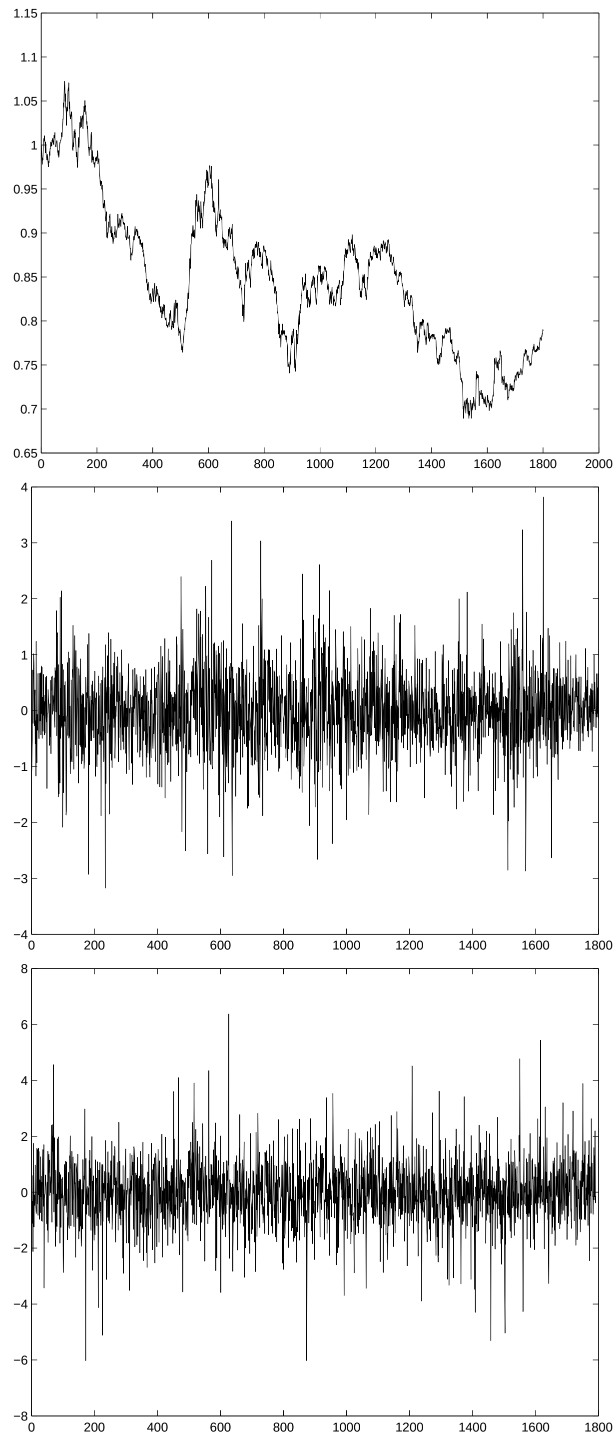


Figure 6.1: *USD/DM data. Top plot: Price process, middle: Log-returns, bottom: Log-returns, devolatized using Variance Window volatility.*

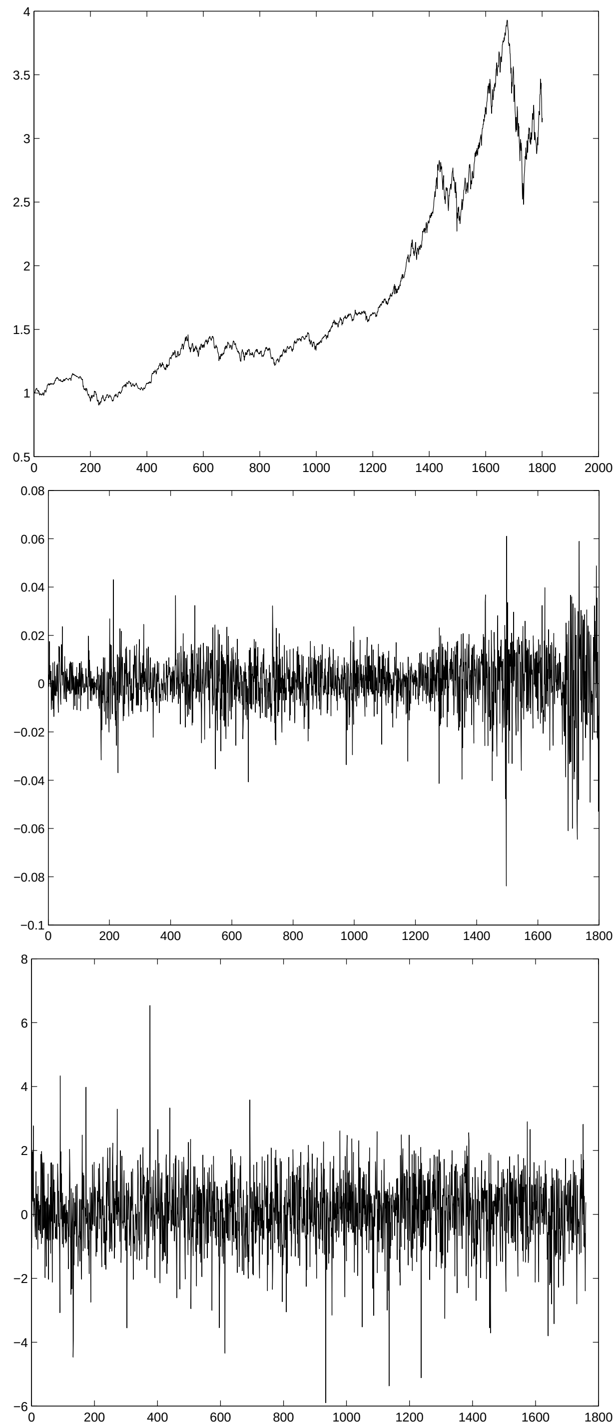


Figure 6.2: *DAX data. Top plot: Price process, middle: Log-returns, bottom: Log-returns, devolatized using Nadaraya-Watson volatility.*

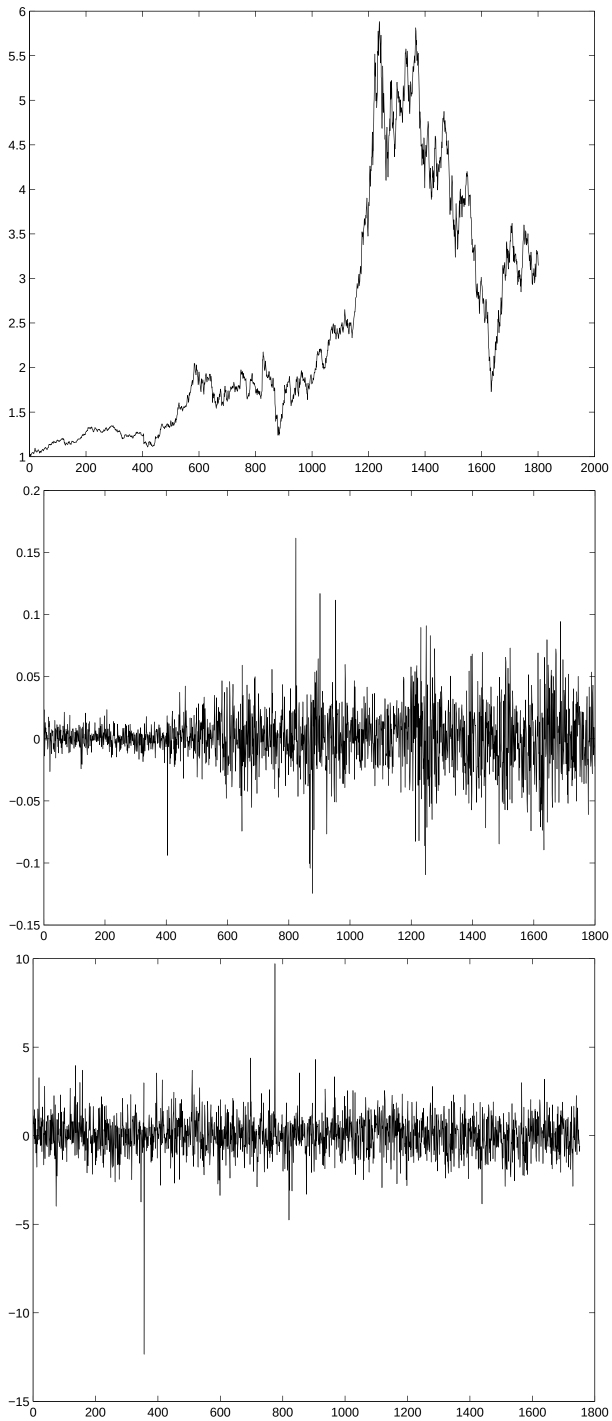


Figure 6.3: *Siemens data. Top plot: Price process, middle: Log-returns, bottom: Log-returns, devolatized using Nadaraya-Watson volatility.*

6.2 Are the Devolatized Log-returns Independent?

Since we do not assume a simultaneous Gaussian law for the log-returns, zero autocorrelation does not necessarily imply independence. We will therefore use the BDS-test (Brock, Dechert & Scheinkman test for independence based on the correlation dimension.), an independence test which reacts to accumulations of similar values in a time series, see [18]. The test is carried out using MatLab code written by Ludwig Kanzler, see also [31]. To present the test, we use notation from [45] and [50]. The BDS statistic, $W_{m,\,n}(\epsilon)$, for fixed parameters $m \in \mathbb{N}$ and $\epsilon > 0$, is defined as

$$W_{m,n}(\epsilon) = \sqrt{n}\,\frac{C_{m,n}(\epsilon) - (C_{1,n}(\epsilon))^m}{\sigma_{m,n}(\epsilon)}.$$

Here n is the number of observation. Further, for $n_m = n - m + 1$, and

$$1_\epsilon(s,t) = 1_{[-\epsilon,\epsilon]}\Big\{\max_{i\in\, 0,\ldots,m-1} |X(t+i) - X(s+i)|\Big\},$$

we have

$$C_{m,n}(\epsilon) = \sum_{1\le t<s\le n_m} 1_\epsilon(s,t)\frac{2}{n_m(n_m-1)}.$$

and

$$\begin{aligned}\sigma^2_{m,n}(\epsilon) &= 4\Big(K_n(\epsilon) + 2\sum_{j=1}^{m-1} K_n(\epsilon)^{m-j}C_{1,n}(\epsilon)^{2j} \\ &+ (m-1)^2 C_{1,n}(\epsilon)^{2m} - m^2 K_n(\epsilon)C_{1,n}(\epsilon)^{2m-2}\Big),\end{aligned}$$

where

$$K_n(\epsilon) = \sum_{1\le t<s<r\le n_m} \frac{2(1_\epsilon(t,s)1_\epsilon(s,r) + 1_\epsilon(t,r)1_\epsilon(r,s) + 1_\epsilon(s,t)1_\epsilon(t,r))}{n_m(n_m-1)(n_m-2)}.$$

Under the null hypothesis, that of independence, we have that $W_{m,n}(\epsilon)$ is asymptotically standard Gaussian distributed. As suggested in [19], we choose the parameter values $m = 4$ and $\epsilon = 1.5s_L$, where s_L is the empirical standard deviation of A_t.

Below is a table with the results from the BDS-test, for each data set, and volatility model. Regarding devolatization, we see that all stochastic volatility models seem to work well. As one would expect, constant volatility fails for all three data sets.

USD/DM	BDS statistic	p-value	i.i.d. hypothesis
Nadaraya-Watson	-0.51	0.61	Not Rejected
Var-Win	-0.47	0.64	Not Rejected
GARCH-AR	-1.07	0.28	Not rejected
Constant	5.56	< 0.001	Rejected

DAX	BDS statistic	p-value	i.i.d. hypothesis
Nadaraya-Watson	0.33	0.74	Not Rejected
Var-Win	-0.47	0.64	Not Rejected
GARCH-AR	0.46	0.64	Not rejected
Constant	-27.37	< 0.001	Rejected

SIEMENS	BDS statistic	p-value	i.i.d. hypothesis
Nadaraya-Watson	-0.32	0.74	Not Rejected
Var-Win	1.69	0.09	Not Rejected
GARCH-AR	-1.37	0.17	Not rejected
Constant	-27.37	< 0.001	Rejected

Table 6.1: *BDS test for independency of devolatized returns.*

6.3 How Well do the Marginal Densities Fit Data?

To test the fit of univariate marginal distributions to data, we use the Kuiper test statistic K, which is given by

$$K = \max_{x\in[0,1]} (F_{\text{emp}}(x) - x) + \max_{x\in[0,1]} (x - F_{\text{emp}}(x)).$$

Here, F_{emp} denotes the empirical distribution of the sequence $(U_t)_{t=1,\ldots,T} = (\hat{F}(X_t))_{t=1,\ldots,T}$, where $\hat{F}$ is the estimated distribution function of the log-returns. Asymptotically, the p-value of the test is given by, see [45],

$$2\sum_{j=1}^{\infty}(4j^2\lambda^2 - 1)\exp\{-2j^2\lambda^2\},$$

where

$$\lambda = K\left(\sqrt{T} + 0.155 + \frac{0.24}{\sqrt{T}}\right).$$

We also use standard graphical procedures, such as PP-plots and QQ-plots, which are presented below. The PP-plots, QQ-plots and Kuiper tests are based only on observations in the lower part of the dataset. The reason for this, is that we, in a VaR setting, are mainly interested in how well the data fits the left tail of a given probability density function.

Remark Before making the PP-plots, we transform $(U_t)_{t=1,\ldots,T}$ to standard Gaussian, and then we present a *normal Probability Plot.* This gives a good indication of how well the lower tail of the distribution fits the smallest observations.

We also display the evolvement of the probability density functions, of the devolatized log-returns over time. Over a 1300 day period, the parameters of the probability density functions are estimated every five days, using a 500 day window. Since parameter estimation for the GH distribution is time consuming we have used the Hyperbolical distribution instead.

It should be noted that, throughout chapter 6, the Pearson VII distribution is referred to as just the Pearson distribution.

6.3.1 USD/DM

NIG	α	β	γ	δ	λ
Constant	1.8253	-0.0000	0.9122	-0.0108	1/2
Hyperbolical	α	β	γ	δ	λ
Nadaraya-Watson	1.7657	0.0193	1.0620	-0.0217	1
Variance Window	1.6005	0.0475	0.9460	-0.0588	1
GARCH-AR	1.9929	0.0673	1.0590	-0.0572	1
Constant	2.4638	0.1117	0.6761	-0.0598	1
GP	ξ	σ	ζ_u	u	
Nadaraya-Watson	0.0047	0.6625	0.2337	0.7	
Variance Window	0.0530	0.6352	0.2481	0.7	
GARCH-AR	-0.0249	0.6066	0.2454	0.6	
Constant	0.0237	0.4317	0.2262	0.5	
Pearson	m_-	c_-			
Nadaraya-Watson	2.1982	3.6318			
Variance Window	2.2038	3.4743			
GARCH-AR	2.1947	4.1099			
Constant	1.5639	3.7844			
Gaussian	μ	σ			
Nadaraya-Watson	-0.0195	1.0685			
Variance Window	-0.0101	1.1242			
GARCH-AR	0.0061	0.9760			
Constant	0.0013	0.7459			

Table 6.2: *Estimated parameters for USD/DM data.*

Hyperbolical	K	*p value*	Hypothesis
Nadaraya-Watson	0.0617	0.6699	Not Rejected
Variance Window	0.0604	0.6973	Not Rejected
GARCH-AR	0.0650	0.5604	Not Rejected
Constant	0.0718	0.3845	Not Rejected
Gaussian			
Nadaraya-Watson	0.1053	0.0202	Rejected
Variance Window	0.1102	0.0105	Rejected
GARCH-AR	0.0988	0.0372	Rejected
Constant	0.1091	0.0111	Rejected
Pearson			
Nadaraya-Watson	0.0518	0.8958	Not Rejected
Variance Window	0.0534	0.8617	Not Rejected
GARCH-AR	0.0609	0.6739	Not Rejected
Constant	0.0676	0.4910	Not Rejected
GP			
Nadaraya-Watson	0.0518	0.9198	Not Rejected
Variance Window	0.0314	1.0000	Not Rejected
GARCH-AR	0.0588	0.7351	Not Rejected
Constant	0.0496	0.9513	Not Rejected
NIG	0.0213	0.7484	Not Rejected

Table 6.3: *Kuiper test for devolatilized log-returns (USD/DM).*

Table 6.3 shows that, the Hyperbolical distribution, the Pearson distribution and the GP distribution, all pass the Kuiper test, regardless of the volatility model used (even the constant). However, the Gaussian distribution fails in all four cases.

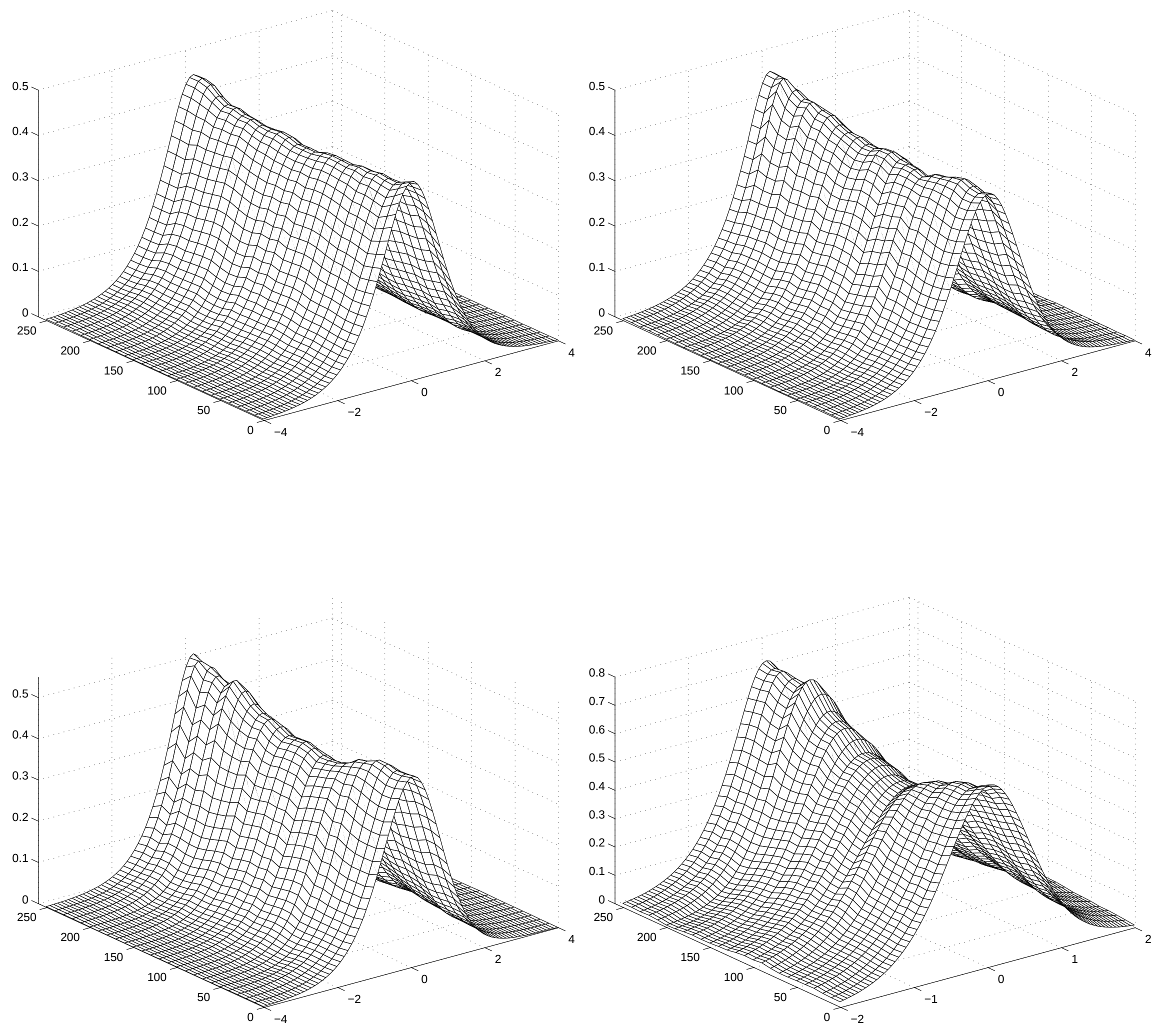


Figure 6.4: *Evolvement through time of the estimated Hyperbolical probability density function for the devolatized log-returns of the USD/DM data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

Looking at figure 6.4, the Nadaraya-Watson volatility model really seems to impose stationarity for the distribution of the devolatized log-returns. The same goes for the Variance Window, though not as clear as for the Nadaraya-Watson. In the constant volatility case, the underlying distribution is clearly changing over time, (i.e. not stationary).

Also the plots, in figure 6.5, show a good fit for the Hyperbolical distribution. As seen, there are a few “extreme”observations that are not captured by the Hyperbolical distribution.

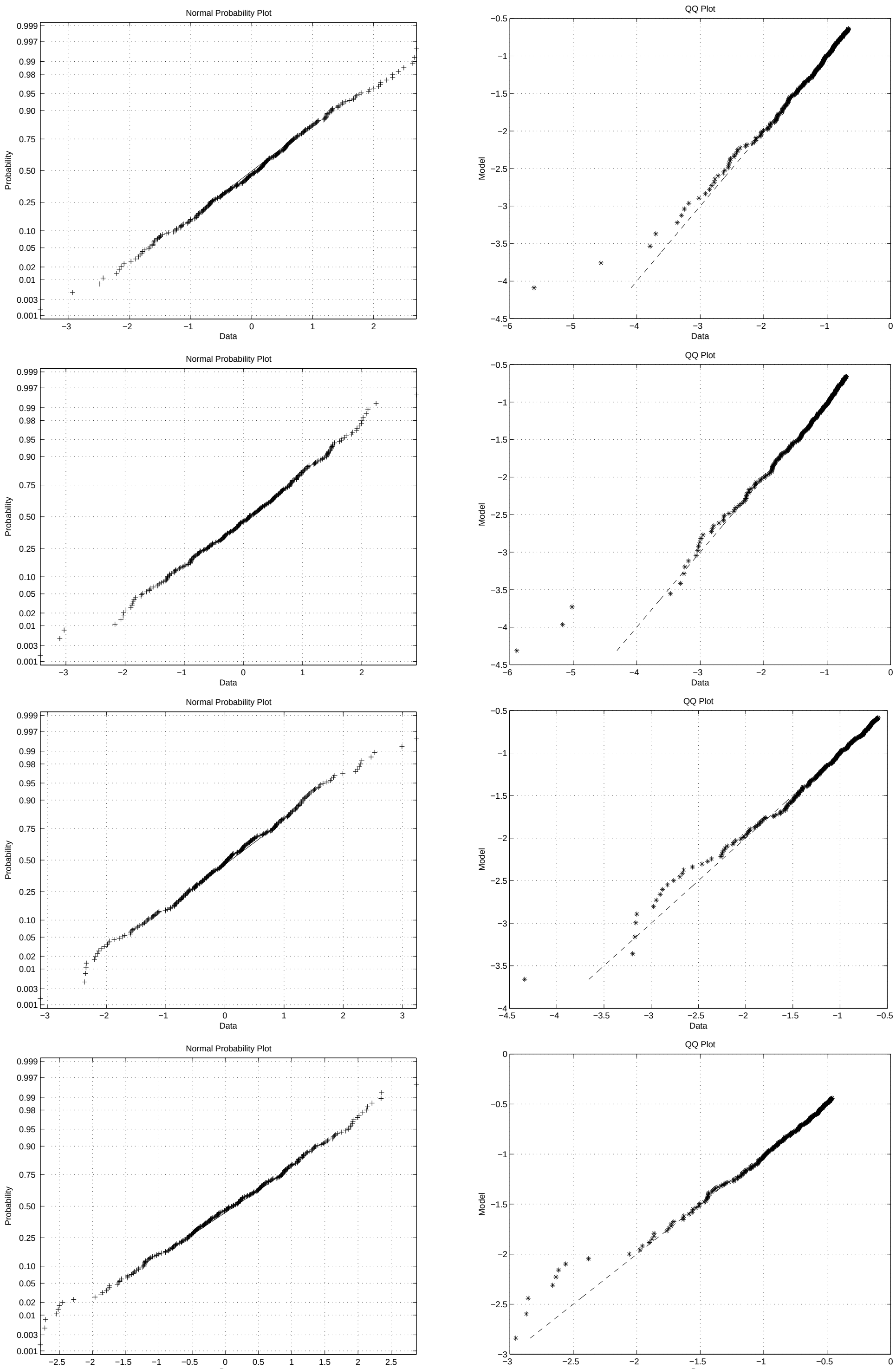


Figure 6.5: *PP- and QQ-plots for USD/DM data, under assumption of Hyperbolical distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

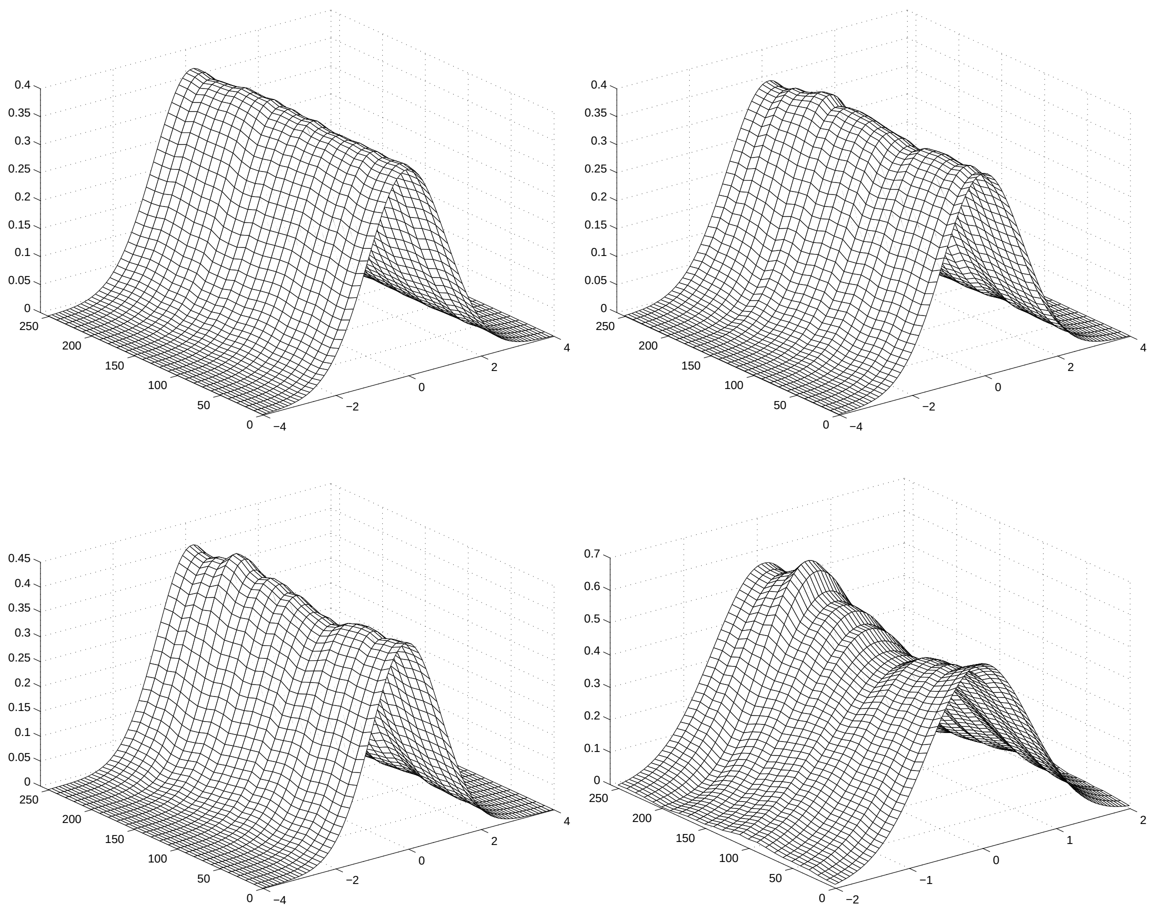


Figure 6.6: *Evolvement through time of the estimated Gaussian probability density function for the devolatized log-returns of the USD/DM data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

As seen in figure 6.6, the Nadaraya-Watson devolatization procedure, seems to make the devolatized log-returns stationary. This is also the case, though not as clear, for both the Variance Window and the GARCH-AR volatility models.

As expected, the plots in figure 6.7, indicate that the Gaussian distribution does not fit the data very well.

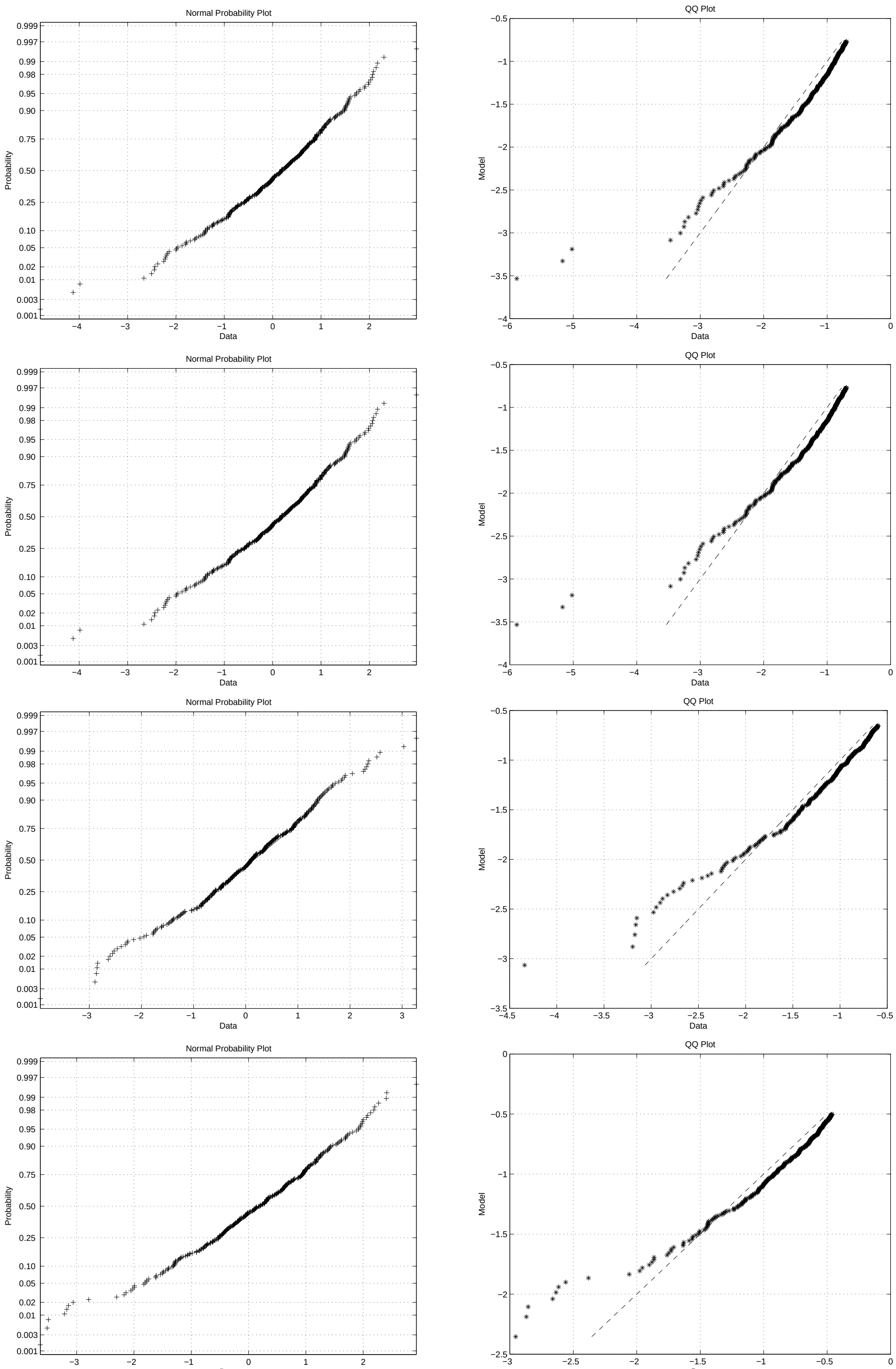


Figure 6.7: *PP- and QQ-plots for USD/DM data, under assumption of Gaussian distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

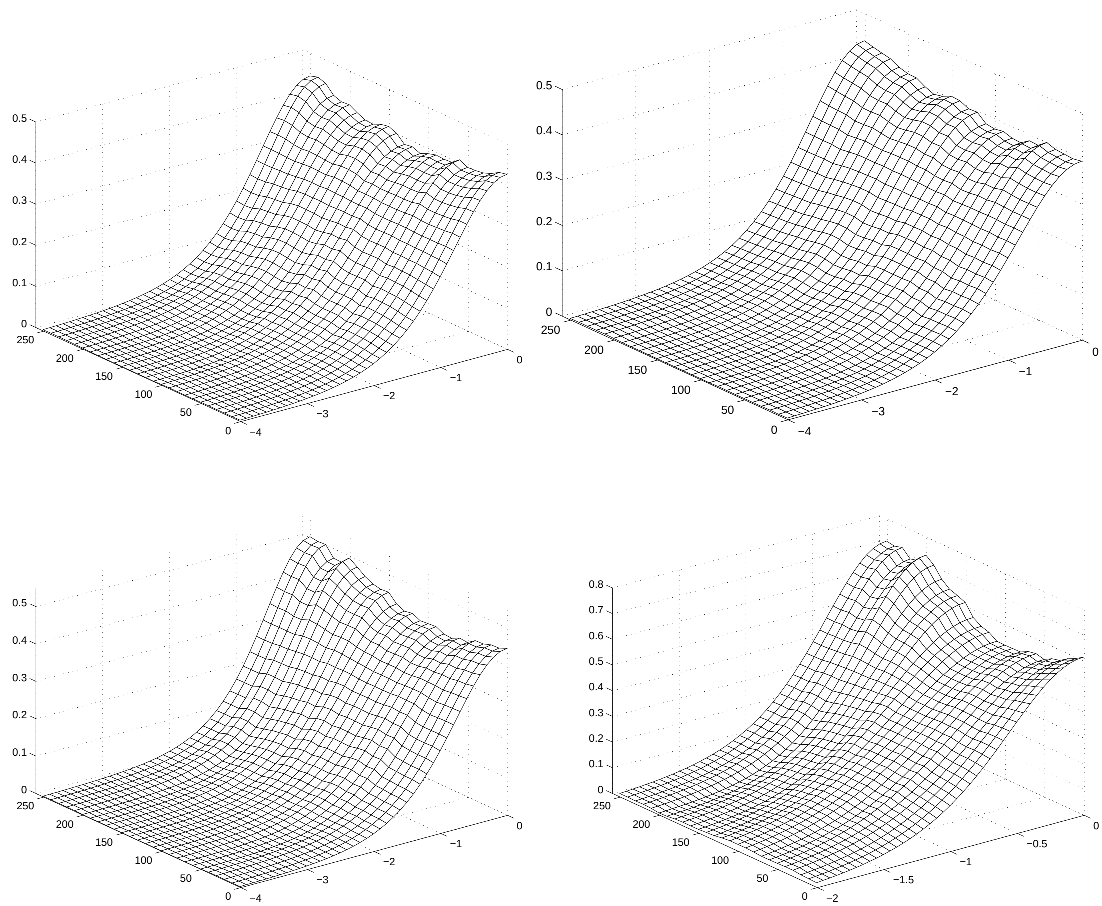


Figure 6.8: *Evolvement through time of the estimated Pearson probability density function for the devolatized log-returns of the USD/DM data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

Considering stationarity of the devolatized log-returns, the Nadaraya-Watson, and the Variance Window seem to be the best volatility models.

The plots, in figure 6.9, indicate that the Pearson density fits the USD/DM data well. However, there are a few extreme observations, that do not seem to be captured by this model. For the USD/DM data, the Nadaraya-Watson seems to be the best volatility model, to use together with the Pearson distribution.

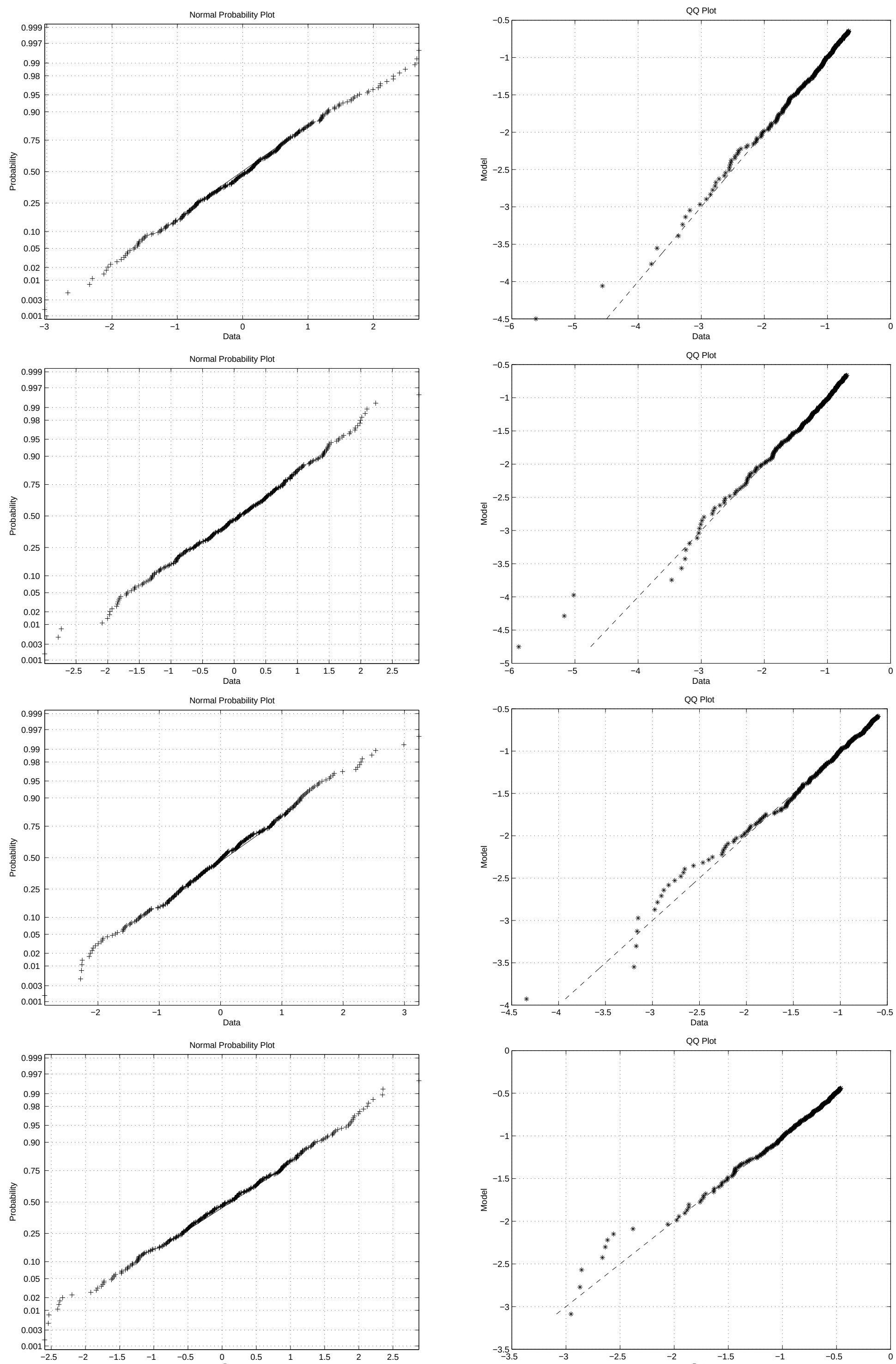


Figure 6.9: *PP- and QQ-plots for USD/DM data, under assumption of Pearson distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

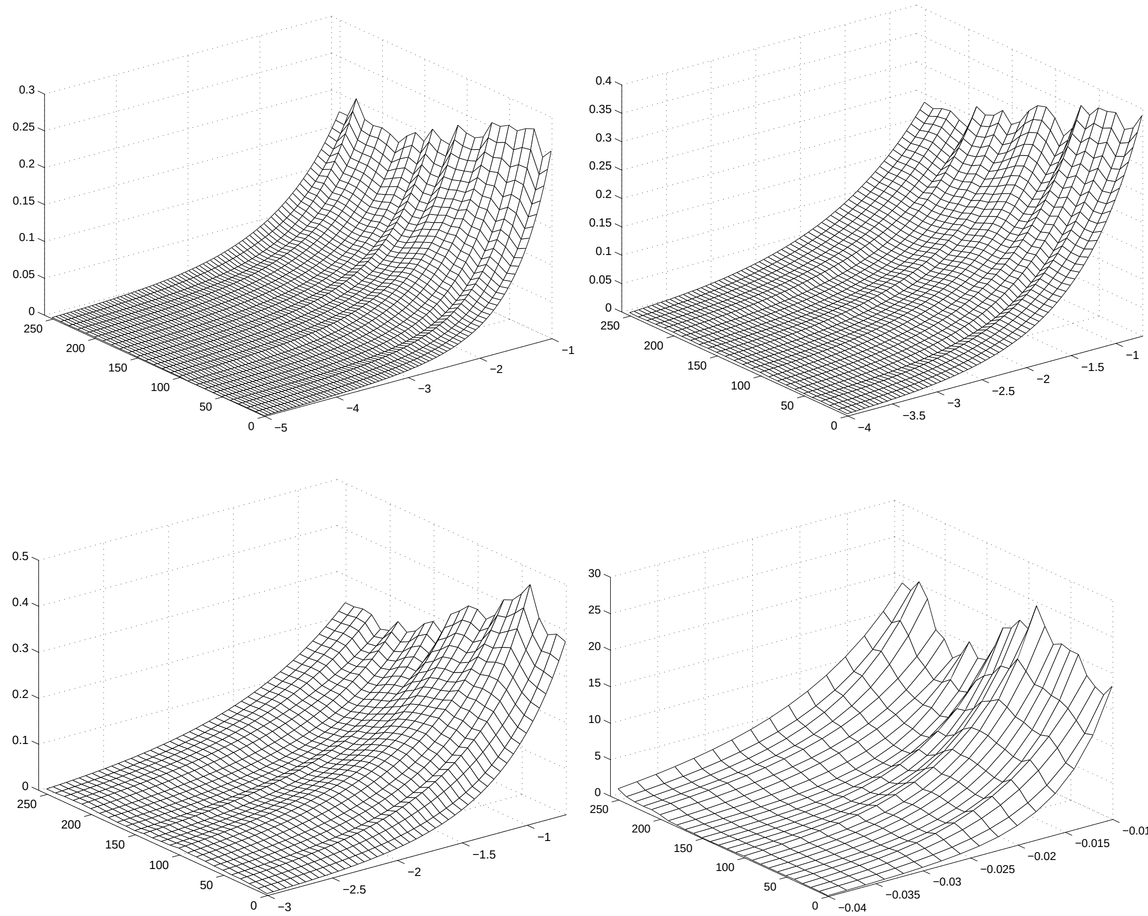


Figure 6.10: *Evolvement through time of the estimated GP probability density function for the devolatized log-returns of the USD/DM data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

Once again, devolatization using the Nadaraya-Watson and the Variance Window volatility models, seems to result in stationary log-returns.

As seen in the top two plots, in figure 6.11, devolatized data using Nadaraya-Watson volatility, seems to fit the GP density very well. Overall, the GP density fits data well, just as the Kuiper test indicates.

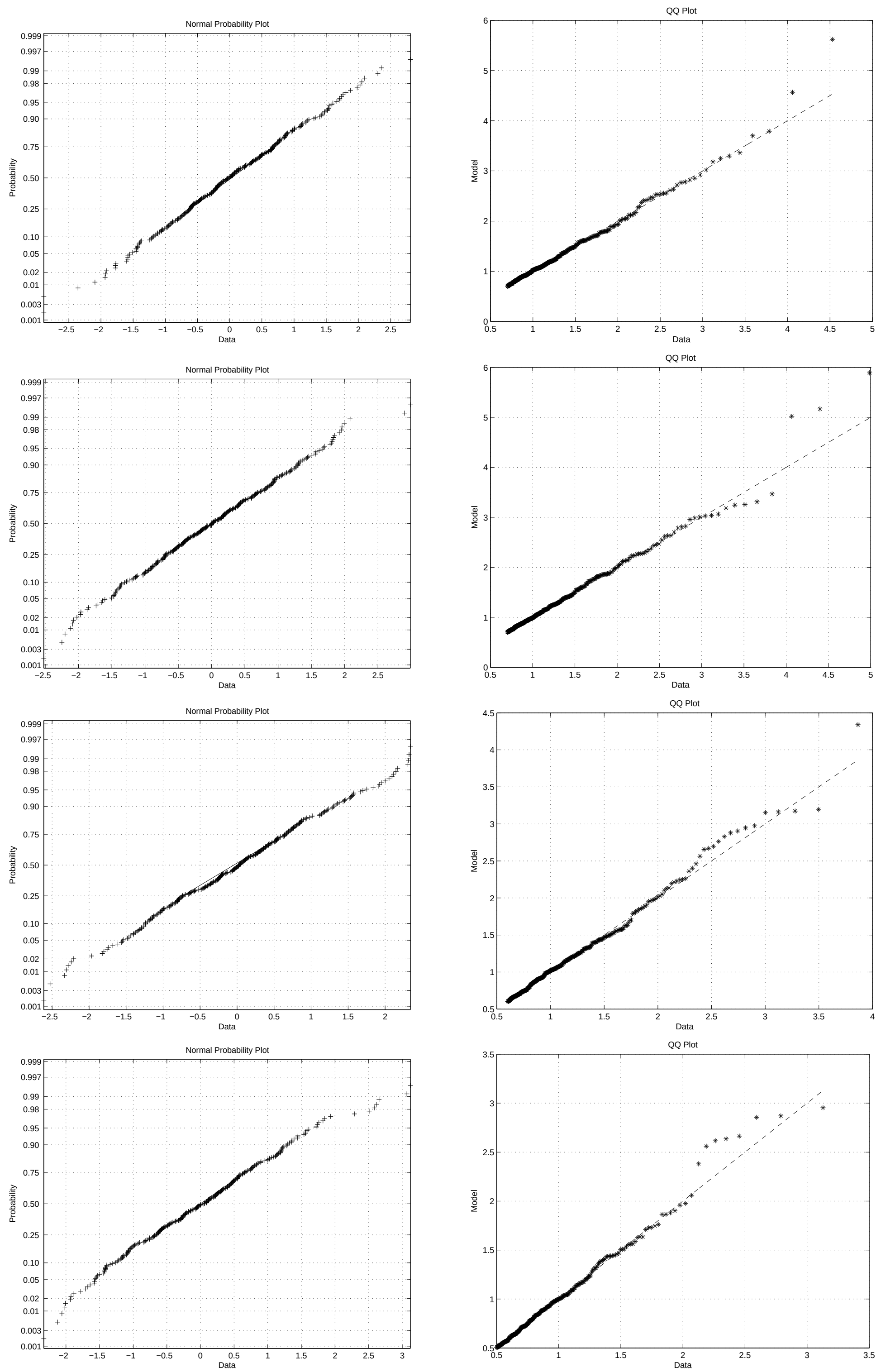


Figure 6.11: *PP- and QQ-plots for USD/DM data, under assumption of GP distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

6.3.2 DAX

Hyperbolical	α	β	γ	δ	λ
Nadaraya-Watson	1.9790	0.0000	1.5238	0.0972	1
Variance Window	1.9680	-0.4753	1.5086	0.6555	1
GARCH-AR	2.1271	-0.5388	1.3480	0.5594	1
Constant	131.1511	-18.7622	0.0071	0.0036	1
GP	ξ	σ	ζ_u	u	
Nadaraya-Watson	0.0647	0.7017	0.1476	1	
Variance Window	0.0457	0.7713	0.1541	1	
GARCH-AR	0.0574	0.7262	0.1554	0.95	
Constant	0.1571	0.0081	0.1623	0.01	
Pearson	m_-	c_-			
Nadaraya-Watson	2.2245	3.1833			
Variance Window	2.1954	2.9549			
GARCH-AR	2.1585	3.3211			
Constant	0.0170	2.1308			
Gaussian	μ	σ			
Nadaraya-Watson	-0.0118	1.0937			
Variance Window	0.0716	1.1619			
GARCH-AR	0.0027	1.0745			
Constant	0.0007	0.0130			

Table 6.4: *Estimated parameters for DAX data.*

Hyp	K	*p value*	Hypothesis
Nadaraya-Watson	0.0643	0.6082	Not Rejected
Variance Window	0.0507	0.9086	Not Rejected
GARCH-AR	0.0622	0.6383	Not Rejected
Constant	0.0494	0.9248	Not Rejected
Gaussian			
Nadaraya-Watson	0.1026	0.0292	Rejected
Variance Window	0.0901	0.0937	Not Rejected
GARCH-AR	0.0931	0.0677	Not Rejected
Constant	0.0996	0.0342	Rejected
Pearson			
Nadaraya-Watson	0.0593	0.7396	Not Rejected
Variance Window	0.0419	0.9891	Not Rejected
GARCH-AR	0.0618	0.6490	Not Rejected
Constant	0.0481	0.9412	Not Rejected
GP			
Nadaraya-Watson	0.0796	0.6308	Not Rejected
Variance Window	0.0602	0.9218	Not Rejected
GARCH-AR	0.0783	0.6708	Not Rejected
Constant	0.0679	0.8011	Not Rejected

Table 6.5: *Kuiper test for devolatilized log-returns (DAX).*

As for The USD/DM data, the Hyperbolical distribution, the Pearson distribution and the GP distribution, all seem to fit the DAX data extremely well. Since the Gaussian distribution (barely) passes the Kuiper test for Variance Window and GARCH-AR volatility, it seems to better fit the DAX data than the USD/DM data. However, the Gaussian distribution, once again, gives a poor overall impression.

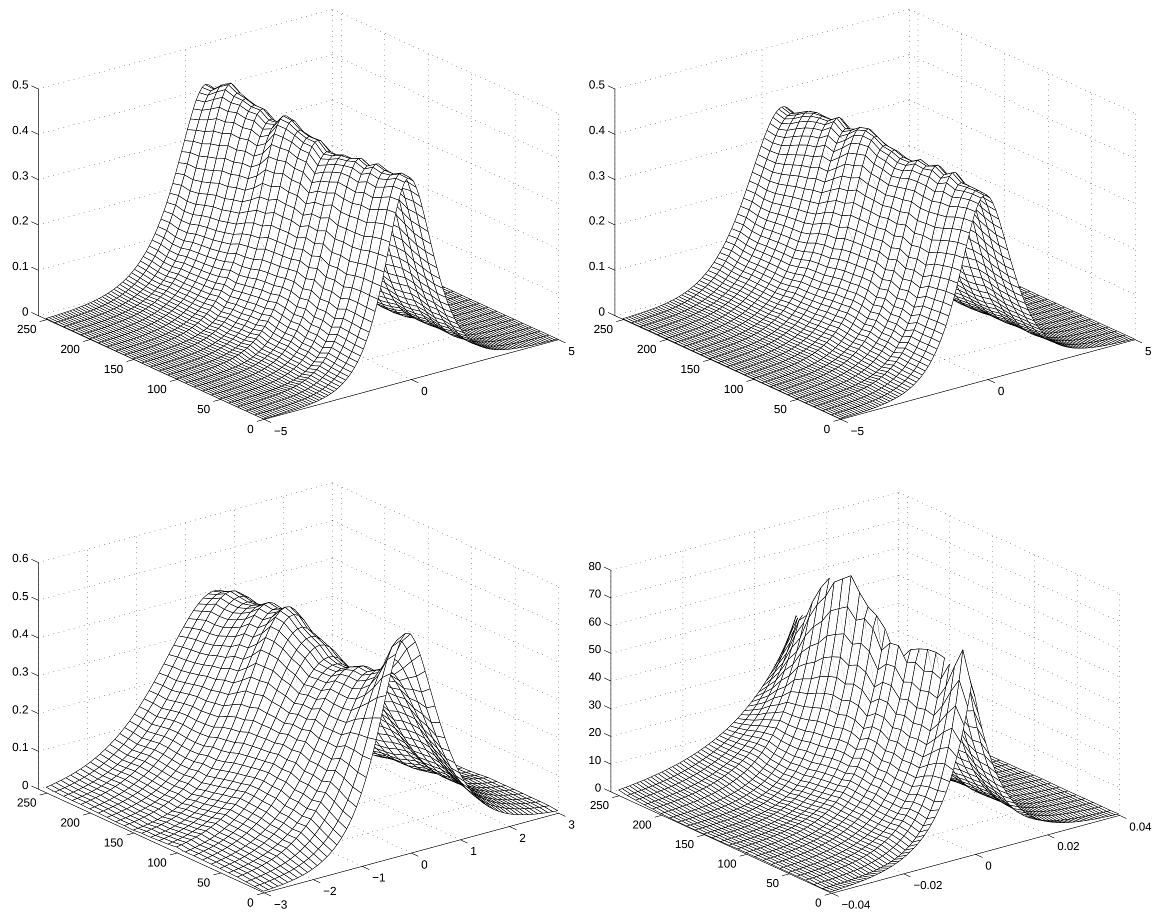


Figure 6.12: *Evolvement through time of the estimated Hyperbolical probability density function for the devolatized log-returns of the DAX data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

Regarding stationarity, the Nadaraya-Watson and the Variance Window volatility models excel. Noticable is the "peakiness" of the estimated Hyperbolical density function, in the constant volatility case.

For the DAX data, looking at figure 6.14, the Hyperbolical distribution does not perform as well as in the USD/DM case. The GARCH-AR volatility model together with the Hyperbolical marginal density, does not seem to work very well.

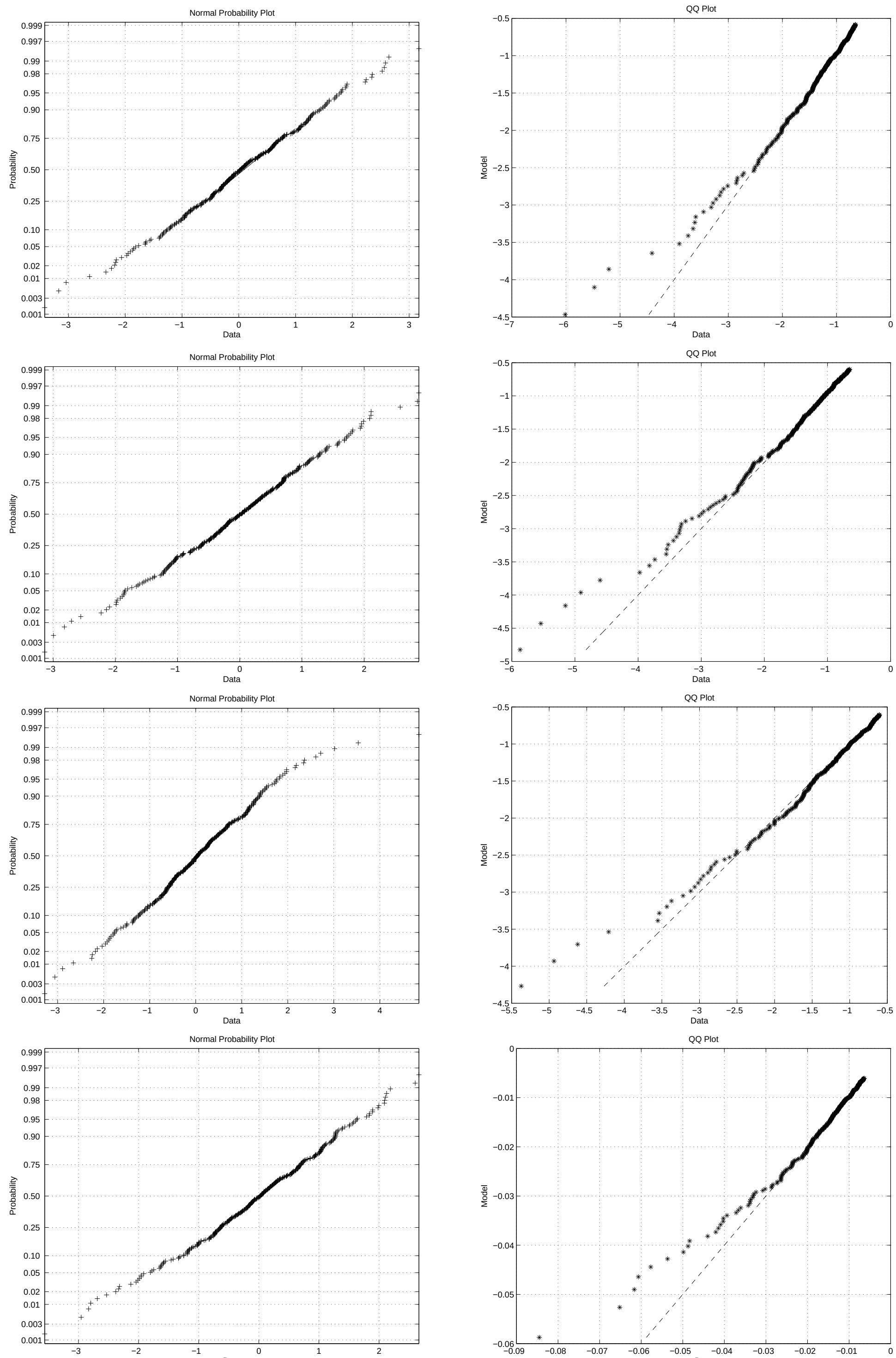


Figure 6.13: *PP- and QQ-plots for DAX data, under assumption of Hyperbolical distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

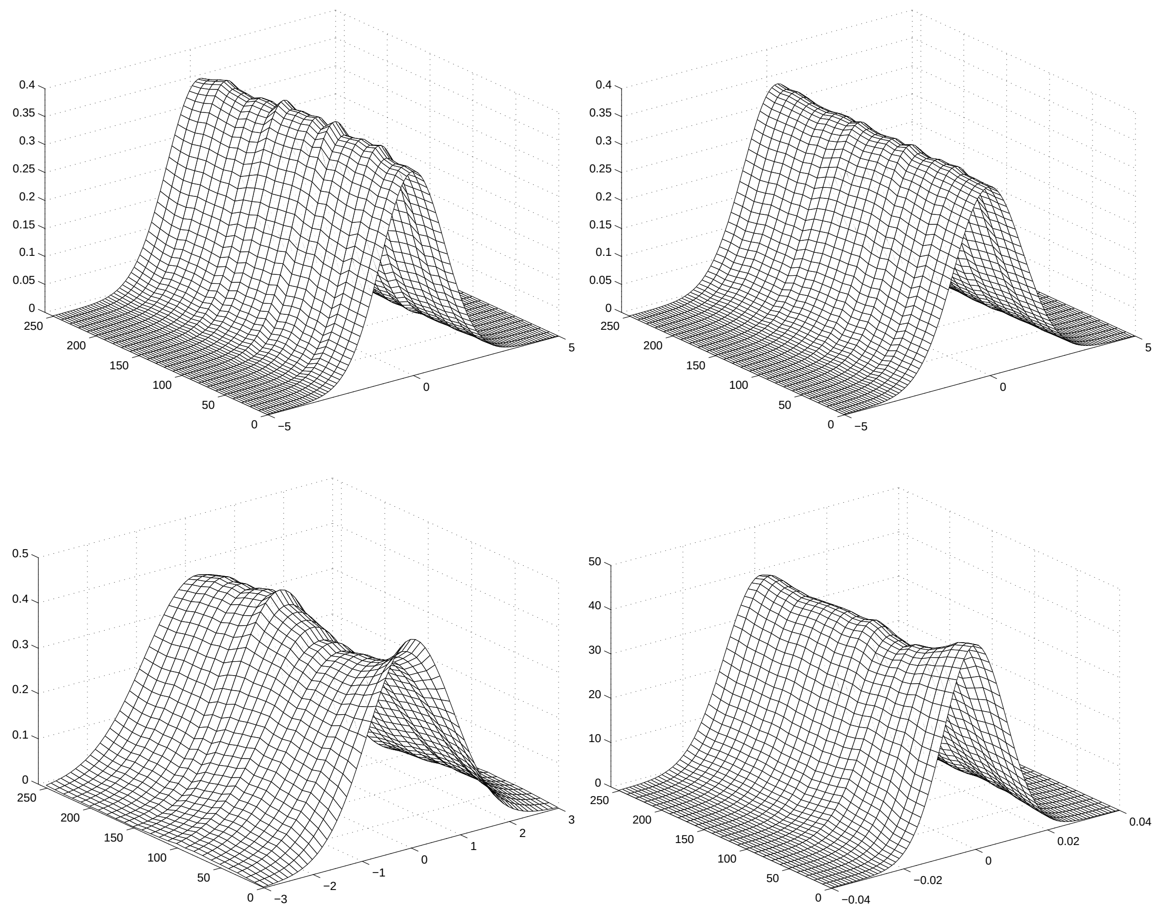


Figure 6.14: *Evolvement through time of the estimated Gaussian probability density function for the devolatized log-returns of the DAX data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

When it comes to stationarity, the Variance Window is the best choice of volatility model. Remarkably, devolatization using the GARCH-AR volatility seems to remove log-returns from stationarity, rather then take them closer to it.

Figure 6.15 shows that, the Gaussian distribution again fails miserably. It is reasonable to believe, that the assumption of Gaussian distribution for devolatized log-returns will underestimate the VaR, no matter what volatility model we use.

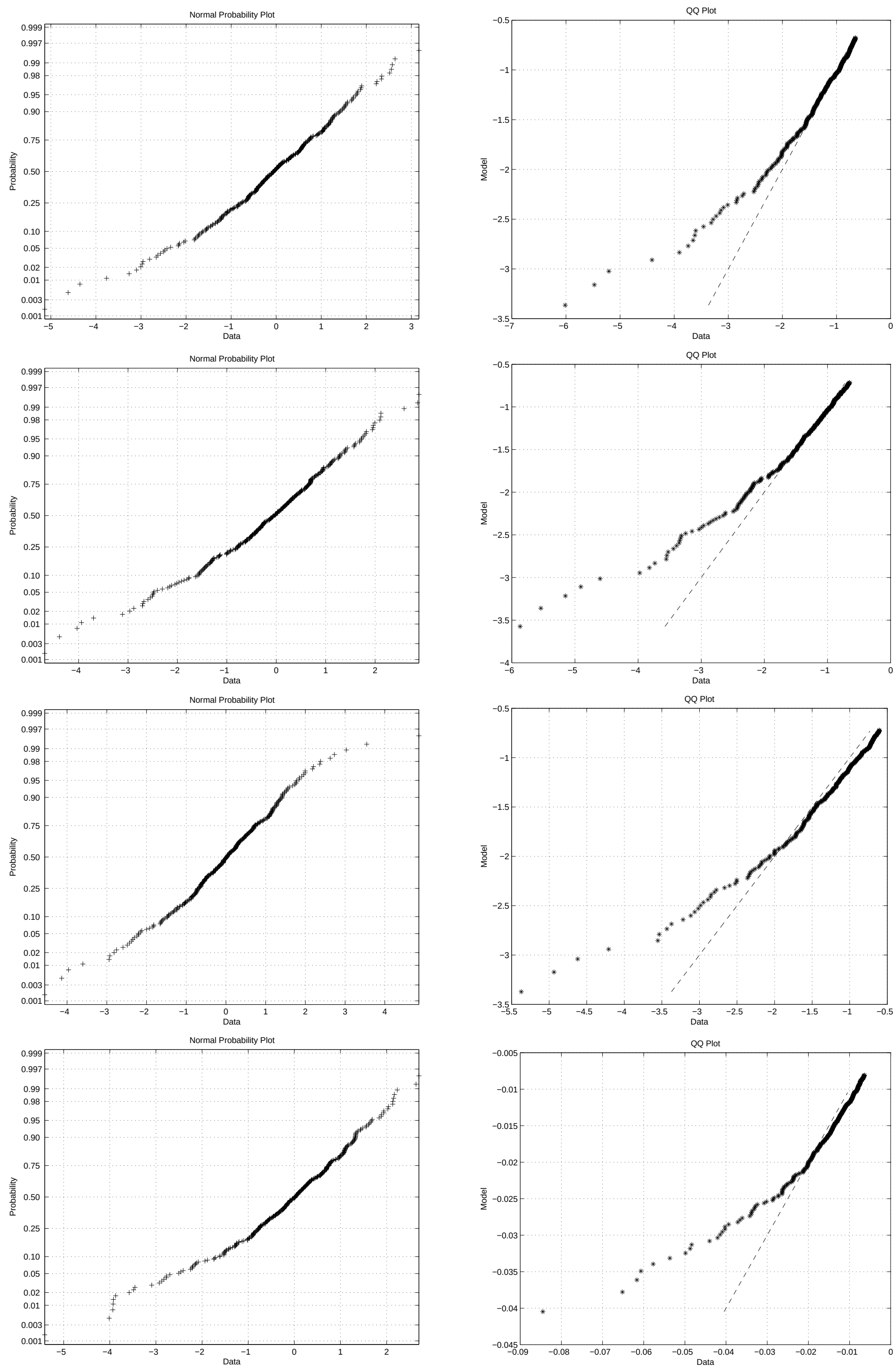


Figure 6.15: *PP- and QQ-plots for DAX data, under assumption of Gaussian distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

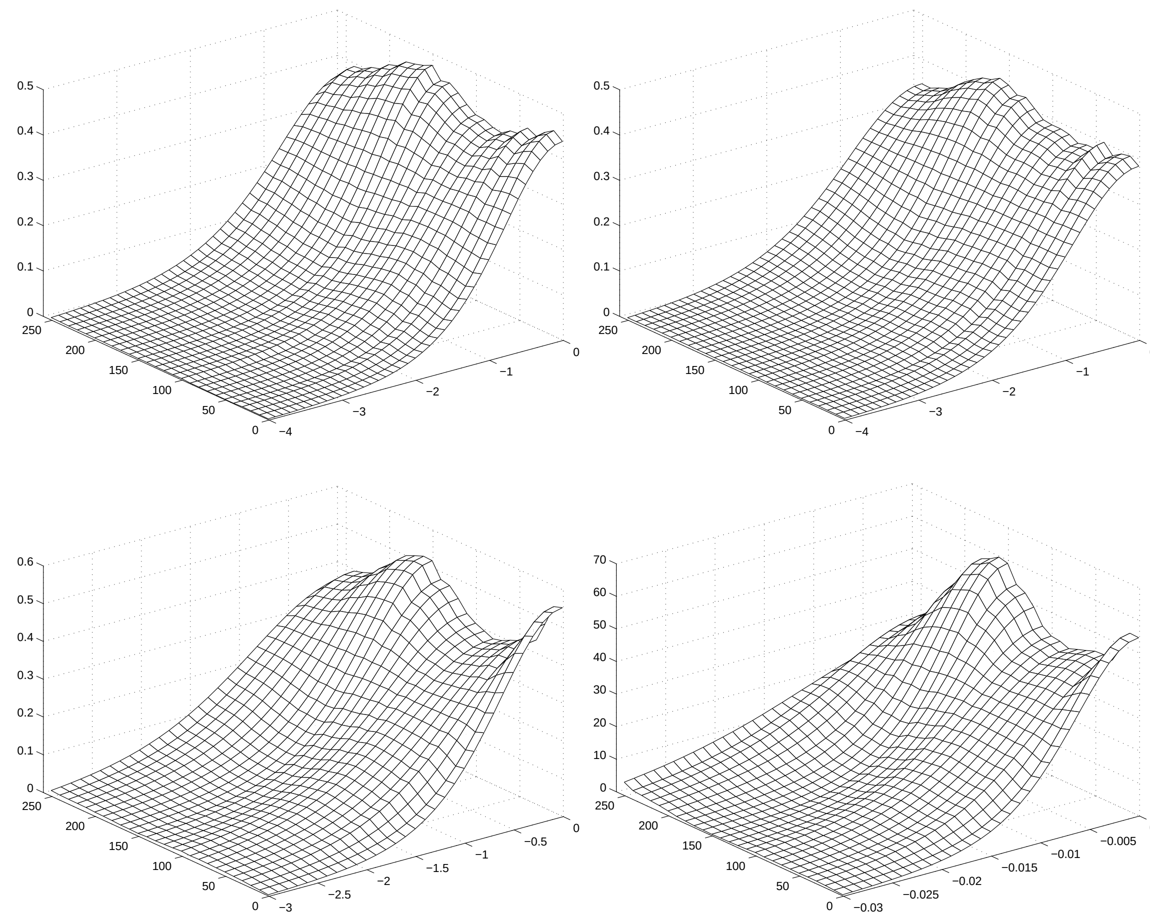


Figure 6.16: *Evolvement through time of the estimated Pearson probability density function for the devolatized log-returns of the DAX data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

The Nadaraya-Watson, and the Variance Window volatility models, makes the devloatized log-returns stationary.

The plots, in figure 6.17, look really good. For the DAX data, the Pearson distribution seems to outperform all the other distributions. This result was also indicated by the Kuiper test. In this case, deciding what volatility model is best, is a close call.

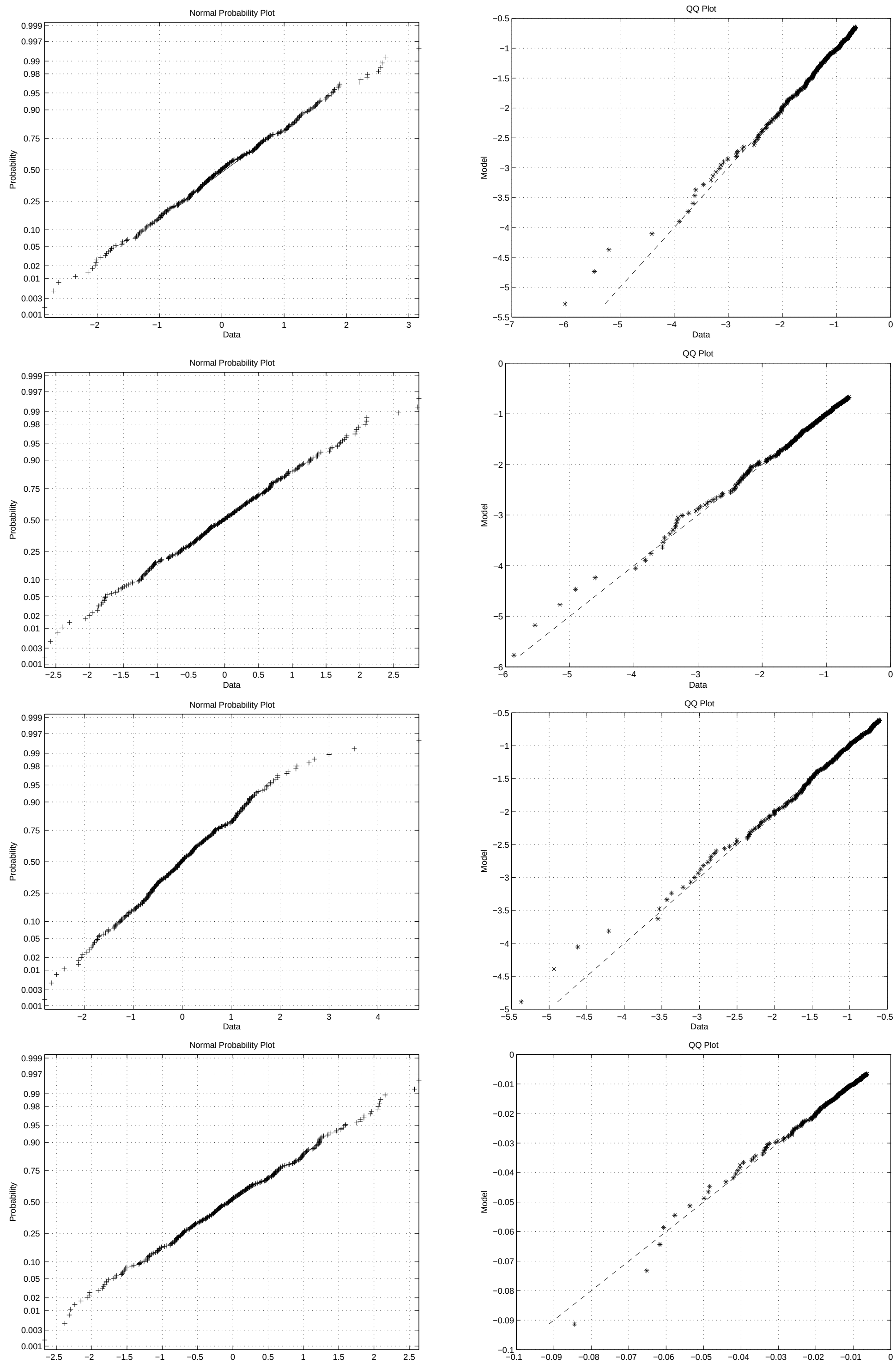


Figure 6.17: *PP- and QQ-plots for DAX data, under assumption of Pearson distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

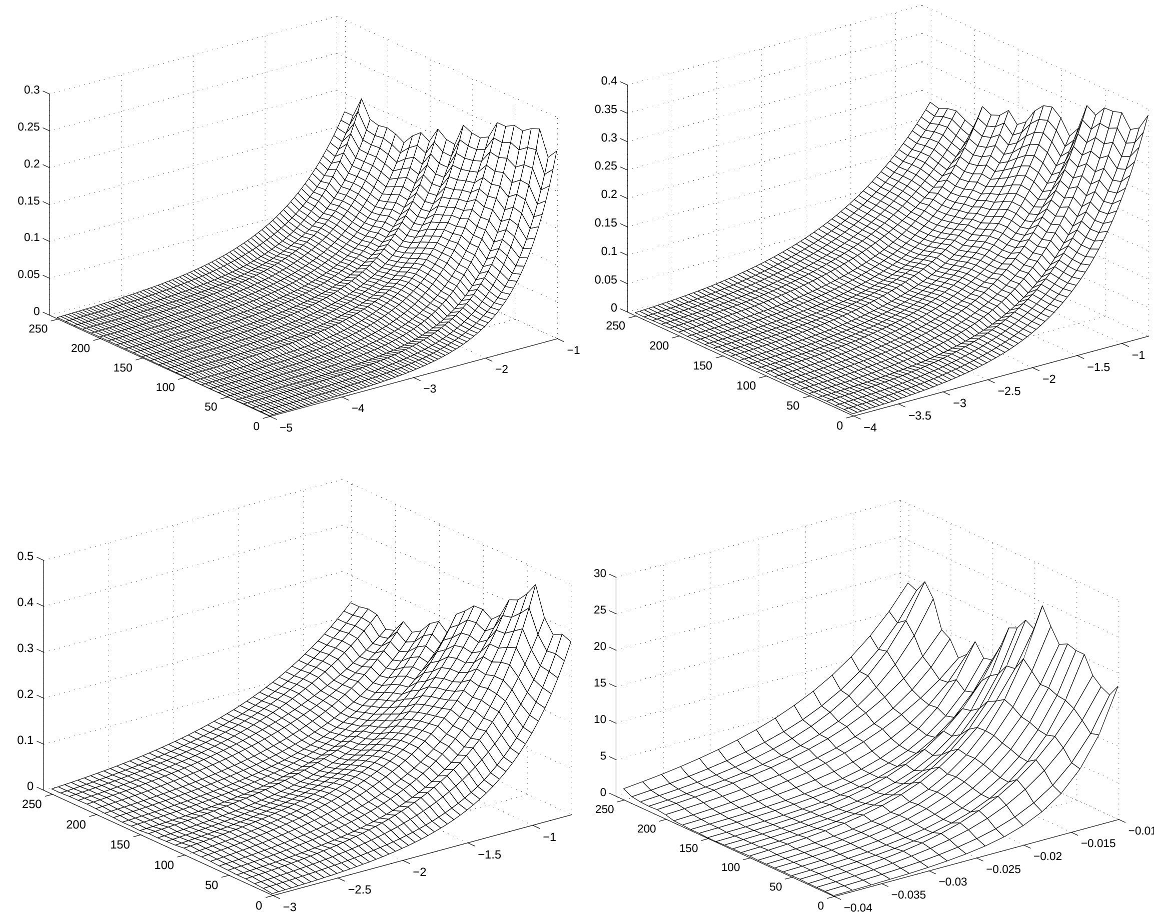


Figure 6.18: *Evolvement through time of the estimated GP probability density function for the devolatized log-returns of the DAX data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

Once again, devolatization using the Nadaraya-Watson, and the Variance Window volatility models makes devolatized log-returns stationary.

As indicated by figure 6.19 ,the GP distribution seems to perform well with the three stochastic volatility models. We also see that, for the DAX data, the GP is the only distribution that seems to work well using constant volatility.

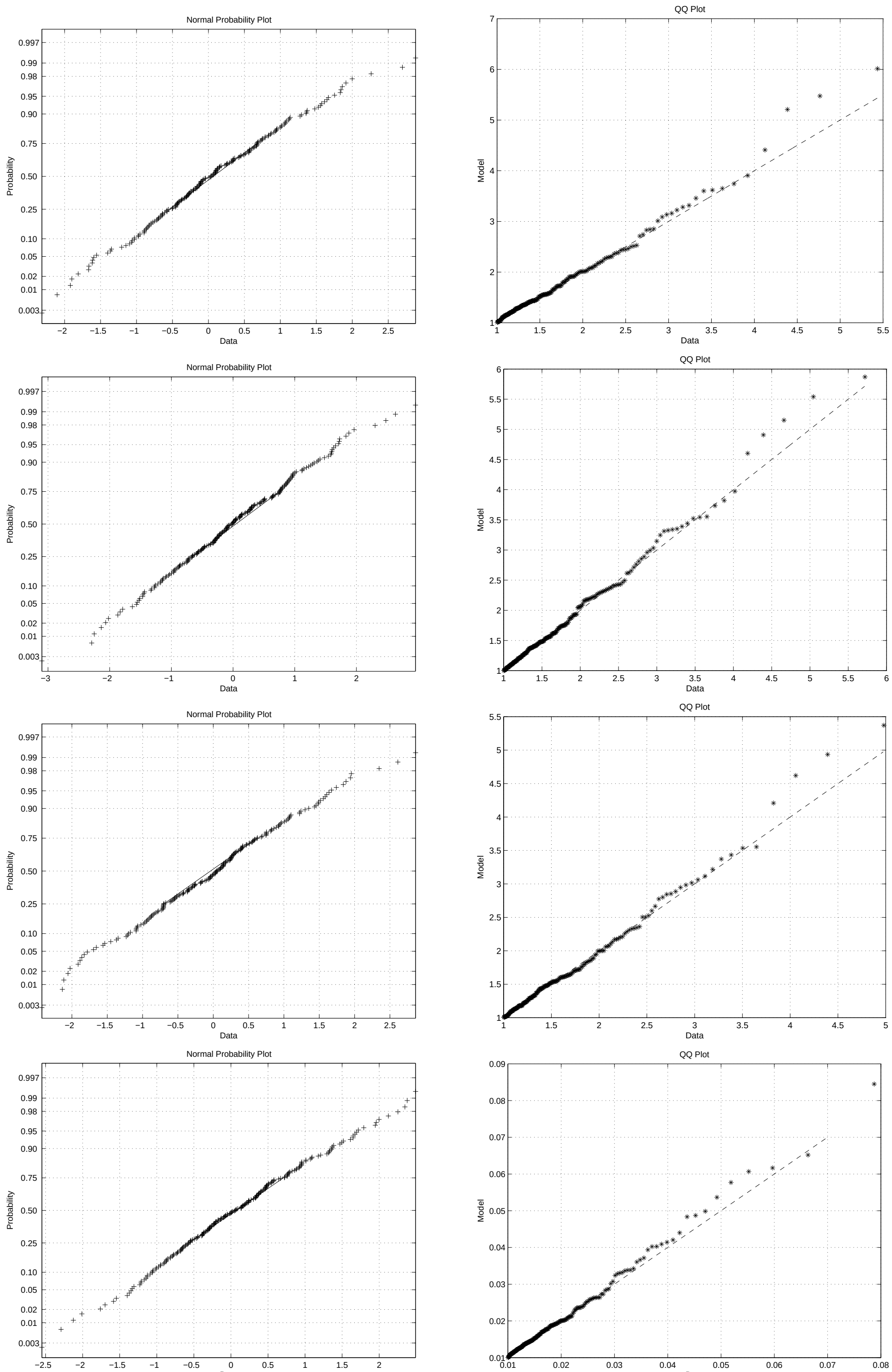


Figure 6.19: *PP- and QQ-plots for DAX data, under assumption of GP distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

6.3.3 SIEMENS

Hyperbolical	α	β	γ	δ	λ
Nadaraya-Watson	1.8980	0.0008	1.3625	0.0251	1
Variance Window	1.7005	0.0019	1.1677	0.0287	1
GARCH-AR	1.7723	-0.0729	1.2377	0.1042	1
Constant	67.2131	-1.0169	0.0251	0.0014	1
GP	ξ	σ	ζ_u	u	
Nadaraya-Watson	0.0260	0.5286	0.1078	1.3	
Variance Window	0.0545	0.5445	0.0966	1.4	
GARCH-AR	0.0334	0.6434	0.1615	1	
Constant	0.0568	0.0153	0.1285	0.03	
Pearson	m_-	c_-			
Nadaraya-Watson	3.1464	5.8391			
Variance Window	2.8622	4.8461			
GARCH-AR	2.3203	3.6711			
Constant	0.0562	3.6033			
Gaussian	μ	σ			
Nadaraya-Watson	0.0261	1.0956			
Variance Window	0.0716	1.1619			
GARCH-AR	0.0155	1.1181			
Constant	0.0006	0.0274			

Table 6.6: *Estimated parameters for the Siemens data*

Hyp	K	*p value*
Nadaraya-Watson	0.1062	0.4322
Variance Window	0.0976	0.5647
GARCH-AR	0.0808	0.8409
Constant	0.0888	0.7129
Gaussian		
Nadaraya-Watson	0.1076	0.4085
Variance Window	0.0927	0.6544
GARCH-AR	0.0859	0.7631
Constant	0.1280	0.1352
Pearson		
Nadaraya-Watson	0.0983	0.5713
Variance Window	0.0911	0.6833
GARCH-AR	0.0651	0.9809
Constant	0.0844	0.7873
GP		
Nadaraya-Watson	0.0800	0.8712
Variance Window	0.0934	0.7190
GARCH-AR	0.0753	0.6676
Constant	0.0752	0.8231

Table 6.7: *Kuiper test for devolatilized log-returns (Siemens).*

In this case, all marginal distributions, including the Gaussian, pass the Kuiper test, with high p-values as well. Except for the case of constant volatility, the Gaussian distribution seems to perform almost as well (or as bad) as the Hyperbolical distribution. Regardless of which volatility model used, both the Pearson distribution and the GP distribution fit the

Siemens data better.

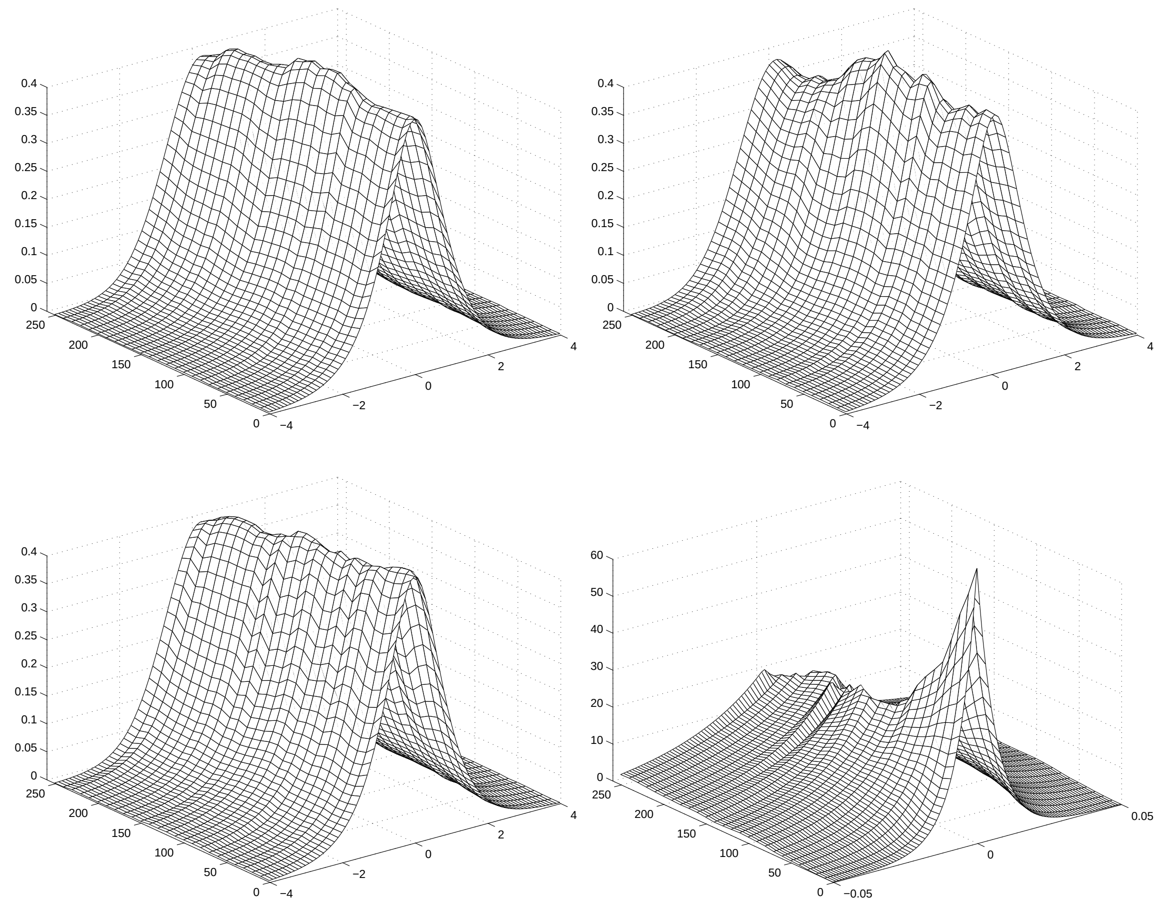


Figure 6.20: *Evolvement through time of the estimated Hyperbolical probability density function for the devolatized log-returns of the Siemens data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

Here, the GARCH-AR volatility model competes with the Nadaraya-Watson for stationarity honours. Noticeable is the "peakiness" in the constant volatility case.

The Hyperbolical distribution does not fit the Siemens data as well as it fits the USD/DM data or the DAX data. The Nadaraya-Watson volatility model seems to be the best in this setting.

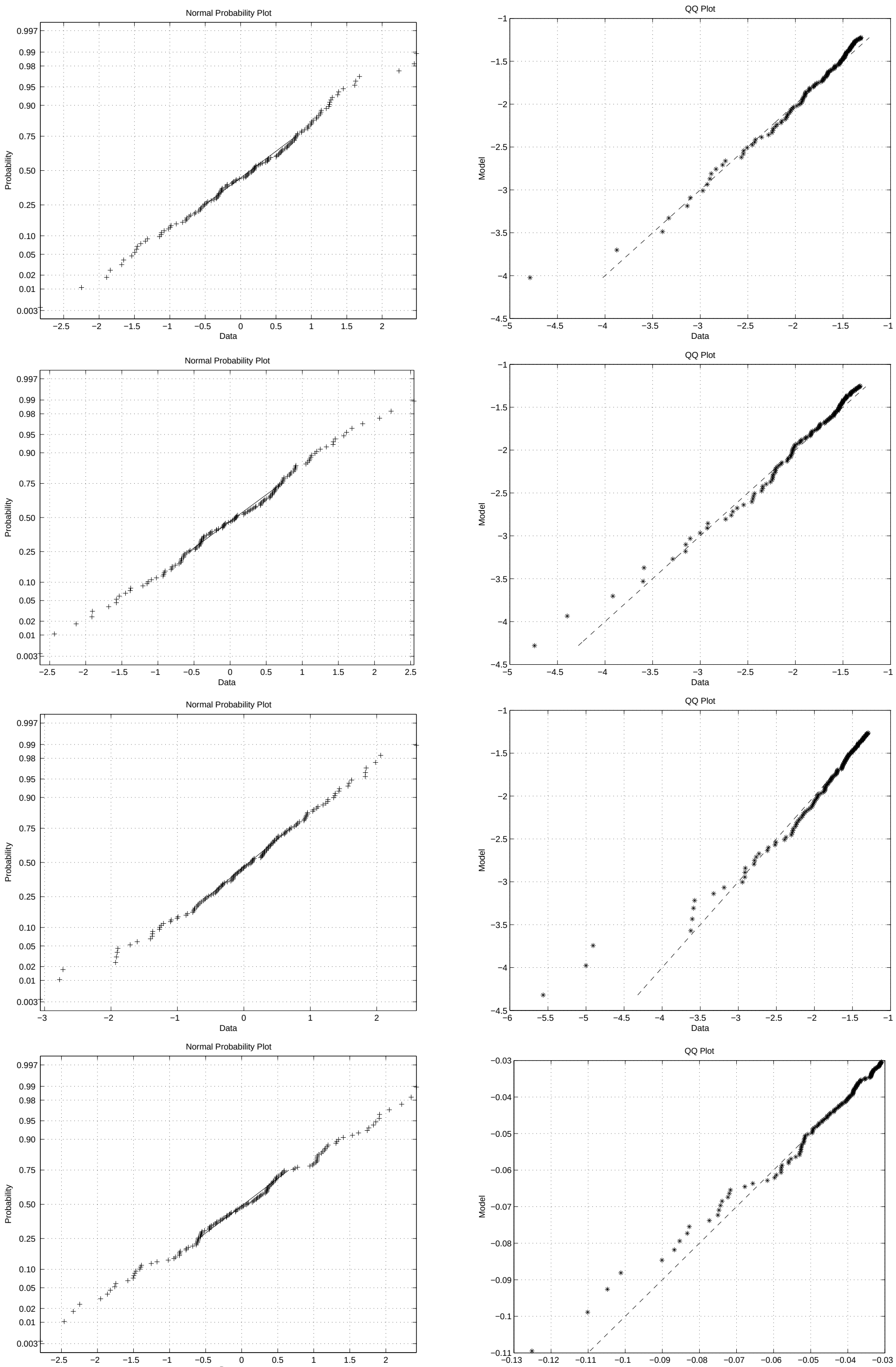


Figure 6.21: *PP- and QQ-plots for Siemens data, under assumption of Hyperbolical distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

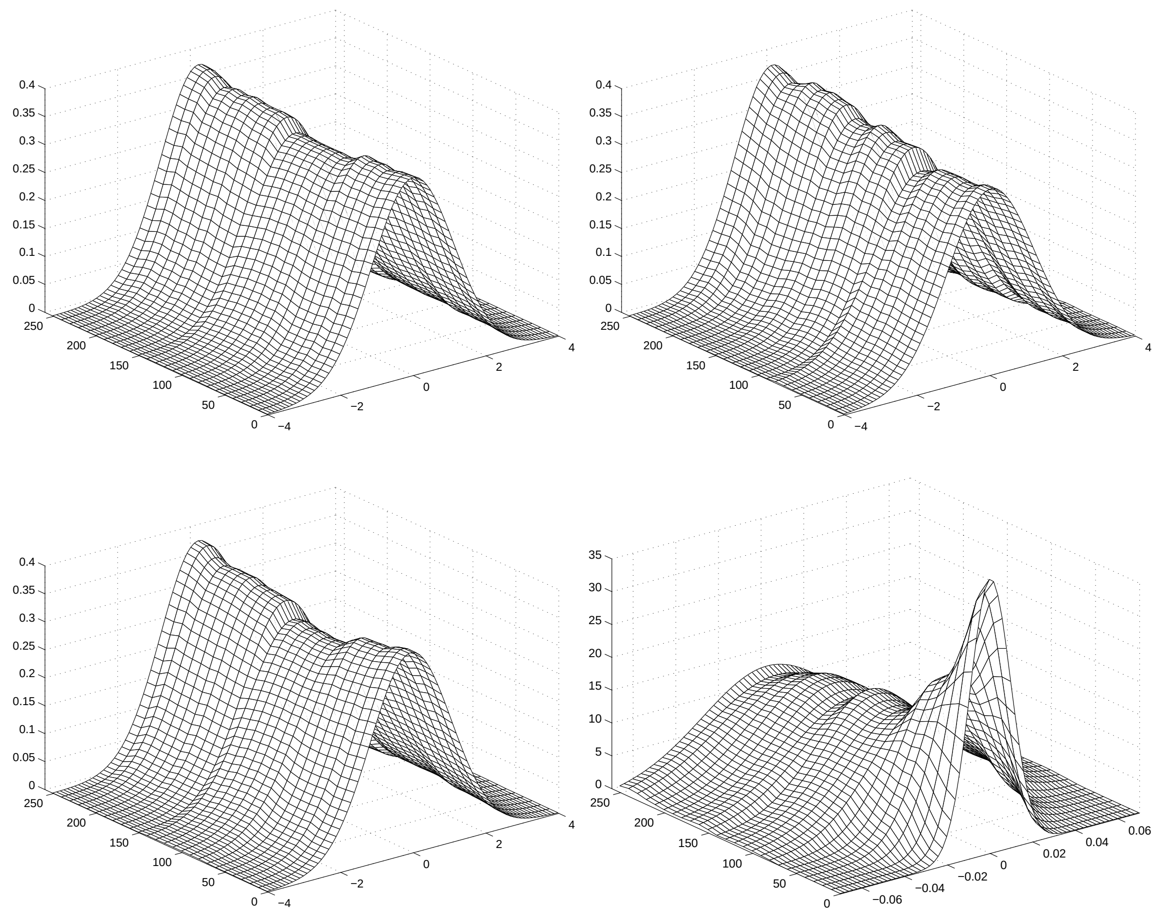


Figure 6.22: *Evolvement through time of the estimated Gaussian probability density function for the devolatized log-returns of the Siemens data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

None of the volatility models seem to make devolatized log-returns stationary.

Considering the result from the Kuiper test, the Hyperbolical and the Gaussian distributions seem to fit the Siemens data equally well. However, comparing the plots in figure 6.23 to the corresponding ones for the Hyperbolical distribution strongly suggests that the Hyperbolical distribution should be chosen over the Gaussian distribution for the Siemens data.

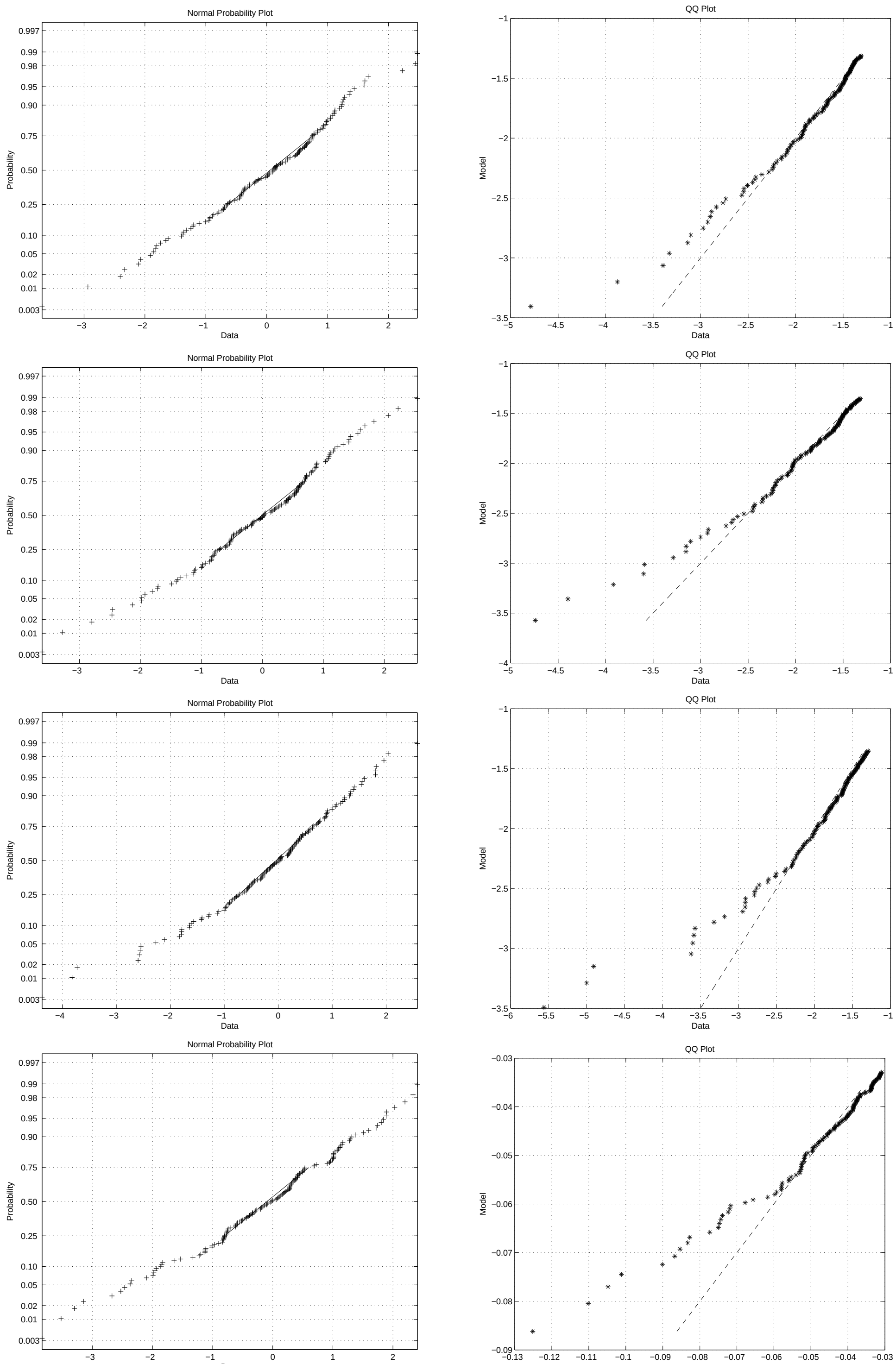


Figure 6.23: *PP- and QQ-plots for Siemens data, under assumption of Gaussian distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

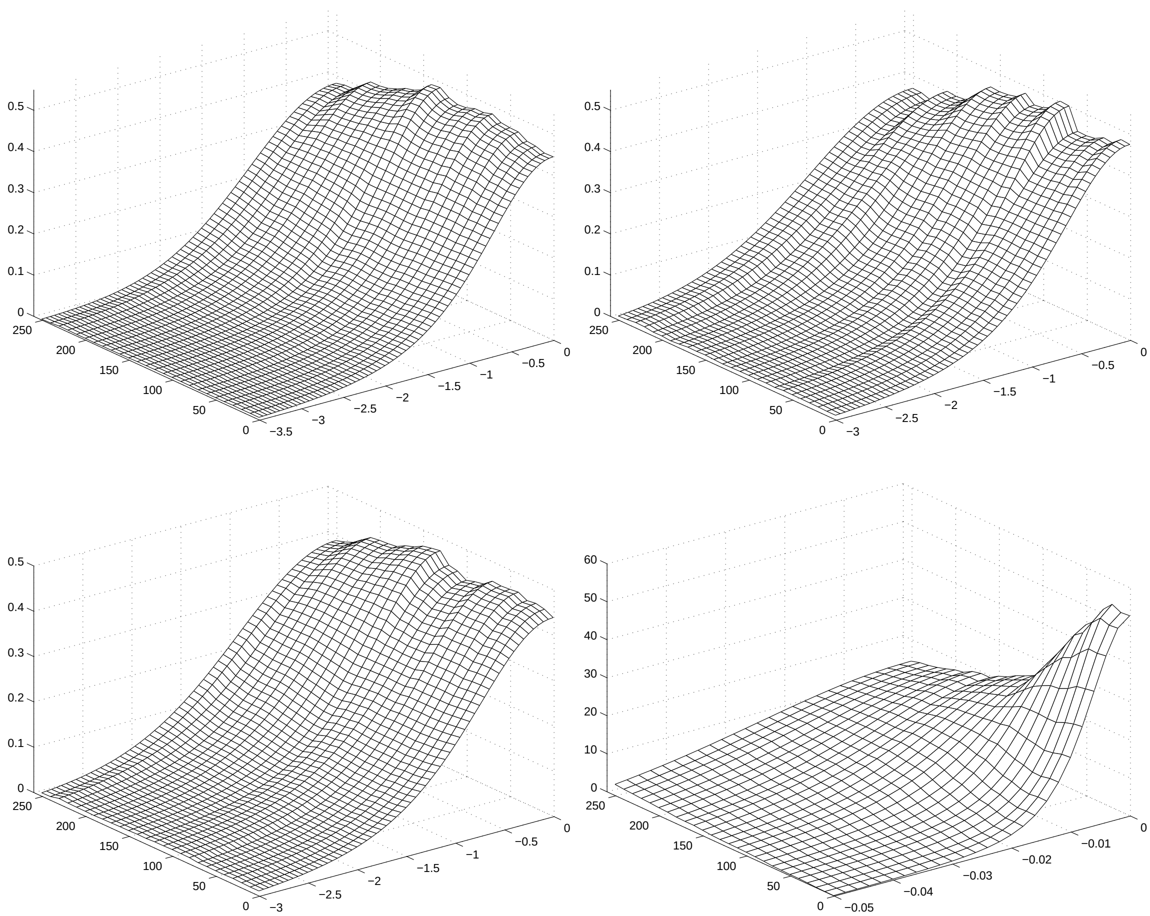


Figure 6.24: *Evolvement through time of the estimated Pearson probability density function for the devolatized log-returns of the Siemens data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

When it comes to stationarity, the Nadaraya-Watson and the GARCH-AR volatility models seem to be the best.

For all volatility models, the Pearson distribution once again performs really well. However, the Nadaraya-Watson seems to excel among volatility models.

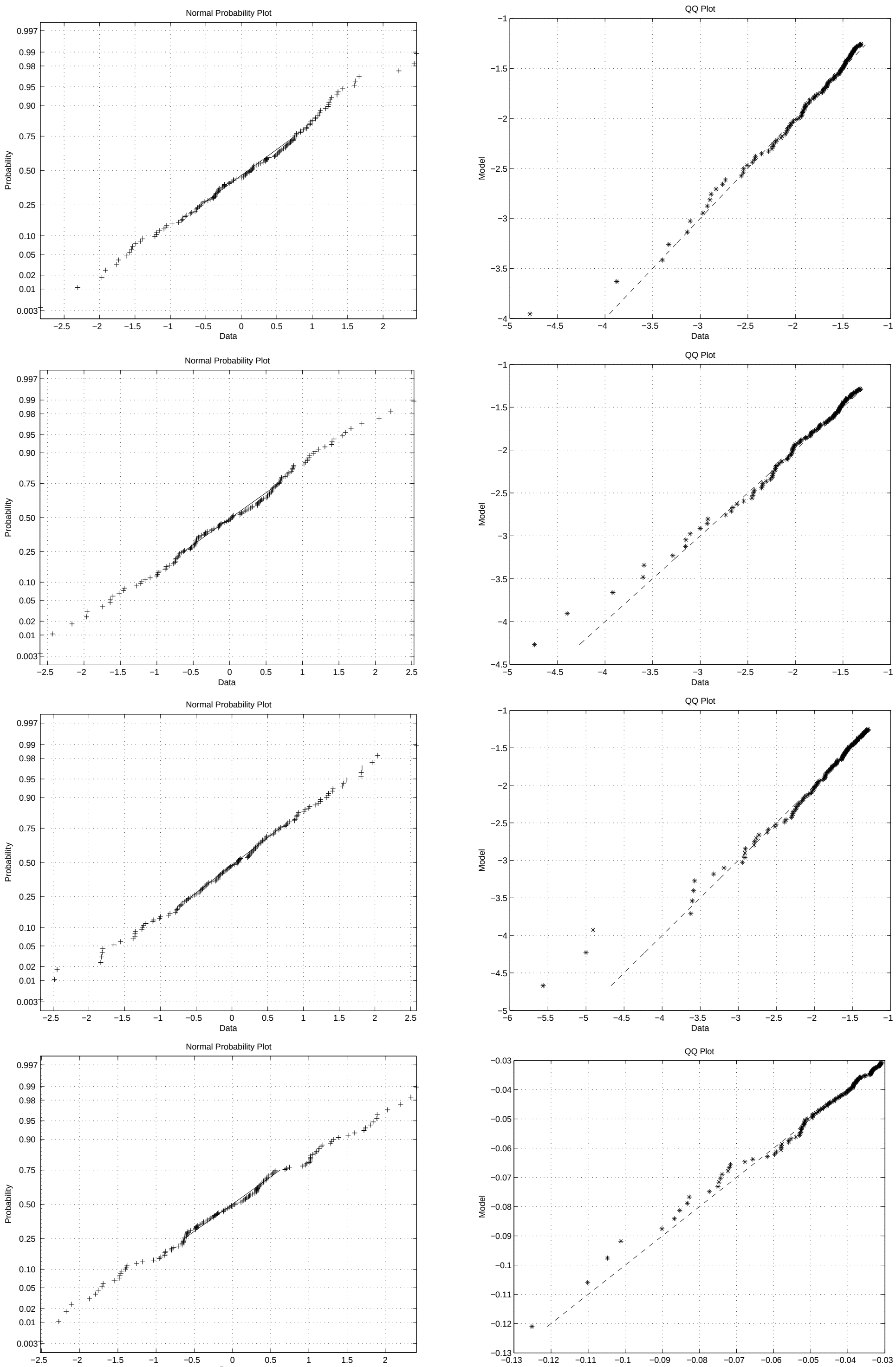


Figure 6.25: *PP- and QQ-plots for Siemens data, under assumption of Pearson distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

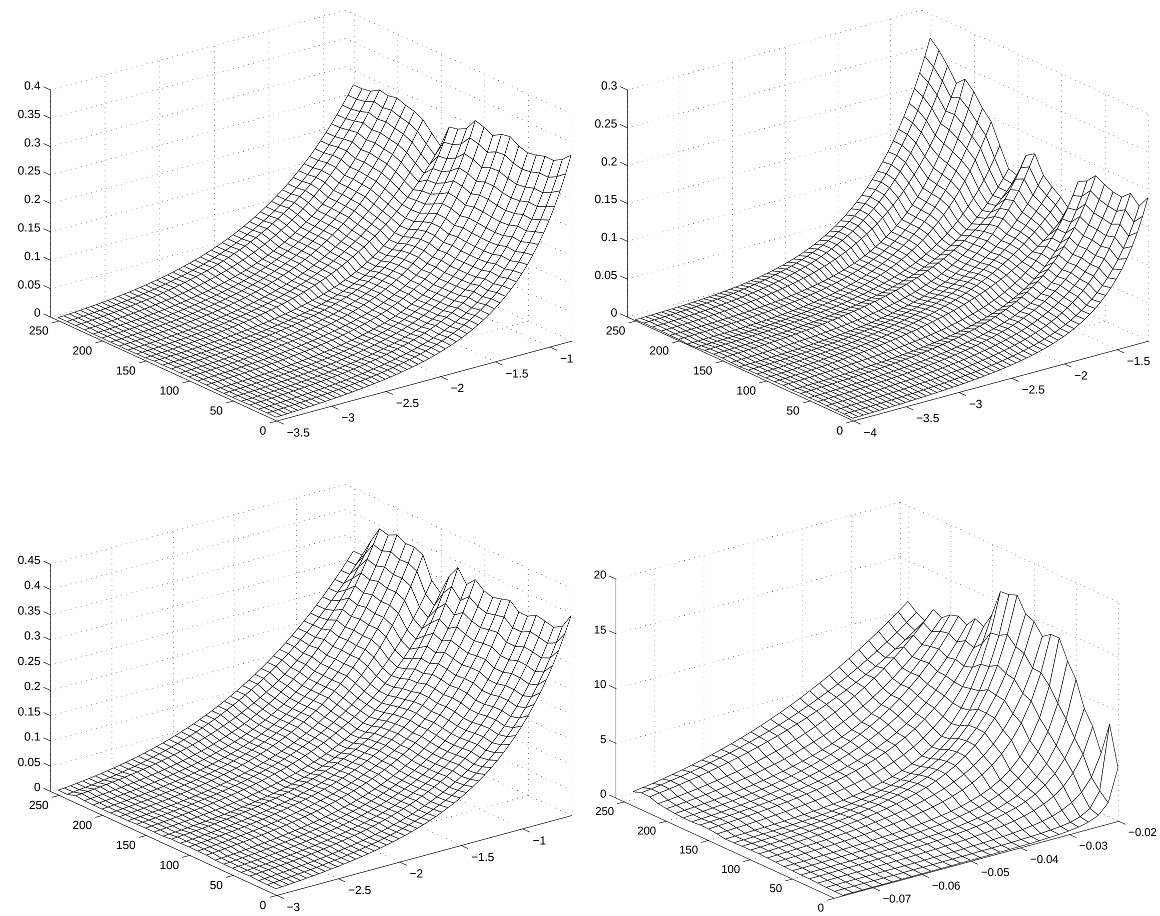


Figure 6.26: *Evolvement through time of the estimated GP probability density function for the devolatized log-returns of the Siemens data. Volatility models used are: Nadaraya-Watson (top left), Variance Window (top right),* GARCH-AR *(bottom left), Constant (bottom right).*

For the GP distribution, as for the Pearson distribution, devolatization using the Nadaraya-Watson, and the GARCH-AR volatility models seems to impose stationarity on the log-returns.

For the Siemens data, as is the case with the Pearson distribution, the GP distribution works really well with all volatility models. Using the plots in figures 6.26 and 6.27 to decide what volatility model to use is not easily done.

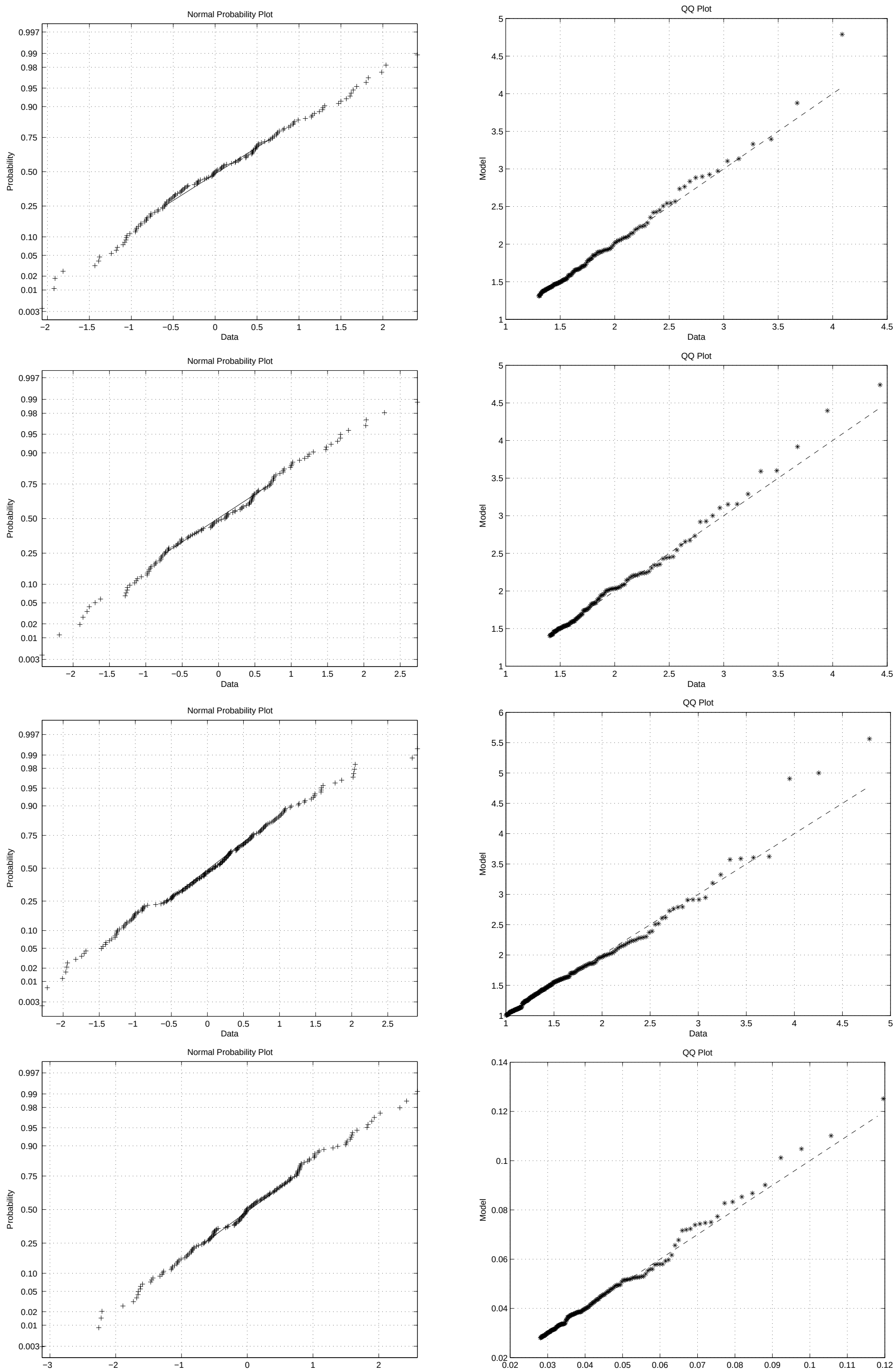


Figure 6.27: *PP- and QQ-plots for Siemens data, under assumption of GP distribution. Volatility models used are: Nadaraya-Watson (first row), Variance Window (second row),* GARCH-AR *(third row), Constant (fourth row).*

6.4 Tails

The *tail* (lower) of a distribution is defined as

$$\mathrm{T}_X^l(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

Remark Equivalently, the upper tail is defined as $\mathrm{T}_X^u(x) = \mathbb{P}(X \geq x) = \int_x^\infty f_X(t)dt$.

In the following plots, we graph the estimated lower tails for each distribution, and each volatility model. The Hyperbolical distribution is represented by a solid thick line (**-**), the GP distribution by a solid line ($-$), the Pearson VII distribution by a dotted line ($-.$), and the Gaussian distribution is represented by a broken line ($--$).

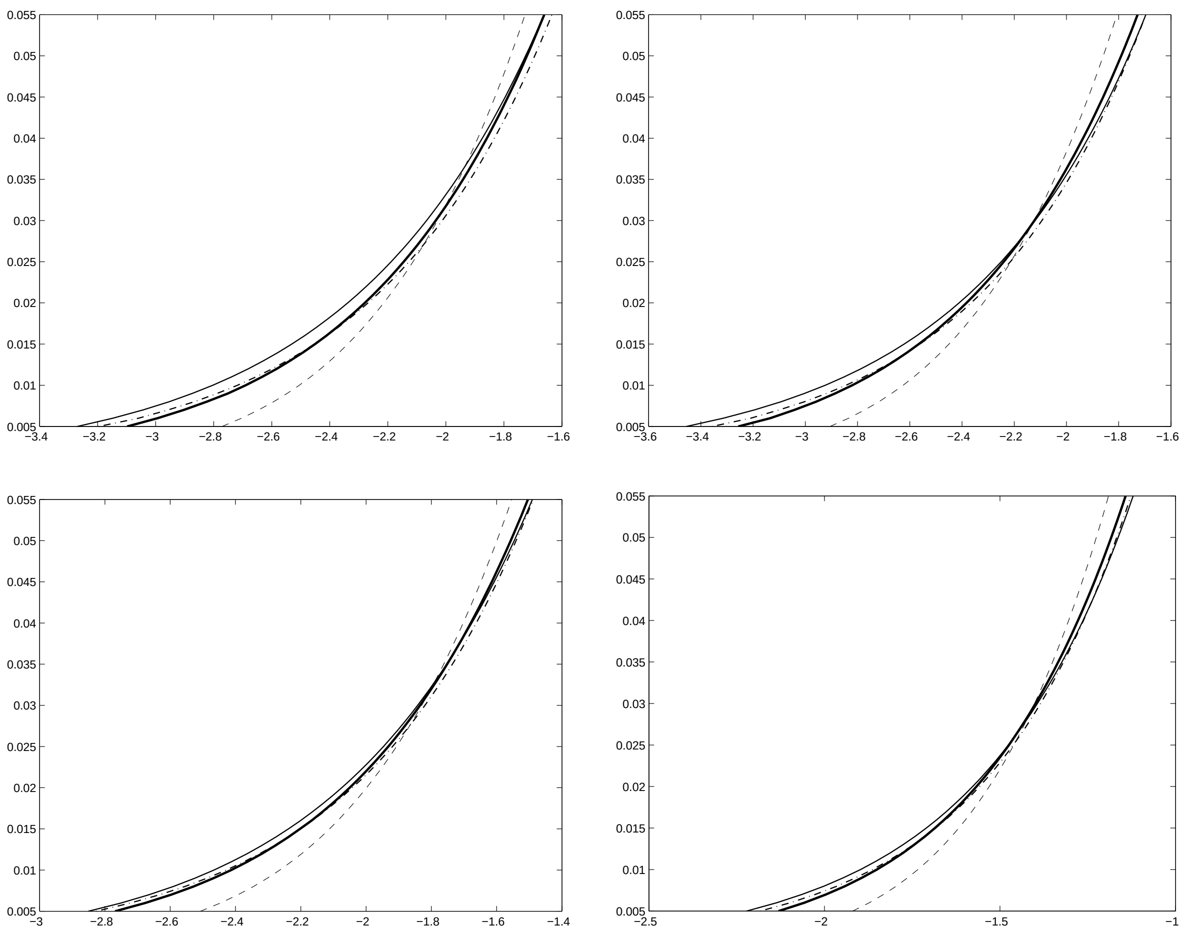


Figure 6.28: *Tail-plots, USD/DM data. Top left: Nadaraya-Watson, top right: Variance Window, lower left:* GARCH-AR *and lower right: Constant volatility.*

It is obvious, that assuming Gaussian distribution for log-returns will underestimate the VaR. Even when devolatized, the Gaussian tail is too light. The tails of the other three distributions are much heavier.

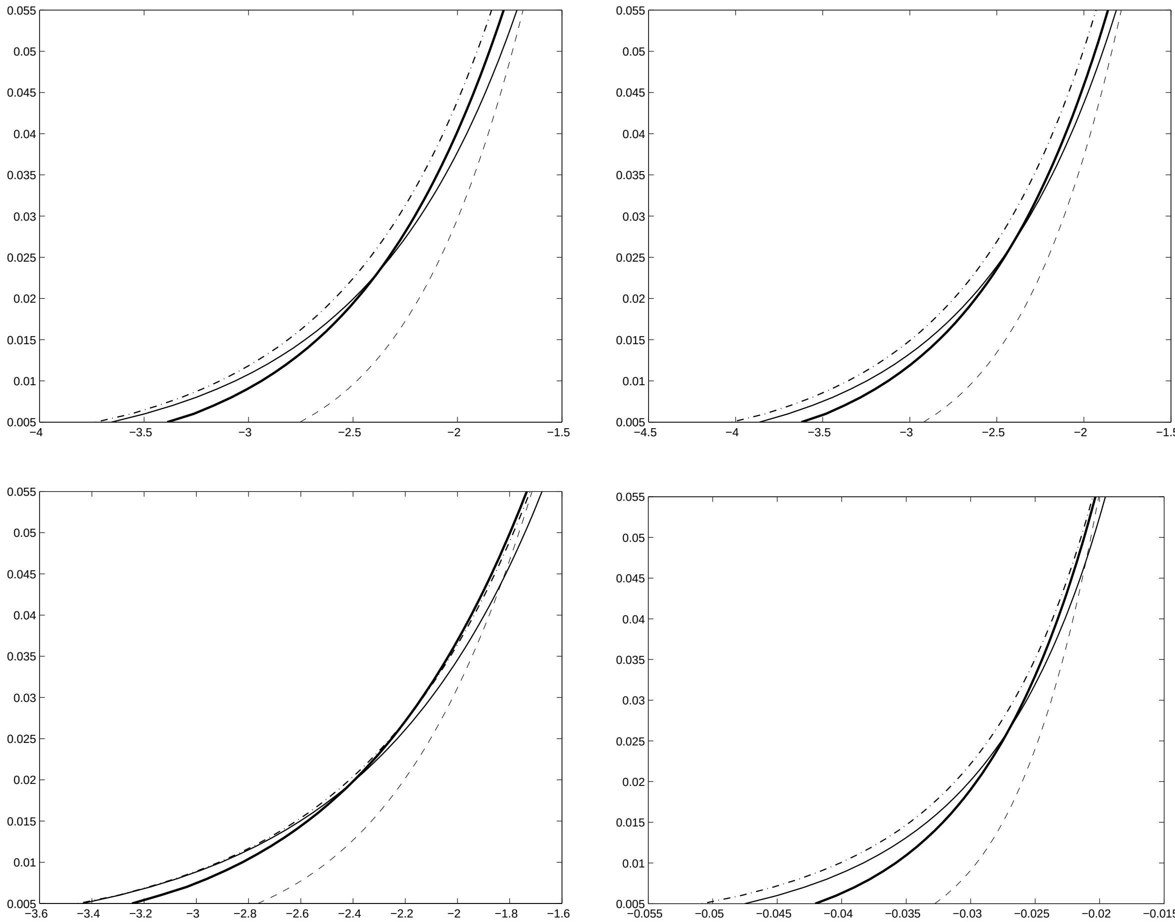


Figure 6.29: *Tail-plots, DAX data. Top left: Nadaraya-Watson, top right: Variance Window, lower left:* GARCH-AR *and lower right: Constant volatility.*

Here, we see even bigger discrepancies between the Gaussian distribution and the other three distributions, than we did for the USD/DM data. For the USD/DM data, the GP gave the heaviest tail, but for the DAX data, the Pearson distribution has a heavier tail than all the three other distributions. These tail plots further strengthen the case against using Gaussian distributions in finance, especially in risk management.

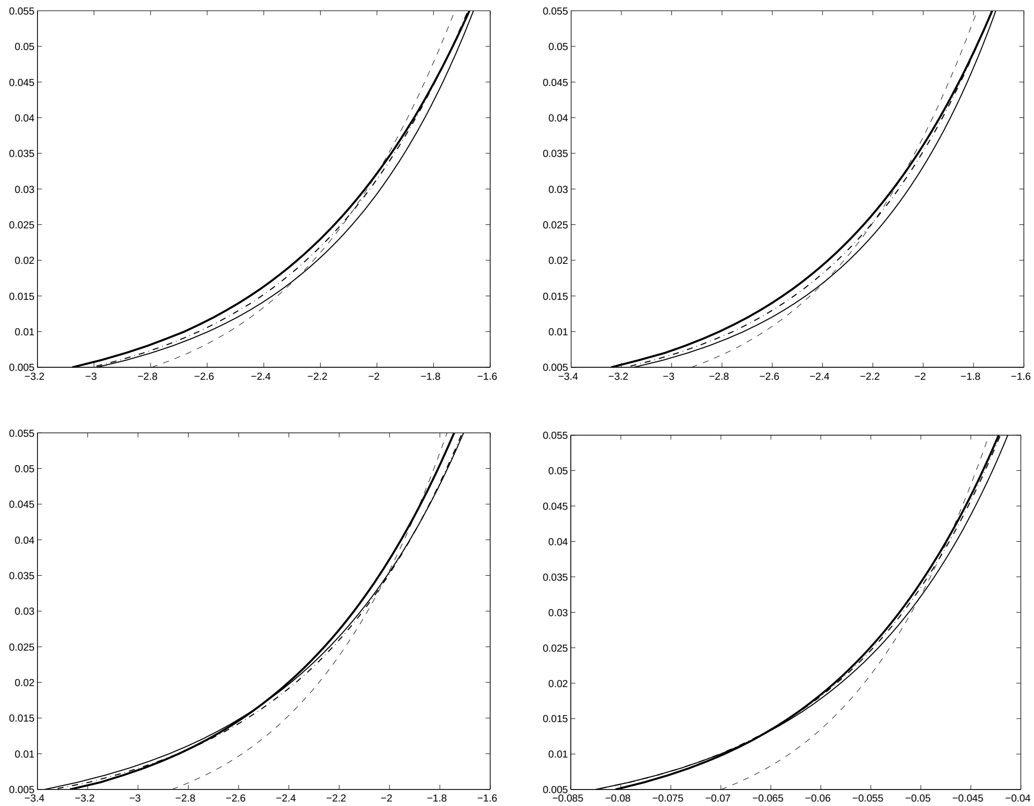


Figure 6.30: *Tail-plots, Siemens data. Top left: Nadaraya-Watson, top right: Variance Window, lower left:* GARCH-AR *and lower right: Constant volatility.*

When using Nadaraya-Watson and Variance Window volatility, the Gaussian tail is considerably closer to the other three. Still, when it comes to the GARCH-AR, and the constant volatility, there are quite big differences. For the constant volatility, the tails of the hyperbolic, the Pearson and the GP distribution are almost the same. This indicates that using these three models will produce almost the same result.

6.5 Backtesting (Performance of VaR Estimates)

One might ask, how good is a VaR prediction? There are a few things that have to be considered. We do not want a model that either overestimates or underestimates the risk of holding a financial asset. This means that the model has to allow for a few excesses of the limit specified by the VaR prediction, while keeping an as good as possible fit to the peaks of the smallest log-returns. In the diagrams below, we see the performance of the various models used. When studying an α percent VaR-plot, ideally one should observe a 100-α percent frequency of excessive losses. Also the times of excessive losses should be independently distributed.

We will use two methods to evaluate VaR estimates, namely, the *Frequency of Excessive Losses* test (FOEL) and the *Lopez Probability Forecasting Approach* (LPFA).

6.5.1 Frequency of Excessive Losses

One way to test the assumption of 100-α percent frequency of excessive losses, is to use a FOEL test. It is a likelihood ratio test, proposed by Kupiec, see also [33]. The FOEL test is based on the idea that if the predicted VaR is correct, then the number of excessive losses follow a binomial distribution, with parameters p_0 and T. Here $1 - p_0$ is the predicted VaR-level (i.e. 95-, 97.5- or 99%), and T is the sample length. The test statistic is defined as

$$R(f, T, p_0) = -2 \log \left[(1 - p_0)^{T-f} p_0^f \right] + 2 \log \left[(1 - \frac{f}{T})^{T-f} (\frac{f}{T})^f \right],$$

where f is the number of days in the sample that exceeds the predicted VaR. Under the null hypothesis, the statistic is χ_1^2-distributed. The assumption of times of exceedances being independent is often violated. This might negatively affect the power of the FOEL test.

The results of the FOEL test are presented in tables. First the VaR level (α) is specified, then the frequency of excessive losses (FOEL), a confidence interval for the FOEL, the value of the statistic (R) and finally the p-value. The test is performed for α-levels 95, 97.5, and 99. Throughout we set $T = 1300$.

6.5.2 Lopez Probability Forecasting Approach

Lopez, see [37], proposes a method, as an alternative to the FOEL test (and to other statistical evaluation methods, see also [22]). The LPFA is a procedure that is meant to gauge the accuracy of VaR models, not by hypothesis testing, but by using a forecast evaluation criterion. The idea is that we specify a forecast loss function and gauge the accuracy of the VaR forecasts, by how well they score, in terms of this loss function. The higher the score, the poorer the model.

We start by specifying the event of interest. In our case, such an event is an excessive loss (i.e. loss that exceeds the VaR estimate). Then, we predict the probability of this event in the next period (1 day). Theoretically, this should be $1 - \alpha$ for an VaR_α estimate. We also keep a record if this event happens in the period, or not. Finally, we choose a loss function, by which to evaluate the goodness of these forecasts, against subsequently realized outcomes. The particular loss function that Lopez uses (which also we will use), is the *Quadratic Probability Score* (QPS), due originally to Brier, see [17].

$$QPS = \frac{2}{T} \sum_{t=1}^{T} \left(p_t^f - I_t \right)^2,$$

where p_t^f is the predicted probability of an excessive loss in period t, and

$$I_t = \begin{cases} 1, & \text{if we observe an excessive loss in period } t; \\ 0, & \text{else.} \end{cases}$$

Lopez compares the accuracy of this approach, against the accuracy of several statistical approaches, by means of a simulation experiment. His results suggest that the power of the statistical procedures are generally rather low, and that such tests are likely to be inaccurate. His results also show that his loss funciton correctly identifies the "true" model in a large majority of simulated cases. A result that seems to confirm that the LPFA is likely to be quite reliable.

The QPS scores are presented in tables. For each VaR level and each volatility model, the density giving the lowest score is indicated.

6.5.3 USD/DM

Non-par.	95%	97.5%	99%	**Var. Win.**	95%	97.5%	99%
Hyperbolic	0.0771	0.0514	0.0248	Hyperbolical	0.0771	0.0561	0.0321
Gaussian	0.0816	**0.0441**	0.0248	Gaussian	0.0747	0.0537	0.0370
Pearson	**0.0656**	0.0561	0.0248	Pearson	**0.0656**	0.0585	**0.0075**
GP	0.0797	0.0513	**0.0150**	GP	0.0773	**0.0514**	0.0175
GARCH-AR	95%	97.5%	99%	**Constant**	95%	97.5%	99%
Hyperbolic	0.0794	0.0514	0.0272	Hyperbolic	0.0794	0.0514	0.0248
Gaussian	0.0770	0.0561	0.0321	Gaussian	0.0816	**0.0465**	0.0272
Pearson	**0.0726**	0.0561	0.0248	Pearson	**0.0657**	0.0514	0.0248
GP	0.0842	**0.0513**	**0.0150**	GP	0.0799	0.0489	**0.0150**

Table 6.8: *Results from the* LPFA*.*

The table shows that, both the Pearson and the GP distributions perform well. It is arguably remarkable that the Gaussian distribution gets the lowest score on two occasions (for the 97.5% level.)

NIG

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Integrated Vol.	95%	5.6%	[4.4%,6.9%]	1.00	0.318	Not Rejected
Integrated Vol.	97.5%	3.5%	[2.5%,4.6%]	4.42	0.036	Rejected
Integrated Vol.	99%	2.1%	[1.3%,2.9%]	11.6	< 0.001	Rejected

Table 6.9: FOEL *test if returns are assumed to have a* NIG*law.*

Though this approach works well at the 95%-level, it fails completely further out (i.e at 97.5 and 99% levels) in the tail. However, this may be due to insufficiently many intra-day observations.

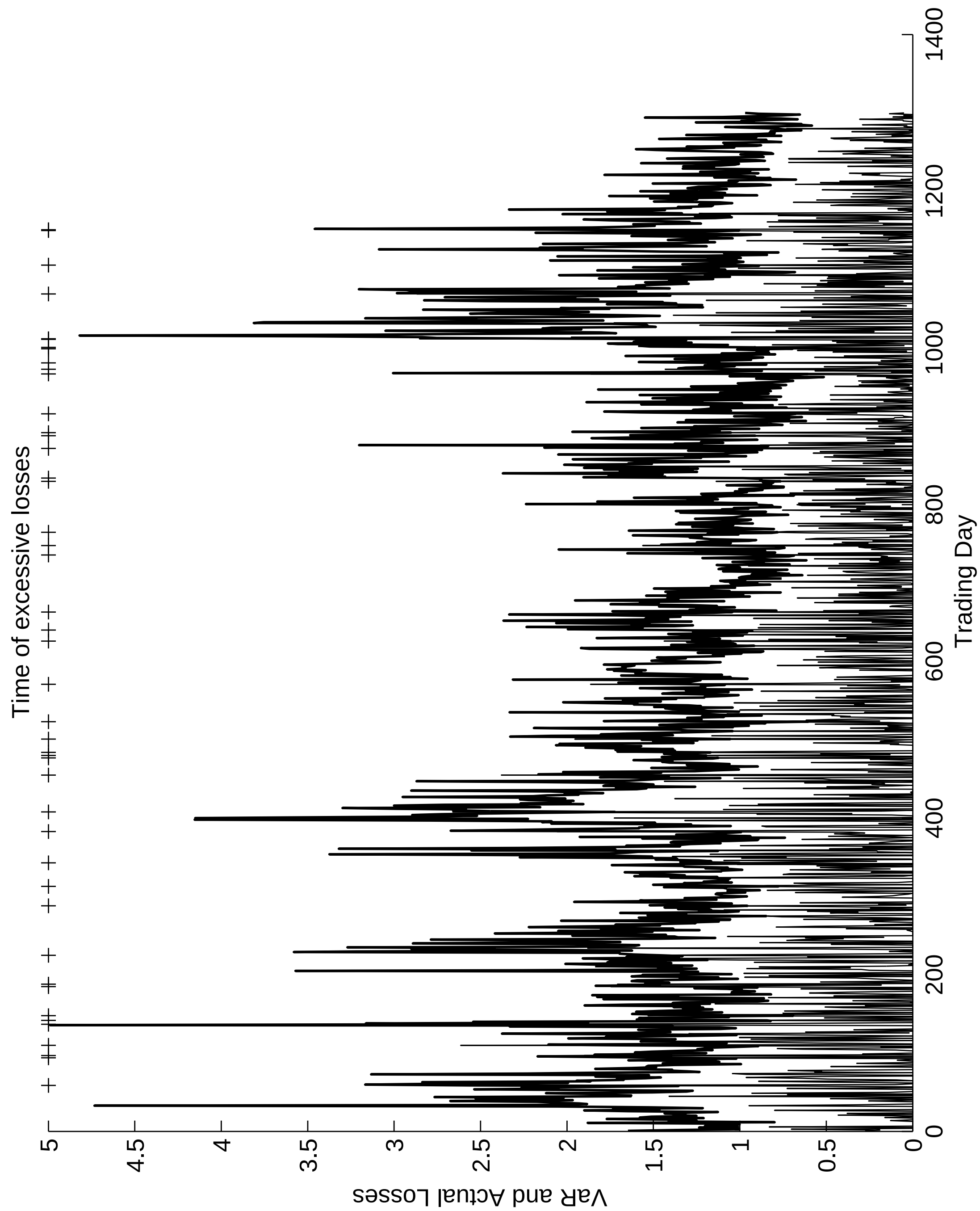


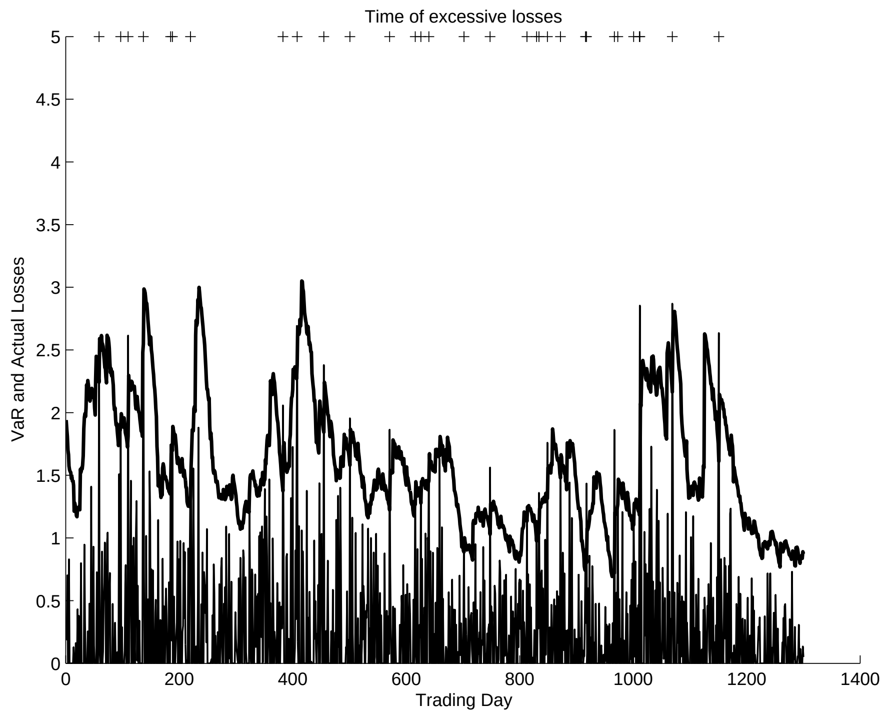
Figure 6.31: VaR *(97.5%) predictions, and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Log-retruns are assumed* NIG *distributed, and the volatility model used is Integrated Volatility.*

Hyperbolical

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.1%	[3.0%,5.2%]	2.48	0.115	Not Rejected
Var-Win	95%	4.6%	[3.5%,5.8%]	0.42	0.519	Not Rejected
GARCH-AR	95%	3.8%	[2.8%,4.9%]	3.95	0.047	Rejected
Constant	95%	4.6%	[3.5%,5.8%]	0.42	0.519	Not Rejected
Nadaraya-Watson	97.5%	2.4%	[1.6%,3.2%]	0.07	0.79	Not Rejected
Var-Win	97.5%	2.2%	[1.4%,2.9%]	0.67	0.413	Not Rejected
GARCH-AR	97.5%	2.5%	[1.7%,3.4%]	0.01	0.929	Not Rejected
Constant	97.5%	2.5%	[1.6%,3.3%]	0.01	0.929	Not Rejected
Nadaraya-Watson	99%	1.0%	[0.5%,1.5%]	0	1.000	Not Rejected
Var-Win	99%	1.1%	[0.5%,1.6%]	0.08	0.783	Not Rejected
GARCH-AR	99%	1.2%	[0.6%,1.8%]	0.65	0.420	Not Rejected
Constant	99%	0.9%	[0.4%,1.4%]	0.08	0.778	Not Rejected

Table 6.10: FOEL *test under assumption of devolatized log-returns being Hyperbolically distributed,* USD/DM *data.*

The Hyperbolical distribution works well with all volatility models. Note that the GARCH-AR fails at the 95% level. For all volatility models, the Hyperbolical distribution seems to overestimate the VaR at the 95% level (the FOEL lies between $3.8 - 4.6\%$). On the other hand, it produces an excellent result at both the 97.5% and the 99% level.



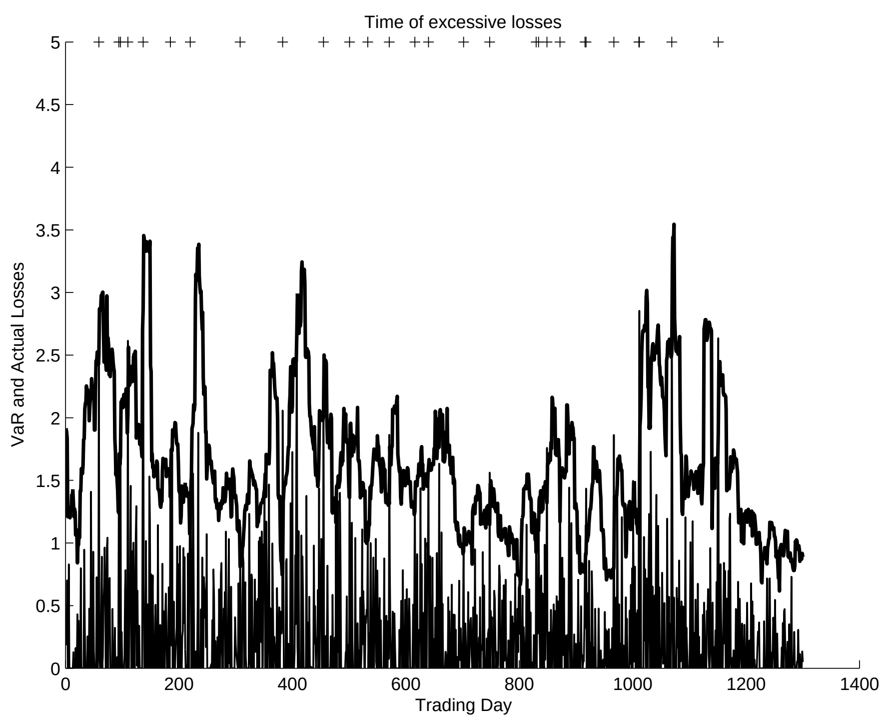


Figure 6.32: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Hyperbolically distributed.*

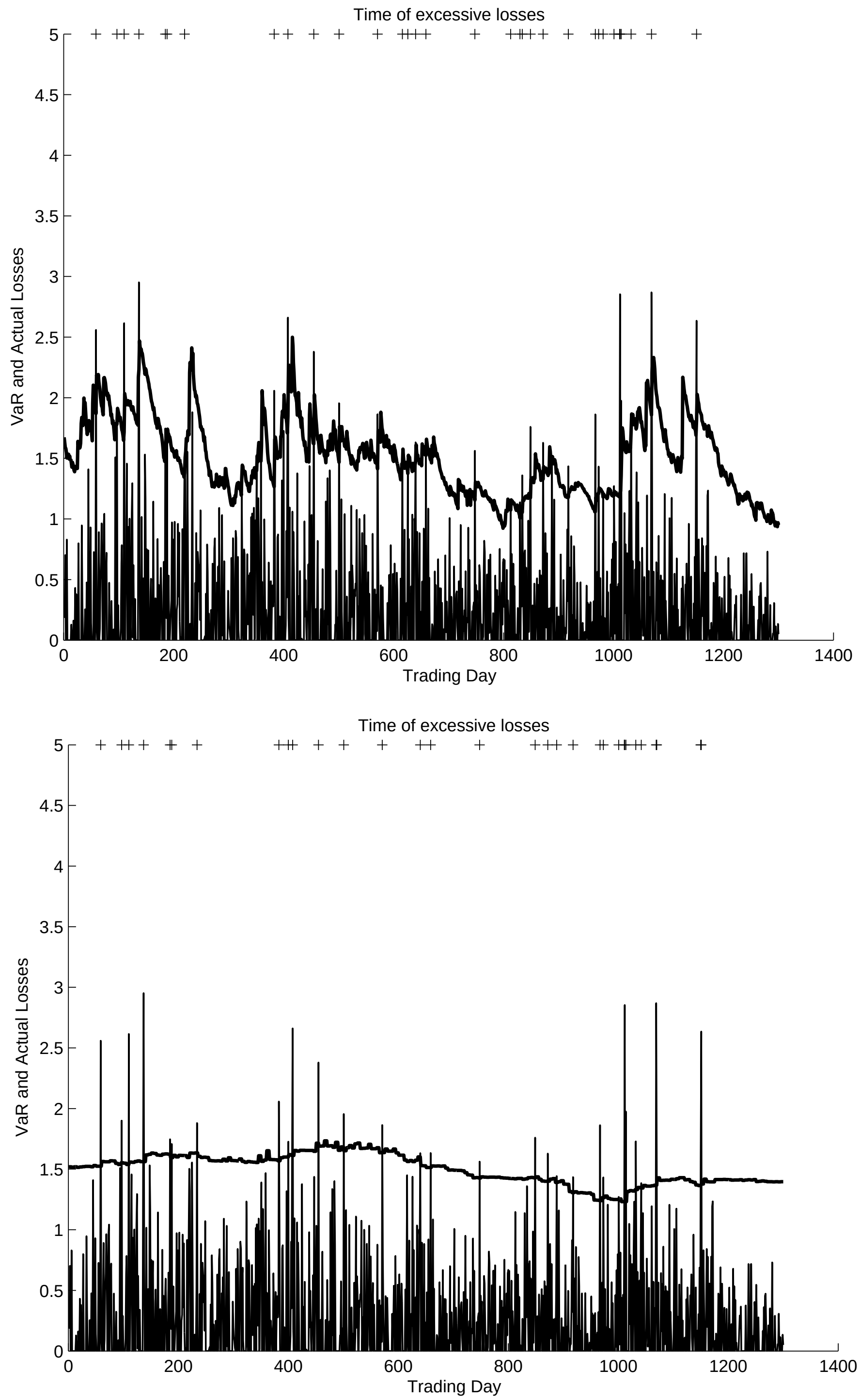


Figure 6.33: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Hyperbolically distributed.*

Gaussian

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.2%	[3.1%,5.2%]	2.07	0.150	Not Rejected
Var-Win	95%	3.9%	[2.8%,5.0%]	3.41	0.065	Not Rejected
GARCH-AR	95%	3.8%	[2.8%,4.9%]	3.95	0.047	Rejected
Constant	95%	4.3%	[3.2%,5.4%]	1.37	0.241	Not Rejected
Nadaraya-Watson	97.5%	2.6%	[1.7%,3.5%]	0.07	0.791	Not Rejected
Var-Win	97.5%	2.6%	[1.7%,3.5%]	0.07	0.791	Not Rejected
GARCH-AR	97.5%	2.7%	[1.8%,3.6%]	0.19	0.661	Not Rejected
Constant	97.5%	2.6%	[1.7%,3.5%]	0.07	0.791	Not Rejected
Nadaraya-Watson	99%	1.5%	[2.2%,3.3%]	3.26	0.071	Not Rejected
Var-Win	99%	1.6%	[0.9%,2.3%]	4.19	0.041	Rejected
GARCH-AR	99%	0.4%	[0.1%,0.7%]	6.50	0.011	Rejected
Constant	99%	1.2%	[0.6%,1.7%]	0.30	0.586	Not Rejected

Table 6.11: FOEL *test if devolatized returns are assumed to be Gaussian distributed,* USD/DM *data.*

As in the Hyperbolical distribution case, the Gaussian distribution, using the GARCH-AR volatility model, fails the FOEL test at the 95%, but not at the 97.5% VaR level. Using Variance Window or GARCH-AR volatility models, it also (as expected) fails, at the 99% VaR level. Remarkably, with constant volatility, the Gaussian distribution passes the FOEL test, at the 99% VaR level.

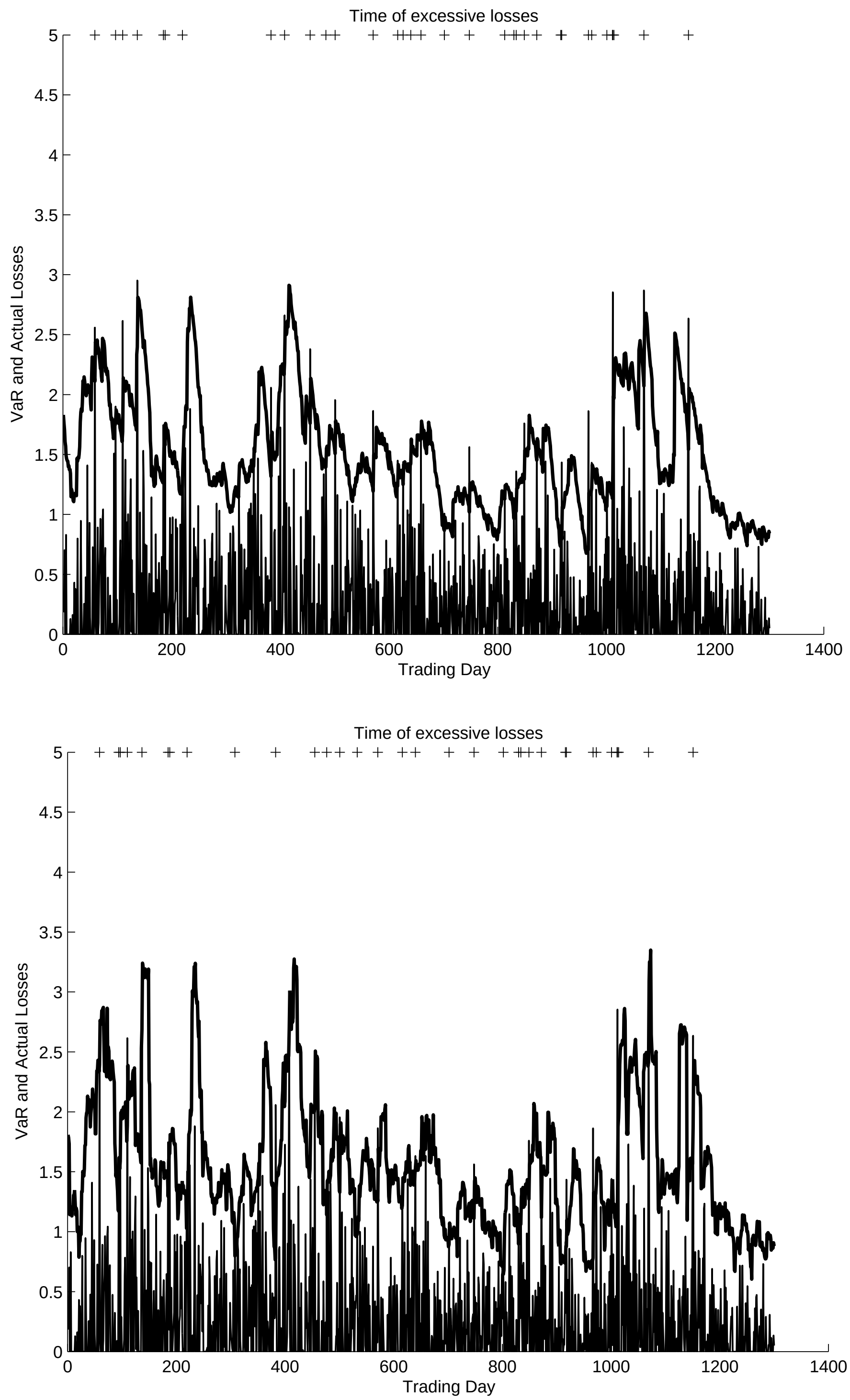
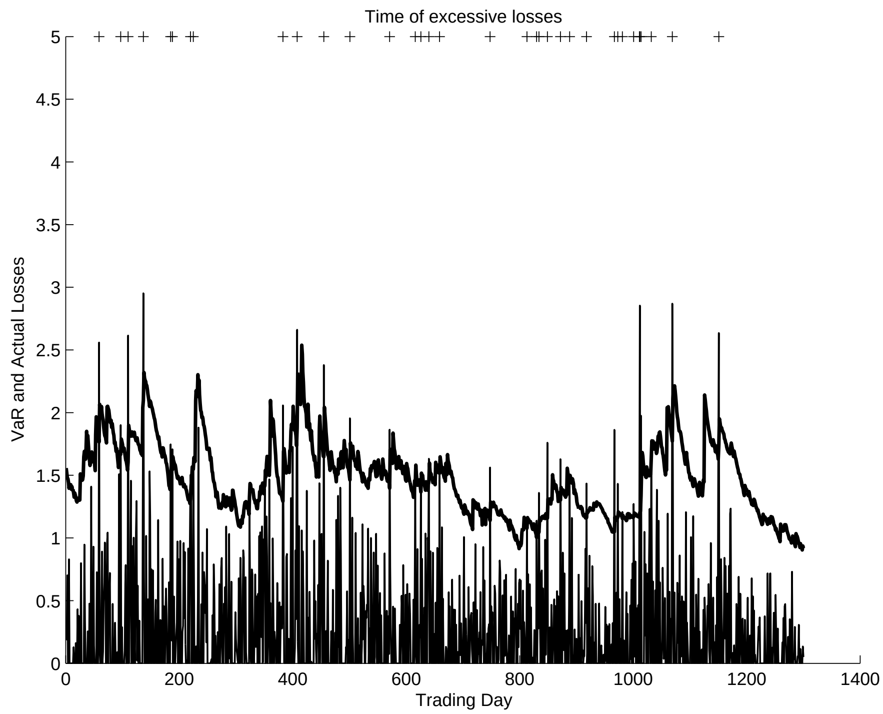


Figure 6.34: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Gaussian distributed.*



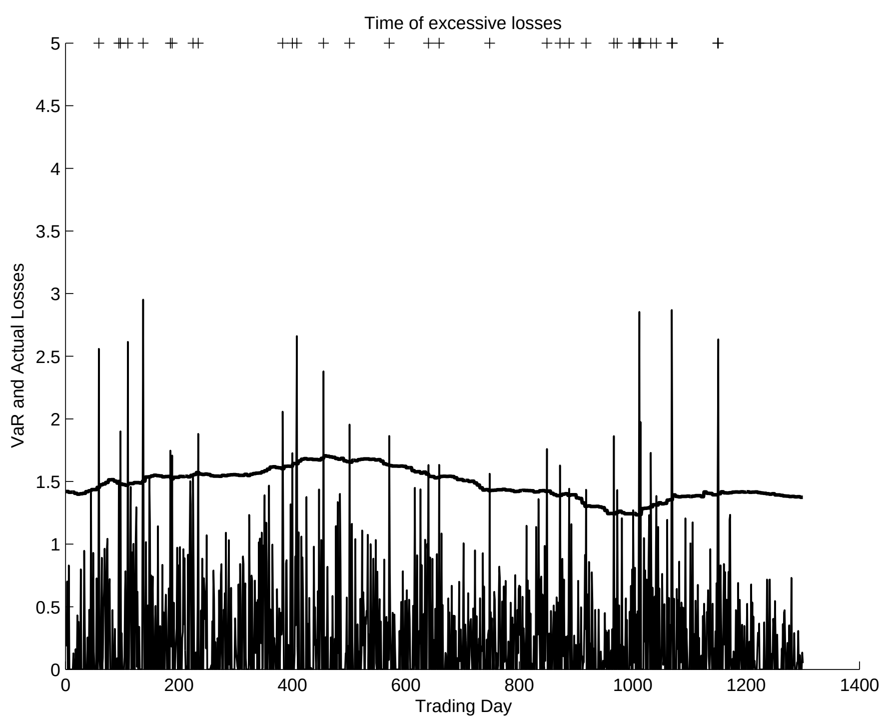


Figure 6.35: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Gaussian distributed.*

Pearson

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.2%	[3.1%,5.2%]	2.07	0.150	Not Rejected
Var-Win	95%	4.6%	[3.5%,5.8%]	0.42	0.519	Not Rejected
GARCH-AR	95%	4.1%	[3.0%,5.2%]	2.48	0.115	Not Rejected
Constant	95%	4.6%	[3.5%,5.8%]	0.42	0.519	Not Rejected
Nadaraya-Watson	97.5%	2.4%	[1.6%,3.2%]	0.07	0.788	Not Rejected
Var-Win	97.5%	2.3%	[1.5%,3.1%]	0.20	0.653	Not Rejected
GARCH-AR	97.5%	2.5%	[1.7%,3.4%]	0.01	0.929	Not Rejected
Constant	97.5%	2.5%	[1.6%,3.3%]	0.01	0.929	Not Rejected
Nadaraya-Watson	99%	1.1%	[0.5%,1.6%]	0.08	0.783	Not Rejected
Var-Win	99%	1.2%	[0.6%,1.8%]	0.65	0.420	Not Rejected
GARCH-AR	99%	1.2%	[0.6%,1.8%]	0.65	0.420	Not Rejected
Constant	99%	0.9%	[0.4%,1.4%]	0.08	0.778	Not Rejected

Table 6.12: FOEL *test if devolatized returns are assumed to be Pearson distributed,* USD/DM *data.*

As seen in the table, the Pearson distribution performs really well. It seems to overestimate the VaR at the 95% level. At both the 97.5% and 99% VaR levels, it looks excellent.

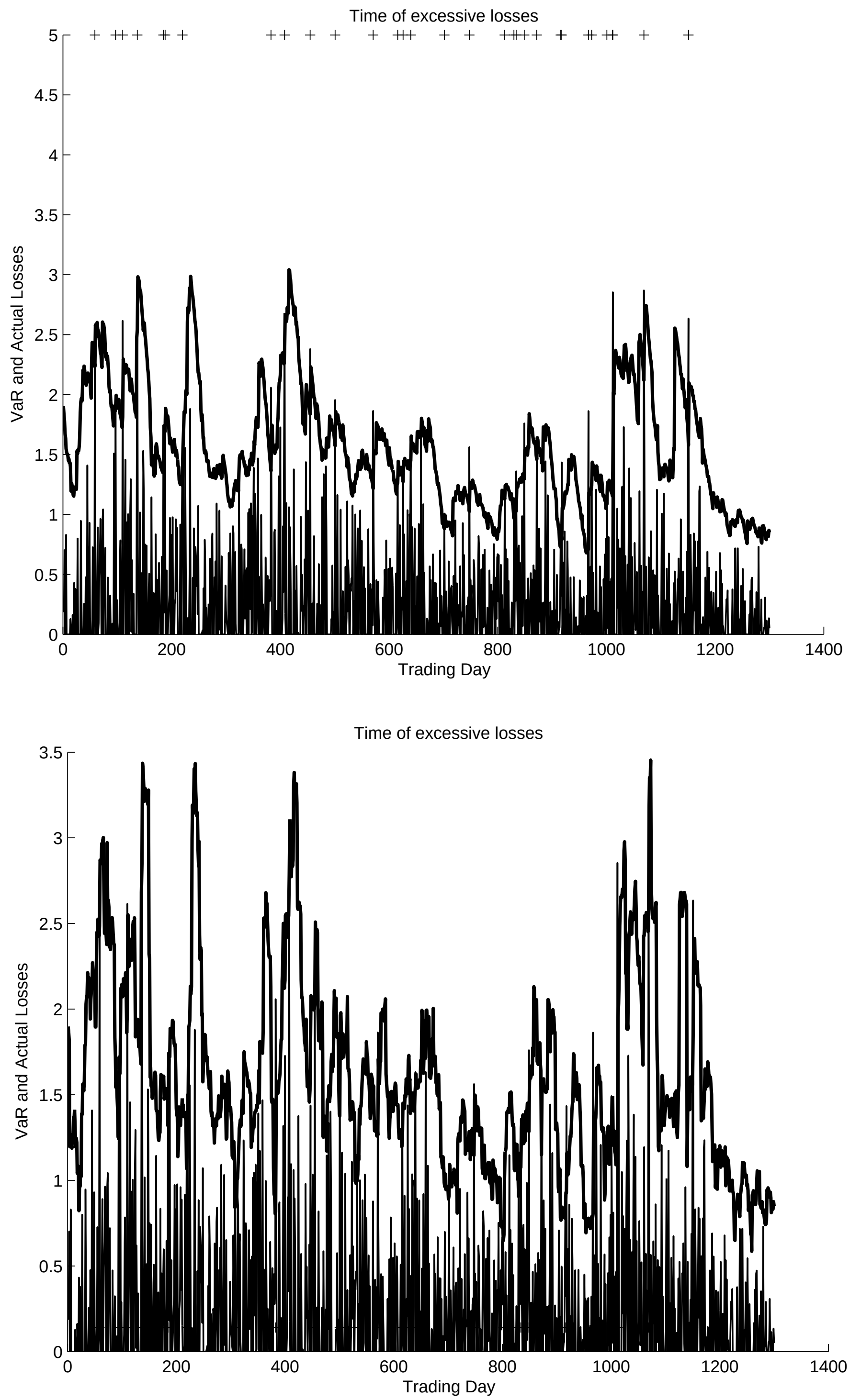


Figure 6.36: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Pearson distributed.*

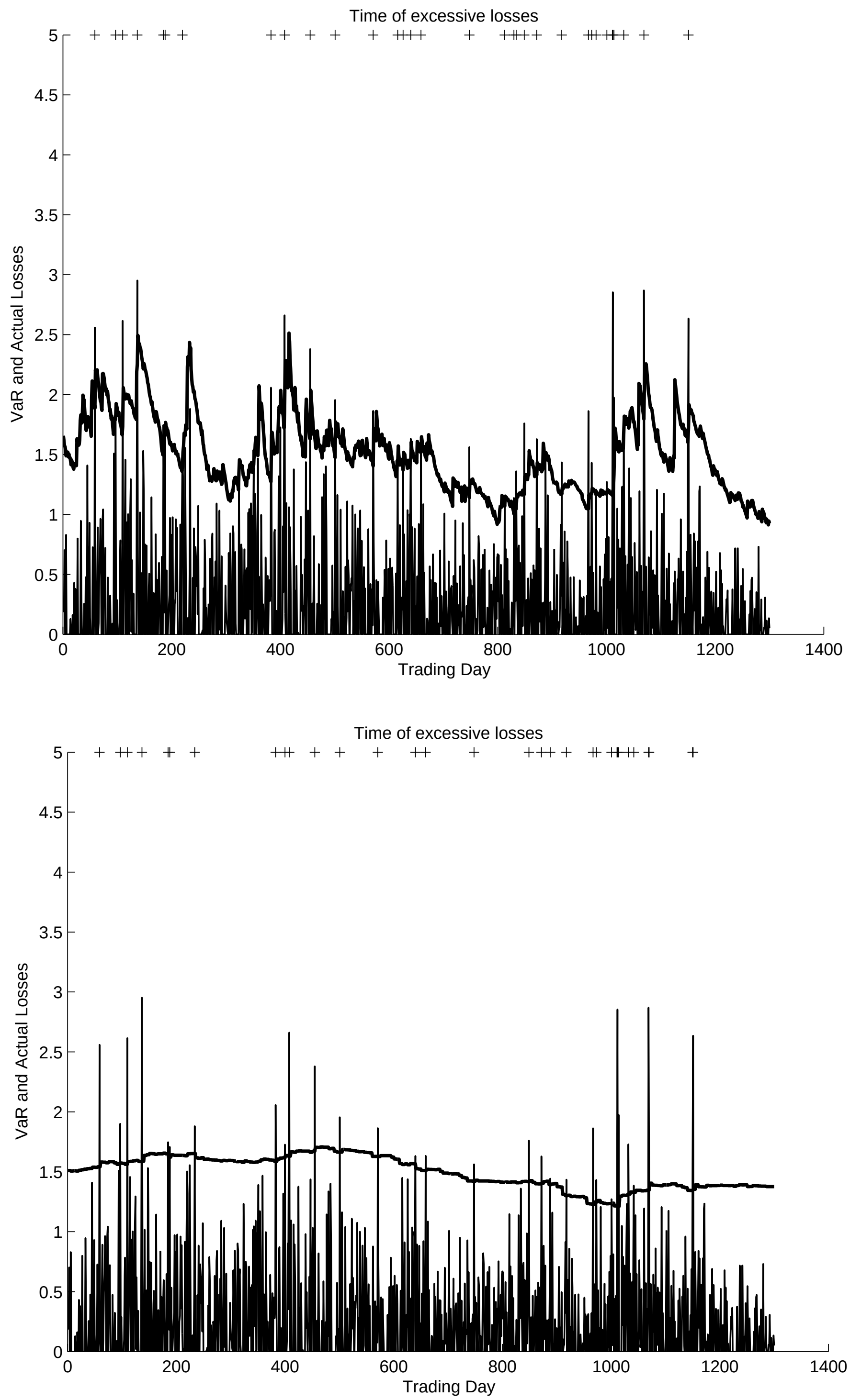


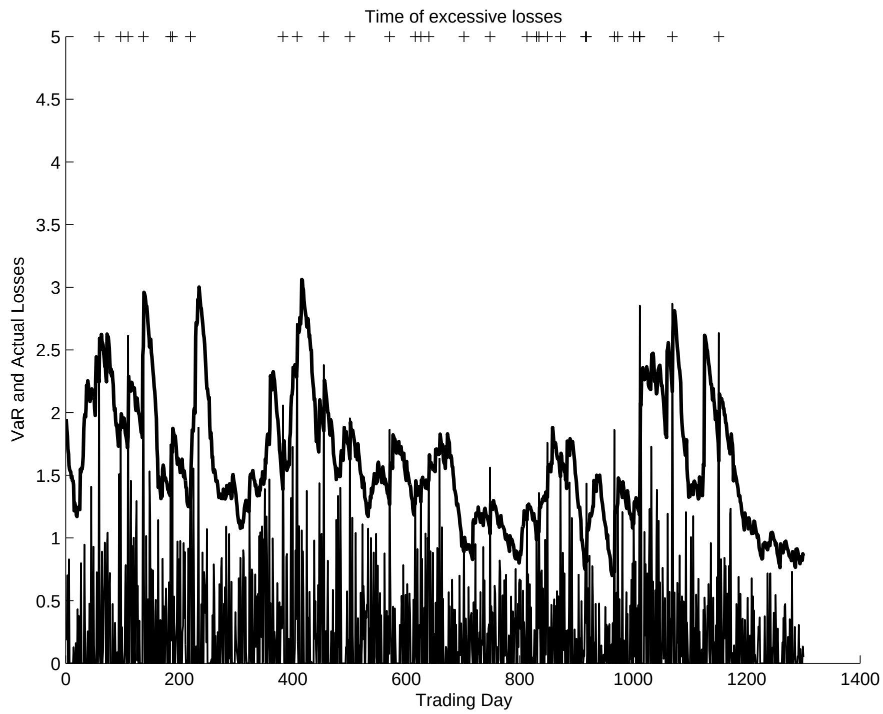
Figure 6.37: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Pearson distributed.*

GP

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.2%	[3.1%,5.3%]	1.70	0.192	Not Rejected
Var-Win	95%	4.1%	[3.0%,5.2%]	2.48	0.115	Not Rejected
GARCH-AR	95%	3.9%	[2.9%,5.0%]	3.42	0.065	Not Rejected
Constant	95%	4.8%	[3.6%,5.9%]	0.15	0.701	Not Rejected
Nadaraya-Watson	97.5%	2.4%	[1.6%,3.2%]	0.07	0.788	Not Rejected
Var-Win	97.5%	2.3%	[1.5%,3.1%]	0.20	0.653	Not Rejected
GARCH-AR	97.5%	2.4%	[1.6%,3.2%]	0.07	0.788	Not Rejected
Constant	97.5%	2.5%	[1.6%,3.3%]	0.001	0.929	Not Rejected
Nadaraya-Watson	99%	1.0%	[0.5%,1.5%]	0	1.000	Not Rejected
Var-Win	99%	1.2%	[0.6%,1.8%]	0.65	0.420	Not Rejected
GARCH-AR	99%	1.2%	[0.6%,1.8%]	0.65	0.42	Not Rejected
Constant	99%	0.9%	[0.4%,1.4%]	0.08	0.778	Not Rejected

Table 6.13: FOEL *test if devolatized returns are assumed to be GP distributed,* USD/DM *data.*

At the 95% VaR level, the GP distribution seems to greatly overestimate the VaR. Going further out in the tail, the GP distribution works really well. As for the other distributions, the GP distribution performs poorly together with the GARCH-AR volatility, at the 95% VaR level. We see that the Nadaraya-Watson volatility model, combined with the GP distribution, excels at the 99% VaR level.



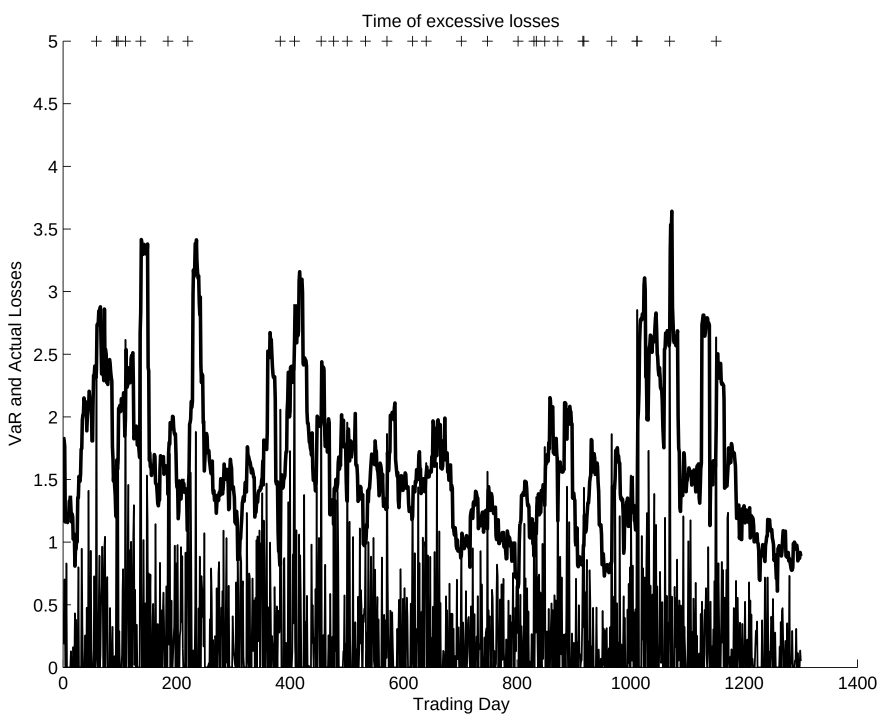


Figure 6.38: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed GP distributed.*

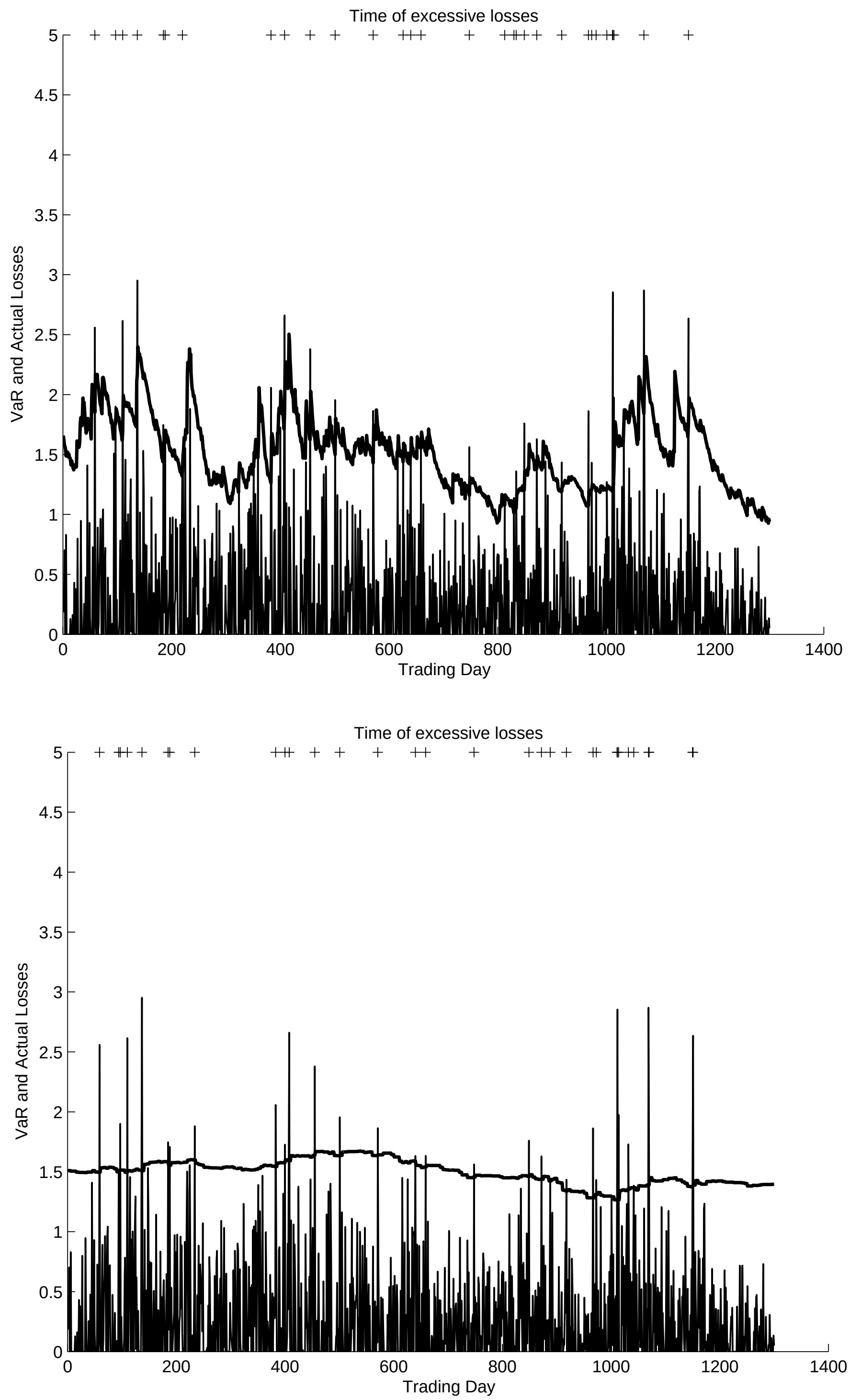


Figure 6.39: VaR *(97.5%) predictions and actual losses,* USD/DM *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed GP distributed.*

6.5.4 DAX

Non-par.	95%	97.5%	99%	**Var. Win.**	95%	97.5%	99%
Hyperbolic	0.0933	0.0466	**0.0199**	Hyperbolical	0.1066	0.0655	**0.0393**
Gaussian	0.1020	0.0489	0.0272	Gaussian	0.1064	0.0702	0.0442
Pearson	**0.0864**	**0.0441**	0.0223	Pearson	**0.0955**	**0.0653**	**0.0393**
GP	0.1283	0.0676	0.0369	GP	0.1538	0.1172	0.0767
GARCH-AR	95%	97.5%	99%	**Constant**	95%	97.5%	99%
Hyperbolic	0.0910	**0.0442**	0.0199	Hyperbolic	0.1023	**0.0442**	0.0199
Gaussian	0.1042	0.0489	0.0199	Gaussian	0.1020	0.0465	0.0199
Pearson	**0.0842**	0.0465	**0.0174**	Pearson	**0.0911**	0.0465	**0.0174**
GP	0.1436	0.0860	0.0345	GP	0.1475	0.0859	0.0440

Table 6.14: *Results from the* LPFA.

The LPFA indicates that the Pearson distribution should be used for the DAX data. Surprisingly, the GP distribution is not a good choice, according to the LPFA.

Hyperbolical

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	5.3%	[4.1%,6.5%]	0.25	0.614	Not Rejected
Var-Win	95%	5.3%	[4.1%,6.5%]	0.25	0.614	Not Rejected
GARCH-AR	95%	4.9%	[3.7%,6.1%]	0.02	0.900	Not Rejected
Constant	95%	6.1%	[4.8%,7.4%]	3.00	0.084	Not Rejected
Nadaraya-Watson	97.5%	2.7%	[1.8%,3.6%]	0.19	0.661	Not Rejected
Var-Win	97.5%	2.5%	[1.7%,3.4%]	0.01	0.929	Not Rejected
GARCH-AR	97.5%	2.5%	[1.7%,3.4%]	0.01	0.929	Not Rejected
Constant	97.5%	3.4%	[2.4%,4.4%]	3.76	0.052	Not Rejected
Nadaraya-Watson	99%	1.1%	[0.5%,1.6%]	0.08	0.783	Not Rejected
Var-Win	99%	1.5%	[0.8%,2.1%]	2.45	0.120	Not Rejected
GARCH-AR	99%	1.5%	[0.8%,2.1%]	2.45	0.120	Not Rejected
Constant	99%	1.5%	[0.8%,2.1%]	2.45	0.120	Not Rejected

Table 6.15: FOEL *test if returns are assumed to be Hyperbolically distributed,* DAX *data.*

At the 95% VaR level, the GARCH-AR volatility model works well. As for the 99% VaR level, the Nadaraya-Watson volatility model, once again, excels.

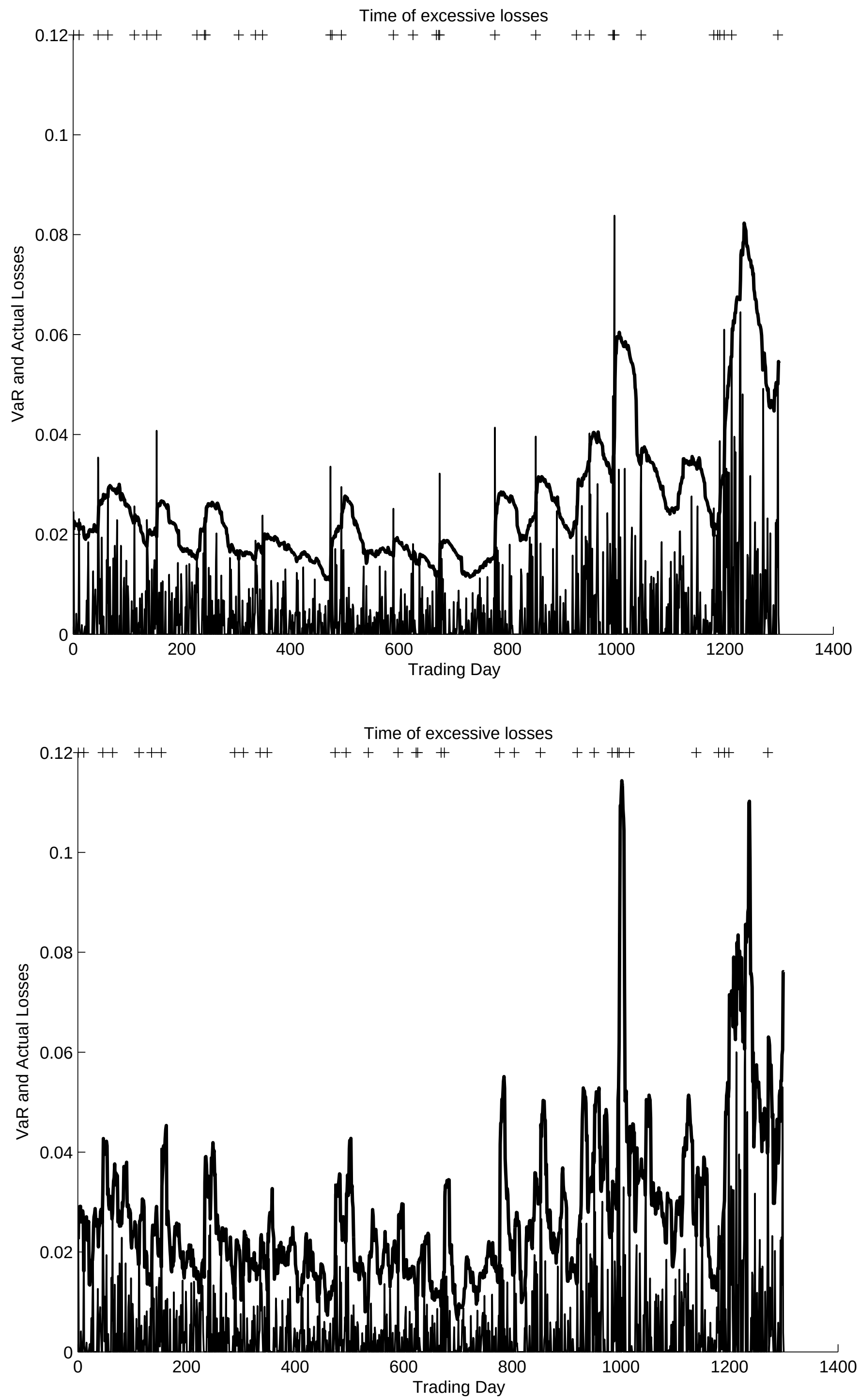


Figure 6.40: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Hyperbolically distributed.*

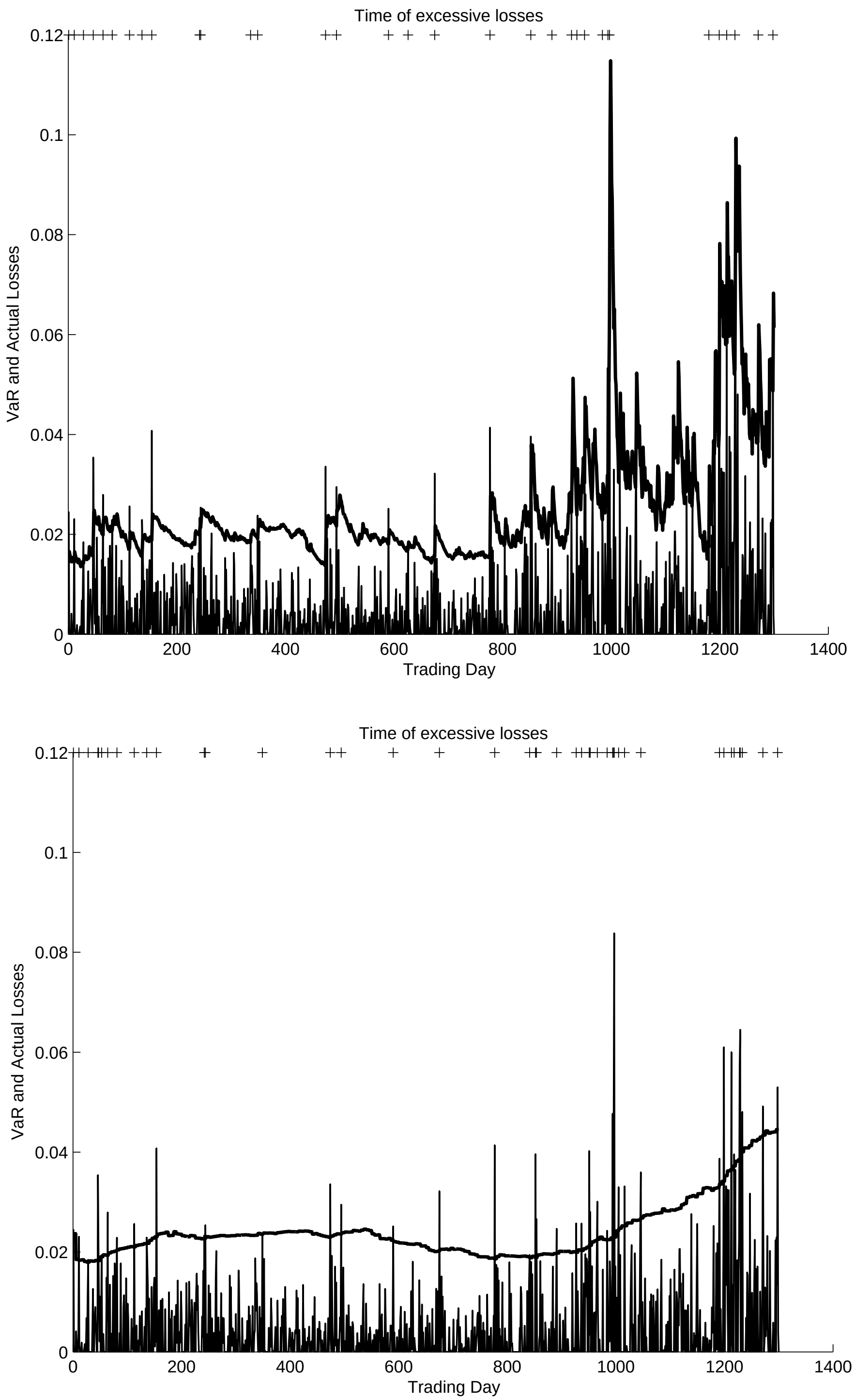


Figure 6.41: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Hyperbolically distributed.*

Gauss

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	5.5%	[4.3%,6.8%]	0.77	0.381	Not Rejected
Var-Win	95%	5.4%	[4.2%,6.6%]	0.40	0.530	Not Rejected
GARCH-AR	95%	5.3%	[4.1%,6.5%]	0.25	0.614	Not Rejected
Constant	95%	7.5%	[6.0%,8.9%]	14.50	<0.001	Rejected
Nadaraya-Watson	97.5%	3.4%	[2.4%,4.4%]	3.76	0.052	Not Rejected
Var-Win	97.5%	3.5%	[2.5%,4.5%]	4.41	0.036	Rejected
GARCH-AR	97.5%	3.3%	[2.3%,4.3%]	3.16	0.075	Not Rejected
Constant	97.5%	5.2%	[4.0%,6.4%]	30.41	<0.001	Rejected
Nadaraya-Watson	99%	1.8%	[1.1%,2.5%]	6.32	0.012	Rejected
Var-Win	99%	2.1%	[1.3%,2.9%]	11.62	<0.001	Rejected
GARCH-AR	99%	2.2%	[1.4%,3.0%]	14.74	<0.001	Rejected
Constant	99%	3.5%	[2.5%,4.5%]	48.56	<0.001	Rejected

Table 6.16: FOEL *test if returns are assumed to be Gaussian distributed,* DAX *data.*

The Gaussian distribution works well at the 95% level, but it is really poor further out in the tail.

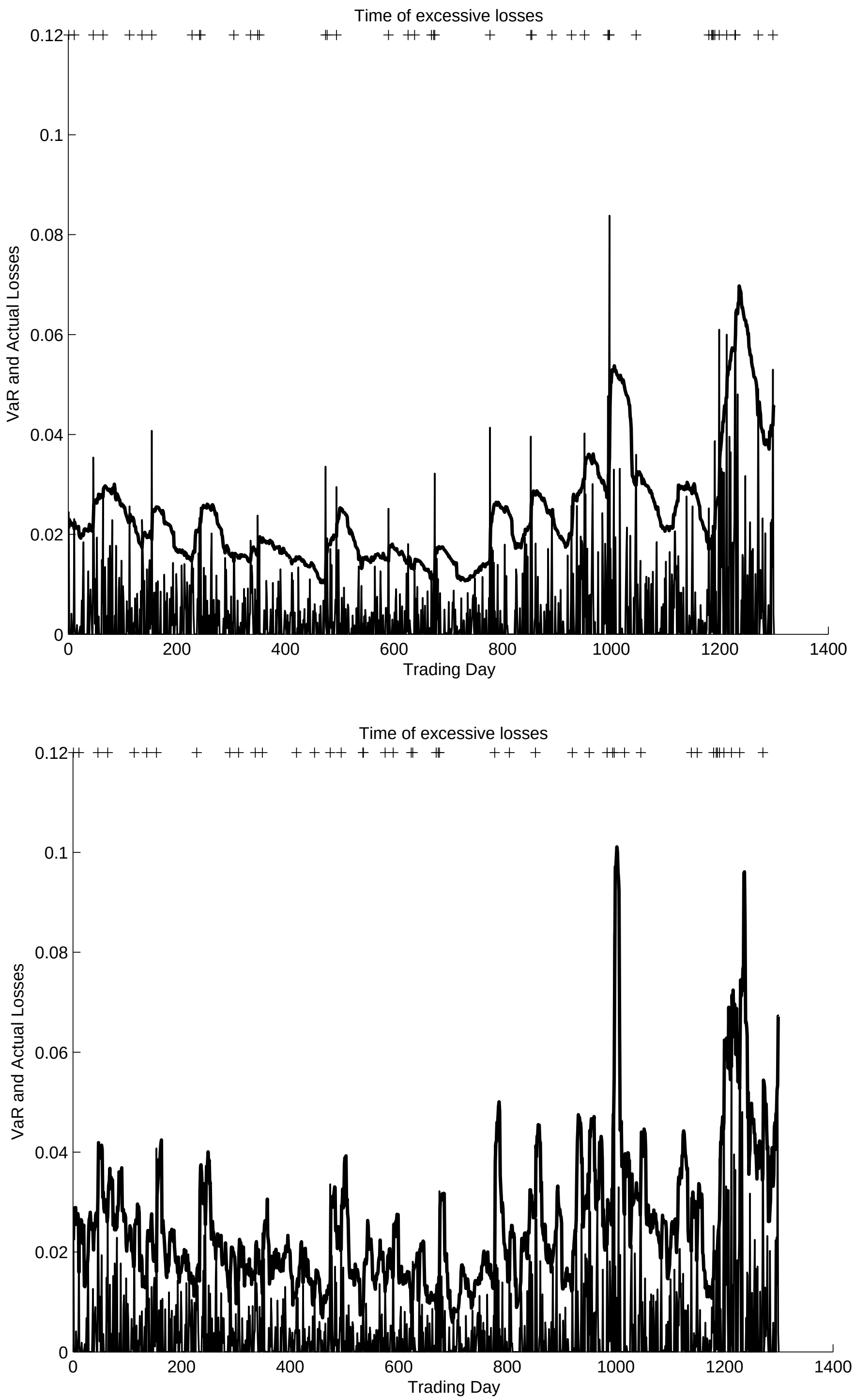


Figure 6.42: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Gaussian distributed.*

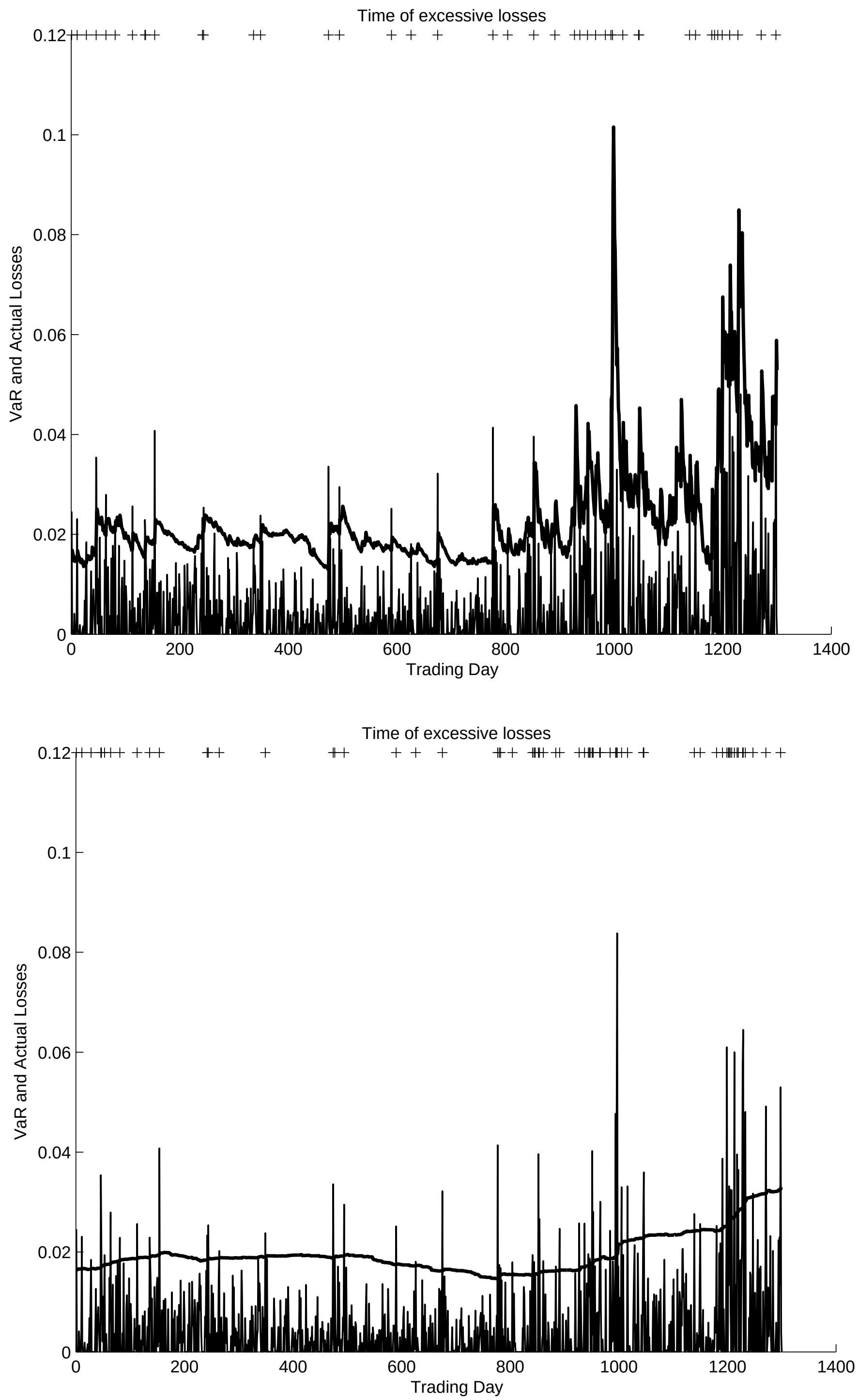


Figure 6.43: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Gaussian distributed.*

Pearson

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	5.2%	[4.0%,6.4%]	0.06	0.800	Not Rejected
Var-Win	95%	5.4%	[4.2%,6.6%]	0.40	0.530	Not Rejected
GARCH-AR	95%	4.9%	[3.7%,6.1%]	0.02	0.900	Not Rejected
Constant	95%	7.2%	[5.8%,8.6%]	11.30	< 0.001	Rejected
Nadaraya-Watson	97.5%	2.6%	[1.7%,3.5%]	0.07	0.791	Not Rejected
Var-Win	97.5%	2.6%	[1.7%,3.5%]	0.07	0.791	Not Rejected
GARCH-AR	97.5%	2.7%	[1.8%,3.6%]	0.19	0.661	Not Rejected
Constant	97.5%	4.1%	[3.0%,5.2%]	11.17	< 0.001	Rejected
Nadaraya-Watson	99%	1.1%	[0.5%,1.6%]	0.08	0.783	Not Rejected
Var-Win	99%	1.2%	[0.6%,1.7%]	0.30	0.586	Not Rejected
GARCH-AR	99%	1.4%	[0.7%,2.0%]	1.73	0.188	Not Rejected
Constant	99%	1.8%	[1.1%,2.5%]	6.32	0.012	Rejected

Table 6.17: FOEL *test if returns are assumed to be Pearson distributed,* DAX *data.*

The table indicates that, the Pearson distribution is a good choice, except for the constant volatility case.

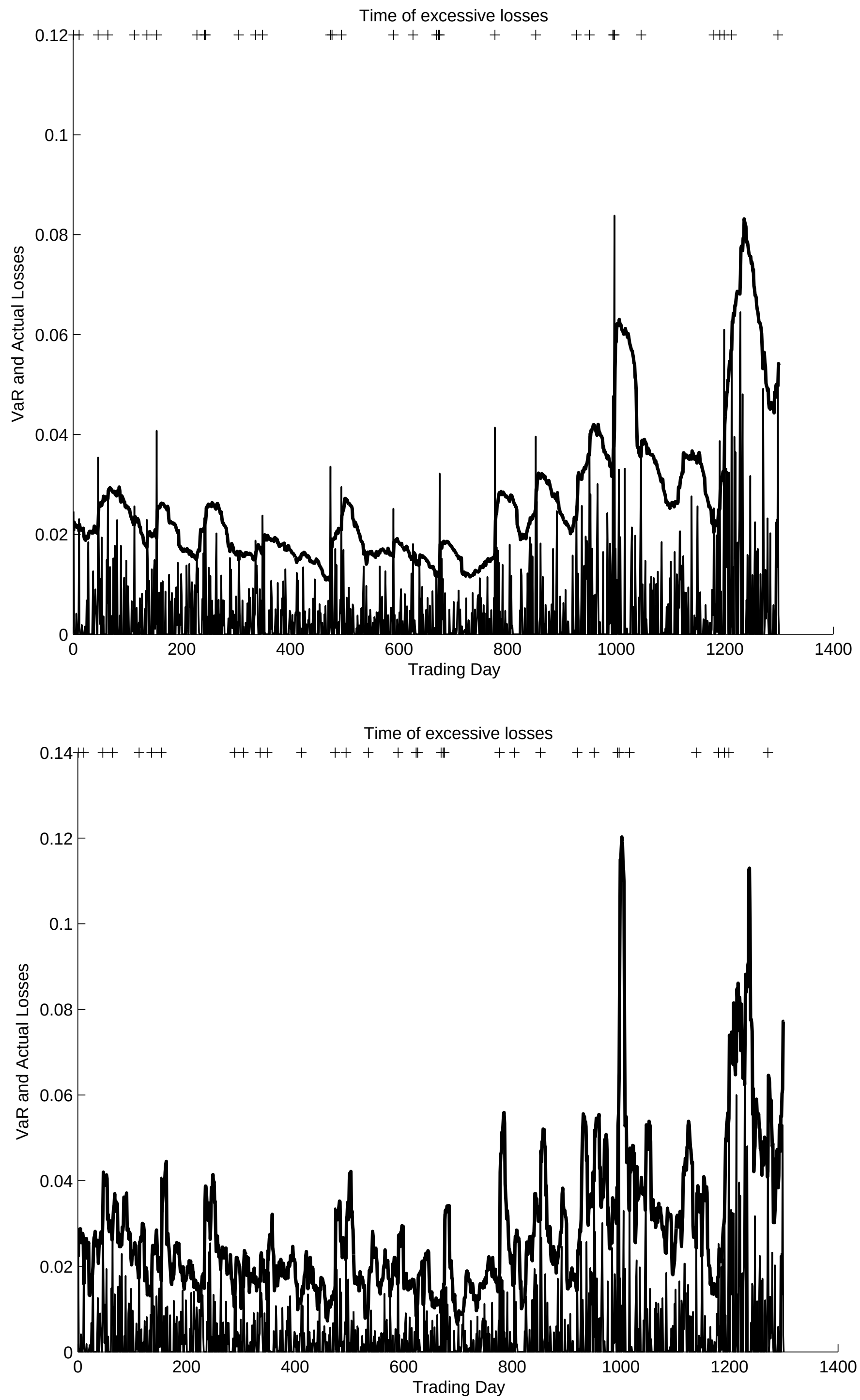
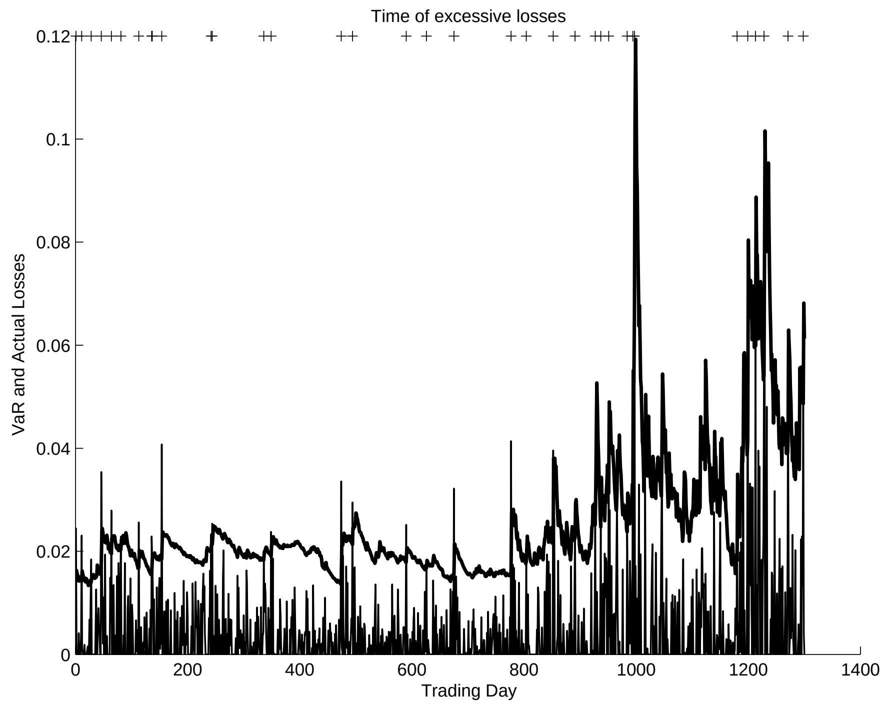


Figure 6.44: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Pearson distributed.*



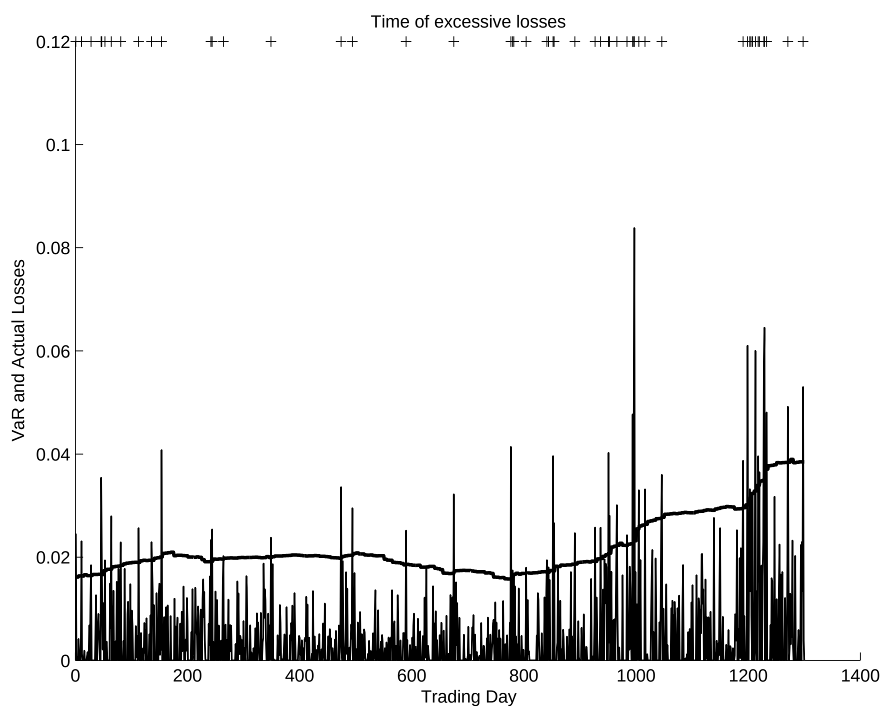


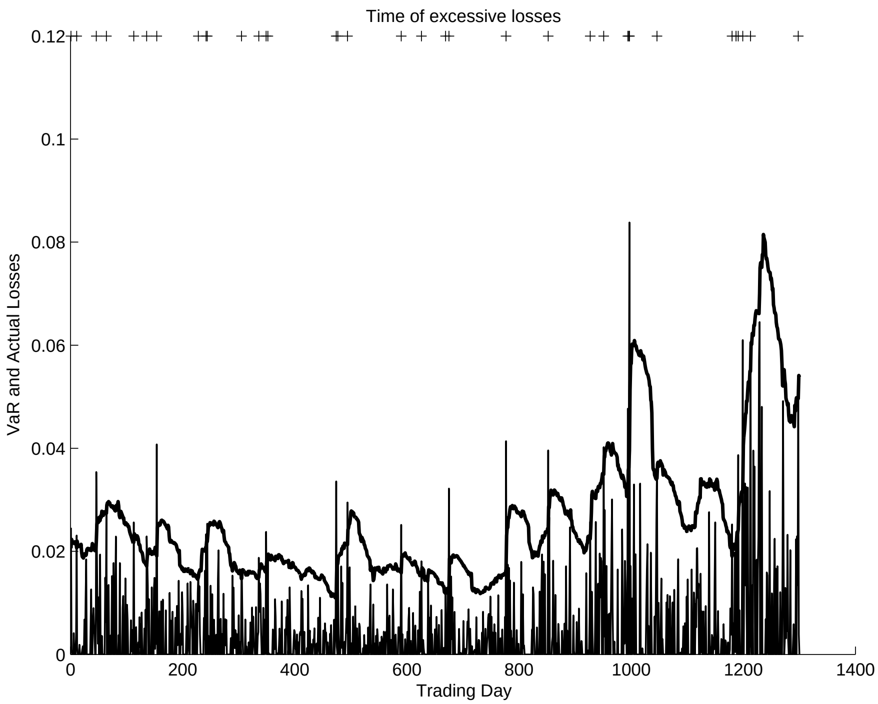
Figure 6.45: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Pearson distributed.*

GP

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	5.8%	[4.6%,7.1%]	1.86	0.172	Not Rejected
Var-Win	95%	5.3%	[4.1%,6.5%]	0.25	0.614	Not Rejected
GARCH-AR	95%	5.3%	[4.1%,6.5%]	0.25	0.614	Not Rejected
Constant	95%	7.2%	[5.8%,8.6%]	12.00	< 0.001	Rejected
Nadaraya-Watson	97.5%	2.7%	[1.8%,3.6%]	0.19	0.661	Not Rejected
Var-Win	97.5%	2.4%	[1.6%,3.2%]	0.07	0.788	Not Rejected
GARCH-AR	97.5%	2.7%	[1.8%,3.6%]	0.19	0.661	Not Rejected
Constant	97.5%	4.1%	[3.0%,5.2%]	11.17	< 0.001	Rejected
Nadaraya-Watson	99%	1.1%	[0.5%,1.6%]	0.08	0.783	Not Rejected
Var-Win	99%	1.2%	[0.6%,1.7%]	0.30	0.586	Not Rejected
GARCH-AR	99%	1.2%	[0.6%,1.8%]	0.65	0.420	Not Rejected
Constant	99%	2.1%	[1.3%,2.9%]	11.62	< 0.001	Rejected

Table 6.18: FOEL *test if returns are assumed to be GP distributed,* DAX *data.*

At the 95% VaR level, the GARCH-AR seems to be the best volatility model, and at the 99% VaR level, the Nadaraya-Watson is the best choice.



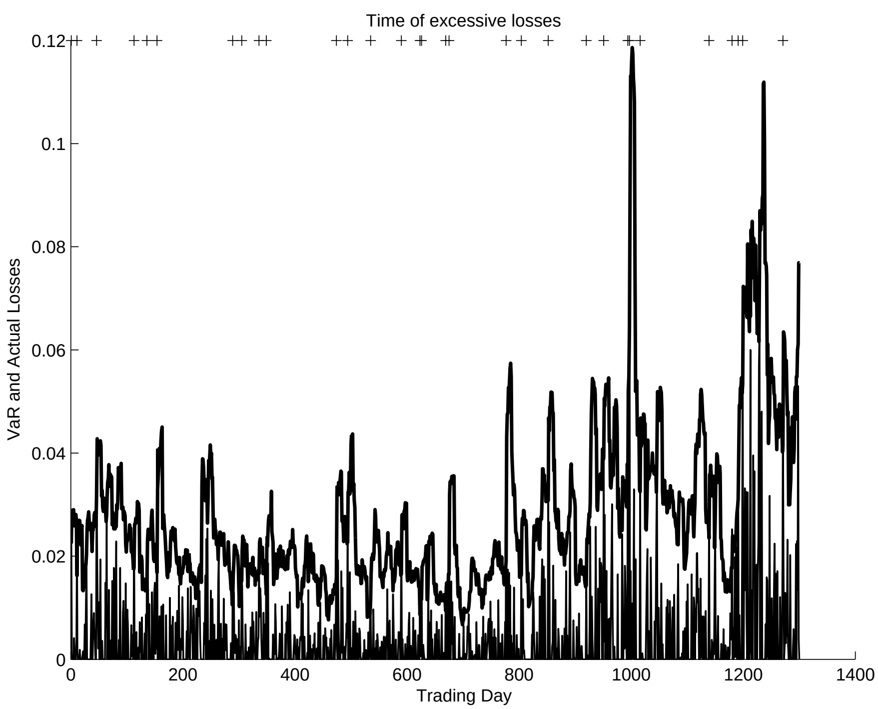


Figure 6.46: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed GP distributed.*

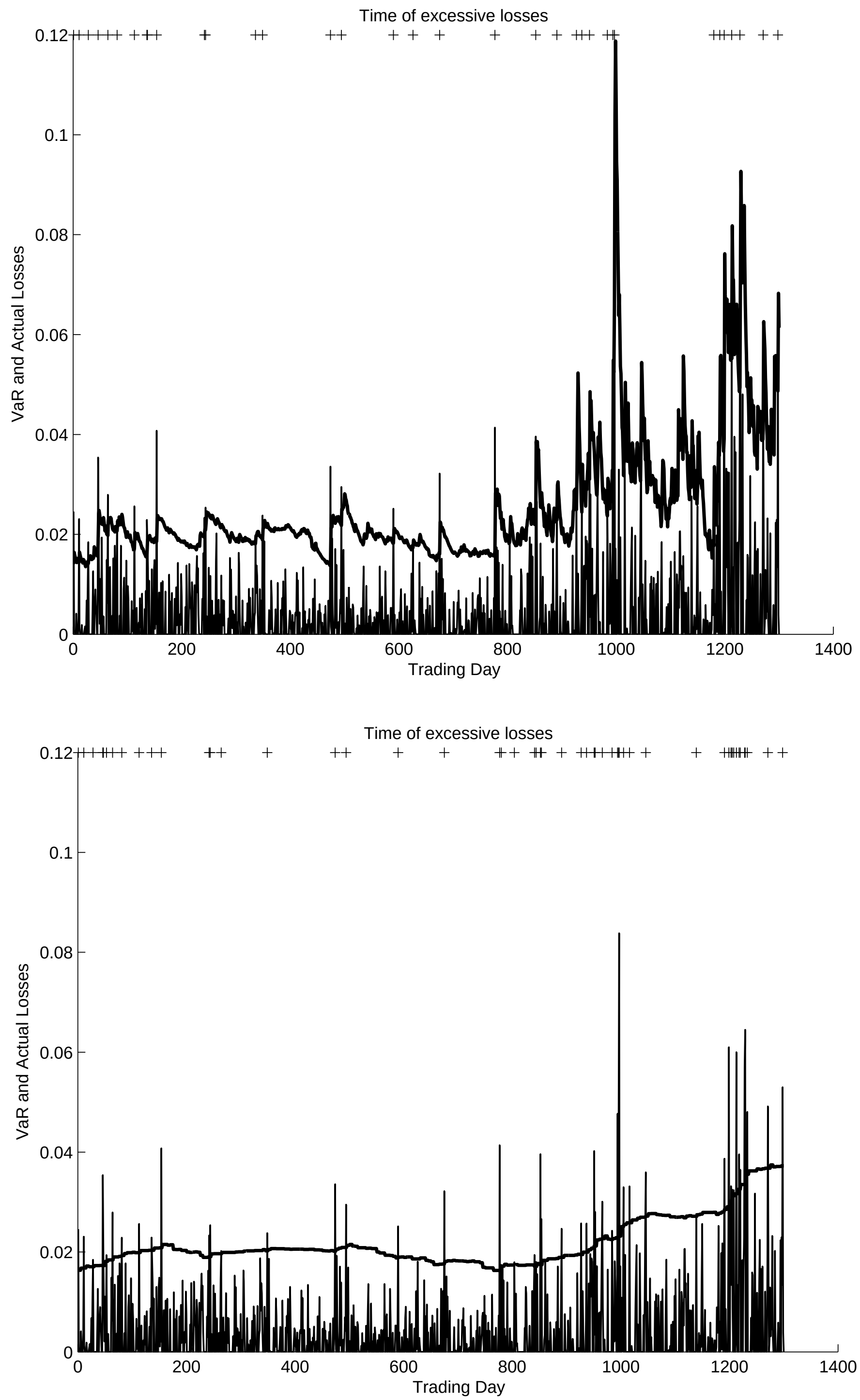


Figure 6.47: VaR *(97.5%) predictions and actual losses,* DAX *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed GP distributed.*

6.5.5 SIEMENS

Non-par.	95%	97.5%	99%	**Var. Win.**	95%	97.5%	99%
Hyperbolic	0.0907	0.0369	0.0150	Hyperbolical	0.0816	0.0465	0.0174
Gaussian	0.0931	0.0441	0.0100	Gaussian	0.0907	0.0559	0.0174
Pearson	0.0795	0.0273	**0.0075**	Pearson	**0.0725**	**0.0370**	**0.0050**
GP	**0.0771**	**0.0234**	0.0103	GP	0.1067	0.0597	0.0302
GARCH-AR	95%	97.5%	99%	**Constant**	95%	97.5%	99%
Hyperbolic	0.0977	0.0417	0.0125	Hyperbolic	0.0886	0.0393	0.0125
Gaussian	0.0908	0.0465	0.0100	Gaussian	0.0976	0.0560	0.0100
Pearson	**0.0841**	**0.0321**	**0.0075**	Pearson	**0.0841**	**0.0273**	**0.0075**
GP	0.1772	0.1170	0.0918	GP	0.1219	0.0523	0.0201

Table 6.19: *Results of the* LPFA.

The LPFA indicates that, at the 95% and 97.5% VaR levels, the GP distibution works well together with the Nadaraya-Watson volatility model. However, considering the other volatility models, and the Nadaraya-Watson at the 99% VaR level, the Pearson distribution is the best choice.

Hyperbolic

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.6%	[3.5%,5.8%]	0.42	0.519	Not Rejected
Var-Win	95%	4.8%	[3.7%,6.0%]	0.07	0.798	Not Rejected
GARCH-AR	95%	4.5%	[3.4%,5.7%]	0.60	0.438	Not Rejected
Constant	95%	6.8%	[5.5%,8.2%]	8.41	0.003	Rejected
Nadaraya-Watson	97.5%	2.0%	[1.2%,2.8%]	1.43	0.232	Not Rejected
Var-Win	97.5%	2.1%	[1.3%,2.9%]	1.01	0.314	Not Rejected
GARCH-AR	97.5%	1.8%	[1.1%,2.5%]	3.17	0.075	Not Rejected
Constant	97.5%	3.4%	[2.4%,4.4%]	3.76	0.05	Not Rejected
Nadaraya-Watson	99%	0.9%	[0.4%,1.4%]	0.08	0.778	Not Rejected
Var-Win	99%	0.7%	[0.2%,1.1%]	1.39	0.238	Not Rejected
GARCH-AR	99%	0.8%	[0.3%,1.2%]	0.76	0.383	Not Rejected
Constant	99%	1.8%	[1.1%,2.5%]	6.32	0.012	Rejected

Table 6.20: FOEL *test if returns are assumed to be Hyperbolically distributed,* Siemens *data.*

The FOEL test indicates that, at lower VaR levels, the Variance Window is the best volatility model to use with the Hyperbolical distribution. At the 99% VaR level, the Nadaraya-Watson volatility model excels once again.

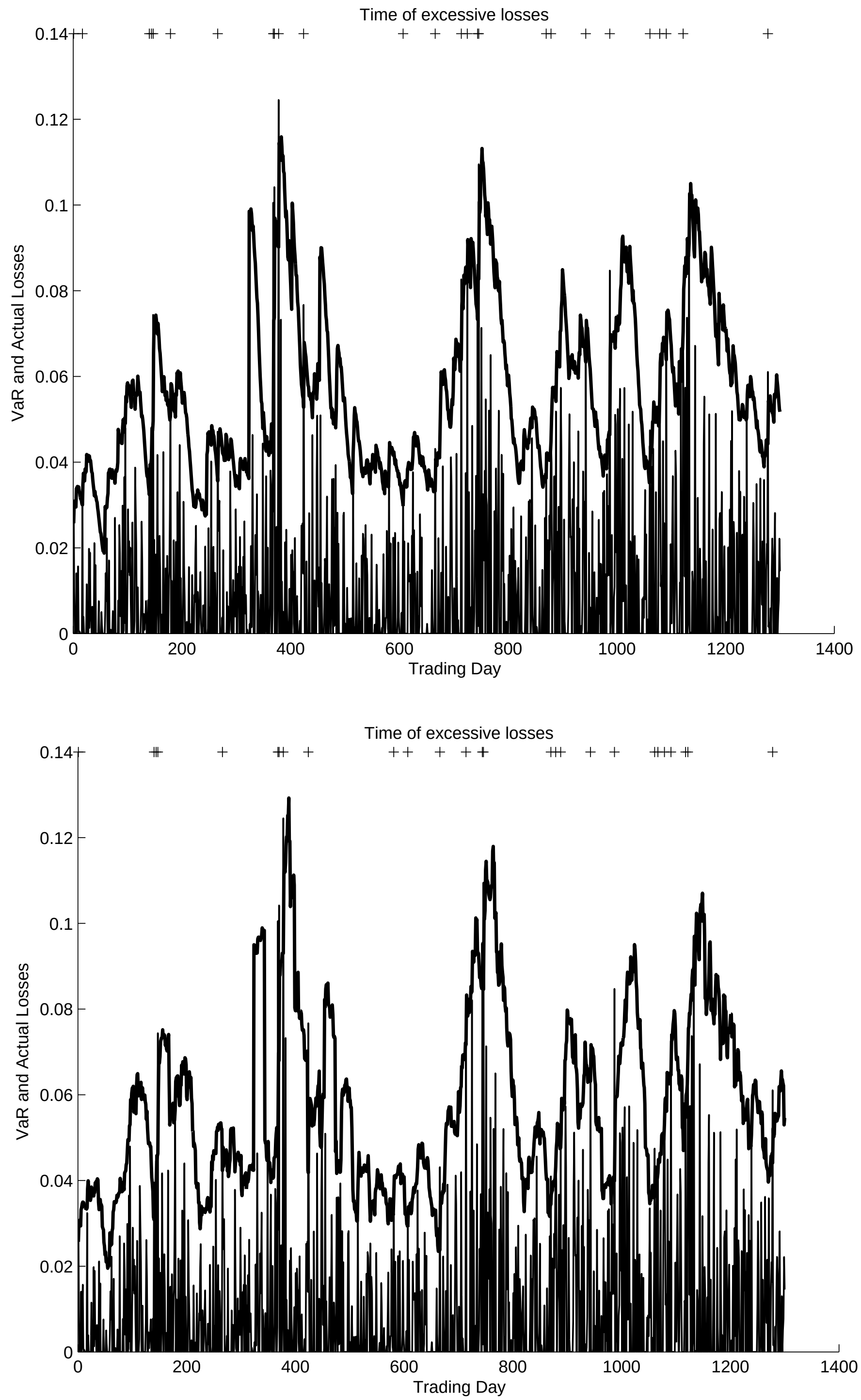
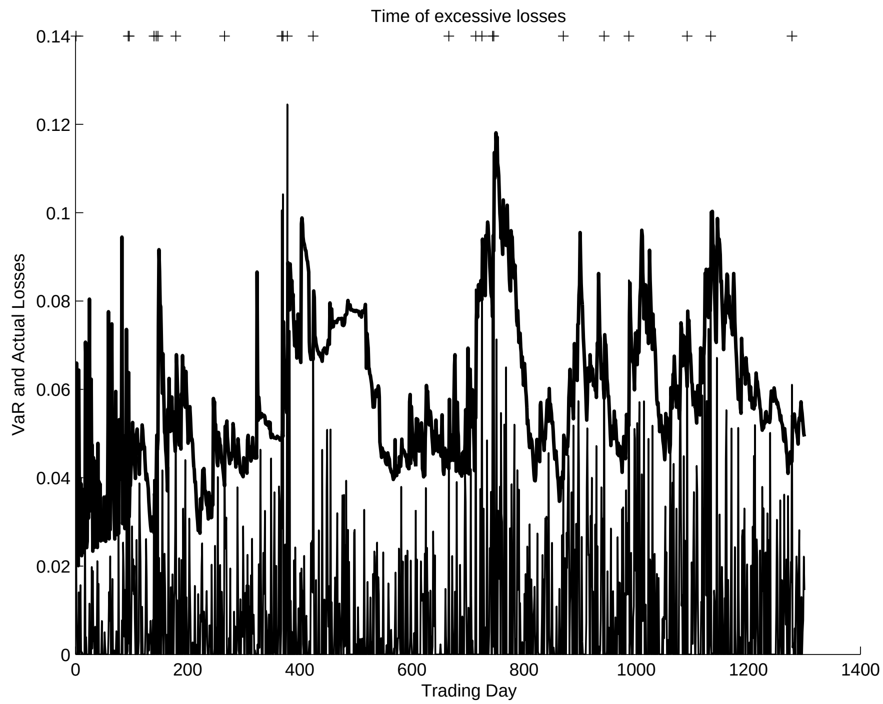


Figure 6.48: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Hyperbolically distributed.*



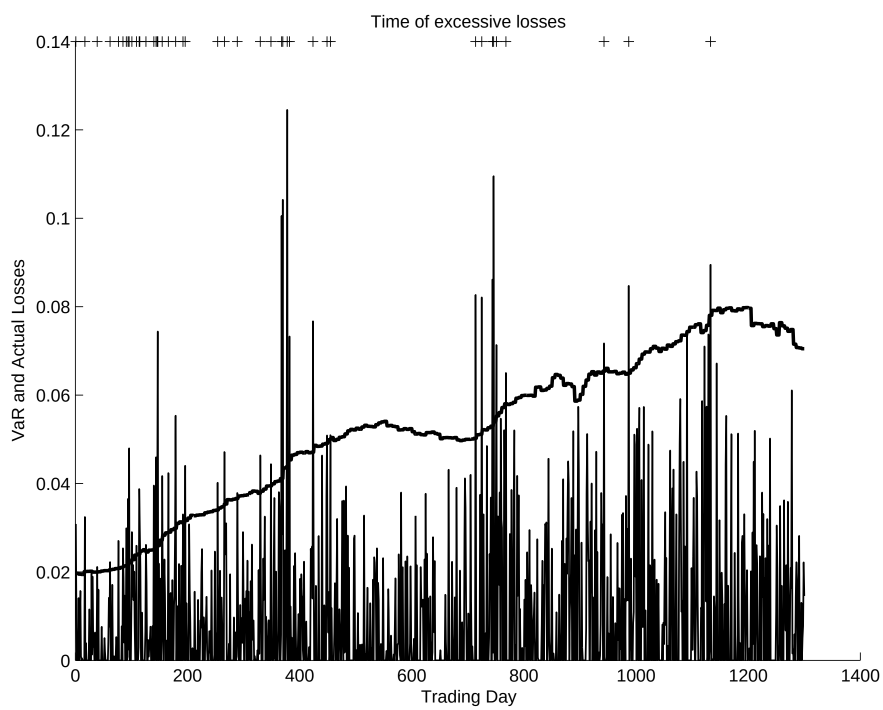


Figure 6.49: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR-*AR (upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Hyperbolically distributed.*

Gauss

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.1%	[3.1%,5.2%]	2.07	0.150	Not Rejected
Var-Win	95%	4.6%	[3.5%,5.8%]	0.42	0.519	Not Rejected
GARCH-AR	95%	4.1%	[3.0%,5.2%]	2.48	0.115	Not Rejected
Constant	95%	7.8%	[6.3%,9.2%]	18.09	< 0.001	Rejected
Nadaraya- Watson	97.5%	2.4%	[1.6%,3.2%]	0.07	0.788	Not Rejected
Var-Win	97.5%	2.5%	[1.6%,3.3%]	0.01	0.929	Not Rejected
GARCH-AR	97.5%	2.2%	[1.4%,3.0%]	0.67	0.413	Not Rejected
Constant	97.5%	4.8%	[3.7%,6.0%]	23.14	< 0.001	Rejected
Nadaraya-Watson	99%	1.2%	[0.6%,1.7%]	0.30	0.586	Not Rejected
Var-Win	99%	1.0%	[0.5%,1.5%]	0	1.000	Not Rejected
GARCH-AR	99%	0.5%	[0.2%,0.9%]	3.36	0.067	Not Rejected
Constant	99%	2.9%	[2.0%,3.8%]	32.00	< 0.001	Rejected

Table 6.21: FOEL *test if returns are assumed to be Gaussian distributed,* Siemens *data.*

The results from the FOEL test for the Siemens data, using the Gaussian distribution, are quite surprising. The Gaussian distribution passes at all VaR levels, using the three stochastic volatility models. Furthermore, the Variance Window volatility model seems to be best at all VaR levels.

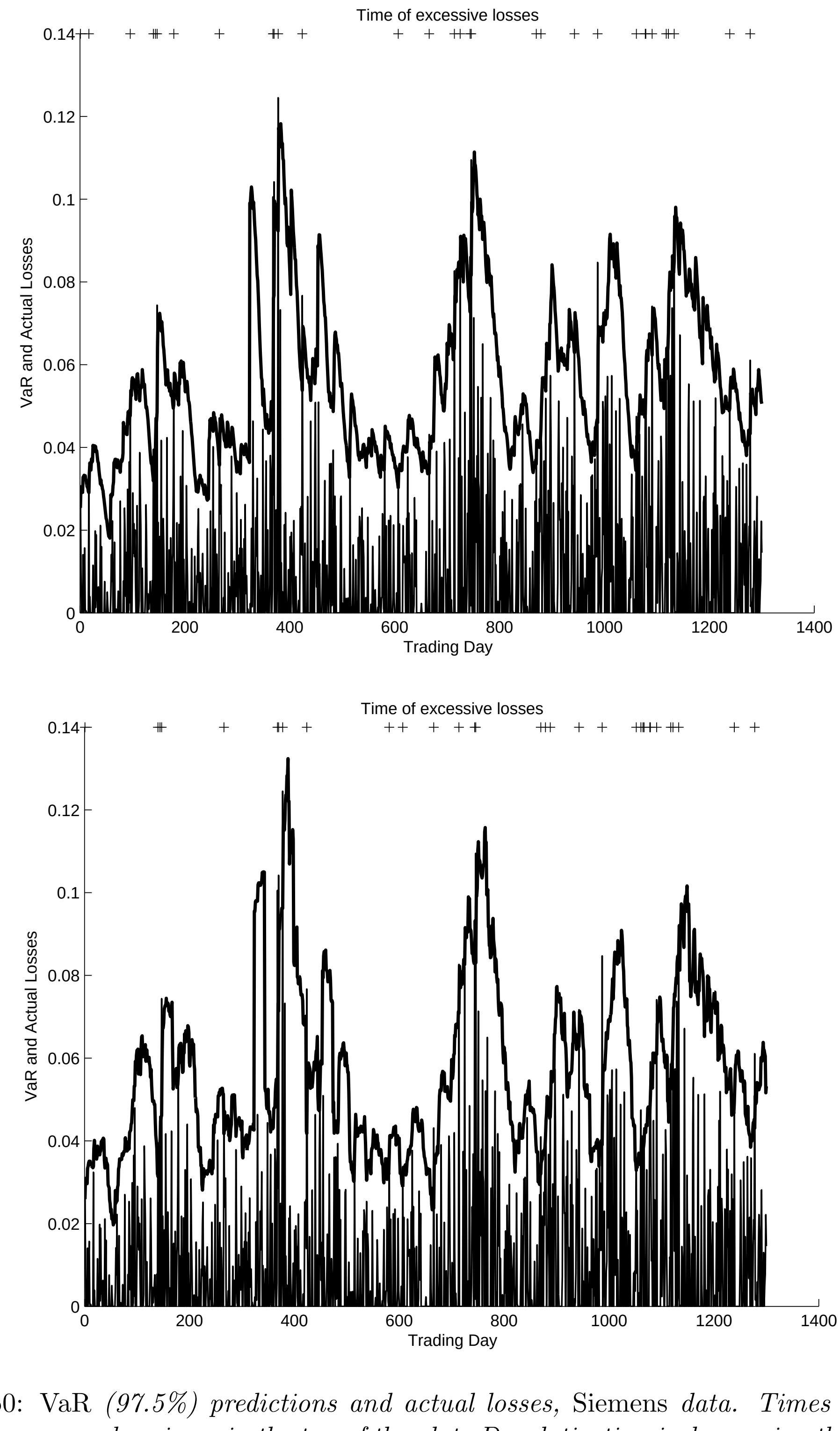
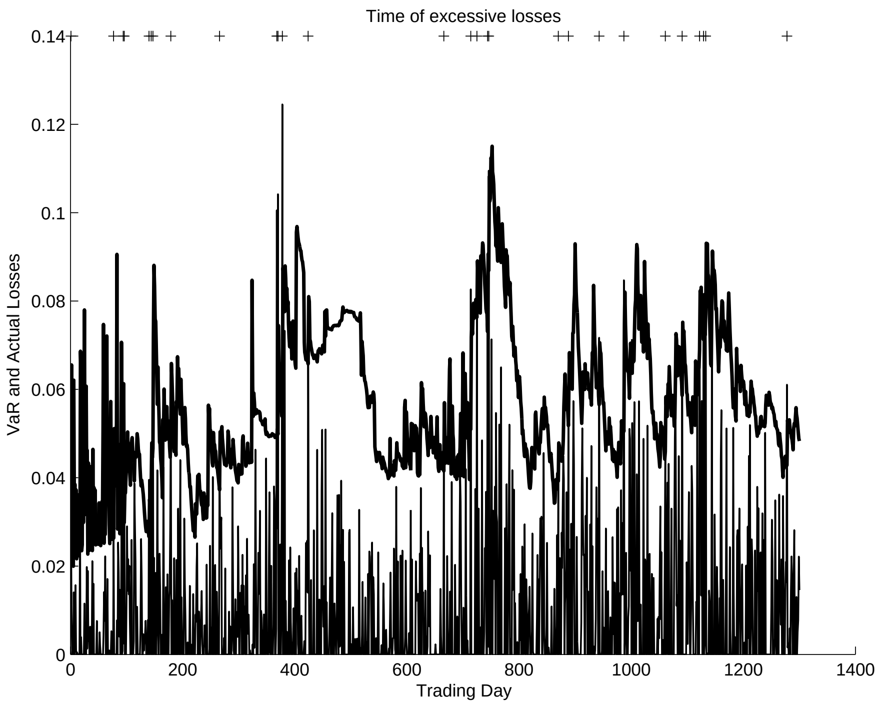


Figure 6.50: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Gaussian distributed.*



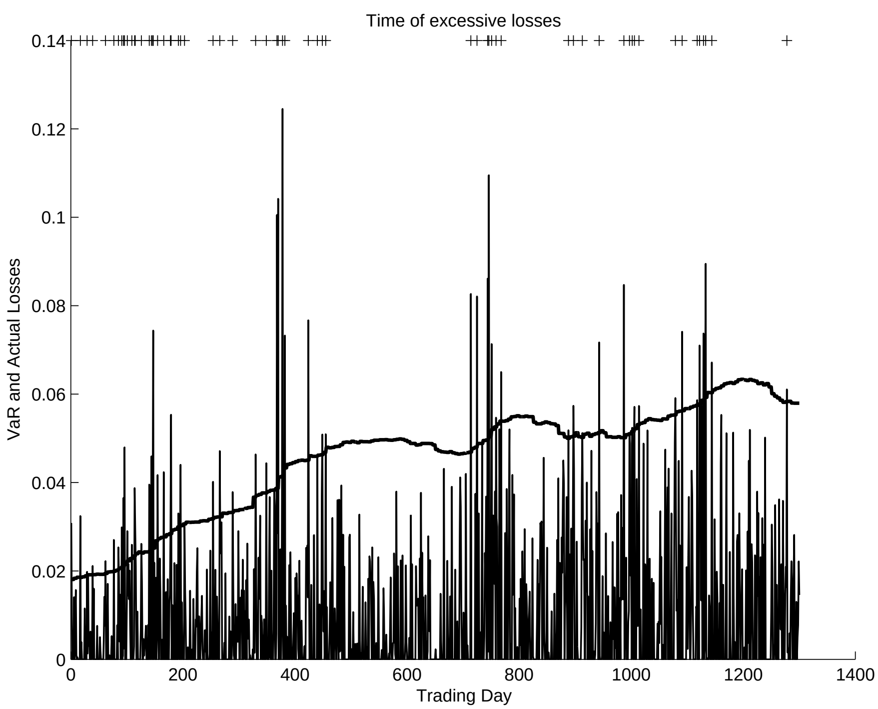


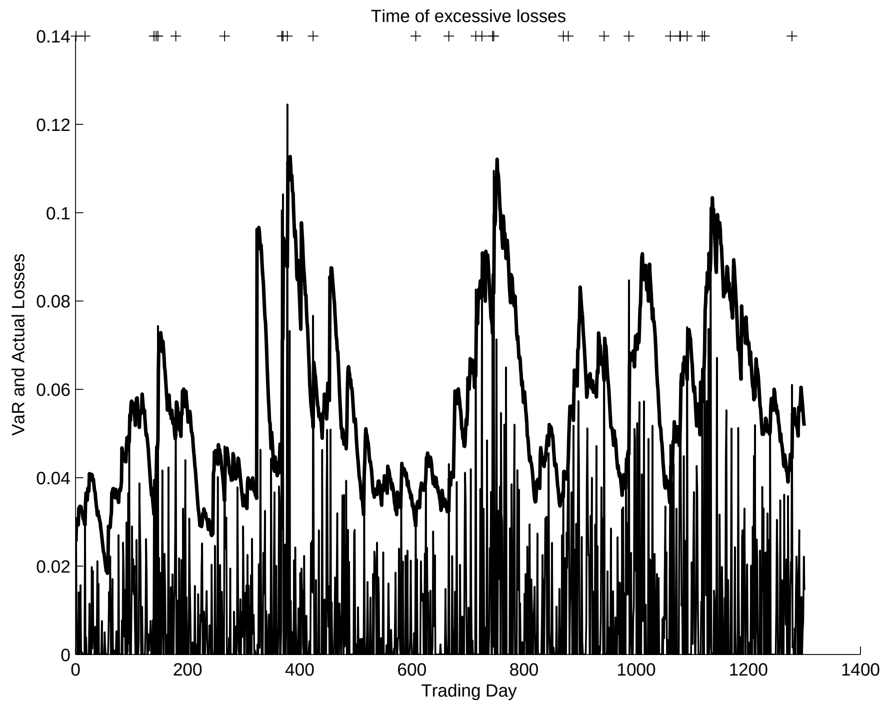
Figure 6.51: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Gaussian distributed.*

Pearson

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	5.2%	[4.0%,6.4%]	0.06	0.800	Not Rejected
Var-Win	95%	4.8%	[3.7%,6.0%]	0.07	0.798	Not Rejected
GARCH-AR	95%	4.8%	[3.7%,6.0%]	0.06	0.798	Not Rejected
Constant	95%	9.8%	[8.2%,11.5%]	50.75	< 0.001	Rejected
Nadaraya-Watson	97.5%	2.2%	[1.4%,2.9%]	0.67	0.413	Not Rejected
Var-Win	97.5%	2.3%	[1.5%,3.1%]	0.20	0.653	Not Rejected
GARCH-AR	97.5%	2.0%	[1.2%,2.8%]	1.43	0.232	Not Rejected
Constant	97.5%	5.2%	[4.0%,6.4%]	30.40	< 0.001	Rejected
Nadaraya-Watson	99%	0.8%	[0.3%,1.3%]	0.33	0.567	Not Rejected
Var-Win	99%	0.6%	[0.2%,1.0%]	2.25	0.114	Not Rejected
GARCH-AR	99%	0.8%	[0.3%,1.2%]	0.760	0.383	Not Rejected
Constant	99%	2.4%	[1.6%,3.2%]	18.13	< 0.001	Rejected

Table 6.22: FOEL *test if returns are assumed to be Pearson distributed,* Siemens *data.*

At the 95% VaR level, it is a close call between the three stochastic volatility models. At the 97.5% VaR level, the Variance Window seems to work best, and at the 99% VaR level, the Nadaraya-Watson volatility model is the best.



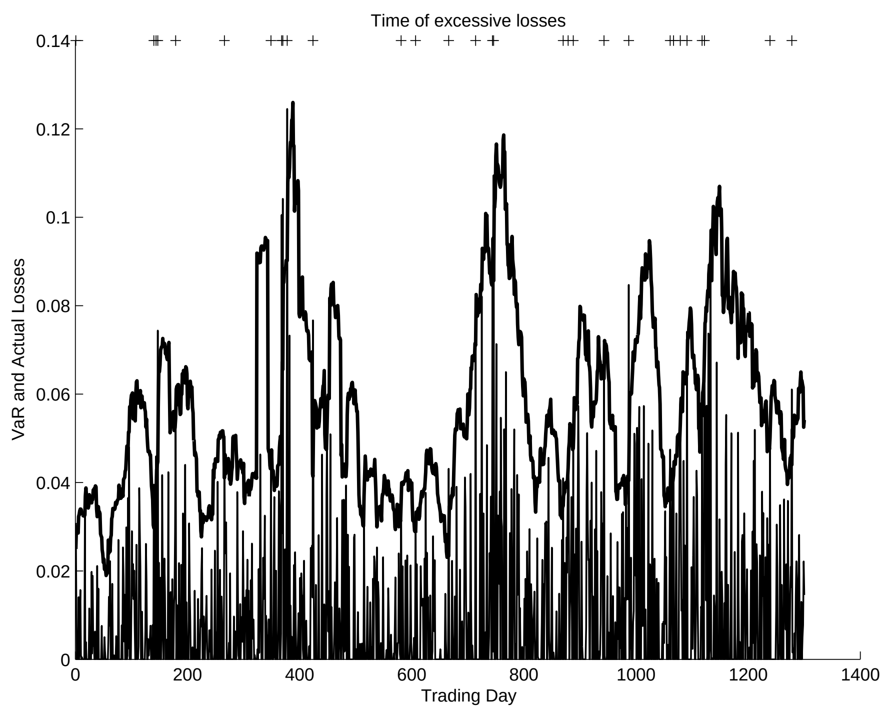


Figure 6.52: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed Pearson distributed.*

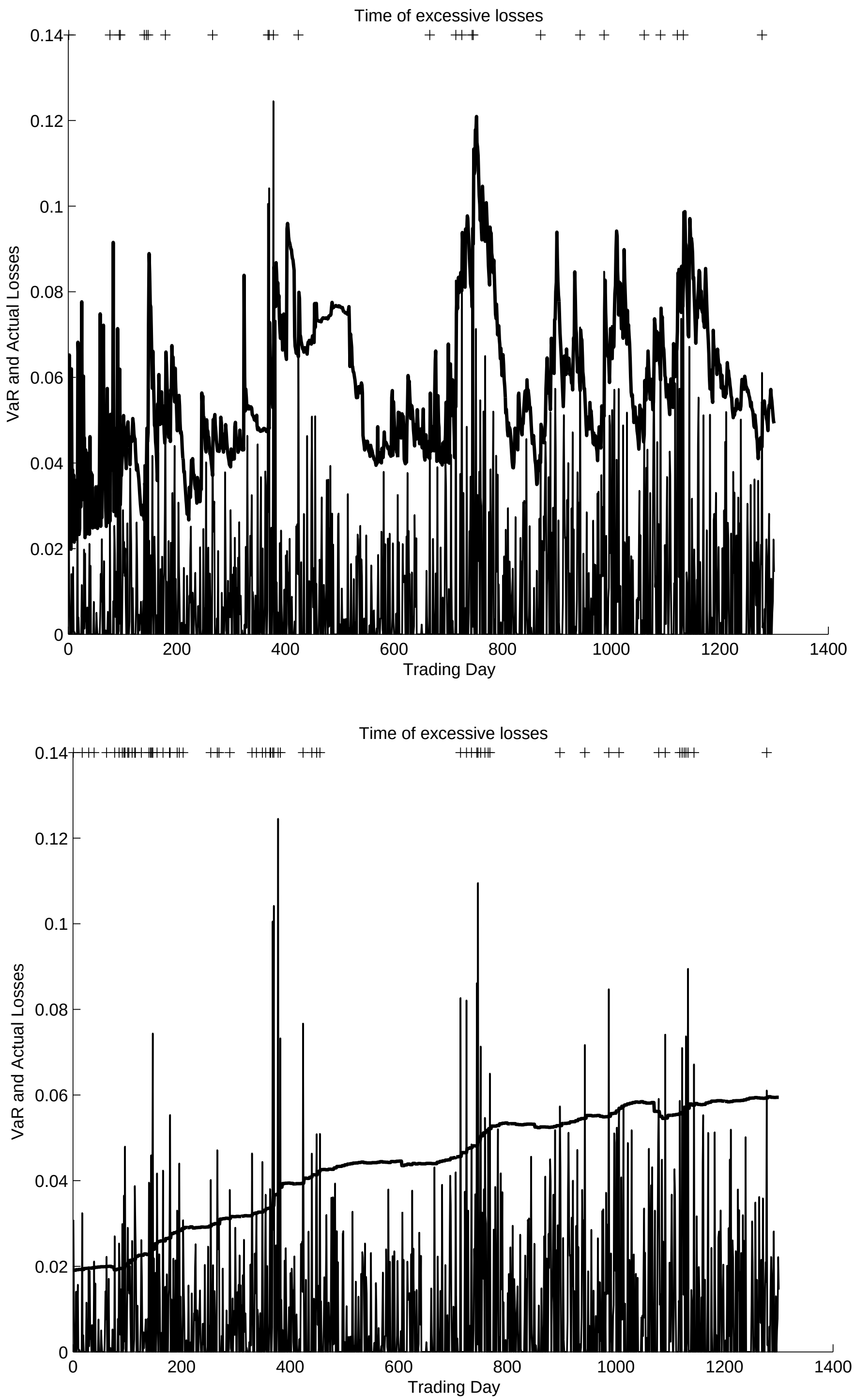


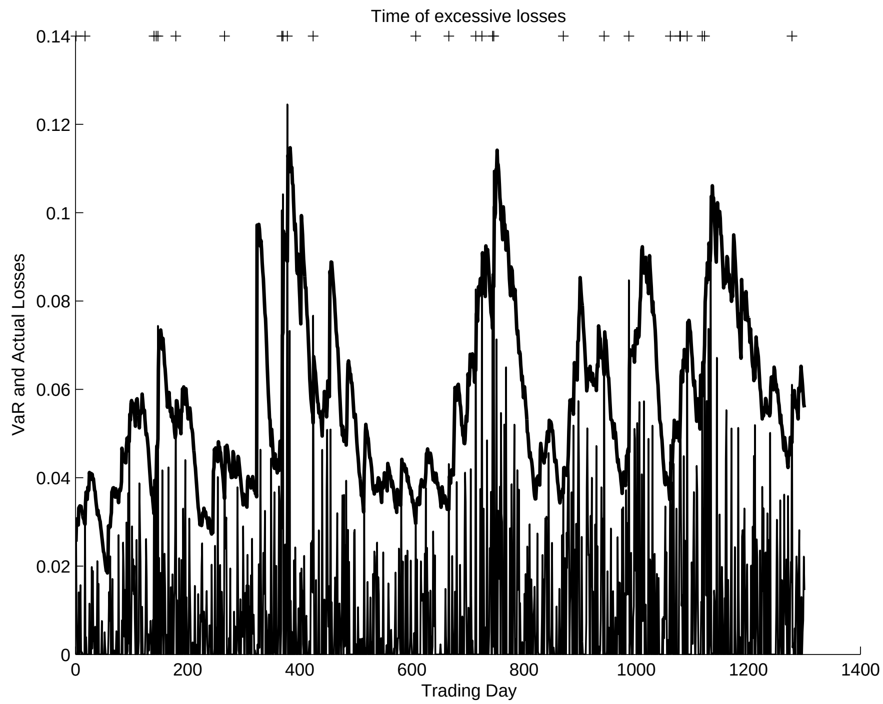
Figure 6.53: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed Pearson distributed.*

GP

Volatility model	VaR Level	FOEL	95% Confidence Interval	R	p-value	VaR Level Hypothesis
Nadaraya-Watson	95%	4.9%	[3.7%,6.1%]	0.02	0.900	Not Rejected
Var-Win	95%	5.2%	[4.0%,6.4%]	0.06	0.800	Not Rejected
GARCH-AR	95%	4.8%	[3.6%,5.9%]	0.15	0.701	Not Rejected
Constant	95%	8.6%	[7.1%,10.1%]	29.69	< 0.001	Rejected
Nadaraya-Watson	97.5%	2.1%	[1.3%,2.9%]	1.01	0.314	Not Rejected
Var-Win	97.5%	2.7%	[1.8%,3.6%]	0.19	0.661	Not Rejected
GARCH-AR	97.5%	1.8%	[1.1%,2.5%]	3.17	0.08	Not Rejected
Constant	97.5%	4.5%	[3.3%,5.6%]	16.70	< 0.001	Rejected
Nadaraya-Watson	99%	0.8%	[0.3%,1.2%]	0.80	0.383	Not Rejected
Var-Win	99%	0.6%	[0.2%,1.0%]	2.25	0.133	Not Rejected
GARCH-AR	99%	0.8%	[0.3%,1.2%]	0.86	0.383	Not Rejected
Constant	99%	2.0%	[1.2%,2.8%]	10.18	0.001	Rejected

Table 6.23: FOEL *test if returns are assumed to be GP distributed,* Siemens *data.*

At the 95% VaR level, as for the Pearson distribution, it is a close call between the three stochastic volatility models. At the 97.5% VaR level, the Variance Window seems to work best, and at the 99% VaR level, the Nadaraya-Watson and the GARCH-AR volatility models are the best.



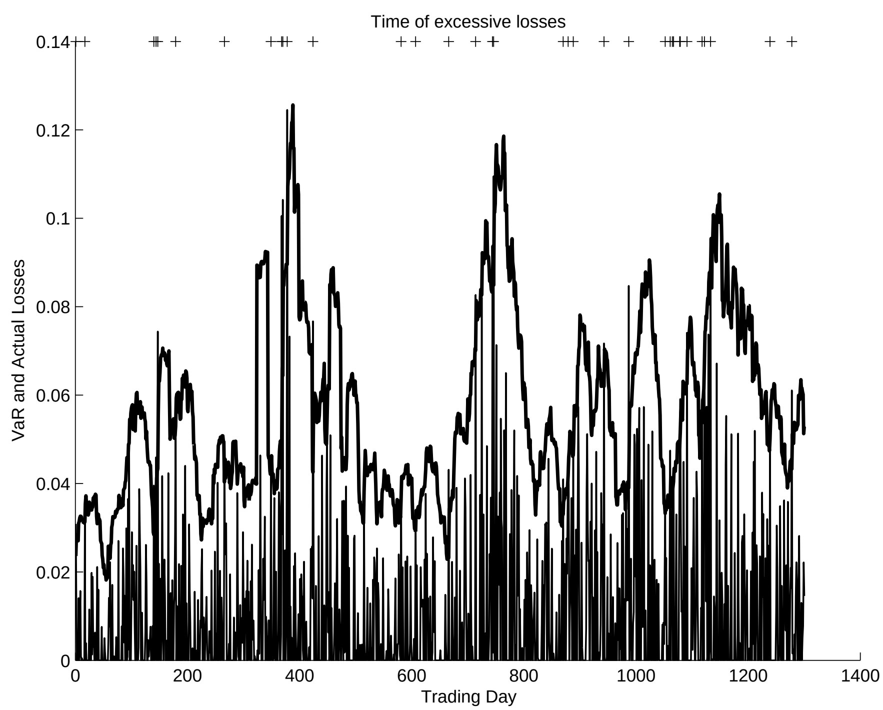
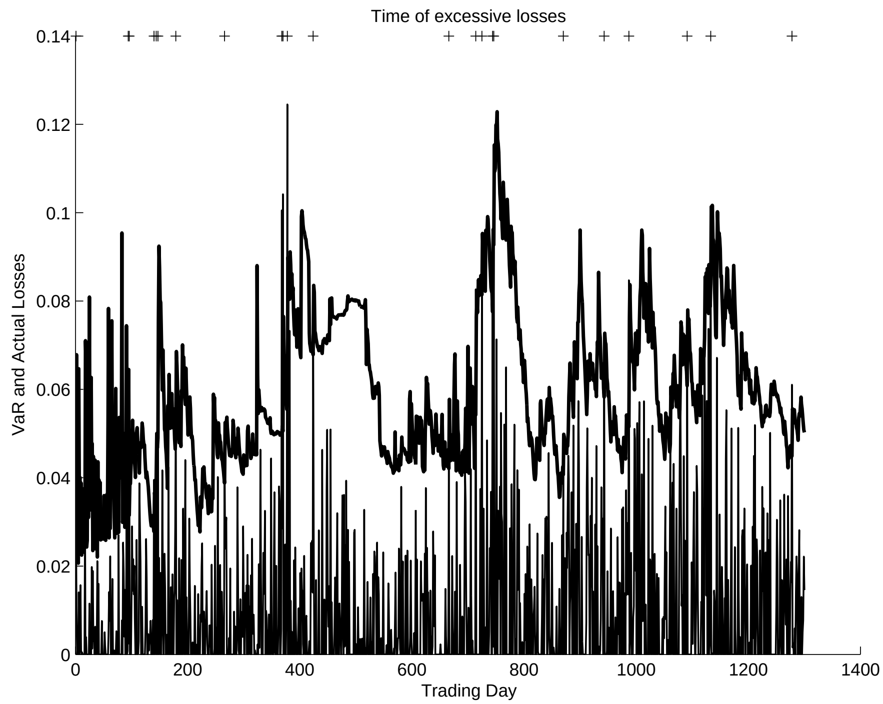


Figure 6.54: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the Nadaraya-Watson (upper), and the Variance Window (lower) volatility models. The devolatized log-returns are assumed GP distributed.*



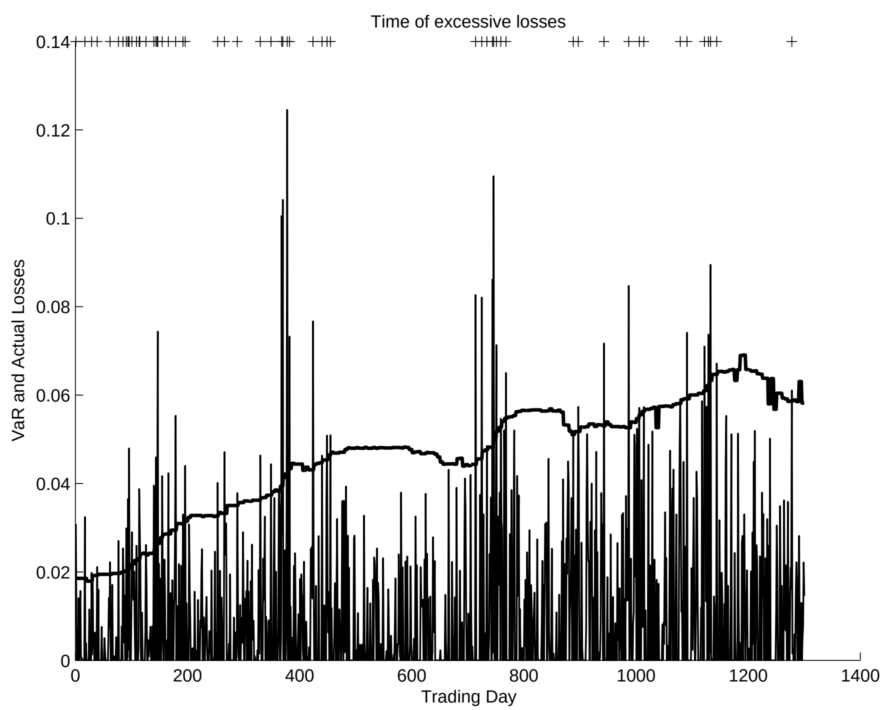


Figure 6.55: VaR *(97.5%) predictions and actual losses,* Siemens *data. Times of excessive losses are seen, as plus signs, in the top of the plot. Devolatization is done using the* GARCH-AR *(upper) volatility model, and constant volatility (lower). The devolatized log-returns are assumed GP distributed.*

Chapter 7

Conclusion

We have shown that stochastic volatility models give far better VaR-estimates, than do models with constant volatility. All three stochastic volatility models considered, seem to devolatize data well.

However, considering the 99% VaR level, the Nadaraya-Watson volatility model is clearly the best. This is attractive, from a practical point of view, since estimations needed to use the Nadaraya-Watson volatility model are easily and rapidly made. For instance, fitting the Nadaraya-Watson model is much less time consuming than, for instance, fitting the GARCH-AR model.

We have seen, that the simple Variance Window volatility model, together with most of the distributions, works very well for all the data sets. This result leads to the question; is it worthwhile using more complex volatility models?

The integrated volatility model, used in the mixing relation, does not seem to work well in a VaR situation. However, it might be too early to rule this model out, since data needed to make thorough enough investigations was not available.

Regarding distributions to be used in a VaR setting, we have seen that the Hyperbolical, Pearson VII and the Generalized Pareto, in most cases, perform better than the Gaussian. What combination of distribution and volatility model to use, depends on the data to which it is applied. Also, different evaluation methods give different answers to this question. The LPFA indicates that the Pearson VII distribution is the best choice in most cases. On the other hand, the FOEL gives no unambiguous answer to which distribution to use.

Further studies A natural extension of this work, would include investigation of multidimensional volatility models, used on a portfolio of financial assets. Another extension, that we would like to have done here, is more extensive studies of the mixing relations, provided availability of tick data. Potentially, as justified by theory and empirical studies, see [5], there are interesting things to be done in this area.

Chapter 8

Background

The founding ideas of this master thesis are based on the paper *Risk Management Based on Stochastic Volatility*, by Ernst Eberlein, Jan Kallsen and Jörn Kristen, see [23]. One of the extensions found in this thesis, is that we investigate other stochastic volatility models than in [23]. The studies in [23] are restricted to comparing the Hyperbolical and the Gaussian distribution, while we also propose both the Pearson VII and the Generalized Pareto distribution as marginal distributions for financial log-returns.

The GARCH-AR and the Variance Window stochastic volatility models are e.g. proposed in Erik Brodin's master thesis *On the Empirical Distribution of Log-returns*, see [20]. Further, we got the idea of using the Nadaraya-Watson stochastic volatility model from the paper, *A simple non-stationary model for stock returns*, see [49], by Cătălin Stărică and Holger Drees. The authors suggest the Nadaraya-Watson stochastic volatility model to be suitable for applications in risk management. It should be mentioned that, we have not seen the Nadaraya-Watson stochastic volatility model used in risk management before.

Further, Stărică and Drees also suggest the Pearson VII distribution, for modelling financial log-returns (i.e. log-returns, devolatized by using the Nadaraya-Watson stochastic volatility model). They also show, concerning devolatization, that the Nadaraya-Watson model is superior to both the GARCH and the EGARCH stochastic volatility models. We believe, though, that the GARCH-AR model works far better than the GARCH model.

The use of extreme value theory (i.e. the Generalized Pareto distribution), in risk management, has been suggested in many articles, papers and books. The Generalized Pareto and the Gaussian distributions are the ones most commonly used in risk management (the Gaussian probably even more so, than the Generalized Pareto). These distributions are often used under the assumption of constant volatility.

The semimartingale model is greatly influenced by the work of Ole Barndorff-Nielsen, and Neil Shephard, see e.g. [5]. We have not seen this model used in risk management before.

Bibliography

[1] P. Albin, *On Basic Stochastic Calculus*, Unpublished Manuscript, 2001.

[2] P. Albin, *Stokastiska Processer*, Studentlitteratur, 2003.

[3] P. Albin, *Fifteen Lectures, From Lévy Processes to Semimartingales*, Unpublished Manuscript, 2002.

[4] L. Bachelier, *Théorie de la Spéculation*, Annales de l'Ecole Normale Superieure, Series 3, 17, pp. 21-86, 1900.

[5] O.E. Barndorff-Nielsen and N. Shephard, *Financial volatility and Lévy based models*, Manuscript, 2002.

[6] O.E. Barndorff-Nielsen and N. Shephard, *Integrated OU processes and non-Gaussian OU-based stochastic volatility models*, Research Report, 14, MaPhySto, 2001.

[7] O.E. Barndorff-Nielsen, *Processes of normal inverse Gaussian type*, Finance and Stochastics, 2, pp. 41-68, 1998.

[8] O.E. Barndorff-Nielsen, *Exponentially decreasing distributions for the logarithm of particle size*, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 353, pp. 401-419, 1977.

[9] O.E. Barndorff-Nielsen, E. Nicolato and N. Shepard, *Some recent developments in stochastic volatility modelling*, Quantitative Finance, 2, pp. 11-23, 2002.

[10] O.E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick, *Lévy Processes: Theory and Applications*, Birkhäuser, 2001.

[11] Basel Committee on Banking Supervision ,www.bis.org, 1998.

[12] P. Best, *Implementing Value at Risk*, John Wiley & Sons, 1998.

[13] F. Black and M. Scholes, *Pricing of Options and Corporate Liabilities*, Journal of Political Economy, 81, pp. 637-659, 1973.

[14] T. Bollerslev, *Generalised Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31 ,pp. 307-327, 1986.

[15] G.E.P. Box and G.M Ljung, *On a Measure of Lack of Fit in Time Series Models*, Biometrica, 65, pp. 297-303, 1978.

[16] G.E.P. Box and D.A. Pierce, *Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Seires Models*, Journals of the American Statistical Associatoin, 65, pp. 1509-1526, 1970.

[17] G.W. Brier, *Verification of forecasts expressed in terms of probability*, Monthly Weather Rievew, 75, pp. 1-3, 1950.

[18] W. Brock, W. Dechert and J. Scheinkman, *A Test for Independence Based on the Correlation Dimension*, Econometric Reviews, 15, pp. 197-235, 1996.

[19] W. Brock, W. Dechert and J. Scheinkman, *Nonlinear Dynamics, Choas and Instability: Statistical theory and Economic Evidende*, Cambridge University Press, 1991.

[20] E. Brodin, *On the Log-returns of Empirical Financial Data*, Master's Thesis, Chalmers University of Technology, 2002.

[21] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, 2001.

[22] K. Dowd, *Beyond Value at Risk*, John Wiley & Sons, 1998.

[23] E. Eberlein, J. Kallsen and J. Kristen, *Risk Management Based on Stochastic Volatility*, FDM Preprint 72, University of Freiburg, 2001.

[24] R.F. Engle, *Autoregressive Conditional Heteroskedasticity with estimates of the variance of U.K inflation*, Econometrica, 50, pp. 987-1008, 1982.

[25] E.J. Elton, M.J. Gruber, S:J. Brown and W.N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*, John Wiley & Sons, 2003.

[26] T. Gasser, J. Engel and B. Seifert, *Nonparametric function estimation*, C.R Rao, ed., *Handbook of statistics, Vol. 9* , Elsevier Science Publishers B.V., 1993.

[27] J.C. Hull, *Options, Futures and other Derivatives.*, Prentice-Hall, 2000.

[28] *Investor Words*, www.investorwords.com.

[29] J. Jacod and A.N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, 1987.

[30] P. Jorion, *Value at Risk*, McGraw-Hill, 1997.

[31] L. Kanzler, http://www2.gol.com/users/kanzler.

[32] F.C. Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, 1998.

[33] P. Kupiec, *Techniques for verifying the accuracy of risk management models*, The Journal of Derivatives, 3, pp. 73-84, 1995.

[34] LO tidningen, 12, pp. 12-13, 2003

[35] R.S. Liptser and A.N. Shiryaev, *Statistics of random processes. I. General theory. II. Applications*, Springer-Verlag, 1977, 1978.

[36] P. Langevin, *Sur la théorie du mouvement brownien*, Comptes Rendus de l'Académie des Sciences, 146, pp. 530-533, 1908.

[37] H. Lopez, *Regulatory valuation of value-at-risk models*, Mimeo. Research and Market Analysis Group, Federal Reserve Bank of New York, 1996.

[38] H. Markowitz, *Portfolio Selection*, Journal of Finance, 7, pp. 77-91, 1952.

[39] P.Y. Moix, *The Measurement of Market Risk*, Springer-Verlag, 2001.

[40] J.P. Morgan and Reuters, *Riskmetric*^{TM}, www.riskmetrics.com

[41] T.C. Mills, *The econometric modeling of financial time series*, Cambridge University Press, 1999.

[42] E.A. Nadaraya, *On estimating regression*, Theory of Probability and its Applications, 9, pp. 141-142, 1964.

[43] P. Penza and V.K. Bansal, *Measuring Market Risk with Value at Risk*, John Wiley & Sons, 2001.

[44] K. Prause, *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*, Ph.D. thesis, University of Freiburg, 1999.

[45] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.

[46] P. Protter, *Stochastic Integration and Differential Equations*, Springer-Verlag, 1990.

[47] K.I. Sato, *Lévy Processes and infinitely divisible distributions*, Cambridge University Press, 1999.

[48] A.N. Shiryaev, *Essentials of Stochastic Finance*, World Scientific, 1999.

[49] C. Starica and H. Drees, *A simple non-stationary model for stock returns*, Manuscript, 2002.

[50] M. Stevens, *Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive tables*, Journal of the Royal Statistical Society, Series B 32, pp. 115-122, 1970.

[51] G.S. Watson, *Smooth regression analysis*, Sankhya, Series A 26, pp. 359-372, 1964.

[52] N. Wiener, *Differential Space*, Journal of Mathematical Physics, 2, pp. 131-174, 1923.