On the logreturns of empirical financial data

Erik Brodin

August 28, 2002

Abstract

In this master thesis we investigate empirical logreturns from the DAX and Olsen DM/Dollar data. We also investigate models which have the ambition to behave as empirical data. We show that the Generalized Hyperbolic model suits the data well and that there exists models which can model the dependence structure of empirical data.

Acknowledgements

I would like to deeply thank my supervisor Patrik Albin for fruitful discussions and support throughout the process of writing this thesis. Patrik Albin has always taken his time for me and inspired me to learn more about higher stochastics. I also want to thank my friends I met during my studies at Chalmers University of Technology and Swiss Federal Institute of Technology in Zürich. You know who you are and I could not have lasted as long as I did without you. I also want to thank the members of Hexaden and of course my family including my adorable girlfriend. Thank you!!

Contents

1 Introduction

In the begining of the 20th century the field of Mathematical finance was born. Louis Bachelier defended his thesis "Théorie de la Spéculation" in march 1900, which was the first work concerning stock and option markets. He described price fluctuations with Brownian motion, which was discovered by the botanist Robert Brown in 1828 and commonly claimed to be first theoretically analyzed by Einstein in 1905[1]. The first mathematically rigourous construction of Brownian motion was made by Wiener in 1923. Observe that Bachelier was earlier then Einstein to use the Brownian motion mathematicly.

More precisely, Bachelier tried to analyze the market of so called "rentes" traded at the Bourse de Paris. Rentes was a compensation for lost property of the aristocrats in the French revolution and could be viewed as a perpetual bond[2] paying a yearly interest rate. Bachelier argued that the price of a rentes moved as a stochastic process with normal distributed and independent increments, a Brownian motion. However, this also implies a positive probability for negative prices on a rentes. Bachelier regarded that as completely negligible due to the values of the standard deviation which he called coefficient of nervousness of a security. Sadly, the work by Bachelier was forgotten and he was not honored until many years after his death. Further information about Bachelier thesis and life can be found at the Bachelier Society[3].

In the 1960th Bachelier was rediscovered and in 1965 Samuelson proposed a geometric Brownian motion as a model for a security, the so-called Samuelson model. This to forbid negative values for the stock model. Although it has its flaws it is the most widely used model for a security. During the 1970th the field of mathematical finance virtually exploded. This mostly due to the work by Fischer Black and Myron Scholes in 1973 where they presented a consistent mathematical treatment of options and thereby allowed full-scale trading. However, by the use of the Samuelson model certain problems arises in risk control and option pricing as the model do not satisfy characteristic behavior of finance data. Our work in this master thesis will concern empirical investigation of models which better fits the empirical observed distribution of securities.

The organization of the thesis is as follows. In Section 2 we describe

[1]Robert Brown observed irregular movement of pollen in water. Einstein explained that the motion was caused by the interaction between the pollen and the water molecules

[2]A bond with no maturity date. Perpetual bonds are not redeemable and pay a steady stream of interest forever.

[3]http://www.bachelierfinance.com

the Samuelson model. Section 3 will investigate if the Samuelson model is a well describing model for empirical financial data. Section 4 and 5 present alternative models. In Section 6 we test the alternative models and Section 7 gives a brief summary of our conclusions.

2 Samuelson Model

The Samuelson model from 1965 of a stock $S(t)$ in continuous time is also known as the Black-Scholes model. The stock is assumed to follow a so called geometric Brownian motion, i.e.

$$S(t) = S(0)e^{\mu t+\sigma B(t)}, t \geq 0 \tag{1}$$

where $\mu \geq 0$, $\sigma > 0$ and $B(t)$ is a standard Brownian motion. The constant μ is called the drift of the stock's logprice and σ is called the volatility, i.e what Bachelier called coefficient of nervousness of a security.

Definition 1 (Standard Brownian Motion) *Recall that a stochastic process $B = B(t), t \geq 0$, is a standard Brownian Motion in law if the following holds*

1. *B has independent increments, i.e. for $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$ we have that $B(t_1) - B(t_0), \ldots, B(t_n) - B(t_{n-1})$ are independent stochastic variables.*
2. *B has stationary increments, i.e. $B(t+h) - B(h) \stackrel{d}{=}$ [4] $B(t) - B(0)$ for $h > 0$.*
3. *For $0 \leq s < t$, $B(t) - B(s) \sim N(0, t-s)$, i.e. $B(t) - B(s)$ is normally distributed with mean 0 and variance $t - s$.*
4. *$B(0) = 0$.*

The Equation 1 is equivalent, through Itô's[5] formula, to the assumption that the stock value develop as

$$dS(t) = S(t)(\mu^* dt + \sigma dB(t)), t \geq 0$$

where $\mu^* = \mu + \frac{\sigma^2}{2}$. However, the Brownian motion has more properties than indicated by the definition. For example, the sample path of a Brownian motion B can be choosen to be continuous. This is very important.

As a consequence we have

$$dX(t) \equiv log(\frac{S(t)}{S(t-1)}) = \mu + \sigma(B(t) - B(t-1)) \approx \mu + \sigma dB(t)$$

[4] $X \stackrel{d}{=} Y$ if X and Y has the same distribution

[5] Tranformation rule from stochastic calculus. See for instance [14]

where $B(t) - B(t-1) \approx dB(t) \sim N(0,1)$ due to the independent increments of the standard Brownian Motion. This implies that the so called logreturns, $dX(t)$, are i.i.d and $dX(t) \sim N(\mu, \sigma^2)$. Notice that via Taylor expansion we have

$$dX(t) = log(\frac{S(t)}{S(t-1)}) \approx \frac{S(t) - S(t-1)}{S(t-1)}$$

In this thesis we study the daytime logreturns $dX(t)$ of the so called performance index DAX[6] from 1992 to 1999, 1876 data points. The reasons for this and not studying stocks is that the behavior of financial data is more traceable for performance indexes and we do not need to consider dividends, as a performance index is dividend modified. We will also study a high frequency dataset, but more about that later. We have manipulated the data sets where we have $dX(t) = 0$ by simple replacing this value with $\pm min(abs(dX(t)) \neq 0)/2$, $t \geq 0$ where $\pm$ is a random sign independent of everything else[7]. Another important assumption we make is that one day is a fine grid. Indeed, the Samuelson model is a continuous time model but we only have a discrete dataset, which we assume is sampled sufficiently often in order not to loose too much information. This also implies that instead of writing $dX(t)$ we should use the notation $\Delta X(t) = X(t) - X(t-1)$

Figure 1 depicts typical pattern for a stock or a performance index. Notice the edginess of $X(t)$. Figure 2 depicts the corrsponding logreturns $\Delta X(t)$.

Figure 1: Figure describes $\frac{S(t)}{S(0)}$ for DAX

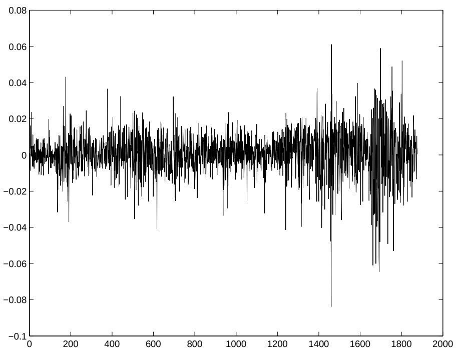

Figure 2: Figure describes $\Delta X(t)$ for DAX

[6]der Deutsche Aktieindex
[7]$P[X = -1] = 0.5$ and $P[X = 1] = 0.5$

3 Testing the Samuelson Model

To be able to test the validness of the Samuelson model we will use standard statistical tools. For more information we recommend the textbooks [8] and [21].

As recollection of the last chapter we have assumed $\Delta X(t) \sim N(\mu, \sigma^2)$ and i.i.d for logreturns. We will test two assumptions that are consequences of our model for the logreturns of financial data.

1. Consequtive logreturns $\Delta X(t)$ are independent.

2. Logreturns $\Delta X(t)$ are $N(\mu, \sigma^2)$ distributed.

The result of this investigation are already known facts. However, to stress the importance of the results we feel it necessary to show the tests to the reader of this paper.

3.1 Dependence structure

There exist numerous techniques for detection of dependence within a stochastic process. To study the ACF (autocorrelation function) is one of them. When plotting the ACF of $\Delta X(t)$, in Figure 3, we get a result which support the independence theory as we cannot detect any significant dependences at a 95% confidence interval. This also gives an explanation to the difficulties of forecasting the stockmarket.

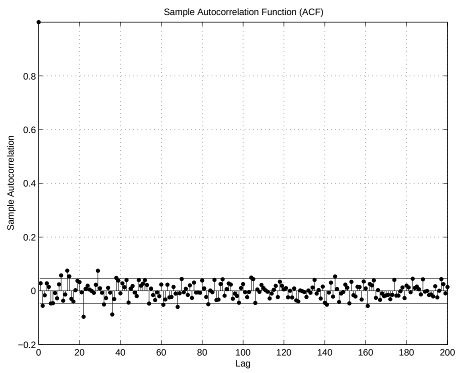

Figure 3: Autocorrelation function for $\Delta X(t)$ from DAX data

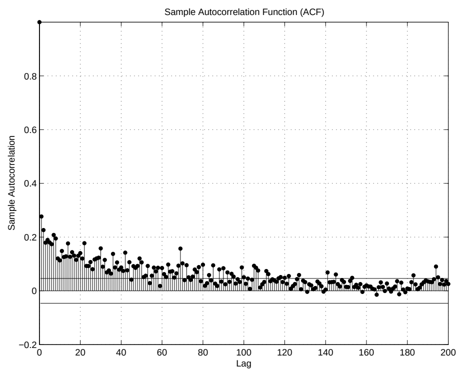

Figure 4: Autocorrelation function for $\Delta X(t)^2$ from DAX data

However, if we instead study $\Delta X(t)^2$, in Figure 4, we have a long-range dependence. This conradicts the independence assumption and therefore we have to relax this assumption. Notice that if ACF really measured independence other than linear, as it indeed would if $\Delta X(t)$

where Gaussian, then $\Delta X(t)^2$ would also have been independent. This contradicts our findings.

3.2 Normal Distribution

Our second issue to investigate is whether $\Delta X(t) \sim N(\mu, \sigma^2)$ or not. To that end we compare a histogram of the logreturns with a normal distribution with parameters $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$.

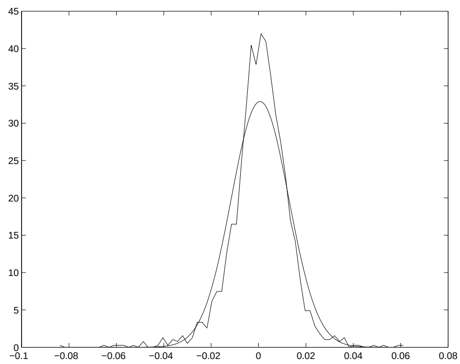

Figure 5: Empirical histogram from the DAX data and estimated normal distribution

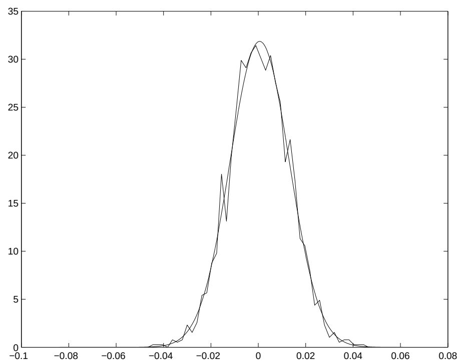

Figure 6: Empirical histogram from normal data and estimated normal distribution

In view of the fact that we have 1876 datapoints, the fit in Figure 5 is not good and should have looked like Figure 6 if $\Delta X(t) \sim N(\mu, \sigma^2)$. However, to find more information we present a quantile-quantile plot (QQ plot) with normal distribution assumption in Figure 7. In the QQ plot we can clearly see that the tails of $\Delta X(t)$'s distribution are much heavier then a normal distribution, so called semiheavy tails [15]. This is very important if focused primarily is on extreme events, as they only depend on the tails. Events as the Black Monday[8] 1987 would not be possible, or at least exceptionally unlikely, if the tails were following a normal distribution.

See Figures 8 and 9 for a comparison of logreturns of our DAX data with a simulated i.i.d process $Y(t) \sim N(\hat{\mu}, \hat{\sigma}^2)$. Notice that extreme values for the simulated data are less extreme then for the actual DAX data.

[8]On October 19, 1987 the Dow Jones Industrial Avarage lost 22.6% of its value. You can compare this with the one day loss of 12.9% which began the famous stockmarket crash of 1929.

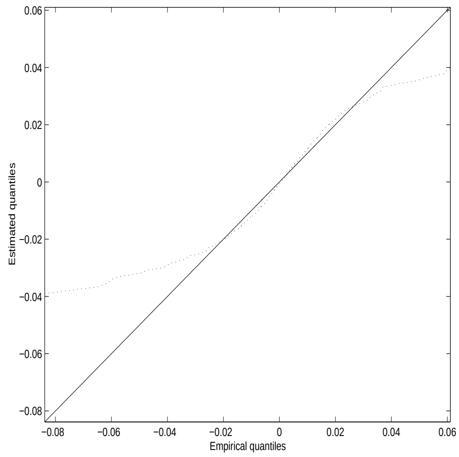

Figure 7: Figure describes a QQ plot for logreturns of the DAX data

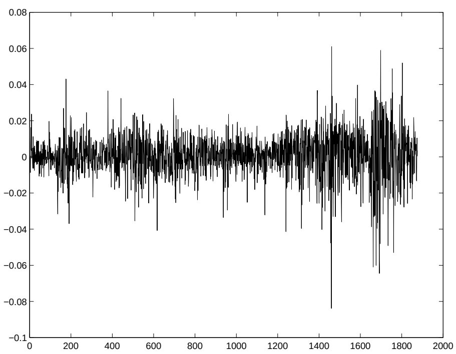

Figure 8: Figure describes $\Delta X(t)$ for DAX

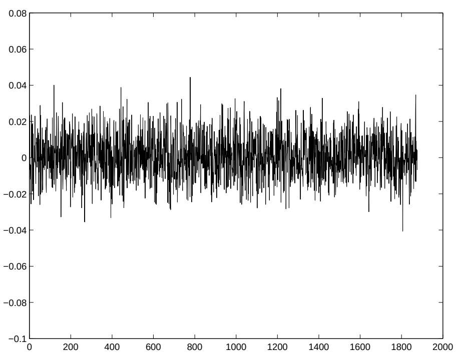

Figure 9: Simulated i.i.d normal distributed process

3.3 Conclusion

We have shown that the Samuelson model does not seem supported by empirical data. The logreturns are neither independent or normal distributed. To be able to make a better model we will have to present more theory. However, the stationary assumption will still be assumed due to dramatically enhanced complexity of theory if it is relaxed. Other investigations have, in addition to failure of the independence and distribution assumptions, reported clustring of extremal events [11] e.g. a large absolute value is often followed by a another large absolute value.

4 Generalized Hyperbolic model

In the previous chapter we came to the conclusion that there are flaws in the Samuelson model. To be able to relax the assumption of a geometric Brownian Motion we will consider a more general class of processes.

Definition 2 (Lévy Process) *Recall that a stochastic process $L = L(t), t \geq 0$, is a Lévy Process in law if the following holds*

1. *L has independent increments.*
2. *L has stationary increments.*
3. *$L(0) = 0$.*

Observe that the Brownian Motion is a Lévy Process. Brownian Motion is also the only continuous Lévy Process (not trivial to show) which makes it easier to use in practice as well as in theory. However, as we have shown that empirical logreturns are not normal distributed and are in fact semiheavy tailed, we cannot use the geometric Brownian Motion as a model for financial data. If we replace that model with another Lévy process based model, it follows that the stock process is not continuous and thus have discontinuities. We present the following model

$$S(t) = S(0)e^{\mu t + \sigma \widetilde{L(t)}} = S(0)e^{L(t)}, t \geq 0 \tag{2}$$

where $\widetilde{L(t)}$ and $L(t)$ are Lévy Processes. With this model we can capture the semiheavy tails if we choose the infinitely divisible distribution [10] of increments of $L(t)$ in a suitable way. A family of very flexible distributions which has turned out to fit empirical data extremely well is the generalized hyperbolic distributions which was introduced by Barndorff-Nielsen in 1977 [2] and first applied to finance by Eberlein and Keller in 1995 [12]. If we assume that the stationary increment distribution of the Lévy process $L(t)$ belongs to the generalized hyperbolic distribution family we get the so called Generalized Hyperbolic model. This will, due to properties of the Generalized Hyperbolic distribution, result in a pure jump process [4]. However, as we only study the process at a discrete grid, this will not really be seen in our applications.

4.1 Generalized Hyperbolic Distribution

Ole E. Barndorff-Nielsen introduced the generalized hyperbolic distribution in 1977 to model grain size distributions of wind blowned sand [2]. We will here present the univariate generalized hyperbolic distribution and a few useful theorems. The theorems have been cited from the Ph. D. thesis by Karsten Prause [17].

Definition 3 (Univariate GH distribution) *A univariate GH distribution is defined by the following Lebesgue density*

$$gh(x; \lambda, \alpha, \beta, \delta, \mu) = a(\lambda, \alpha, \beta, \delta)(\delta^2 + (x-\mu)^2)^{\frac{(\lambda-\frac{1}{2})}{2}} K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)^2})e^{\beta(x-\mu)}$$

where $a(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2-\beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})}$, K_λ *is a modified Bessel function of third kind and* $x \in \mathbb{R}$. *The range of the parameters is* $\mu \in \mathbb{R}$, $\delta \geq 0, |\beta| < \alpha$ *if* $\lambda > 0$, $\delta > 0, |\beta| < \alpha$ *if* $\lambda = 0$ *and* $\delta > 0, |\beta| \geq \alpha$ *if* $\lambda < 0$.

A useful theorem, from [5] Theorem I, is the linearity of the GH distribution.

Theroem 4 (Linearity of GH) *For a linear transformation* $Y = aX + b$, *we have* $Y \sim GH$ *if* $X \sim GH$. *We also have that* $\lambda_Y = \lambda_X$, $\alpha_Y = \frac{\alpha_X}{|a|}$, $\beta_Y = \frac{\beta_X}{|a|}$, $\delta_Y = \delta_X|a|$ *and* $\mu_Y = \mu_X|a| + b$

The parameter μ is a location parameter, δ is a scale parameter, β is a skewness parameter and α affects the size of the kurtosis. Another important aspect of GH is that a vast amount of distributions such as normal distribution and student-t are special cases or limiting distributions of GH. We also notice, which is a more or less general feature of infinitely divisibly distributions [10], that the GH distribution can be represented as a mixture of a normal distriution[9]. This mixture takes the following form

$$gh(x; \lambda, \alpha, \beta, \delta, \mu) = \int_0^\infty N(x; \mu + \beta w, w)gig(w; \lambda, \delta^2, \alpha^2 - \beta^2)$$

where N is a normal density function and $gig(x; \lambda, \chi, \psi)$ denotes the density function of the general inverse gaussian distribution.

[9]See [10] chapter 6.4 on mixing

Definition 5 (GIG distribution) *A univariate GIG distribution is defined by the following Lebesgue density*

$$gig(x; \lambda, \chi, \psi) = \frac{(\frac{\chi}{\psi})^{\frac{\lambda}{2}}}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} e^{-\frac{\chi x^{-1}+\psi x}{2}}, \; x > 0$$

where $\lambda \in \mathbb{R}$ *and* $\psi, \chi \in \mathbb{R}_+$

For those not familiar with the mixing concept we present the following clarification. If $\sigma^2 \sim GIG$ distributed and $\epsilon \sim N(0,1)$ then we have $x = \mu + \beta\sigma^2 + \sigma\epsilon$ GH distributed [2].

As stated above, the expression for the GH density includes a modified Bessel function with index λ. If we fix special values of λ and use calculation rules for the modified Bessel function [1] we can reduce the GH distribution to particular more well-known distributions.

Definition 6 (Hyperbolic distribution) *A univariate HYP distribution, i.e. a GH distribution with* $\lambda = 1$*, is defined by the following Lebesgue density*

$$hyp(x; \alpha, \beta, \delta, \mu) = \frac{\sqrt{\alpha^2-\beta^2}}{2\delta\alpha K_1(\delta\sqrt{\alpha^2-\beta^2})} e^{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)}$$

where $x, \mu \in \mathbb{R}$*,* $\delta \geq 0$ *and* $|\beta| < \alpha$

Definition 7 (Normal Inverse Gaussian distribution) *A univariate NIG distribution, i.e a GH distribution with* $\lambda = -\frac{1}{2}$*, is defined by the following Lebesgue density*

$$nig(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)} \frac{K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\sqrt{\delta^2+(x-\mu)^2}}$$

where $x, \mu \in \mathbb{R}$*,* $\delta \geq 0$ *and* $0 \leq |\beta| < \alpha$

Theorem 4 and Equation 2 allow us to do the following calculation in the generalized hyperbolic model

$$\Delta X(t) = log(\frac{S(t)}{S(t-1)}) = \mu + \sigma(\tilde{L}(t) - \tilde{L}(t-1)) = \mu + \sigma \, \Delta\tilde{L}(t) = \Delta L(t)$$

where $\Delta X(t) = \Delta L(t) \sim GH$. We are now sufficiently prepared to model a semiheavy tailed distribution for our logreturns.

4.2 Maximum Likelihood Estimator

To estimate the distribution of logreturns under Samuelson's Model we simply use the unbiased estimators $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$. However, the GH distribution are more complex and we do not have closed formulas for unbiased estimators of GH parameters. Therefore we need a different approach. The method of maximum likelihood is a common method for parameter estimation and curve fitting in statistics.

Given observed i.i.d data $x_1, x_2, \ldots, x_n$ we define the likelihood function of parameter θ as

$$lik(\theta) = f(x_1, x_2, \ldots, x_n|\theta)$$

where f is the frequency function. Note that if the distribution is discrete the likelihood function gives the probability of observing the given data as a function of the parameter θ. With maximum likelihood estimator (MLE) we maximize the probability. Since $x_1, x_2, \ldots, x_n$ are assumed i.i.d and the natural logarithm is a monotonic function we may instead maximize the log likelihood function

$$l(\theta) = \sum_{1}^{n} log[f(x_i|\theta)]$$

The MLE also have good theoretical properties such as asymptotical efficiency according to Cramer-Rao Inequality. Of course, we can also use this approach to estimate the parameters for a normal distribution. In fact, then we end up with the unbaised estimators $\hat{\mu}$ and $\hat{\sigma}^2$.

For the GH distribution we have the following log likelihood function to maximize

$$\begin{gathered} l(\theta) = l(\lambda, \alpha, \beta, \delta, \mu) = \\ nlog(a(\lambda, \alpha, \beta, \delta) + (\tfrac{\lambda}{2} - \tfrac{1}{4})\sum_{i=1}^{n} log(\delta^2 + (x_i - \mu)^2) + \\ \sum_{i=1}^{n}[log(K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x_i - \mu)^2} + \beta(x_i - \mu))] \end{gathered}$$

In Figure 10 below, we can see a fitted GH distribution and Normal distribution for the logreturns of the DAX data. We have also, in Figure 11, zoomed the left tail to be able to visualize that we have captured the semiheavy tailness of logreturns.

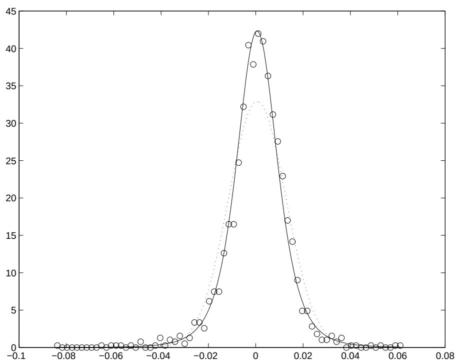

Figure 10: Fitted GH distribution (-) and normal distribtuion (. . .) for empirical data.

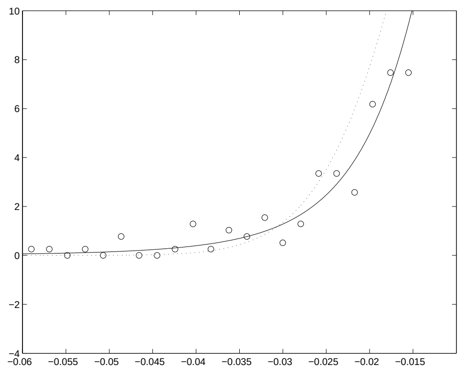

Figure 11: Zoomed in left tail from Figure 10.

The computer time consumed for MLE calculation of GH is quite high. This is mostly due to the apperance of the Bessel function K_λ in the model. This can be circumverted by fixing λ. We will of course loose one degree of freedom and hence receive a less good fit. There is also a problem with the flatness of the log likelihood function. This may be somewhat improved by good initial values. However, due to the same flatness, one may expect the fitted GH model to be quite robust for small deviations of the parameters. As we have a closed formula for the log likelihood function we can enhance the computation speed by using derivatives in closed formulas.

The Generalized Hyperbolic Model seems to model the logreturns distribution very well but we still have to investigate independence. In Section 3 we learned that the assumption about independent logreturns can not be justified. To be able to deal with a "memory" among the logreturns we now turn to a discussion of stochastic volatility.

5 Stochastic Volatility

The volatility was described as the coefficient of nervousness by Bachelier and it can also be regarded as the "temperature" of the stockmarket. As we have shown above the logreturns are not independent and their squares or absolute values are visibly correlated. To model this we can use a stochastic volatility process with dependent increments. Intuitively we can think of this as if it was warm weather yesterday it seems reasonable that the temperature is high today too. However, this process will have to move quite slowly compared to the fluctuations of the logreturns and can therefore not model extremal events, i.e., it should vary very little between consecutive integer time points. We use the following model [11]

$$S(t) = S(0)e^{\sigma(t)L(t)}, t \geq 0 \tag{3}$$

where $\sigma(t) > 0$ is a stochastic stationary process independent of the Lévy process $L(t)$. Moreover, if we study the logreturns and use the assumption that $\sigma(t)$ is a "slow" process we land on the following logreturn model.

$$\Delta X(t) \equiv log(\frac{S(t)}{S(t-1)}) = \sigma(t)L(t) - \sigma(t-1)L(t-1) = \sigma(t)\Delta L(t) \tag{4}$$

Notice that the idea is to capture the dependence in $\sigma(t)$ and the extrem events in $\Delta L(t)$. To motivate the need for a stochastic rather than constant volatility we simply run a window with size n over the DAX data and calculate the variance in each window. This is depiced in Figure 12. Observe that the size n of the window is affecting the variance of that variance and if n is small the variance will be large. We have also plotted the 95 % confidence interval as point lines.

For a second motivation we can take the logarithm of the squared logreturns which leads to the following equation

$$log(\Delta X(t)^2) = log(\sigma(t)^2) + log(\Delta L(t)^2) \tag{5}$$

This can be interpreted, from an engineering point of view, as a signal disturbed by noise. Just for visualisation of the stochastic volatility we take, in Figure 13, a running-mean smoothing backward window of the dataset, i.e.

$$log(\hat{\sigma}(t)^2) = \frac{1}{k}\sum_{i=0}^{k-1} log(\Delta X(t-i)^2) \tag{6}$$

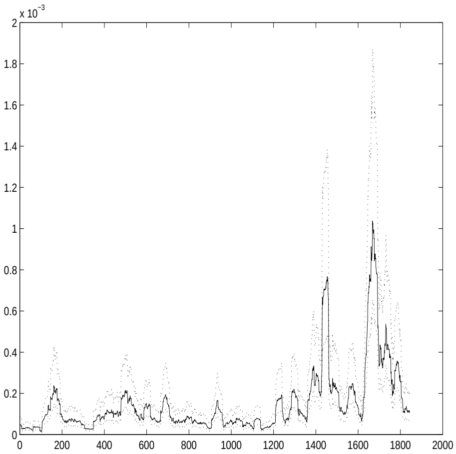

Figure 12: Volatility(-) with window size $n = 30$ for DAX data. The confidence interval is represented by $(\dots)$.

In Figure 13 we can clearly see how the volatility, which is white, varies with time.

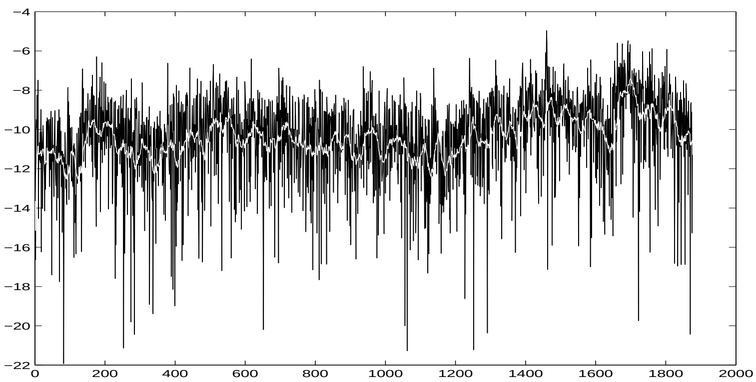

Figure 13: Volatility (white) backwindow with $n = 20$ on $log(\Delta X^2(t)$, for DAX data (-)

If we succeed in getting the devolatilized logreturns, $\frac{\Delta X(t)}{\sigma(t)}$, independent we assume we can model them with a Lévy process. However, to accquire the devolatilized logreturns we have to estimate the stochastic volatility. This is done within a parametric framework.

5.1 Recursive models

To be able to estimate the stochastic volatility $\sigma(t)$, and hence the dependence structure, we assume that we can calculate the future volatility values based on the values of the past. This results in a recursive model for the stochastic volatility.

Numerous models have been assumed over the years, but most of them are based on then underlying noise process being a Brownian motion. This is mainly due to the existence of extensive theory in parameter estimations with Gaussian noise. However, as we want to model semiheavy tails, we have to be a bit more daredevilish. We will present three different recursive volatility models.

5.1.1 Heteroscedastic models

A widely used approach are the so called conditional heteroscedastic models which have the for us convenient property that a large squared volatility value is likely to be followed by another large value. This fits with the empirical dependence structure of the logreturns[10].

Definition 8 (GARCH Model) *Consider $X(t) = \sigma(t)L(t)$. Generalized Autoregressive Conditional Hetroscedastic is defined as*

$$GARCH(p,q) : \sigma(t) = \beta + \sum_{j=1}^{p} \lambda_j X(t-j)^2 + \sum_{k=1}^{q} \delta_k \sigma(t-k)^2$$

where $\beta, \lambda_j, \delta_k > 0, \forall j, \forall k$. Condition for second order stationarity[11]: $\sum_{j=1}^{p} \lambda_j + \sum_{k=1}^{q} \delta_k < 1$.

For $q = 0$ we write GARCH(p,0)=ARCH(p)

It turns out that $\sigma(t)$ in Definition 8 moves rapidly and therefore it is common to slow it down by autoregression.

Definition 9 (AR Process) *Given $p > 0$, a process $X(t)$ is a Autoregressive process of order p if $X(t)$ is stationary and $\forall t \in \mathbb{Z}$*

$$AR(p) : X(t) - \sum_{k=1}^{p} \alpha_k X(t-k) = L(t)$$

where $L(t)$ is noise.

A good and simple model is the combined AR(1) and GARCH(1,1) model

[10]Especially the clustring phenomena [11], that could result from a fast moving stochastic volatility

[11]Also known as weak stationarity, covariance stationarity, stationarity in the wide sense or just stationarity.

$$X(t) = \mu(t) + \sigma(t)L(t)$$

$$\mu(t) = \alpha X(t-1)$$

$$\sigma(t)^2 = \beta + \lambda(X(t-1) - \mu(t-1))^2 + \delta\sigma(t-1)^2$$

with $\beta, \lambda, \delta > 0$ and $|\alpha| < 0$. Observe that $X(t)$ is stationary. However, in this model we have not yet assumed a distribution for $L(t)$. It is standard practice to assume $L(t) \sim N(0,1)$ and then use MLE to estimate parameters. This is not convenient for us due to the semi-heavy tails, but if we run an MLE estimator under normal assumption and then study the residuals $e(t) = X(t) - \hat{X}(t)$ we can try to learn about the true distribution of $L(t)$. This is called pseudo-MLE.

In more detail, we assume that $\mu(t) = E[\Delta X(t)|\Delta X(t-1), \theta(t-1)]$, $\sigma(t)^2 = var(\Delta X(t)|\Delta X(t-1), \theta(t-1))$ and $\Delta X(t)|\mathcal{F}(t-1) \sim N(\mu(t), \sigma(t))$. Here $\theta(t-1)$ are the parameters in the GARCH(1,1)-AR(1) model and $\mathcal{F}(t-1)$ is the information known at time $t-1$. With $\mu(t)$ and $\sigma(t)$ we can form the pseudo-MLE function

$$l(\theta) = l(m(t), v(t)) = -\frac{1}{2}\sum_{t=1}^{n} log(\sigma(t)^2) - \frac{n}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\frac{(\Delta X(t) - \mu(t))^2}{\sigma(t)^2}$$

and investigate the residuals.

5.1.2 Nonparametric models

If we remember the engineering approach in Equation 5 and try to extract the hidden volatility signal $log(\sigma(t)^2)$ from the noise $log(\Delta L(t)^2)$ it is natural to apply a smoothing function as a backward[12] window of the data. Eberlein, Kallsen and Kristen [11] have done this with the same running-mean smoother as in Equation 6. The model can be classified as nonparametric. These authors choose the smoothing parameter k by crossvalidation, i.e minimizing

$$CV(k) = \frac{1}{T}\sum_{t=1}^{T}(log(\Delta X(t)^2) - log(\hat{\sigma}(t)^{-t}))^2$$

where T is the length of the dataset and $log(\hat{\sigma}(t)^{-t} = \frac{1}{k}\sum_{i=1}^{k} log(\Delta X(t-i)^2$.

[12]We do not know the future

To make this model more flexible we can insert a weight function in the smoothing function

$$log(\hat{\sigma}(t)^2) = \frac{1}{k}\sum_{i=0}^{k-1} w_j log(\Delta X(t-i)^2)$$

and try to minimize the BDS[13] statistic of $\Delta L = \frac{\Delta X(t)}{\hat{\sigma}(t)}$. However, we should be very careful when minimizing over a statistic and therefore we also test the model under crossvalidation.

5.1.3 Variance window

The variance window is a simple and transparent model but quite crude. We assume that the volatility $\sigma(t)$ is the variance of the last n days of the dataset. When we choose n small we will get a faster moving volatility but the confidence interval of $\sigma(t)$ will be large. In the variance window case we will also minimize over the BDS statistic.

[13]see section 6.2

5.2 Process model

Instead of calculating the variance based on a recursive model we can assume $\sigma(t)$ to be a stochastic process driven by noise independent of $L(t)$. When chosing a suitable process for $\sigma(t)$ we let us be inspired by Barndorff-Nilesen and Shepard who suggested a process of Ornstein-Uhlenbeck type. This is mainly due to the particular simple Markov dependence structure of a Ornstein-Uhlenbeck processes.

5.2.1 Ornstein-Uhlenbeck Processes

Definition 9.1 (Ornstein-Uhlenbeck Processes) *A stochastic process $\tau(t)$, $t \in \mathbb{R}^+$ is a Ornstein-Uhlenbeck process if it satisfies the stochastic differential equation*

$$d\tau(t) = -\alpha\tau(t)dt + dZ(t) \tag{7}$$

where $\alpha > 0$ and $Z(t)$, $t \geq 0$ is a Lévy Process, which we refer to as the Background Driving Lévy Process (BDLP).

If we take a càdlàg version[14] of the Lévy Process $L(t)$ and $\alpha > 0$ we can solve Equation 7 with

$$\tau(t) = e^{-\alpha t}\tau(0) + \int_0^t e^{-\alpha(t-s)}dZ(s)$$

where the BDLP and $\tau(0)$ are independent. If $\tau(t)$ is stationary and square integrable and $E[\tau(0)] = E[L(1)] = 0$ we will have an autocorrelation function of the form $acf(t) = e^{-\alpha t}$ for $\tau(t), t \geq 0$ [17]. To motivate this we recall the autocorrelation for the squared logreturns $X(t)^2$, assume that $\sigma(t)^2$ is an OU process and try to fit the dependence structure of the OU process to the squared logreturn autocorrelation function. However, to be more flexible we follow Barndorff-Nielsen and also fit a superposition of two[15] independent and stationary OU processes. For $1 > m > 0$ this leads to the following autocorrelation structure

$$acf(t) = me^{-\alpha_1 t} + (1-m)e^{-\alpha_2 t} \tag{8}$$

Observe the excellent fit in Figure 14. Notice that it would be practical to have the marginal distribution for the OU process closed under convolution because of the addition of such random subjects involved.

[14] Continue a droite, limite a gauche. See [14]

[15] Actually, we can fit as many as we like

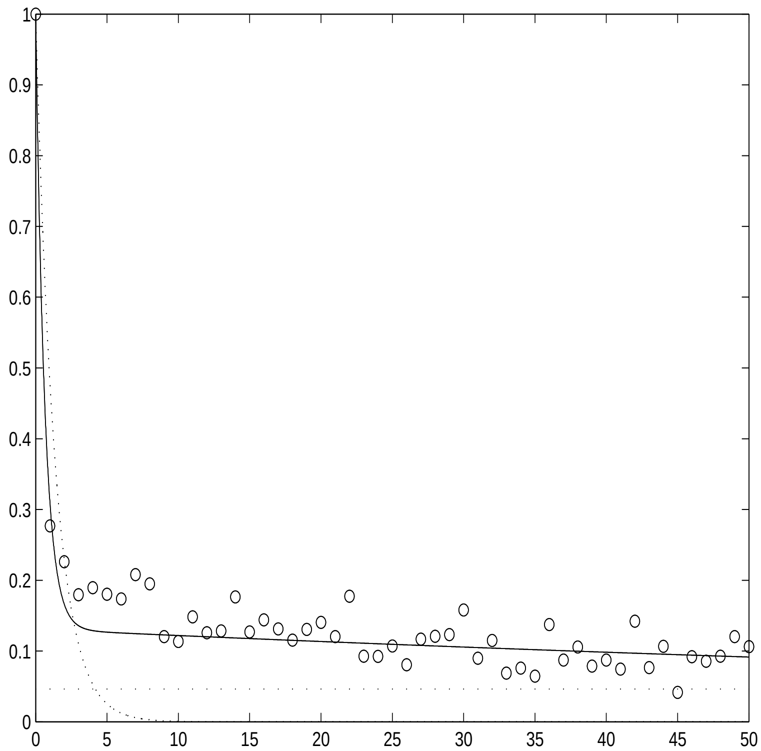

Figure 14: One (...) and two (-) OU processes fitted to correlation structure of squared logreturns (o), DAX data

However, to be able to estimate a stochastic volatility process we will face the problem to separate the processes $\sigma(t)$ and $\Delta L(t)$ in Equation 4. This is unfortunately not trivial. However, if we assume the Generalized Hyperbolic model and use the mixing relationship of the GH distribution we can attach the mixing distribution, the GIG distribution, with an OU dependence structure and try to find the BDLP which makes the marginals of the OU process GIG distributed.

This would lead to a model with logreturns of type

$$dX(t) = \sigma(t)dB(t) + (\mu + \beta\sigma(t)^2)dt \tag{9}$$

$$d\sigma(t)^2 = -\alpha\sigma(t)^2 dt + dZ(t) \tag{10}$$

The processes $\sigma(t)^2$ and $B(t)$ are assumed to be independent. $\sigma(t)^2 \sim$ GIG distributed and $\alpha > 0$.

To be able to use the flexibility of more than one OU process we can restrict us to a fixed λ. For $\lambda = -\frac{1}{2}$ we have a NIG distribution instead of GH distribution and this results in a mixing distribution which is closed under convolution, the Inverse Gaussian distribution.

Definition 10 (The IG distribution) *A univariate IG distribution is defined by the following Lebesgue density*

$$ig(\delta,\gamma) = gig(x; -\tfrac{1}{2}, \chi, \psi) = \tfrac{\delta}{\sqrt{2\pi}} e^{\delta\gamma} x^{-\frac{3}{2}} e^{-\frac{1}{2}(\frac{\delta^2}{x} + \gamma^2 x)},\ x > 0$$

where $\delta, \gamma \in \mathbb{R}^{+}$[16]

Theorem 11 (IG convolution) *Assume* $\gamma > 0$ *and* δ_i *for* $i = 1, \ldots, n$. *The IG distribution is closed under convolution, i.e.*

$$IG(\textstyle\sum_{i=1}^{n} \delta_i, \gamma) = *_{i=1}^{n} IG(\delta_i, \gamma)$$

If we model a superposition of two independent and stationary OU processes which are IG distributed and with parameters $(\delta_1, \gamma, \alpha_1)$ and $(\delta_2, \gamma, \alpha_2)$ respectively, we will have the following autocorrelation function

$$acf(t) = \frac{\delta_1}{\delta} e^{-\alpha_1 t} + \frac{\delta_2}{\delta} e^{-\alpha_2 t}$$

where $\delta = \delta_1 + \delta_2$. It follows from Equation 8 that we can, after estimation of α_1, α_2 and m, construct two independent OU processes IG distributed and parameters $(\delta m, \gamma, \alpha_1)$ and $(\delta(1-m), \gamma, \alpha_2)$ respectively.

Theorem 12 (BDLP distribution of IG OU process) *The BDLP* $Z(t)$ *of the IG distributed OU processes is of the form* $Z(t) = Q(t) + P(t)$ *where* $Q(t)$ *is* $IG(\frac{\delta}{2}, \gamma)$ *distributed and independent of* $P(t)$ *which is of the form* $P(t) = \gamma^{-2} \sum_{i=1}^{N(t)} u_i^2$. $N(t)$ *is a Poisson process with rate* $(\frac{\delta\gamma}{2})^{-1}$ *and* u_i *is standard normal distributed and independent of* $N(t)$ *[4].*

To be able to have a OU process $\sigma^2(t)$ with IG(δ, γ) distribution whatever the value of α [4], we can change the time scale and the BDLP process according to

$$d\sigma^2(t) = -\alpha\sigma(t)dt + d\hat{Z}(t) = -\alpha\sigma(t)dt + dZ(\alpha t) \qquad (11)$$

[16] $\chi = \delta^2$ and $\psi = \gamma^2$

5.2.2 Quadratic Variation

Another approach is to study the so called quadratic variation. This is a very important path property of stochastic processes in stochastic calculus.

Definition 13 (Quadratic variation) *Choose a partition* $\Pi = \{0 = t_0 < t_1 < \ldots < t_n = 1\}$. *The quadratic variation of* $\{X(t), 0 \leq t \leq 1\}$ *over* Π *is defined as*

$$Q(\Pi) = \sum_{k=1}^{n}(X(t_k) - X(t_{k-1}))^2$$

If the limit of $Q(\Pi)$ *exists as* $\max_{1\leq k\leq n}|t_k - t_{k-1}| \rightarrow 0$ *we call the limit the quadratic variation of* X *on* $[0, 1]$

If we instead of $[0, 1]$ had taken $[0, t]$, $t > 0$ the quadratic variation of X over $[0, t]$ would be a function of t and hence a stochastic process which is called the quadratic variation process, $[X](t)$.

The solution of the stochastical differential equation of $dX(t)$ in Equation 9 is

$$X(t) = \int_0^t \sigma(u)dB(t) + \mu t + \beta\sigma(t)^{2*}$$

where the notation $\sigma(t)^{2*} = \int_0^t \sigma(u)^2 du$ is the so called integrated volatility. If we take $\mu = 0$ and $\beta = 0$ or assume that $\sigma(t)^{2*}$ behaves nicely[17] we will have[18] $[X](t) = \sigma(t)^{2*}$.

A more general model to receive $[X](t) = \sigma(t)^{2*}$ would be

$$X(t) = \int_0^t \sigma(u)dB(t) + \alpha(t)$$

where $\alpha(t)$ is a process of bounded variation with continous paths.

This implies that we could approximate the volatility process if we study the quadratic variation of the logreturns. This would indeed be practical as we can be more flexible with the choice of BDLP[19]. However, to be able to calculate an estimator of the stochastic volatility our day to day data is not fine enough. We will have to use intraday data, so call tic data. Thuse we follow Barndorff-Nielsen and Shephard and replace the volatility by an estimator called realised variance or realised volatility defined as

[17] e.g. $\sigma(t)^{2*}$ is a process of bounded variation with continuous paths [14]

[18] This due to the nice properties of the Brownian motion, i.e it has continuous paths

[19] We can also choose another process than OU

$$[X_M](t)^{[2]} = \sum_{j=1}^{M} \Delta X_j(t)^2$$

where M is the number of intraday measurements and

$$\Delta X_j(t) = X((t-1)\Delta + \frac{\Delta j}{M}) - X((t-1)\Delta + \frac{\Delta(j-1)}{M})$$

is the j-th intra-Δ return. The quadratic variation implication is that as $M \to \infty$ we have

$$[X_M](t)^{[2]} - \sigma^2(t) \xrightarrow{p} 0$$ [20]

The ticdata is fairly expensive[21] but Prof. Neil Shepard was very kind to give me the realised variance dataset for the Olsen Dollar/DM data[22], from 1 December 1985 and 2447 tradingdays ahead, for $M = 8$ and $M = 144$. Below, in Figure 15 and Figure 16, we show the realised variance.

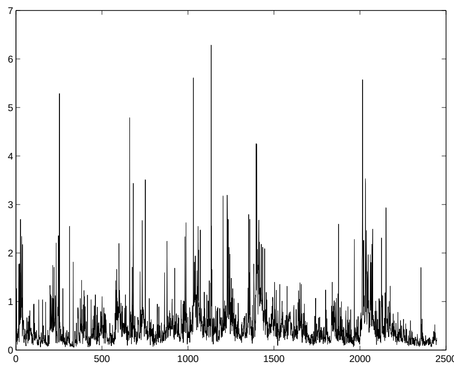

Figure 15: $\sigma^2(t)$ from quadratic variation $M = 144$

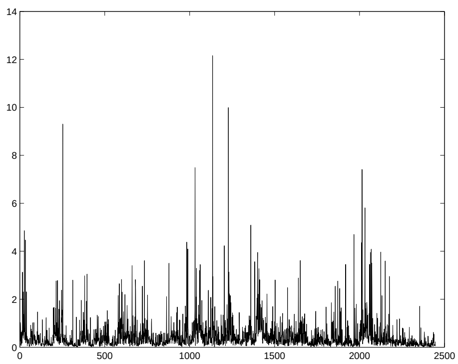

Figure 16: $\sigma^2(t)$ from quadratic variation $M = 8$

To get a feeling for the assumed volatility we also present a histogram of the increments in Figure 17 and Figure 18.

Alternatively we can assume a more complex model then Equation 9-10 with $\lambda = -\frac{1}{2}$, as that in Equation 3. Here $dL(t)$ is a general Lévy Process and if we calculate the quadratic volatility process we

[20] $X_n \xrightarrow{p} Y$ if $P[|X_n - Y| < \epsilon] \to 1$ for all $\epsilon > 0$

[21] Mainly due to Olsen bankrupcy

[22] Originally released in conjunction with the hosting by the Olsen Group of the High Frequency Data in Finance conferences HFDF-I and HFDF-II.

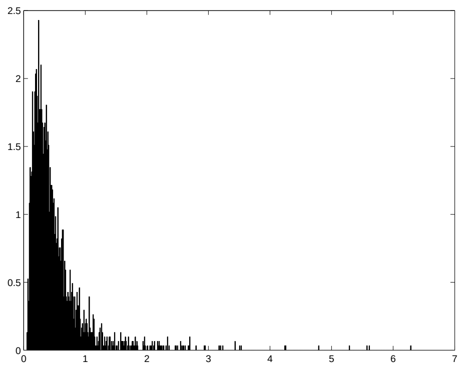

Figure 17: Histogram of $\sigma^2(t)$ increments from quadratic variation $M = 144$

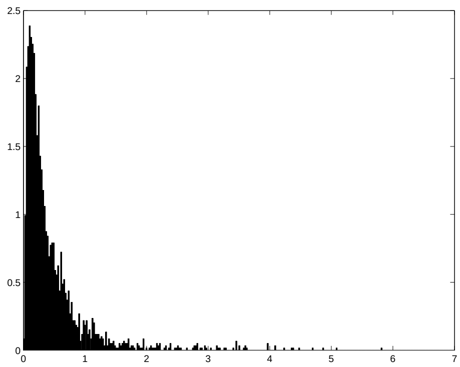

Figure 18: Histogram of $\sigma^2(t)$ increments from quadratic variation $M = 8$

only have $\sigma(t)^{2*} = \int_0^t \sigma(t)^2 du$ if $L(t)$ is a Brownian motion. Otherwise we would have to take the discontinuous part[23] of the process into consideration and have a more complex formula for the quadratic variation.

[23] The jumps

6 Model test

In this section we will evaluate the models previously mentioned. To control how well the models fit empirical data we will test two things. Firstly, if the devolatilized residuals are independent. Secondly, if the estimated distribution is similar to the empirical distribution. Moreover, we will model the noise $\Delta L(t)$ both with increments of Normal distribution and increments of Generalized Hyperbolic distribution. This is to stress the difference between semiheavy tails and non-heavy tails. To be able to evaluate the quadratic variation approach we will model both the DAX data and the Olsen Dollar/DM data. The different models are named after the volatility structure and labeled as follows.

1. Constant volatility
2. Nonparametric
3. Cross volatility
4. Variance window
5. GARCH(1,1)-AR(1)
6. Quadratic variation

The model test follows Eberlein, Keller and Kristen [11] but they used the Hyperbolic distribution as noise increments and a non stationary approach. We will use the Generalized Hyperbolic distribution as a noise increments. The nonparametric model with weight function, the variance window model and the Quadratic variation model have never, by the author of this thesis, been seen statistical tested. We have also used a greater search span in the cross volatility model.

6.1 Modeling

1. With constant volatility we have either the Samulson model, see Section 2, or the Hyperbolic model, see Section 4.1. This, of course, depends on the distribution of $\Delta L(t)$.

2. The estimation of the Nonparametric model is straight forward, see Section 5.1.2. However, it will be computationally demanding due to the "brute force" approach. To be able to have a faster approach we use our assumption about the stochastic volatility and search for a solution in the range $n = 5, \ldots, 15$ which is on a one to two weeks of data base. We also have $w = 0.1, \ldots, 1$ with 0.1 intervalls.

Our results are $n = 5$ and $w = 0.7$ for the DAX data and $n = 6$ and $w = 0.6$ for the Olsen Dollar/DM data.

3. Cross volatility is the nonparametric model with n chosen by crossvalidation, see Section 5.1.2. The crossvalidation was minimized by $n = 22$ for the DAX data and $n = 58$ for the Olsen Dollar/DM data. We notise the high value of n for the Olsen Dollar/DM data but accept it.

4. Variance window from Section 5.1.3 is, as the nonparametric model, also straight forward. We stress the fact about the confidence interval also mentioned in Section 5.1.3.

Our results, for the window size, are $n = 9$ for the DAX data and $n = 15$ for the Olsen Dollar/DM data.

5. The GARCH(1,1)-AR(1) approach, see Section 5.1.1, is modeled by the pseudo-MLE approach. With the residuals we try to fit a GH distributed $\Delta L(t)$ in the hyperbolic case or a Normal distribution in the Samuelson case.

Our results are $\alpha = 0.0334$, $\beta = 0.000003$, $\lambda = 0.0978$ and $\delta = 0.8855$ for the DAX data respective $\alpha = -0.0466$, $\beta = 0.0088$, $\lambda = 0.0539$ and $\delta = 0.9296$ for the Olsen Dollar/DM data.

6. For the Quadratic variation approach, see Section 5.2.2, we use parameters $M = 8$ and $M = 144$. It is devolatilized under normal assumption and as in the mixing model we will also here be inspired by pseudo MLE. However, for simplicity, the adapted process of bounded variation is asumed $\alpha(t) = 0$.

6.2 Independence

As mentioned above the autocorrelation is not a good measure of independence if we are outside the Guassian world. We will use the BDS-test [7], an independence test which reacts sensitively to accumulations of similar values in time series. The matlab code we use has been made by Dr. Ludwig Kansler[24]. To present the test we will use a mathematical description of BDS from [11]. The BDS statistic for fixed parameters $m \in \mathbb{N}$ and $\epsilon > 0$ is defined as

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

where n is the number of observation. For $n_m = n - m + 1$ and $1_\epsilon(s,t) = 1_{[-\epsilon,\epsilon]}(\max_{i\in 0,\ldots,m-1} |X(t+i) - X(s+i)|)$ we have

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

[24]http://www.cs.man.ac.uk/ dbree/Stathis/code.htm

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

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 of independence we have $W_{m,n}(\epsilon)$ asymptotically standard normal distributed. The parameters m and ϵ are chosen after a rule of thumb according to which $m = 4$ and $\epsilon = 1.5\sqrt{s_L^2}$ where s_L^2 is the empirical variance of $\Delta\hat{L}$ [6].

Vol model, DAX	BDS Statistic	p-value	iid hypothesis
Constant Volatility	15.222	0	rejected
Nonpara Volatility	0.3483	0.7276	pass
Cross Volatility	4.3455	0	rejected
Variance Window	-0.3342	0.7382	pass
GARCH-AR	-0.2471	0.8048	pass

Table 1: BDS test for devolatilized DAX residuals

Vol model, Olsen	BDS Statistic	p-value	iid hypothesis
Constant Volatility	6.1062	0	rejected
Nonpara Volatility	-0.0035	0.9972	pass
Cross Volatility	3.9274	0.0001	rejected
Variance Window	0.0918	0.9268	pass
GARCH-AR	-0.9126	0.3615	pass
QVariation $M = 8$	-2.2798	0.0226	rejected
QVariation $M = 144$	-4.1167	0	rejected

Table 2: BDS test for devolatilized Olsen residuals

From the tables above we can see the results from the BDS test at a 5% significance level. It is clear that constant volatility model and cross volatility model fails for both the DAX and the Olsen Dollar/DM data. We also notice that the quadratic variation model fails and for $M = 144$ we have a poorer result than for $M = 8$ although it should be better, if the model was correct, due to tighter sampling. The other models, nonparametric, variance window and GARCH-AR, passed the test.

We also, for illustrative reasons, present a plot on the autocorrelation of the squared devolatilized residuals of the Olsen Dollar/DM data. The quadratic volatility model is represented by $M = 144$.

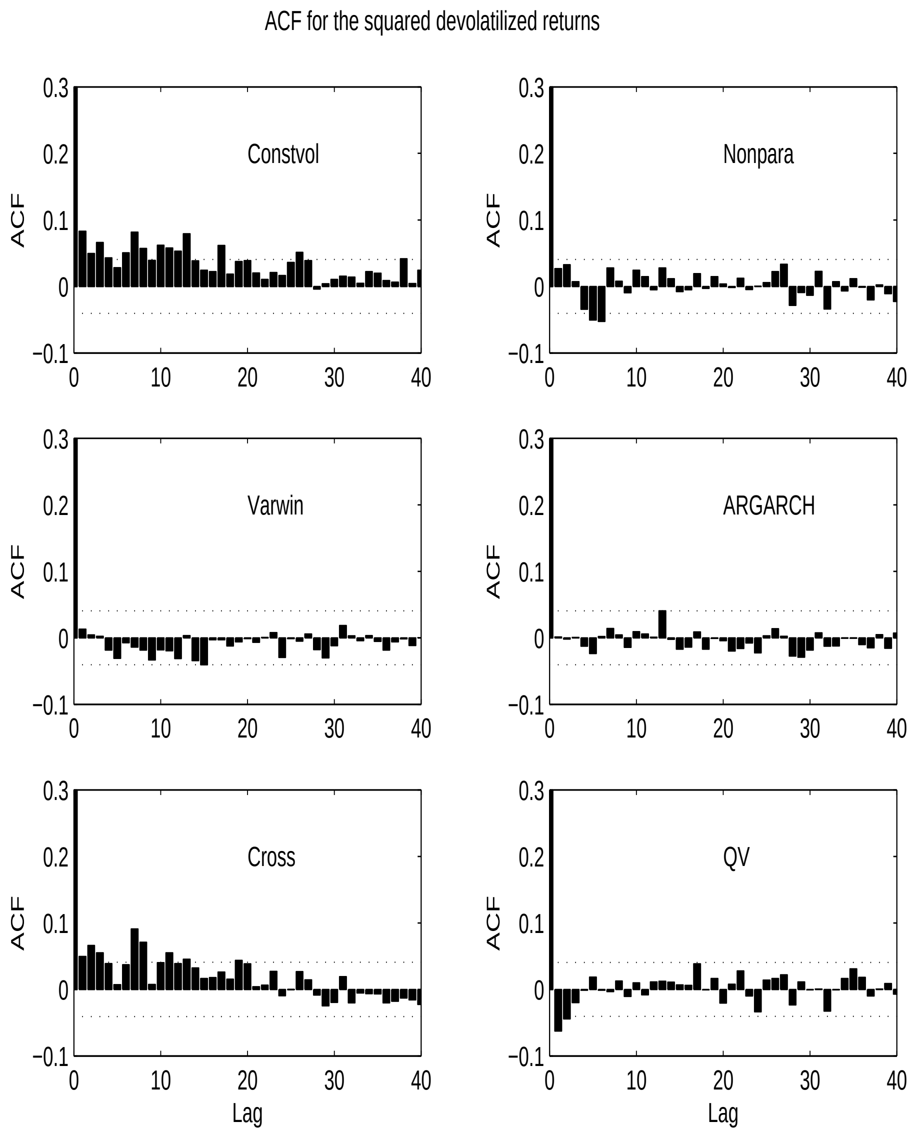

Figure 19: ACF for squared estimated residuals, Olsen Dollar/DM data

6.3 Distribution

In the previous section we studied if the residuals, $\Delta\hat{L}$, from our models had independent increments. However, we need to estimate a distribution to the residuals and statisticly test the estimated distribution to determine if the models are good.

If we consider $U = \hat{F}(\Delta\hat{L})$, where $\hat{F}$ is the estimated distribution from a residual. U should, under null hypothesis, be an i.i.d time serie uniformly distributed on $[0, 1]$. The most common test is the

Kolmogorov-Smirnov test which analyzes if the empirical distribution of U, denatod $\hat{F}_U$, significantly differs from a uniform distribution. However, we will instead use a Kuiper test as it is more sensitive to the tails.

The Kuiper statistic K is defined as [18]

$$K = \max_{x\in[0,1]}(\hat{F}_U(x) - x) + \max_{x\in[0,1]}(x - \hat{F}_U(x))$$

and the p-value for the Kuiper statistic is asymptotically

$$p = 2\sum_{k=1}^{\infty}(4k^2\lambda^2 - 1))e^{-2k^2\lambda^2}$$

where $\lambda = K(\sqrt{n} + 0.155 + \frac{0.24}{\sqrt{n}})$.

To calculate $U = \hat{F}(\Delta\hat{L})$ under GH distribution asumption on the devolatilized logreturns we use a numerical method based on Simpson's rule [22]. To be able to speed up the calculation we sorted the $\Delta\hat{L}$ in size and used 500 intervals between each sorted $\Delta\hat{L}$. $\hat{F}_U$ was calculated by dividing $[0,1]$ in 1000 intervals.

6.3.1 Samuelson model

We start our Kuiper test with assuming our devolatilized logreturns to be normal distributed. The distribution parameters are estimated according to $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$.

Vol model	$\hat{\sigma}_{DAX}$	$\hat{\mu}_{DAX}$	$\hat{\sigma}_{Olsen}$	$\hat{\mu}_{Olsen}$
Constant Volatility	0.0121	0.0006	0.7113	-0.0078
Nonpara Volatility	0.4424	0.0276	1.3740	-0.0251
Cross Volatility	2.0285	0.1102	2.0387	-0.0334
Variance Window	1.1578	0.0649	1.1158	-0.017
GARCH-AR	1.0234	0.0594	0.9687	-0.0159
QVariation $M = 8$	-	-	1.0214	-0.0025
QVariation $M = 144$	-	-	0.9544	-0.0092

Table 3: Estimated normal distribution for devolatilized returns

As we used a QQ plot in section 2 to motivate a different model from the Samuelson model, we present QQ plots for the Olsen data to visualize the Kuiper test. The quadratic variation model is represented by $M = 144$.

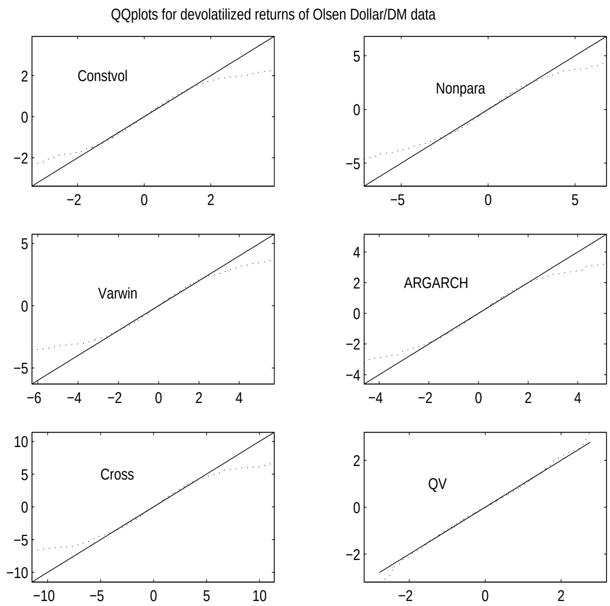

Figure 20: QQ plots under normal assumption,Olsen Dollar/DM data

We notice that no model seems to be well modeled with normal assumption as each QQ plot shows deviation from the straight line.

Vol model	K_{DAX}	p_{DAX}	K_{Olsen}	p_{Olsen}
Constant Volatility	0.1059	0	0.0692	0
Nonpara Volatility	0.0807	0	0.0823	0
Cross Volatility	0.0791	0	0.0682	0
Variance Window	0.0592	0	0.0688	0
GARCH-AR	0.0515	0.0017	0.0612	0
QVariation $M = 8$	-	-	0.1021	0
QVariation $M = 144$	-	-	0.0440	0.0026

Table 4: Kuiper test for devolatilized residuals under GH distribution assumtion

The Kuiper test summaries the QQ-plots in statistical numbers. No model passed the test. However, one of our assumptions in the Kuiper test is that U has i.i.d increments. We follow [11] and test this assumption.

Vol model	BDS_{DAX}	p_{DAX}	BDS_{Olsen}	p_{Olsen}
Constant Volatility	13.7633	0	5.5306	0
Nonpara Volatility	-2.3314	0.0197	-2.9608	0.0031
Cross Volatility	1.5238	0.1276	2.3894	0.0169
Variance Window	-0.4341	0.6642	0.5791	0.5625
GARCH-AR	-1.0664	0.2863	-1.5831	0.1134
QVariation $M = 8$	-	-	-1.7407	0.0817
QVariation $M = 144$	-	-	-3.1453	0.0017

Table 5: BDS test for Kuiper data under Normal distribution assumtion

Under normal assumption of the devolatilized logreturns we can conclude that no model passed both the statistical tests. Only the variance window model and the GARCH-AR model passed the BDS test for each data set.

6.3.2 Generalized Hyperbolic model

We now assume the devolatilized logreturns to be GH distributed and estimate the distribution parameters for each volatility model and data set. We use the MLE method described in section 4.

Vol model	λ	α	β	δ	μ
Constant Volatility	-2.1189	0.9469	0.5932	0.0183	0.0005
Nonpara Volatility	-1.8559	1.3798	-0.2279	0.6839	0.0720
Cross Volatility	-2.9827	0.1711	-0.1135	4.0158	0.5542
Variance Window	-2.1383	0.8956	-0.2321	2.1896	0.3637
GARCH-AR	-2.0201	1.4157	-0.38	2.1174	0.4362

Table 6: Estimated GH parameters for devolatilized DAX residuals

Vol model	λ	α	β	δ	μ
Constant Volatility	-2.7281	0.3195	-0.0106	1.3416	-0.0027
Nonpara Volatility	-1.7655	0.4652	-0.0194	2.0939	0.0090
Cross Volatility	-1.9820	0.3679	-0.0205	3.4822	0.0489
Variance Window	-2.5088	0.3920	-0.0315	2.0458	0.0225
GARCH-AR	-2.7068	0.6065	-0.0377	1.9288	0.0197
QVariation $M = 8$	0.3984	14.2684	-0.2027	14.8379	0.2079
QVariation $M = 144$	-1.0491	11.7552	-0.1656	10.7518	0.1410

Table 7: Estimated GH parameters for devolatilized Olsen residuals

We also present the MLE values.

Vol model	MLE_{DAX}	MLE_{Olsen}
Constant Volatility	5733.1	-2555.5
Nonpara Volatility	-1047.7	-4141.6
Cross Volatility	-3868.6	-5016.5
Variance Window	-2874.1	-3637.7
GARCH-AR	-2667.8	-3335.9
QVariation $M = 8$	-	-3523.5
QVariation $M = 144$	-	-3357.3

Table 8: MLE for GH assumtion

In table 9 we can clearly see that the assumption with GH distributed $\Delta \hat{(L)}$ outperforms the normal assumtion. However, the Quadratic

Vol model	K_{DAX}	p_{DAX}	K_{Olsen}	p_{Olsen}
Constant Volatility	0.0385	0.0748	0.0201	0.8277
Nonpara Volatility	0.0216	0.8918	0.0233	0.6023
Cross Volatility	0.0308	0.3510	0.0226	0.6711
Variance Window	0.0195	0.9593	0.0171	0.9576
GARCH-AR	0.026	0.6427	0.0206	0.7939
QVariation $M = 8$	-	-	0.0995	0
QVariation $M = 144$	-	-	0.0418	0.0418

Table 9: Kuiper test for devolatilized returns under GH distribution assumtion

Variation model, both $M = 8$ and $M = 144$, do not perform significantly better. This is explained by a study of the histogram of $\Delta\hat{L}$, where it seems as the center of the empirical distribution is camelshaped. This severely affects the Kuiper test.

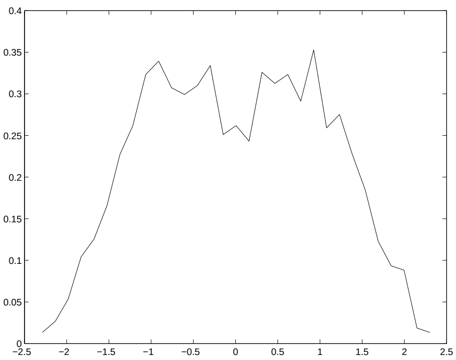

Figure 21: Histogram for devolatilized logreturns under Quadratic Variation model, $M = 8$

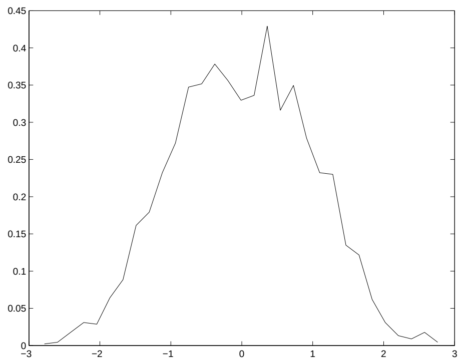

Figure 22: Histogram for devolatilized logreturns under Quadratic Variation model, $M = 144$

However, if we only measure the tails e.g.

$$\max_{x \in [0,0.2:0.8,1]}$$

the Quadratic volatility model with $m = 144$ passes the test with $p - value = 0.0994$ on a 95% confidence interval. The reason for the camelshape is not known to the author of this thesis.

In contrast of the normal assumption, for the devolatilized logreturns under the assumption of GH distributions, the variance window, the cross and the GARCH(1,1)-AR(1) volatility models passed

Vol model	BDS_{DAX}	p_{DAX}	BDS_{Olsen}	p_{Olsen}
Constant Volatility	10.9112	0	5.4175	0.0006
Nonpara Volatility	-3.1383	0.0017	-3.5124	0.0004
Cross Volatility	0.8384	0.4018	1.8228	0.8322
Variance Window	-0.6899	0.4902	0.5778	0.5634
GARCH-AR	-1.4263	0.1538	-1.5317	0.1256
QVariation $M = 8$	-	-	-1.7492	0.0803
QVariation $M = 144$	-	-	-3.1284	0.0018

Table 10: BDS test for Kuiper data under GH distribution assumtion

all statistical test at a 95% confidence interval. However, we should remember that the cross volatility model failed the BDS test for the devolatilized residuals. Despite its nice mathematical properties the Quadratic volatility model did not perform well.

6.4 Mixing Model

The mixing model, with two noise processes, is a bit different than the other models. We refresh the reader of the model and the mixing concept. If $\sigma^2 \sim GIG$ distributed and $\epsilon \sim N(0,1)$ then we have $x = \mu + \beta\sigma^2 + \sigma\epsilon$ GH distributed [2]. However, to get the flexibility of two OU processes we instead assume $\sigma^2 \sim IG$ and as a consequence we have $x \sim NIG$. It is easy to simulate such a process because we can simulate an IG distributed random number with an algorithm from Michael, Haas and Schucany [16]. The author of this thesis has never seen simulated results of the validity of distribution and dependence assumption of OU processes.

The estimated NIG parameters can be found in table 11. We got $MLE = 5730.7$ for the DAX data and $MLE = -2558.2$ for the Olsen Dollar/DM data.

Data	λ	α	β	δ	μ
DAX	-0.5	83.334	0.6437	0.0120	0.0005
Olsen Dollar/DM	-0.5	1.8130	-0.0131	0.9071	-0.0017

Table 11: Estimated NIG distribution parameters for DAX and Olsen data

This implies that the parameters from the IG distribution should be

Data	$\delta_{IG} = \delta_{NIG}$	$\gamma_{IG} = \sqrt{\alpha_{NIG}^2 - \beta_{NIG}^2}$
DAX	0.012	83.3315
Olsen Dollar/DM	0.9071	1.8130

Table 12: Estimated IG distribution parameters for DAX and Olsen data

We plot the IG distribution for the Olsen Dollar/DM data and compare it with the histogram from the realized volatility in figure 23 and figure 24.

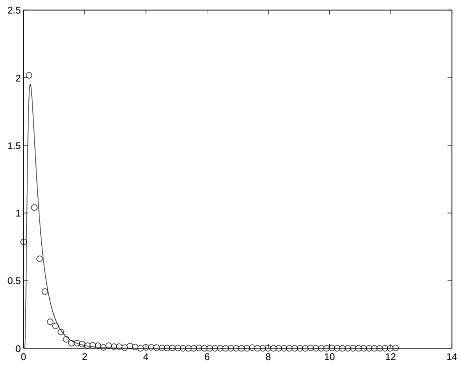

Figure 23: Estimated IG distribution and realized volatility for $M = 8$

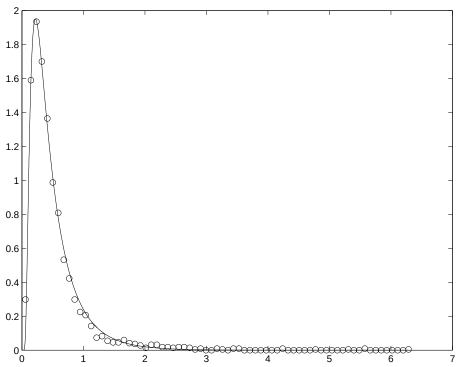

Figure 24: Estimated IG distribution and realized volatility for $M = 144$

Notice how well the realized volatility with $M = 144$ fits the IG distribution.

However, we want to use stochastic volatility to model dependence and therefore we will have to attach an artificial dependence structur in $\sigma^2(t)$ of $\Delta X^2(t)$ type. We recall from section 5 that Ornstein-Uhlenbeck processes had nice theoretical properties for the correlation structure which fits empirical data well. Following Equation 8 we estimat the parameters and use Equation 9 and Equation 10 to model our data.

Data	m	α_1	α_2
DAX	0.1309	0.0072	1.5744
Olsen Dollar/DM	0.0572	0.0244	3.5552

Table 13: Estimated OU parameters for DAX and Olsen data

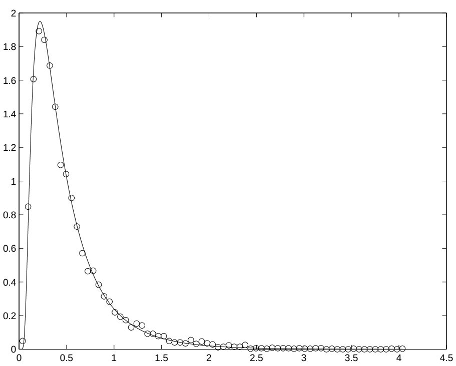

Figure 25: Estimated IG

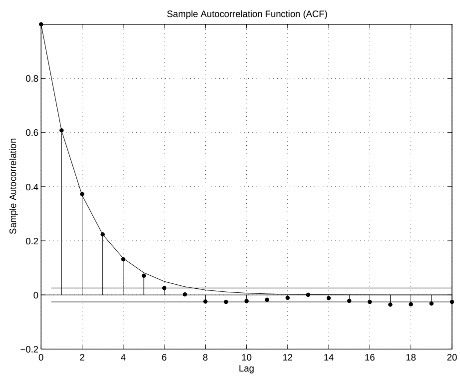

Figure 26: Dependence

If we follow Theorem 12 we can simulate IG distributed OU processes independent of α via Equation 11. We choose timesteps of $dt = 0.1$ and as $\sigma^2(0)$ we use the mean of the IG distribution. In Figure 25 we have simulated an IG distributed OU process with parameters $\delta = 0.9071$, $\gamma = 1.813$ and $\alpha = 0.5$ in 6000 timesteps and plotted the histogram with the true distribution. In Figure 26 we have plotted the ACF for the simulated OU process as well as the theoretical values.

We now simulate the superposition of two IG distributed OU processes with parameters estimated from the DAX and Olsen data. We choose to neglect the first 1000 values of superpositioned OU processes to avoid dependence of the initial value. In Figure 27 and Figure 28 we have plotted the histograms for 9000 simulated values with the true distributions from Table 12.

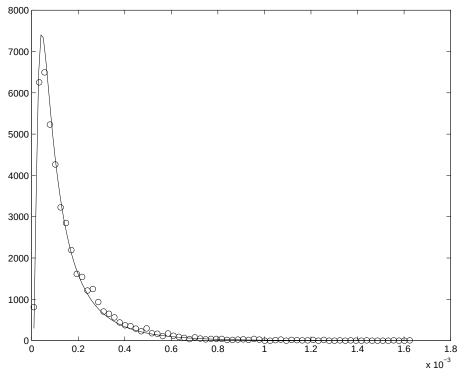

Figure 27: DAX distribution

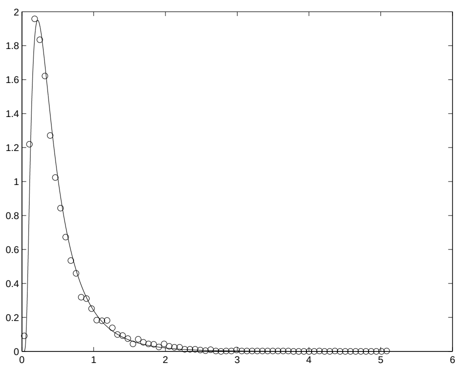

Figure 28: Olsen distribution

The dependence structures, for 9000 simulated values, via ACF can be seen in Figure 29 and Figure 30. The circles are the ACF and the lines are the theoretical values from Equation 8

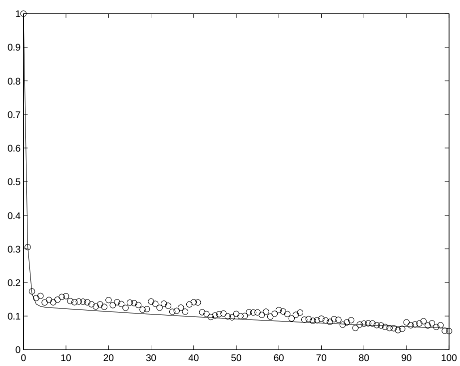

Figure 29: DAX

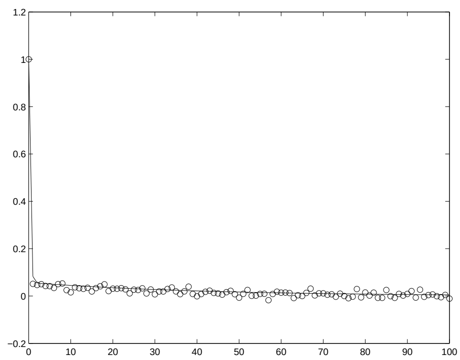

Figure 30: Olsen

Unfortunately, if we choose the size of the simulated data to 2000, which is in the same range as our DAX and Olsen data, the ACF seems to overestimate the dependence for the DAX data and underestimate the Olsen data. This can be seen in Figure 31 and Figure 32.

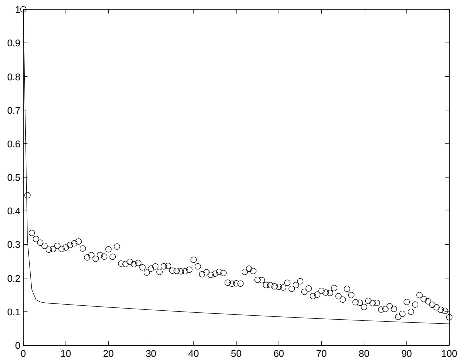

Figure 31: DAX

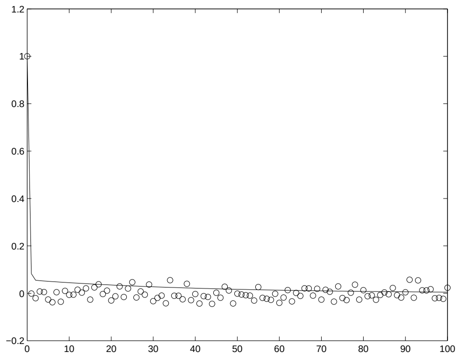

Figure 32: Olsen

However, we have found that the σ^2, modeled by a superposition of two OU processes, in the mixing model captures the dependence structure of the empirical data if we use a large amount of simulated datapoints.

7 Conclusion

We have shown that the most common model for a security, the Samuelson model, deviates significantly from real financial data, by means of empirical facts presented in the thesis. We have also shown that it is necessary to use stochastic volatility to model dependence in the time. We tested an alternative noise process distribution to the traditional Gaussian, the generalized hyperbolic distribution. The GH distribution combined with stochastic volatility made it possible to model logreturn data from the DAX and Olsen Dollar/DM very well. For instance, the AR-GARCH stochastic volatility model combined with the GH distribution passed all statistical tests. We also notice that the variance window model passed all statistical tests. However, due to the high variance of the variance in this model we take this result lightly.

Although this thesis is a study in finance, it illustrates an important aspect of general stochastic theory, independence. We have shown that the BDS test could reject independence under one transfromation of a dataset and pass the same dataset under another transformation. This is not preferable.

We have also shown that the mixing model with volatility modeled by a superposition of two OU processes has promising statistical features.

References

[1] M. Abramovitz, I. A. Stegun *Handbook of Mathematical Functions*, Dover Publications, New York, 1968.

[2] O. E. Barndorff-Nielsen, *Exponentially decreasing distributions for the logarithm of particle size*, Proceedings of the Royal Society London, 353, 401-419, 1977.

[3] O. E. Barndorff-Nielsen, E. Nicolato, N. Shepard *Some recent developments in stochastic volatility modelling*, To appear in Quantitative Finance 2002, Nuffield, 2001.

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

[5] P. Blæsild, B. *The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data*, Biometrica 68, 251-263, 1981.

[6] W. Brock, W. Dechert, J. Scheinkman, *Nonlinear Dynamics, Choas and Instability: Statistical theory and Economic Evidende*, Campebridge, 1991.

[7] W. Brock, W. Dechert, J. Scheinkman, *A Test for Independence Based on the Correlation Dimension*, Econometric Reviews 15(3), Campebridge, 1996.

[8] P. J. Brockwell, R. A. Davis, *Time Series: Theory and Methods*, Springer, Berlin, 1996.

[9] J-M. Courtault et al *Louis Bachelier´s life and work*, Mathematical Finance, Vol 10 nr 3, 2000.

[10] R. Durett, *Probability: Theory and Examples*, Duxbury Press, Belmont, 1995.

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

[12] E. Eberlein, U. Keller *Hyperbolic distribution in finance*, Bernoulli, 1, 281-299, 1995.

[13] H. Geman, D. Madan, S.R. Pliska, T. Vorst *Bachelier and his times: A conversation with Bernard Bru*, Mathematical Finance - Bachelier Congress 2000, Springer, 2001.

[14] I. Karatzas, S. E. Shreve *Brownian Motion and stochastic Calulus*, Springer, New York, 2000.

[15] B. B. Mandelbrot, *The variation of certain speculative prices*, Journal of Business 36, 519-530, 1963.

[16] J. R. Michael, R. W. Haas W. R. Schucany *Generating Random Variates Using Transformations with Multiple Roots*, The American Statistician 30 , 88-90, 1976.

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

[18] W. Press, S. Teukolsky, W. Vetterling, B. FLannery *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1992.

[19] S. Raible, *Lévy Processes in Finance: Theory, Numerics, and Empirical Facts*, Ph. D. thesis, University of Friburg, 2000.

[20] S. I. Resnick, *Adventures in Stochastic Processes*, Birkhäuser, Boston, 1994.

[21] J. A. Rice, *Mathematical Statistic and Data Analysis*, Duxbury Press, Belmont, 1995.

[22] L. Råde, B. Westergren, *Mathematics Handbook for Science and Engineering BETA*, Studentlitteratur, LUND, 1995.