A Comparison Of GARCH And MSM Volatility Models.

J R Boer

October 2009

Student number: 296707jb
Email: jesper_boer@hotmail.com
Supervisor: Prof. dr. D van Dijk

Erasmus University Rotterdam
Master thesis Quantitative Finance

Abstract

Volatility is probably the biggest subject in finance. There has been done a lot of research on the subject and even today there is still a lot of research going on in order to find the true volatility model. This thesis describes a comparison of two volatility models that rely on completely different assumptions. On the one hand we have the popular GARCH model. On the other hand the relatively new Markov Switching Multifractal model. We are interested in the properties of the models and whether the models are able to produce simulated return series that are in consensus with the stylized facts on financial returns. The models are both estimated on historical volatility and on expected future volatility. Historical volatility in the form of historical returns on a stock index and expected future volatility in the form of observed option prices. The parameter estimations from both estimation procedures are used to produce volatility forecasts, which are evaluated by measuring the value at risk performance for various time horizons with different confidence intervals. We find that the GARCH model performs better than the MSM model in forecasting volatility on a stock index, even though the performance off the GARCH model is also far off when the forecast horizon increases. The main reason why the MSM model performs poor in forecasting volatility is because the Gaussian draws to construct the returns are unable to replicate the distribution of the observed returns, with fat tails and negative skewness. The GARCH model with Gaussian shocks did a better job in replicating the distribution of the observed returns. However, the MSM model seems a lot better in replicating another stylized fact on volatility of financial return series. The autocorrelation functions of some transformations of the simulated returns from the MSM model looks real similar to the autocorrelation functions of the same transformations of the observed returns. The GARCH model performed slightly better in the out of sample test on the valuation of option prices. All the models have a lot of difficulties with valuating options that are far out-of-the-money. They do a better job in valuating in-the-money options. Finally, both the models failed in forecasting the expected future volatility, since the expected future volatility changes quickly over small time periods.

Index

1. Introduction

Volatility is one of the biggest topics in finance today. Volatility refers to the degree of unpredictable change over time of a certain variable[1]. In finance, volatility is the standard deviation of the return series of a financial instrument. Volatility is the most important measure of risk and plays a crucial role in the valuation of derivatives. Therefore , volatility estimations are essential for most financial decisions. It has proven extremely difficult to model and forecast the volatility we witness in time series . Usually volatility exhibits non linearity's and clustering, which makes the modeling of volatility difficult. The fact that volatility is not directly observable makes modeling even more difficult. Over the years there have been different attempts to model volatility.

The most popular models today are the models that belong to the GARCH[2] family. However, the performance of this model is not yet completely satisfying. 'Whilst GARCH models are able to capture the observed clustering effect in asset price volatility in-sample, they appear to provide relatively poor out-of-sample forecasts'[3]. As the traditional models seem to fail predicting volatility correctly, even within large confidence intervals, the search for alternative volatility models increased over the years.

One of the alternative and relatively new volatility model is the Markov Switching Multifractal[4] or MSM model. The MSM model is the result of further research on multifractal models[5]. Where the GARCH model describes volatility as the weighted sum of the previous levels of, the MSM model sees volatility as the product of multiple volatility components or layers that all switch value with different frequencies but that have the same distribution. In that process the frequencies depend on the scaling in the time series.

The MSM model assumes that volatility shocks have the same magnitude at all time scales and that volatility has a multifractal structure. There exists a lot of evidence in favor of this assumptions, numerous studies confirm the assumptions underlying the MSM model[6,7,8,9]. Furthermore research has shown that the MSM model can outperform the GARCH type models in a point forecast comparison over various horizons[10]. All this evidence suggests that the MSM model can prove to be a good addition to the existing collection of volatility models. Fact is that the GARCH and the MSM model rely on completely different assumptions.

Because of the different assumptions it is interesting to take a closer look at both models and to compare the different properties. In this thesis we start out with looking at the stylized facts on financial returns and how the different volatility models manage to replicate those stylized facts and the actual

---

[1] Wikipedia definition.
[2] Bollerslev, T. 1986.
[3] McMillan, D. G. and Speight, A. E. H. 2004.
[4] Calvet, L. E. and Fisher , A.J. 2002.
[5] Mandelbrot, B. B., Fisher, A. J. and Calvet, L. E. 1997.
[6] Galluccio, S., Calderelli, G., Marsili, M. and Zhang, Y.C. 1997.
[7] Richards, G. 2000.
[8] Malmini, R. 2007.
[9] Yamasaki, K., Wang, F., Havlin, S. and Stanley, H. E. 2006.
[10] Calvet, L. E. and Fisher , A.J. 2008.

financial time series. Think about fat tails, skewness, autocorrelation in lags of the squared and absolute return and volatility clustering. We continue by estimating the models from historical volatility in the form of historical return series. We investigate the simulated series produced by both models and use the estimated parameters to construct volatility forecasts. Those volatility forecasts are evaluated in a Value at Risk setup. Besides looking at the performance over the whole period we also look at the performance over different sub periods.

The GARCH and the MSM model parameters are also estimated from expected future volatility. The expected future volatility is obtained from option prices as option prices contain the expectation of the market on the development of the underlying. The estimated models are evaluated out of sample where there is special focus on the moneyness and the time to maturity of the option contracts. Again the estimated parameters are used to construct volatility forecasts. These volatility forecasts are also evaluated in a Value at Risk setup.

The major contribution of this work is that it is in fact a step by step example on how to estimate various volatility models and how to use the resulting estimates for derivative pricing. This should come in handy for practitioners and others interested in modeling volatility. Furthermore this is the first work in which the MSM model is estimated from option prices and used in an option pricing application and the first work where the option pricing performance of the GARCH model and the MSM model is compared. As the option market is a huge part of the financial world and as the current volatility models are still unable to explain the option prices we observe in the market sufficiently, it is of crucial importance that the financial world keeps looking for other volatility models or for ways to improve their current models.

2. Volatility and Volatility models.

Volatility refers to the degree of unpredictable change over time of a certain variable. Volatility is an important measure of risk as someone who is exposed to a volatile variable has a big chance of losing a lot. In finance we usually focus on the volatility of return series of stocks, indexes or derivatives as the returns determine whether we will earn or lose money. Volatility is often measured as the standard deviation or as the variance of the returns of the time series. Modeling volatility remains the biggest challenge in risk management and the search for the right volatility model is still in progress. There have been several approaches, starting from the Gaussian constant volatility to GARCH models to Markov-switching models and multifractal models. Fact remains that it seems very difficult to replicate the volatility process we witness in financial time series. Especially the long term dependencies in volatility series and the sudden jumps in magnitude of volatility.

In Finance, there exist a couple of well known 'stylized facts' on returns. Stylized facts are empirical findings in datasets that we come across in most of the datasets and therefore hold in general[11]. For financial return series we have the following stylized facts:

1. Distribution of returns is not normal
   - (a) Large (and small) returns occur more often than expected under normality: Excess kurtosis (fat-tailed and peaked distribution).
   - (b) Large negative stock returns occur more often than large positive ones: Negative skewness.
2. (Almost) No significant autocorrelations in returns.
3. Small, but very slowly declining autocorrelations in squared and absolute returns; Periods of large returns alternate with periods with small returns, suggesting that volatility is not constant over time. In other words, asset returns are heteroskedastic instead of homoskedastic. This feature is called "volatility clustering".

In order to model volatility correctly, it is obvious that the models we use should be able to generate returns that are consistent with these stylized facts. The most important cause for the stylized returns we witness is that volatility is time varying. Therefore we will look at models with time varying volatility. In this thesis we will look at GARCH models and at the MSM model. The models will be compared in their ability to replicate the characteristics of financial time series. We will continue by introducing those models.

2.1. The constant volatility model

The constant volatility model is one of the most simple models used to simulate financial time series. And even though it has already proven to be inconsistent with the stylized facts we like to add the constant volatility model in some parts of this thesis as a benchmark model. That way we are not only comparing 'complex' volatility models with each other, but we also keep an eye on the added value of these more complex models on a more simple model.

The constant volatility model only has one parameter, σ. Returns are modeled as

[11] These stylized facts have been found in numerous studies, for example; 'Mandelbrot, B. 1963.' and 'Fama, E. 1965.'

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t = z_t \sigma$$

With $z_t \sim \mathrm{N}(0,1)$. As σ is not dependent on the time t, σ stays constant over time.

An advantage of the constant volatility model is that there are analytical solutions available for the option prices[12].

2.2. GARCH models

The generalized autoregressive conditional heteroskedasticity (GARCH) model is probably the most popular volatility model. The models define current conditional volatility as a function of past conditional volatility and past squared error terms.

The GARCH model[13] extends the ARCH[14] model by including the lagged conditional variance as a regressor for h_t.

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t = z_t \sqrt{h_t}$$
$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \alpha_2 \varepsilon_{t-2}^2 + \beta_2 h_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_p h_{t-p}$$

Most commonly used is the GARCH(1,1) model, where the conditional variance is modeled as

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

To guarantee that $h_t \geq 0$ for all t, $\omega \geq 0, \alpha_1 \geq 0, \beta_1 \geq 0$. For covariance stationary, $\alpha_1 + \beta_1 < 1$. Then the unconditional variance $\sigma^2 = \frac{\omega}{1-\alpha_1-\beta_1}$.
The autocorrelation of the squared returns in a GARCH(1,1) model is given by

$$\rho_1 = \alpha_1 + \frac{\alpha_1^2 \beta_1}{1 - 2\alpha_1\beta_1 - \beta_1^2}$$

for the first order autocorrelation and

$$\rho_k = (\alpha_1 + \beta_1)^{k-1} \rho_1$$

for the k'th order autocorrelation. If α_1 is small and $\alpha_1 + \beta_1$ is close to 1 the model will generate an ACF that starts out with a low correlation that slowly decreases. In that case the GARCH model is consistent with the stylized facts on the autocorrelation of return series.

In the GARCH model, positive and negative shocks have the same effect. The model is symmetric. We observe that in financial time series negative returns usually have a bigger effect on the volatility than

[12] See Appendix A.
[13] All the GARCH type models covered in this thesis are described in 'Bollerslev, T. 2008.'.
[14] See Appendix B.

positive returns. The GJR GARCH model[15] allows the conditional variance to respond differently to the lagged positive and negative innovations. The GJR model is equivalent to the popular TGARCH or Threshold GARCH model. We can write the GJR(1,1) model as

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0) + \beta_1 h_{t-1} \qquad (1)$$

Where $I(.)$ is an indicator function (1 if true and 0 if false) and where γ is the so called leverage parameter. Usually we find γ to be positive. For a positive γ the volatility increases more after a negative shock than after a positive shock of the same magnitude.

In this thesis we work with the GJR-GARCH model because it allows for asymmetry in the influence of positive and negative shocks on volatility. This makes the model more flexible and brings the model closer to reality than the standard GARCH model. From here on, when we refer to the GARCH model we mean the GJR-GARCH model. Furthermore, because we make use of demeaned series we fix μ at zero, which leaves us with four parameters to estimate.

Because of their easy structure and because they are related and compatible with other financial concepts the GARCH models are very popular. There is however, an important drawback of the GARCH models. GARCH models are parametric specifications that operate best under relatively stable market conditions. Although GARCH is explicitly designed to model time-varying conditional variances, GARCH models often fail to capture highly irregular phenomena, including wild market fluctuations and other highly unanticipated events that can lead to significant structural change[16]. The irony is that the events that GARCH has difficulties with modeling, are the events that will do the most damage to investors, risk managers and other market participants.

Another drawback of the GARCH models is that the volatility in those models does not change that rapid as the volatility in real life does. As the volatility in the model is a function of past conditional volatility and past squared error terms, the volatility process is a relatively smooth process. In reality the volatility of return series can jump rapidly from low to high and in reversed direction. GARCH models have no problem with producing the jump from low volatility to high volatility because this will happen with a large ε_{t-1}^2 when in a low volatility state, see equation (1). GARCH models do have a problem with the jump from high to low volatility, as the β_1 parameter is usually quite large. When in a high volatility state with a large β_1, decline in volatility will only go gradually. Therefore, researchers are also looking in other directions and for other techniques to simulate volatility and especially the randomness in volatility.

2.3. The Markov Switching Multifractal model (MSM)

The idea behind multifractal models is that volatility consists of multiple components that change with different frequencies. Multifractal models are of added value to econometrics and finance as they are able to generate different degrees of long term dependencies. This is especially interesting for the modeling of volatility. Multifractal models are capable of simulating time series that replicate the stylized facts we witness in financial return series.

[15] Glosten, L. R., Jagannathan, R. and Runkle, D. E. 1993.
[16] Gourieroux, C. 1997.

The most important assumption on the MSM model is that "volatility is determined by components that have different degrees of persistence. These components randomly switch over time, generating a volatility process that can be both highly persistent and highly variable. The transition probabilities are heterogeneous across components and follow a tight geometric specification. When a component switches its new value is drawn from a fixed distribution that does not depend upon the component, for parsimony. The model assumes that volatility shocks have the same magnitude at all time scales. These assumptions appear broadly consistent with financial data."[17] To summarize, the model leans on the believe that volatility has a multifractal structure, which means that volatility over a certain time period will have a similar pattern as volatility over any other time period. These multifractal structures have been found in financial time series in numerous studies.

For the fixed distribution for the value of the components we can make use of a binomial or a multinomial distribution. The binomial distribution is the simplest, as it gives each component two possible values.

The Markov-Switching Multifractal (MSM) model in discrete time:
The volatility is driven by $\bar{k}$ components with values $M_{k,t}$.

The dynamics of $M_{k,t}$ can be summarized as;
$M_{k,t}$ drawn from distribution *M* with probability γ_k
$M_{k,t} = M_{k,t-1}$ with probability $1 - \gamma_k$

With independent switching events across *k* and *t*. A requirement for the distribution *M* is that it has a positive support, $M \geq 0$ and unit mean $E[M] = 1$. The distribution *M* can for example be a binomial or a lognormal distribution. Then stochastic volatility is modeled by

$$\sigma(M_t) \equiv \bar{\sigma}\left(\prod_{k=1}^{\bar{k}} M_{k,t}\right)^{\frac{1}{2}} \qquad (2)$$

With $\bar{\sigma}$ a positive constant, the unconditional standard deviation of the returns r_t. Returns r_t are then modeled as $r_t = \sigma(M_t)\varepsilon_t$, with ε_t i.i.d. standard Gaussian $N(0,1)$.

The transition probabilities γ_k are specified as $\gamma_k = 1 - (1 - \gamma_1)^{(b^{k-1})}$, where $\gamma_1 \in (0,1)$ and $b \in (1, \infty)$, as these restrictions make sure that γ_k is always a probability, $\gamma_k \in (0,1)$. The MSM model assumes power law scaling in financial time series, which means that financial time series are scale invariant. The parameters of $\gamma_{\bar{k}}$ and b depend on the scaling in the financial data we like to model. The parameter γ_1 is the transition probability of the component with the lowest transition frequency. The parameter $\gamma_{\bar{k}}$ is the transition probability of the component with the highest transition frequency and the parameter b determines the spacing between the transition frequencies γ_k. The structure assures that the transition probabilities follow a geometric sequence.

The model is called a Markov-Switching model because due to the structure with the $\bar{k}$ volatility-components we end up with $2^{\bar{k}}$ volatility states or in Markov terms: regimes. With the transition probabilities it is possible to construct the corresponding Markov transition matrix. Because of the richness of different regimes, the model gives the possibility to produce conditional volatility that

[17] For the representation of the MSM model, and later on for the estimation of the MSM model we would like to refer to 'Calvet, L. E. and Fisher , A.J. 2008.'

evolves smoothly for some of the times and on the other hand can jump quickly from a high regime to a low regime or vice versa, just as we observe in real time series. Within this structure the probability of moving from one regime to another is the same as moving the other way around.

The main drawback from the MSM model results directly from the main advantage of the MSM model. The model incorporates a huge number of states that grows exponentially with the number of volatility components $\bar{k}$ and the number of elements in the transitions matrices grows even faster. The underlying idea is that the closer the number of regimes approaches infinity, the more the model will approach reality. But as the number of regimes increases, the computational time needed to estimate and execute the models increases as well. The choice of $\bar{k}$ is regarded a model selection problem and will in most cases be a consensus concerning the model with the maximum number of regimes that is still feasible in a appropriate amount of time.

In this thesis we choose to use the simplest version of the MSM model, the binomial model where the variables M_t can only attain two values, m_0 and m_1. With the assumption that these outcomes occur with equal probability (so $m_1 = 2 - m_0$), the model is parameterized by $\psi \equiv (m_0, \bar{\sigma}, b, \gamma_{\bar{k}})$. Outcomes from the binomial MSM model is found similar to outcomes of a MSM model with lognormal distributions for the volatility components in recent publications on the subject. For parsimony we choose therefore for the model with the binomial distribution.

2.4. Estimation of volatility models

Estimation of GARCH

We estimate the GARCH model by maximum likelihood. In a maximum likelihood optimization you try to find the parameter vector θ that maximize the likelihood function. Where $\theta = (\mu, \omega, \alpha_1, \beta_1)$. The optimal parameter vector is

$$\hat{\theta} = argmax\ \mathcal{L}(\theta)$$

where $argmax$ is the value of the argument θ that results in the maximum value of the function $\mathcal{L}$.
The likelihood function is the joint probability density function of the returns.

$$\mathcal{L}(\theta) = f(r_1, r_2, \dots, r_T; \theta)$$

According the law of conditional probability, $f(A|B) = f(A \wedge B)/f(B)$, the joint pdf can be written in terms of the conditional densities.

$$f(r_1, r_2, \dots, r_T; \theta) = f_1(r_1; \theta) \cdot f_2(r_2|r_1; \theta) \cdot f_3(r_3|r_2, r_1; \theta) \cdot \dots \cdot f(r_T|r_{T-1}, \dots, r_1; \theta)$$

By taking logarithms we can sum over the conditional densities. We end up with:

$$\hat{\theta} = argmax\ \mathcal{L}(\theta) = argmax \sum_{t=1}^{T} l_t(\theta)$$

With $l_t(\theta) = \ln\big(f(r_t|r_{t-1}, \dots, r_1; \theta)\big)$ as the log likelihood function for observation t. And $f(.)$ as the distribution of the error term of the GARCH process.

Estimation of MSM

Just like the GARCH model, the MSM model can also be estimated by Maximum likelihood, although the approach is a little different. The aim is to optimize the likelihood of the observed returns by finding the optimal parameter set $\psi \equiv (m_0, \bar{\sigma}, b, \gamma_{\bar{k}})$. In the same way as with the GARCH model, the likelihood function $\mathcal{L}(\psi) = f(r_1, r_2, \dots, r_T; \psi)$ is the joint probability density function of the returns and can be written in terms of the conditional densities.

$$f(r_1, r_2, \dots, r_T; \psi) = f_1(r_1; \psi) \cdot f_2(r_2|r_1; \psi) \cdot f_3(r_3|r_2, r_1; \psi) \cdot \dots \cdot f(r_T|r_{T-1}, \dots, r_1; \psi)$$

By taking the natural logarithm of the likelihood function, we obtain the log likelihood function which is the sum over the conditional densities. Because the volatility is unobserved, the conditional density can be written as

$$f(r_t|r_{t-1}, \dots, r_1; \psi) = f(r_t|\sigma_t^2) \cdot f(\sigma_t^2|r_{t-1}, \dots, r_1) \quad (3)$$

The likelihood we want to optimize is the sum of the densities of r_t conditional on the volatility times the densities of the volatility at time t conditional on the observations up to time $t-1$. In the remainder of this chapter we will show step by step how to determine these conditional densities.

When M has a discrete distribution there exist a finite number of volatility states, this number depends on the number of components $\bar{k}$. The number of volatility states is $d = 2^{\bar{k}}$. The state vector M_t takes finitely many values $m^1, m^2, \dots, m^d$ and the dynamics of M_t are characterized by transition matrix $A = (a_{i,j})$ where $1 \leq i, j \leq d$.

Recall that $M_t = \bar{\sigma}\big(\prod_{i=1}^{\bar{k}} M_{k,t}\big)^{\frac{1}{2}}$, see equation (2). The elements $a_{i,j}$ are given by $a_{i,j} = P\big(M_{t+1} = m^j | M_t = m^i\big)$, the probability that the next volatility state will be m^j given that the current state is m^i. Then the density of the return r_t conditional on the volatility state is $density_{r_t}\big(r|M_t = m^i\big) = N\Big(r; \sigma^2\big(m^i\big)\Big)$. This gives the probability that the return we witness has volatility $\sigma^2\big(m^i\big)$, given that the returns are normally distributed.

We do not observe M_t but we can compute the conditional probabilities $\prod_t^j \equiv P\big(M_t = m^j | r_1, \dots, r_t\big)$, the probability that the we are currently in volatility state m^j given all previous observations of the return. Then we stack them in the row vector $\Pi_t = \big(\Pi_t^1, \dots \Pi_t^d\big)$. Π_t is the probability density function of the volatility state of size $1 \times d$. It shows what the probability is for the volatility to be in state m^i, given the returns up to time t. As Π_t is a probability density function, the sum over the columns of Π_t will be exactly 1.

The conditional probabilities are computed recursively by Bayesian Updating. By Bayes' rule[18] Π_t (the posterior) can be expressed as a function of the previous belief (the prior) Π_{t-1} and the innovation r_t: $\Pi_t = \frac{\omega(r_t) * (\Pi_{t-1} A)}{[\omega(r_t) * (\Pi_{t-1} A)] 1'}$, with $1 = (1,,,1)$, $x * y$ denotes the Hadamard product $(x_1 y_1, \dots, x_d y_d)$ and

[18] See Appendix C.

$\omega(r_t) = \left(n[r_t; \sigma^2(m^1)], \ldots, n\left[r_t; \sigma^2\left(m^d\right)\right]\right)$. $\omega(r_t)$ is the density $f(r_t|\sigma^2)$ from equation (3) we need to estimate the MSM model. The product of Π_{t-1} with the transition matrix A gives the conditional density for the volatility, $f(\sigma_t^2|r_{t-1}, \ldots, r_1)$ from equation (3).

In empirical applications the initial vector Π_0 is chosen to be the ergodic distribution of the Markov process. Since the multipliers are mutually independent, the ergodic distribution is given by $\prod_0^j = \prod_{l=1}^{\bar{k}} P\left(M = m_l^j\right)$, which means that every volatility state is just as likely as the other, with a probability of $\frac{1}{2^{\bar{k}}}$.

Starting out with the neutral prior believe that every possible volatility state has the same probability as another at time $t = 0$, we consult the data for every t up till $\bar{t}$, the latest observation we have. In this process our posterior believe will change every step on the way. The posterior at time t will serve as the prior at time $t + 1$.

The log likelihood function to be maximized is $\ln \mathcal{L}(r_1, \ldots, r_t; \psi) = \sum_{t=1}^{T} \ln[\omega(r_t) \cdot (\Pi_{t-1}\mathrm{A})]$.

Once again we are looking for the parameter settings that maximize the log likelihood, for the values for $m_0, \gamma_{\bar{k}}, b, \bar{\sigma}$ that lead to the largest value of the likelihood function. The choice of the parameter $\bar{k}$ can be considered a model selection problem. In this thesis $\bar{k}$ is fixed at 8 at all time. With $\bar{k} = 8$ there exist a large number of possible volatility states, $2^{\bar{k}} = 2^8 = 256$, and at the same time the number of components is not yet too large to estimate the model relatively quick.

3. Illustration; volatility and estimation of models on index

3.1. Stylized facts on volatility and the DJIA

Let's take a closer look on volatility, on the MSM model and on a GARCH model by introducing an actual financial time series. We have the Dow Jones Industrial Average (DJIA) starting at the beginning of the year 1990 until the end of the year 2006[19]. Figure 1 shows the course of the Dow Jones. Figure 2 shows the daily returns from the Dow Jones Industrial Average.

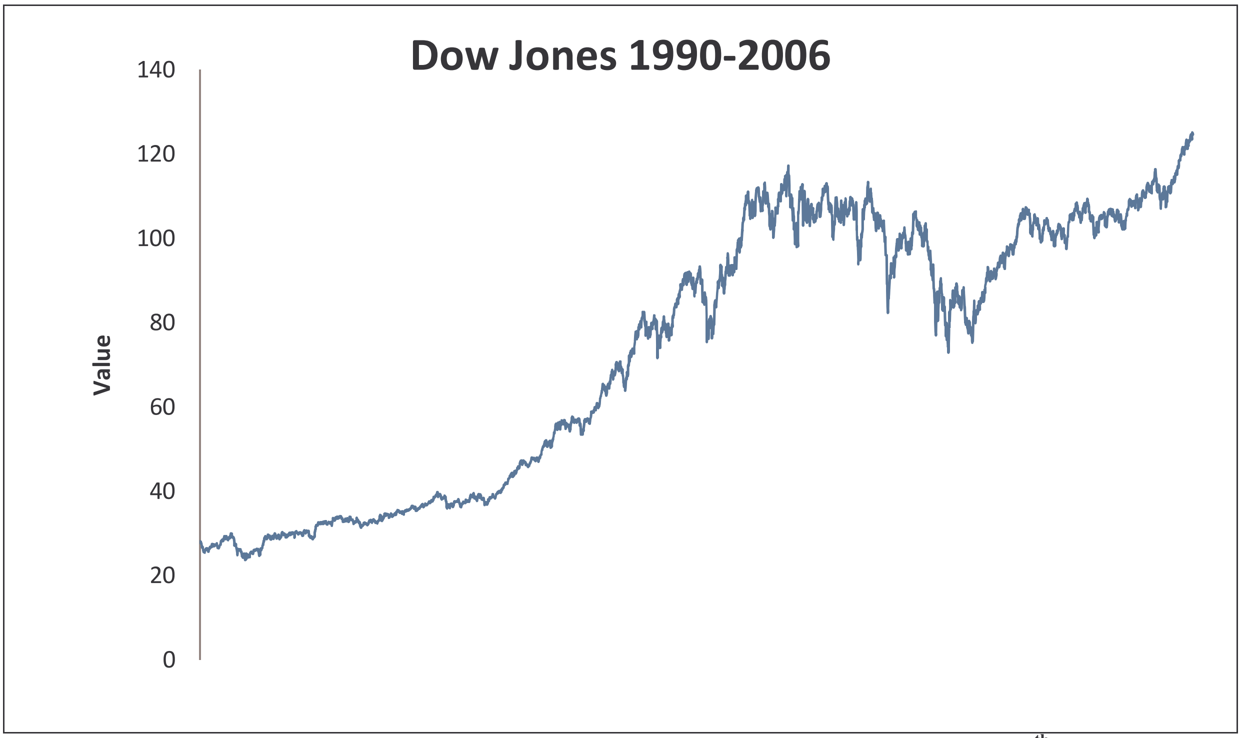


Figure 1 – The Dow Jones over the period January 3th 1990 up until December 29th 2006.

[19] Source: OptionMetrics.

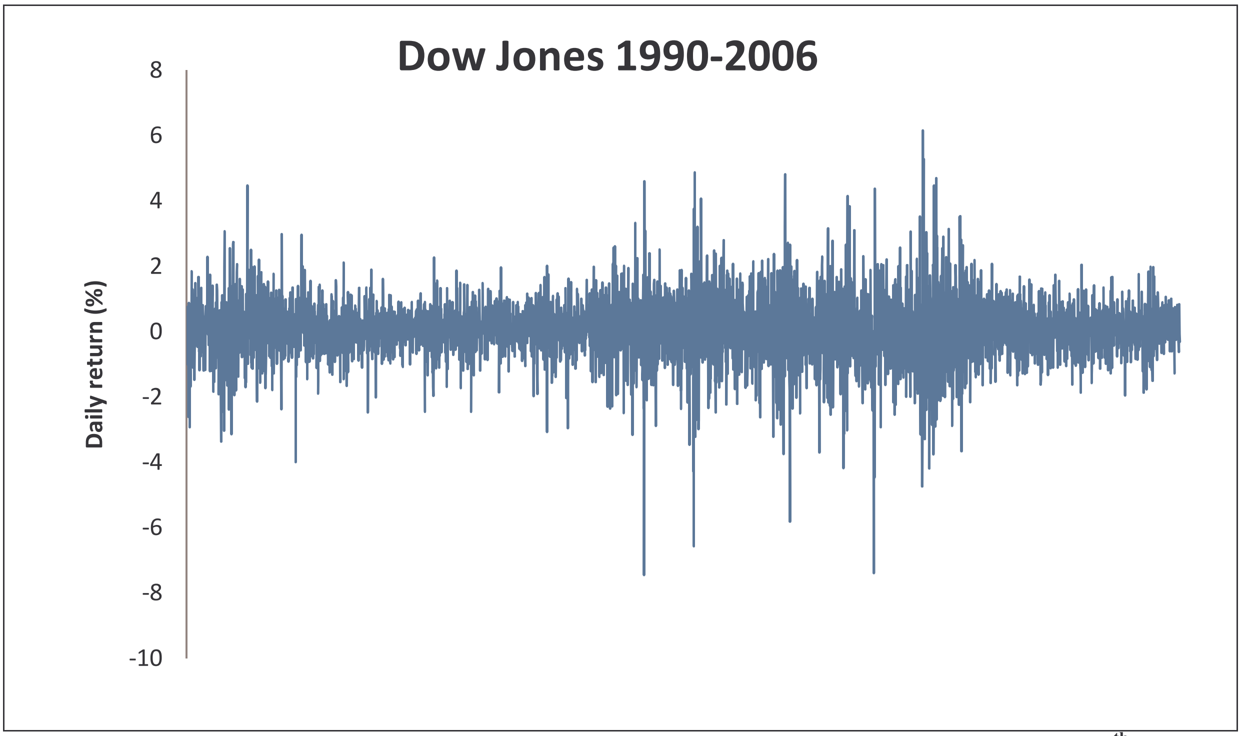
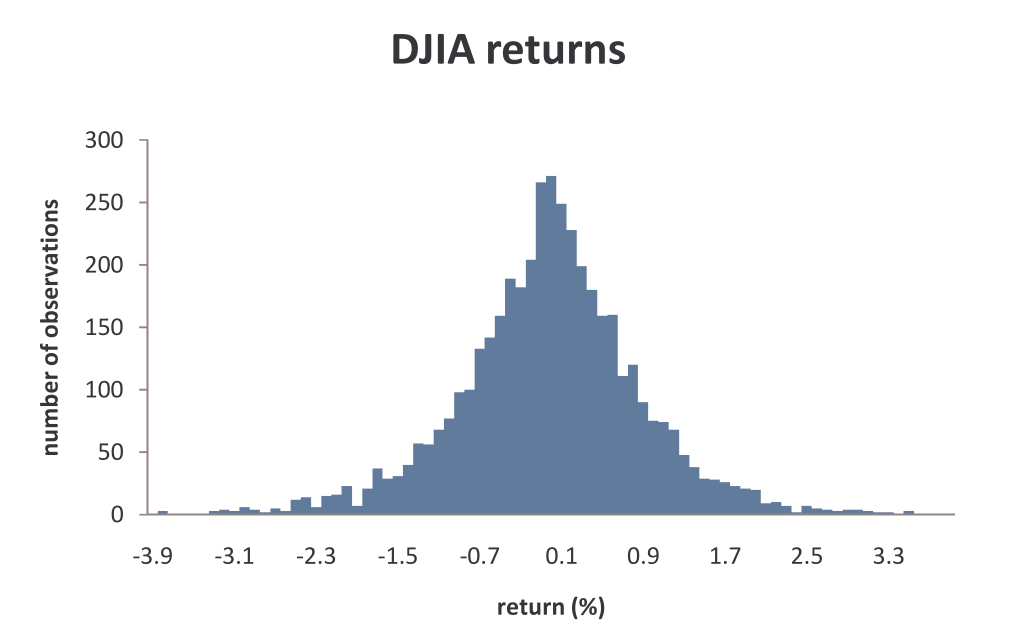


Figure 2 – The returns of the Dow Jones over the period January 3th 1990 up until December 29th 2006.

First we will check whether the stylized facts for financial returns are also present in the DJIA.
1. Distribution of returns is not normal

> *(a) Large (and small) returns occur more often than expected under normality: Excess kurtosis (fat-tailed and peaked distribution)*
> *(b) Large negative stock returns occur more often than large positive ones: Negative skewness;*

For the first stylized fact we will look at some of the statistics of the DJIA, these are shown in Figure 3.



observations	4324
mean (%)	0.03
stdev	0.98
min (%)	-7.45
max (%)	6.15
skewness	-0.23
kurtosis	4.78

Figure 3 – DJIA returns and statistics.

These statistics confirm the first stylized fact. The kurtosis is a lot larger than the kurtosis under normality, where the kurtosis would be around 3 and we calculate some negative skewness instead of a skewness of zero under normality. A Jarque-Bera test on normality rejects that the returns are normally distributed. Lets continue with the other stylized facts.

2. (Almost) No significant autocorrelations in returns;

3. Small, but very slowly declining autocorrelations in squared and absolute returns; Periods of large returns alternate with periods with small returns, suggesting that volatility is not constant. This is called "volatility clustering".

To examine the second stylized fact we look at the autocorrelation function of the returns, the squared returns and the absolute returns.

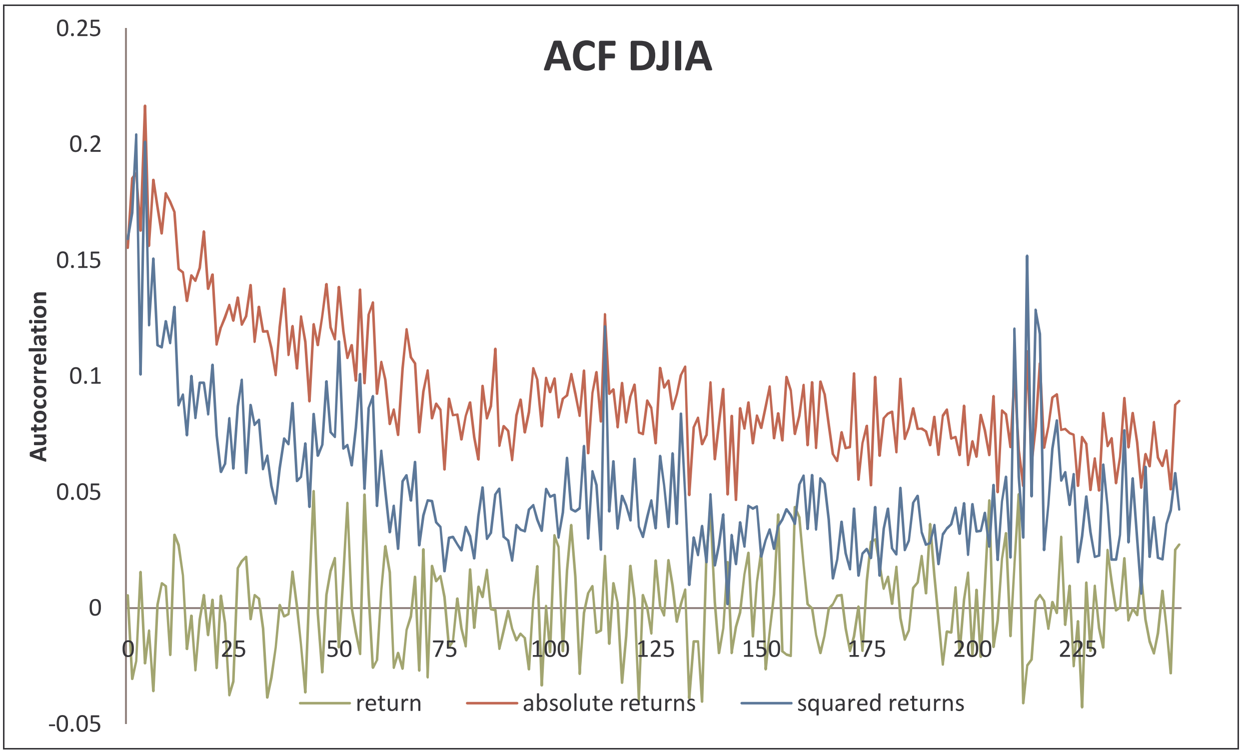


Figure 4 – The DJIA autocorrelation function.

Figure 4 shows that there is no autocorrelation in the return and a small (around 0.15-0.20 for the first 250 lags), slowly declining autocorrelation in the absolute and squared returns. Figure 2 clearly shows the volatility clusters. This confirms the stylized facts.

3.2. The Stylized facts on returns and the GARCH and MSM model

We confirmed the presence of the stylized facts on volatility in the DJIA time series. The next step is to examine whether the volatility models introduced in the previous chapter are capable of imitating the stylized facts we observe in actual time series.

We start out with estimating the models on the 4324 observations of DJIA returns we have from the beginning of 1990 until the end of 2006. Table 1 and Table 2 show the results of the parameter estimates for the volatility processes:

GARCH	ω	α_1	γ	β_1
	0.01	0.01	0.09	0.93

Table 1 – GARCH parameter estimates.

MSM($\bar{k} = 8$)	m_0	$\gamma_{\bar{k}}$	b	$\bar{\sigma}$
	1.27	0.98	3.66	1.02

Table 2 – MSM parameter estimates.

Now it is interesting to look how volatility behaved in our data sample according to the GARCH and the MSM model.

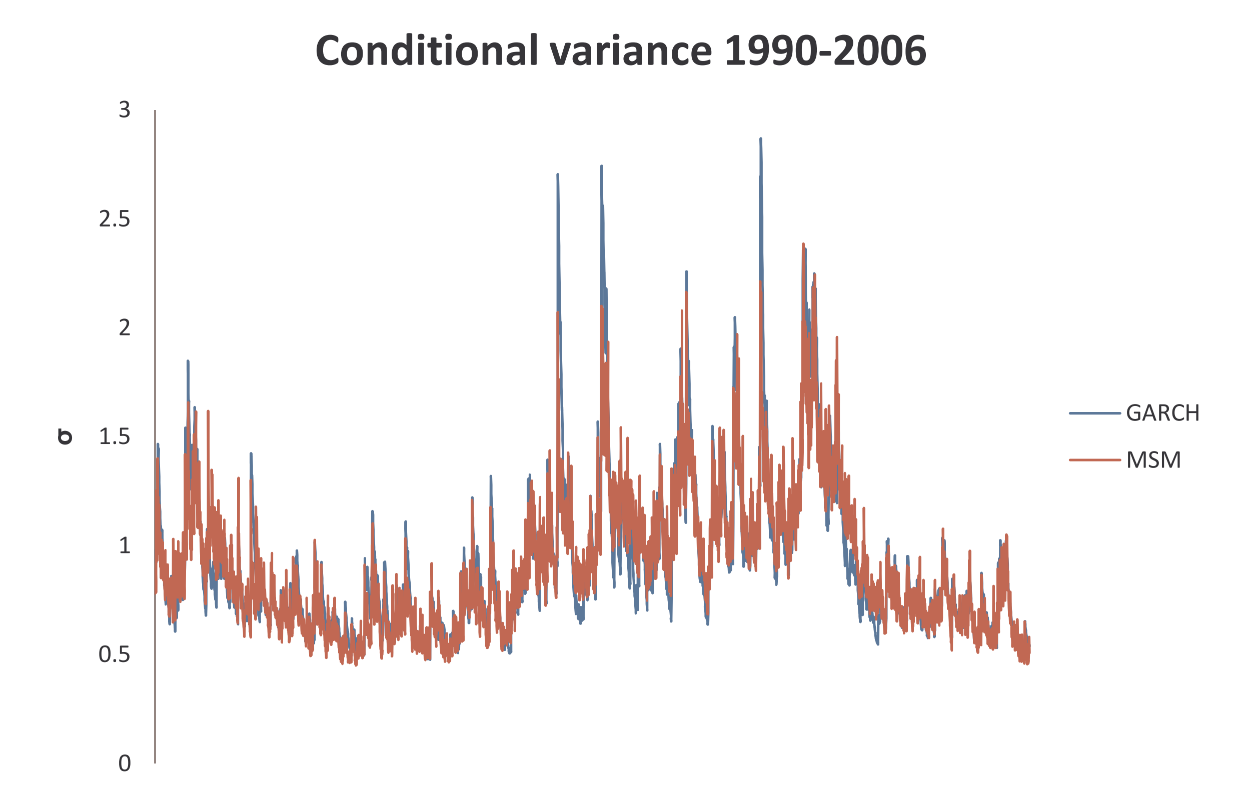


Figure 5 – Volatility in DJIA 1990-2006 according to the volatility models.

Figure 5 shows the volatility in terms of σ over the period 1990 up until the 2006. The construction of the conditional volatility, the volatility conditional on the observations, of the GARCH model is straightforward. The MSM model on the other hand consists of multiple states that all have a certain probability of being the real state at every point in time. The red line plotted for the MSM model in Figure 5 therefore is the expected volatility at time t. This expected volatility is obtained by a matrix multiplication of Π_t with its vector of corresponding values of volatilities, a vector of σ's. Π_t being the probability density function of the volatility state conditional on the observations up to time t and the vector of σ's consists of the different values for σ that belong to the volatility states in Π_t.

The two graphs look very similar. The most noticeable difference is that the GARCH model has a few larger peaks in volatility than the MSM model does. When we compare this figure with Figure 2 we see that the estimates of the GARCH and the MSM model follow the same pattern as the DJIA on which we estimated the models. When the magnitude of the returns is small in our data sample the models volatility estimates are low and when the magnitude of the returns is high, the models both produce high volatility estimates.

With the estimated parameters we simulate volatility paths and returns. The models supply us with ε_t , $\varepsilon_t = z_t\sqrt{h_t}$. Adding the mean gives us the return r_t, $r_t = \mu + \varepsilon_t$.

First, let's take a closer look at the volatility paths the GARCH and the MSM model produced. Figure 6 displays a subsample of the simulated volatility at consecutive points in time for both the models.

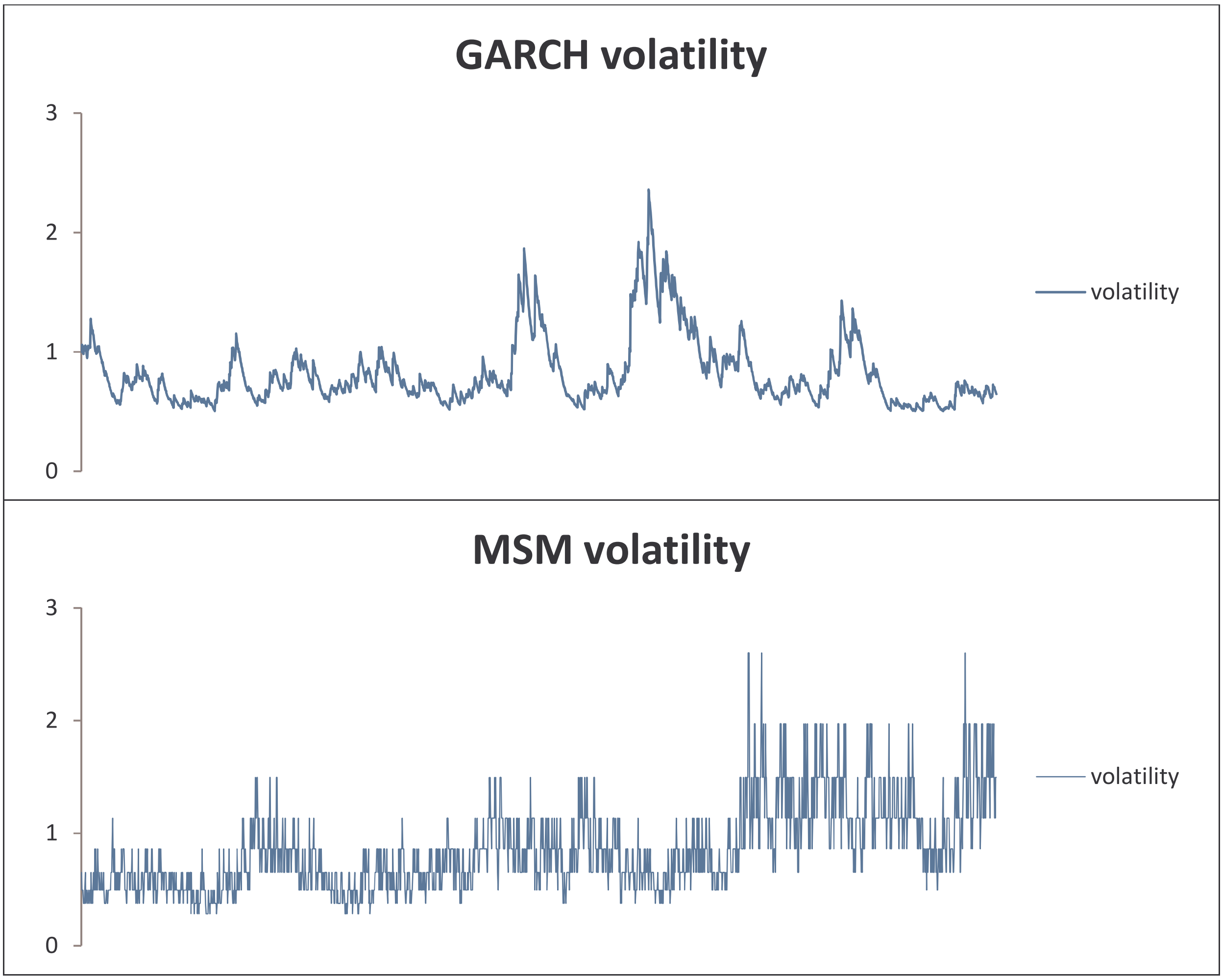


Figure 6 – GARCH and MSM simulated volatility.

First of all, the GARCH volatility path and the MSM volatility path are a product of a random simulation. They are unrelated and for that reason don't have to exhibit the same pattern of peaks (and in this figures they certainly don't). As expected the MSM path shows sudden jumps from state to state where the GARCH path follows a more continuous course, although it contains some rapid increases in volatility as well. Figure 6 – GARCH and MSM simulated volatility. also shows that the GARCH model has some difficulties with rapid decreases in volatility, as its volatility decreases more gradually from high to low.

Figure 7 illustrates how the volatility of the MSM process is determined.

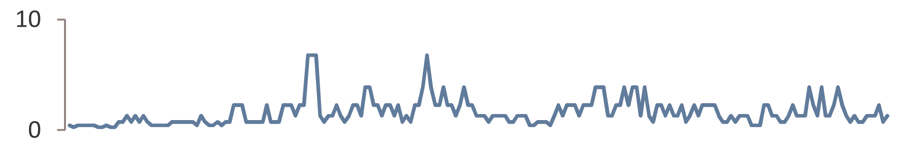
Component k	γ_k*10^2	Graphical illustration
1	0.05	
2	0.17	
3	0.61	
4	2.22	
5	7.88	
6	25.94	
7	66.66	
8	98.20	
All together (product over the k components)	-	

Figure 7 – Illustration multifractal volatility.

Each of the components can take on two different values. In the figure we see how the simulated components change over a period of 200 trading days. We see that the 'low' components have small transition probabilities. Recall that in the binomial model the possibility that we end up with the same value of your previous component after a transition is still fifty percent as there are only two possible values we can draw from the distribution of the components, the current value and the other possible value, both with the same probability. This means that on average we should see $\frac{1}{2}\gamma_k \cdot \bar{t}$ transitions over a period of $\bar{t}$ trading days for component k. The product of all volatility components at time t times the unconditional volatility gives us the volatility at time t.

In the simulated return series we look for the stylized facts on volatility the same way we did for the DJIA returns. The statistics of the simulated series are summarized in Table 3. For both the GARCH and the MSM model we simulated 50000 consecutive returns. We choose to simulate a series of 50000 returns in order to obtain more stabile and converged estimates of the properties of the simulated series.

50000 obs.	GARCH	MSM
mean (%)	0.04	0.03
Stdev.	0. 83	1.05
min (%)	-10.76	-8.80
max (%)	7.37	7.34
skewness	-0.06	0.01
kurtosis	3.03	2.53

Table 3 – Statistics of simulated series.

Table 3 shows that both the GARCH and the MSM model don't display (large) negative skewness and fat tails (A possible way to fix this is to make use of a skewed t-distribution for z_t instead of a Gaussian distributed z_t. In this thesis we won't make use of skewed distributions, but we reckon that this is certainly an interesting direction of future research, if not the most interesting). Although the GARCH model is somewhat closer to reality based on the reported skewness and kurtosis, we have to conclude that both the GARCH and the MSM model don't comply with the first stylized fact on volatility in financial return series.

From the 50000 simulated returns Figure 8 shows a subsample of 5000 returns for both the volatility models.

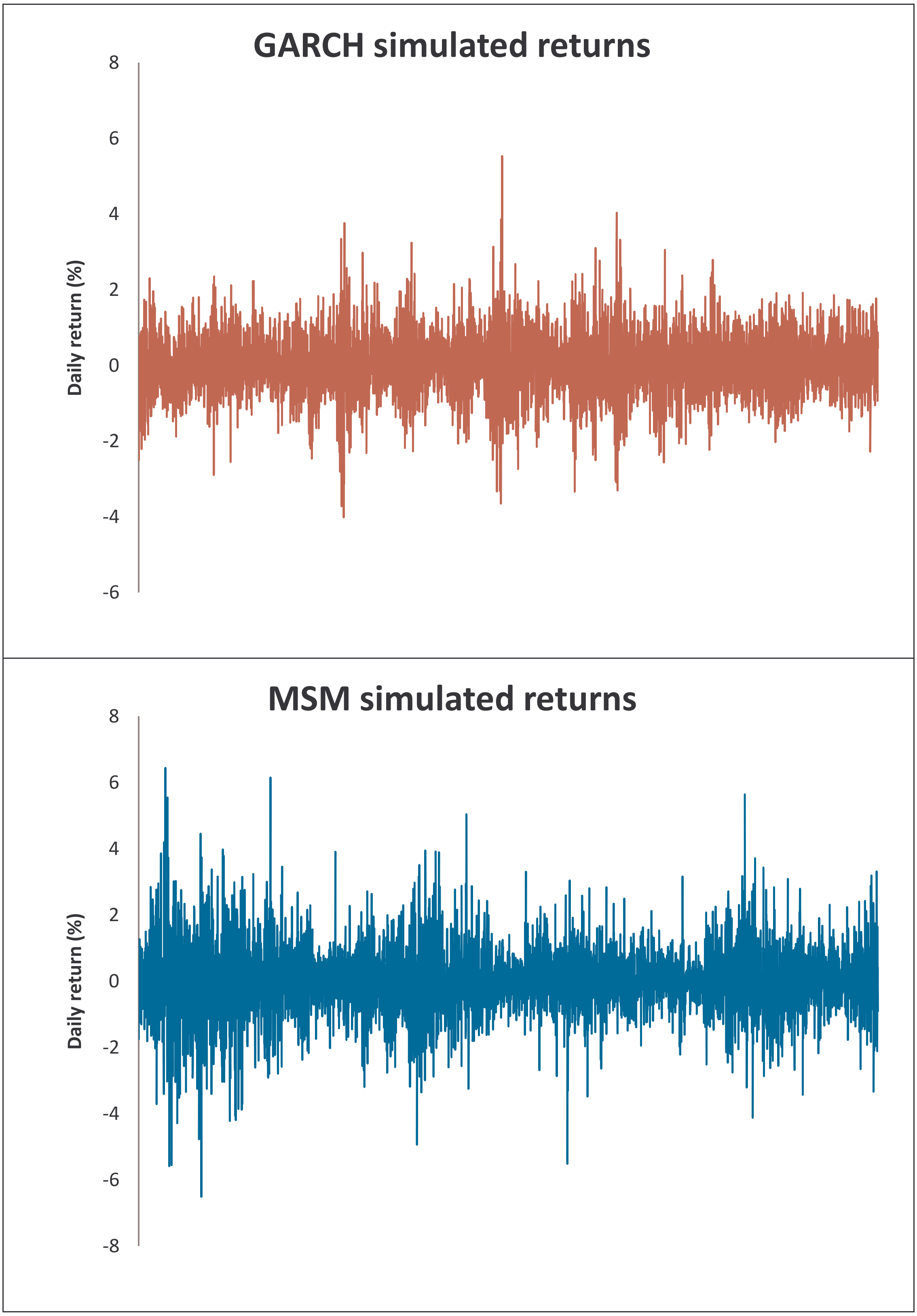
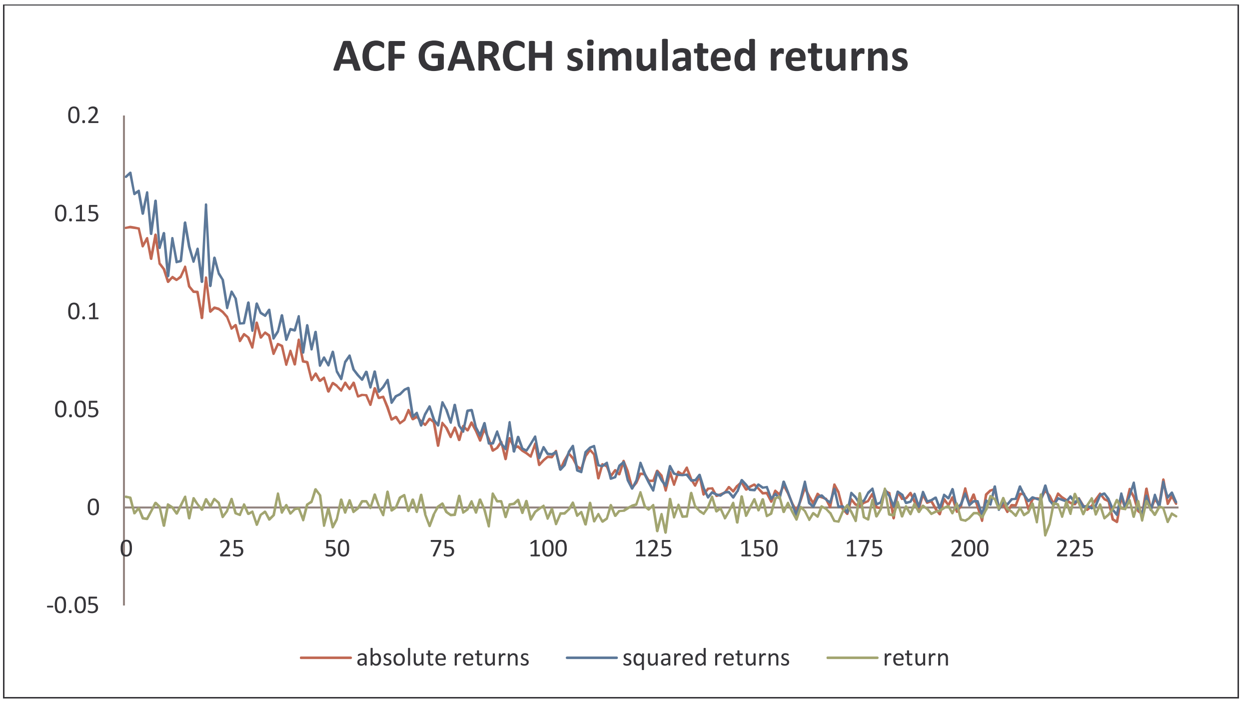


Figure 8 – simulated GARCH and MSM returns.

Figure 8 shows that both the models display heteroskedasticity in the returns over time, like we witness in the actual data in Figure 2. So the models are in consensus with the stylized fact on heteroskedasticity.

Because of the different structures and assumptions, it is very interesting to look at the autocorrelation functions of the simulated returns from both models. Figure 9 contains graphs of the autocorrelation functions of the transformed simulated return series.



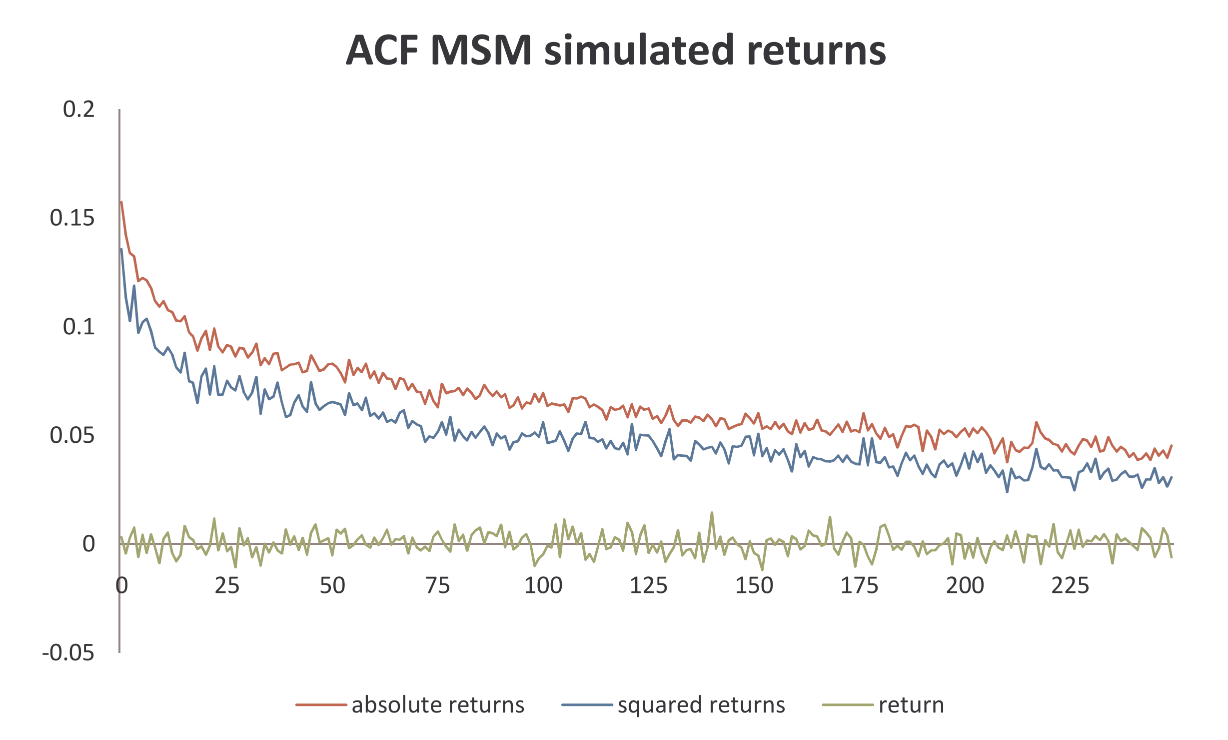


Figure 9 – Autocorrelation functions of simulated GARCH and MSM series.

This figure shows, just as with the observed Dow Jones returns in Figure 4, no autocorrelation in lags of the simulated returns. Furthermore, we do observe autocorrelation in lags of the absolute and in lags of the squared simulated returns from the GARCH and the MSM model. So we can conclude that both models are in consensus with the second stylized fact on volatility. However, there is a clear difference between the decay in the autocorrelation functions of the returns simulated by both models. Because of their structure, the decay of the GARCH autocorrelation function is exponential, where the decay of the MSM autocorrelation function follows a power law[20]. Furthermore, in the MSM model the autocorrelations in the absolute returns are larger than the autocorrelations in the squared returns. This observation can also be made for the actual Dow Jones series. In the GARCH model however, we observe the opposite.

When we compare the ACF of the observed series with the ACF's of the simulated series (The ACF of the observed Dow Jones series follows a much rougher pattern. This is because the ACF of this series is calculated on about 4300 returns, where the ACF's of the simulated series are calculated on 50000 returns.) we notice that the ACF of the MSM simulated series looks more like the ACF of the observed series than the ACF of the GARCH simulated series does. The ACF of the GARCH series decays to zero quickly where the ACF of the MSM series still incorporates a considerable autocorrelation after 250 lags, or one year, just as we witness in the real time series. This observation suggests that the MSM model should do a better job handling long term volatility than the GARCH model does. It also suggests that the ACF of the DJIA follows a power law decay.

To investigate this further, we make a logarithmic transformation of the axes of graphs of the autocorrelation functions to check whether we can detect a power law relationship. In a log-log plot (a

[20] See Appendix D.

plot where the axes are displayed in logarithms), a power law relationship is recognized by a straight line.

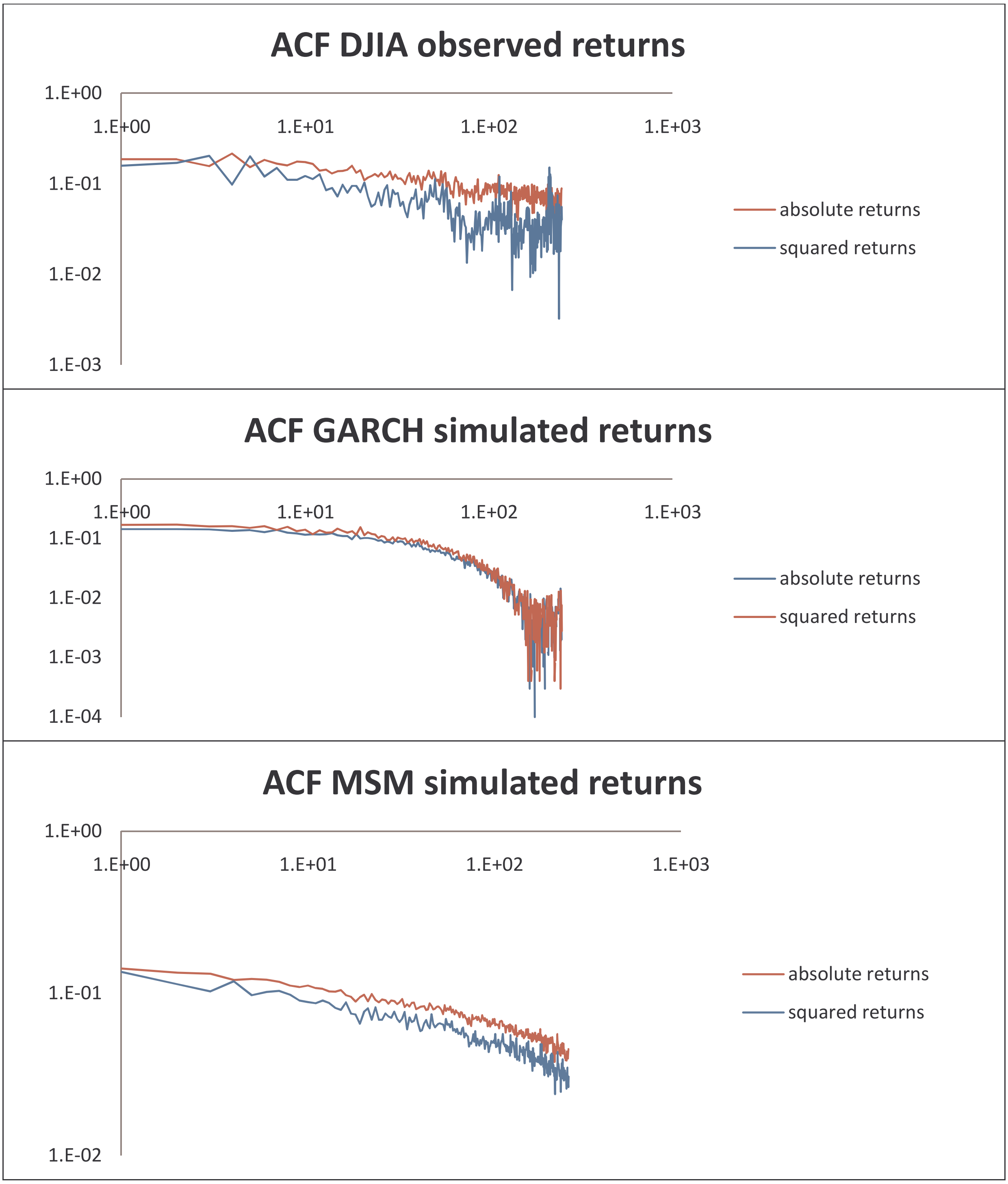


Figure 10 – Logarithmic plots of autocorrelation functions.

Figure 10 shows exactly what we already expected. It is not that hard to visualize a straight line through the plotted ACF of the squared and absolute returns of the observed returns. And it is even easier for the ACF of the MSM simulated returns. The relationship in the ACF of the GARCH simulated returns however, cannot be approximated by a straight line. We can conclude that the MSM model does a better job in replicating the autocorrelation function of the observed time series.

3.3. Application: VaR forecast on the Dow Jones index

The forecast performance of the volatility models is tested by making and evaluating Value at Risk (from now on VaR) forecasts. VaR is defined as the minimum return that could occur over a given holding period (x) with a specified confidence level (α). The VaR measure is a popular risk management tool because it is easy to compute and to interpret. A single value provides you with a lot of useful information. However, it is dangerous to rely on a VaR only because it gives no additional information about the remainder of the tail of the distribution. This is bad because it gives no indication how worse it is going to be when the VaR is violated. Therefore it is useful to imply additional risk measures, like the expected shortfall (ES) for example. But that is beyond the scope of this thesis. We like to evaluate volatility forecasts and for that purpose the Value at Risk is sufficient.

We have DJIA returns starting January 3th 1990 until December 29th 2006. The first window for parameter estimation is January 3th 1990 – January 4th 1999. With the observations of the first window we estimate the parameters for January 4th 1999 that we will use to make VaR forecasts for the first $\bar{t}$ days from January 4th 1999. Then the window expands with one observed return and we estimate the parameters for the models for the first trading day after January 4th 1999 which we use to make the $\bar{t}$ VaR forecasts and so on. This way the first observation for parameter estimation is always January 3th 1990. We use an expanding window because we are looking at long VaR horizons as well. For long term volatility dynamics and long forecast horizons 'old' information is likely to stay valuable.

For every day in the subset January 4th 1999 until December 31th 2004 the parameters for the GARCH and the MSM model are estimated. This subsample contains 1527 trading days. With each of those days as a starting point VaR forecasts for the first $\bar{t}$ following trading days are made. So every x day VaR, $x <= \bar{t}$, has 1527 consecutive forecasts. This number seems large enough to evaluate the forecasting performances of the volatility models on various horizons.

Table 4 contains the mean parameter estimates and the corresponding standard deviations of the models.

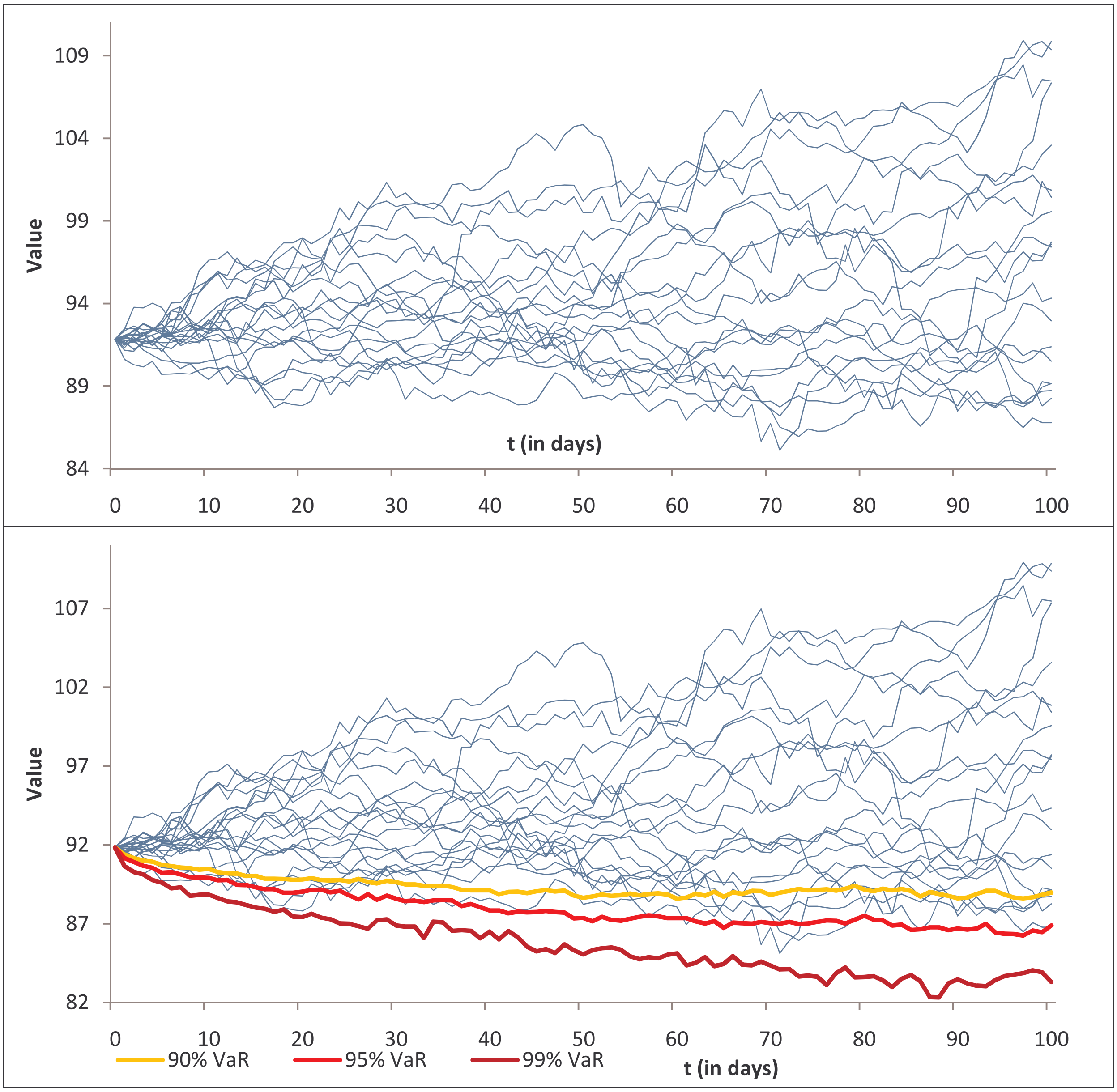
GARCH	ω	α_1	γ	β_1
Mean	0.014	0.013	0.089	0.926
Standard deviation	0.001	0.002	0.010	0.003

MSM($\bar{k} = 8$)	m_0	$\gamma_{\bar{k}}$	b	$\bar{\sigma}$
Mean	1.281	0.988	3.547	1.005
Standard deviation	0.006	0.004	0.299	0.072

Table 4 – Parameter estimates for the GARCH and the MSM model.

To construct the VaR forecast, we simulate 10,000 price paths of length $\bar{t}$ with both the GARCH and the MSM model. The prices at t, $1 \leq t \leq \bar{t}$, are the probability density of the prices t days from now according to the relevant model. Then we take the α'th percentile of the 10,000 prices for every t, these percentiles are our VaR forecasts. The choice of α depends on the confidence level we want to achieve. We will look at $\alpha's$ of 0.1, 0.05 and 0.10. Those $\alpha's$ correspond with a 90%, 95% and 99% VaR. All the VaR forecasts are evaluated by comparing the forecasts with the actual prices. The number of violations of the VaR forecasts will be counted.

Figure 11 shows how the VaR procedure works for the estimates of one single day.



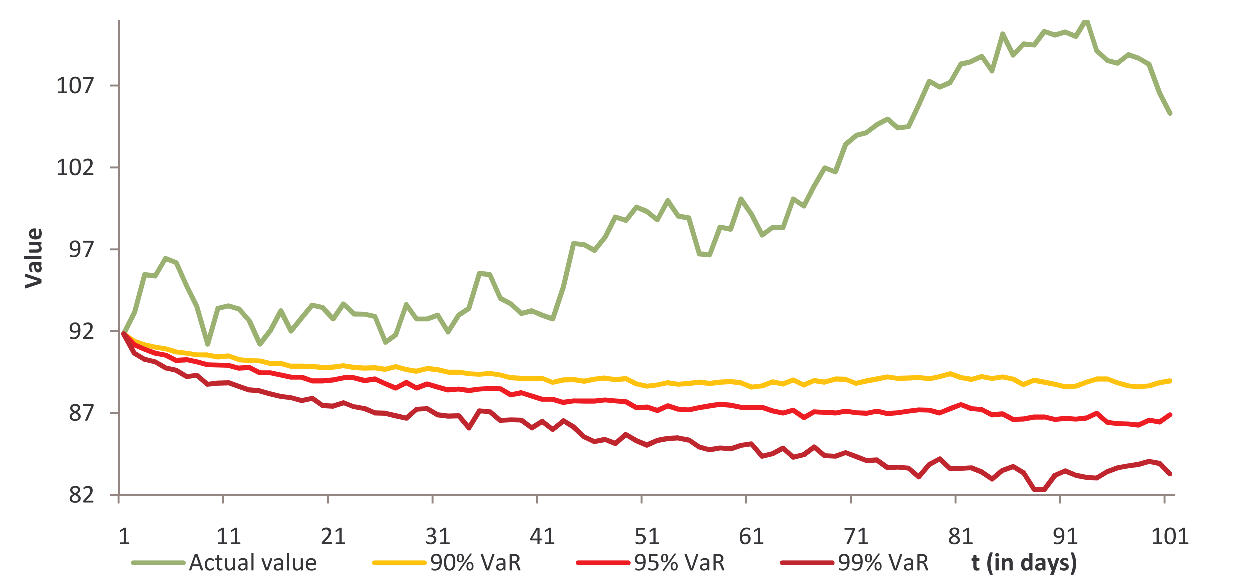


Figure 11 – Illustration of the Value at Risk measure.

The single day in this example is January 4th 1999. The first graph shows 20 of the 10,000 price paths of a length of 100 trading days simulated with the MSM model. The second graph shows the same price paths with the accompanying VaR estimates, which are estimated from the 10,000 price paths. The third graph shows the VaR estimates and the Actual course of the Dow Jones index. In order to evaluate the *x* day VaR estimate, all we have to do is to count for all days in our estimation sample how many times the line of the Dow Jones is below the line of the VaR we are interested in t days from now.

Recall that the definition of VaR is 'the minimum return that could occur over a given holding period with a specified confidence level'. We are now looking at the minimum price of the Dow Jones that could occur over a given holding period with a specified confidence level. This will yield the same results as the price of the Dow Jones in *x* days and the return on the Dow Jones over *x* days correspond one to one to each other.

	x day VaR									
	1	2	3	...	25	26	27	...	50	51	52
GARCH 90% VaR	0.115	0.115	0.115	...	0.139	0.145	0.147	...	0.149	0.151	0.157
GARCH 95% VaR	0.054	0.064	0.062	...	0.081	0.083	0.081	...	0.065	0.064	0.066
GARCH 99% VaR	0.016	0.017	0.018	...	0.023	0.022	0.024	...	0.017	0.017	0.014
MSM 90% VaR	0.261	0.259	0.261	...	0.270	0.267	0.263	...	0.271	0.263	0.268
MSM 95% VaR	0.203	0.197	0.206	...	0.207	0.204	0.199	...	0.208	0.213	0.210
MSM 99% VaR	0.104	0.113	0.110	...	0.111	0.112	0.113	...	0.103	0.097	0.101

Table 5 – Results for short and medium horizon VaR evaluation.

Table 5 shows the results for the VaR evaluation for short and medium term VaR forecasts. The GARCH model performs reasonably well on the short term VaR forecasts up to a week or 10 days for the 90% VaR. When the forecast horizon increases, the performance get worse and we see too many violations.

For the MSM model, all the *x* day VaR forecasts are far off for all confidence levels. We observe far too many violations of the VaR forecasts. The VaR forecasts of the MSM model are far worse than the VaR forecasts produced by the GARCH model. A possible explanation for this can be found in Table 3 that displays the statistics of the simulated series of both models. The kurtosis of the observed series is 4.78, see Table 2, where the kurtosis of the GARCH and the MSM model is 3.03 and 2.52 respectively. Note that the parameter estimates for the period January 4th 1999 – December 31th 2004 differ slightly from the parameters used to produce the simulated series so the descriptive statistics of the returns simulated for the VaR forecasts will not differ that much from the descriptive statistics reported in Table 3 – Statistics of simulated series.. For the GARCH model the failure of forecasting over longer horizons could be due to the simple result from the kurtosis of the GARCH model, which is a lot smaller than the kurtosis of the actual series. The GARCH returns are just not fat tailed enough. The same holds even more for the MSM returns, which have an even smaller kurtosis. The kurtosis of the MSM is so low that even the 90% VaR forecasts turn out bad.

The results in Table 5 can be analyzed further by looking at the violations over the different years as the violations do not have to be equally distributed over the whole period.

	x day VaR								
	1	2	3 …	25	26	27 …	50	51	52
GARCH 90% - 1999	0.121	0.137	0.117 …	0.089	0.107	0.099 …	0.095	0.091	0.101
GARCH 90% - 2000	0.133	0.137	0.122 …	0.213	0.214	0.229 …	0.235	0.231	0.242
GARCH 90% - 2001	0.108	0.133	0.150 …	0.267	0.272	0.274 …	0.380	0.402	0.399
GARCH 90% - 2002	0.141	0.153	0.130 …	0.209	0.246	0.247 …	0.315	0.307	0.343
GARCH 90% - 2003	0.100	0.077	0.093 …	0.062	0.054	0.058 …	0.075	0.080	0.076
GARCH 90% - 2004	0.096	0.061	0.093 …	0.098	0.098	0.099 …	0.035	0.045	0.051
GARCH 95% - 1999	0.060	0.073	0.085 …	0.053	0.058	0.063 …	0.020	0.020	0.025
GARCH 95% - 2000	0.072	0.069	0.069 …	0.129	0.134	0.112 …	0.130	0.116	0.131
GARCH 95% - 2001	0.080	0.089	0.089 …	0.182	0.179	0.179 …	0.185	0.181	0.162
GARCH 95% - 2002	0.044	0.065	0.053 …	0.138	0.143	0.144 …	0.150	0.161	0.182
GARCH 95% - 2003	0.036	0.036	0.036 …	0.036	0.040	0.040 …	0.010	0.010	0.005
GARCH 95% - 2004	0.040	0.052	0.049 …	0.009	0.013	0.018 …	0.000	0.000	0.000
GARCH 99% - 1999	0.028	0.024	0.020 …	0.009	0.005	0.009 …	0.000	0.000	0.000
GARCH 99% - 2000	0.020	0.028	0.032 …	0.044	0.045	0.054 …	0.030	0.035	0.030
GARCH 99% - 2001	0.012	0.024	0.028 …	0.080	0.076	0.076 …	0.045	0.045	0.030
GARCH 99% - 2002	0.020	0.012	0.012 …	0.022	0.022	0.022 …	0.055	0.050	0.046
GARCH 99% - 2003	0.008	0.004	0.000 …	0.000	0.000	0.000 …	0.000	0.000	0.000
GARCH 99% - 2004	0.008	0.012	0.016 …	0.000	0.000	0.000 …	0.000	0.000	0.000

Table 6 - GARCH results on VaR evaluations by year.

Table 6 shows the same evaluation as Table 5 does, but now the six year period is split up over the single years. This way it is possible to look whether there were some years with a lot more violations than other years and subsequently, what the characteristics of the volatility in that years are. Note that for example ‘GARCH 90% - 1999’ means the percentage of violations in the VaR forecasts for the year 1999. For the 90% GARCH VaR we see the most violations in the forecasts for the year 2002 for the short term forecasts. For longer term forecasts we see the most violations in the year 2001. For the 95% VaR we also notice that the forecasts for the year 2001 show way more violations than the other years do. Looking at the 99% VaR we see too many violations for the 1 to 3 day forecasts for the first four years and for the forecasts with a longer horizon in 2001 up to 2003. In the other years we observe way to few or even no violations.

	x day VaR										
	1	2	3	…	25	26	27	…	50	51	52
MSM 90% - 1999	0.237	0.242	0.235	…	0.200	0.188	0.188	…	0.185	0.181	0.187
MSM 90% - 2000	0.281	0.307	0.320	…	0.382	0.380	0.368	…	0.435	0.422	0.429
MSM 90% - 2001	0.309	0.282	0.283	…	0.413	0.429	0.417	…	0.590	0.593	0.601
MSM 90% - 2002	0.337	0.351	0.368	…	0.453	0.451	0.457	…	0.550	0.528	0.551
MSM 90% - 2003	0.261	0.246	0.235	…	0.164	0.161	0.166	…	0.225	0.221	0.222
MSM 90% - 2004	0.173	0.153	0.162	…	0.200	0.196	0.202	…	0.085	0.075	0.076
MSM 95% - 1999	0.173	0.182	0.166	…	0.129	0.134	0.117	…	0.110	0.106	0.116
MSM 95% - 2000	0.217	0.238	0.255	…	0.302	0.304	0.309	…	0.310	0.322	0.313
MSM 95% - 2001	0.229	0.230	0.243	…	0.356	0.339	0.359	…	0.485	0.493	0.500
MSM 95% - 2002	0.309	0.290	0.332	…	0.378	0.357	0.345	…	0.480	0.487	0.475
MSM 95% - 2003	0.209	0.177	0.178	…	0.124	0.138	0.135	…	0.180	0.186	0.182
MSM 95% - 2004	0.104	0.085	0.093	…	0.116	0.116	0.099	…	0.025	0.040	0.035
MSM 99% - 1999	0.072	0.105	0.089	…	0.058	0.063	0.072	…	0.020	0.020	0.025
MSM 99% - 2000	0.141	0.157	0.142	…	0.160	0.156	0.161	…	0.115	0.106	0.096
MSM 99% - 2001	0.137	0.145	0.150	…	0.236	0.223	0.233	…	0.280	0.256	0.268
MSM 99% - 2002	0.181	0.182	0.190	…	0.240	0.241	0.233	…	0.285	0.292	0.303
MSM 99% - 2003	0.084	0.069	0.081	…	0.062	0.080	0.067	…	0.085	0.070	0.086
MSM 99% - 2004	0.024	0.036	0.028	…	0.000	0.000	0.005	…	0.000	0.000	0.000

Table 7 – MSM results on VaR evaluations by year.

Table 7 displays the results for the VaR evaluation with the MSM model for the single years. We already established that there are way too many violations on the VaR produced by the MSM model but still it can be useful to see how the violations are distributed over the different periods. In the same way as with the GARCH model, the forecasts for the years 2001 and 2002 contain way more violations than the other years. To get a better understanding why this is happening we plot the daily returns for the Dow Jones index for the period 1999 up to 2004 in Figure 12.

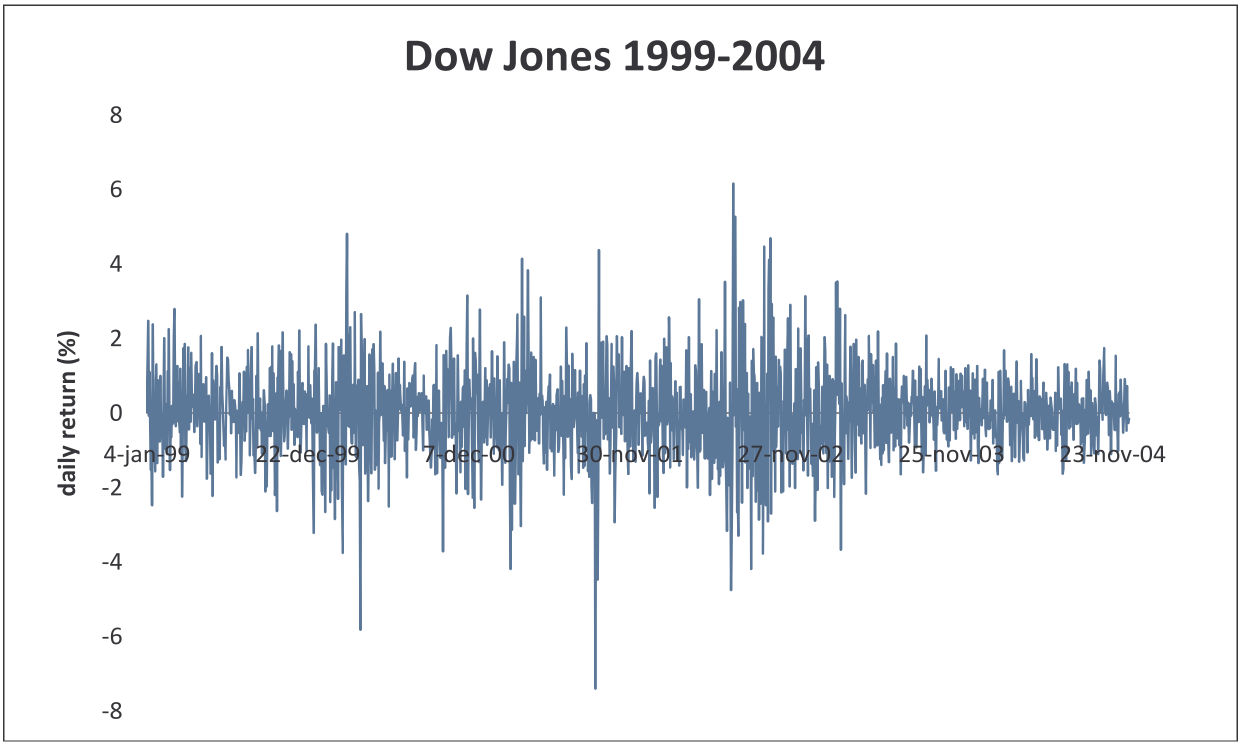


Figure 12 – Dow Jones daily returns 1999-2004.

When we look at the years 2001 and 2002 we notice that the returns in these periods are very volatile, especially in the year 2002. Furthermore, the year 2001 contains an enormous decrease in September 2001. This huge decrease is the reason for a lot of violations in consecutive VaR forecasts. The volatile period in 2002 is also a big reason for a lot of violations. A possible solution to deal better with small periods with large volatility is to make use of a smaller estimation period, in order to have models that are more sensitive to a (recent) change in volatility. Or in other words, to make use of a moving window estimation period instead of a expanding window.

4. Option pricing applications with volatility models

In this thesis we compare the performance of different option pricing models. We will start out with the option pricing basics. In previous research by others, it was found that the GARCH model was the best option pricing model in a comparison between the constant volatility model, the Hull-White model and the GARCH model[21]. We are interested in how the MSM model performs compared to the GARCH model.

4.1. Option pricing basics

Options exist in many varieties, but there are two basic types of options. A call option gives the holder the right to buy the underlying asset by a certain date for a certain price. A put option gives the holder the right to sell the underlying asset by a certain date for a certain price. We will call the price in the contract the strike price and the date in the contract the maturity. An American option can be exercised at any time up to the maturity where an European option can only be exercised at the maturity itself. In this thesis the focus lies on the European options[22]. The most important feature of options is that the option gives the holder the right to buy (or sell, in case of a put option) the underlying, but not the obligation to do so. The holder does not have to exercise his right.

A simple illustration, let's suppose that the unknown stock price (let's assume that the underlying is a stock) at maturity $S_{\bar{t}}$ can take on all values between 95 and 105 and that the strike price *K* in both the put and call contracts is 100. The holder of the call option can buy the stock at maturity for *K* by exercising the option or he can buy the stock in the market for S_t. So he only exercises when $K < S_{\bar{t}}$. With the same reasoning, the holder of a put option only exercises the contract when $K > S_{\bar{t}}$. This results in the following typical option payoff patterns:

	Call option			Put option		
S_t	$S_t - K$	exercise?	payoff	$K - S_t$	exercise?	payoff
95	-5	no	0	5	yes	5
96	-4	no	0	4	yes	4
97	-3	no	0	3	yes	3
98	-2	no	0	2	yes	2
99	-1	no	0	1	yes	1
100	0	yes/no	0	0	yes/no	0
101	1	yes	1	-1	no	0
102	2	yes	2	-2	no	0
103	3	yes	3	-3	no	0
104	4	yes	4	-4	no	0
105	5	yes	5	-5	no	0

[21] Lehar, A., Scheicher, M. and Schittenkopf, C. 2002.
[22] Hull, J. 2006.

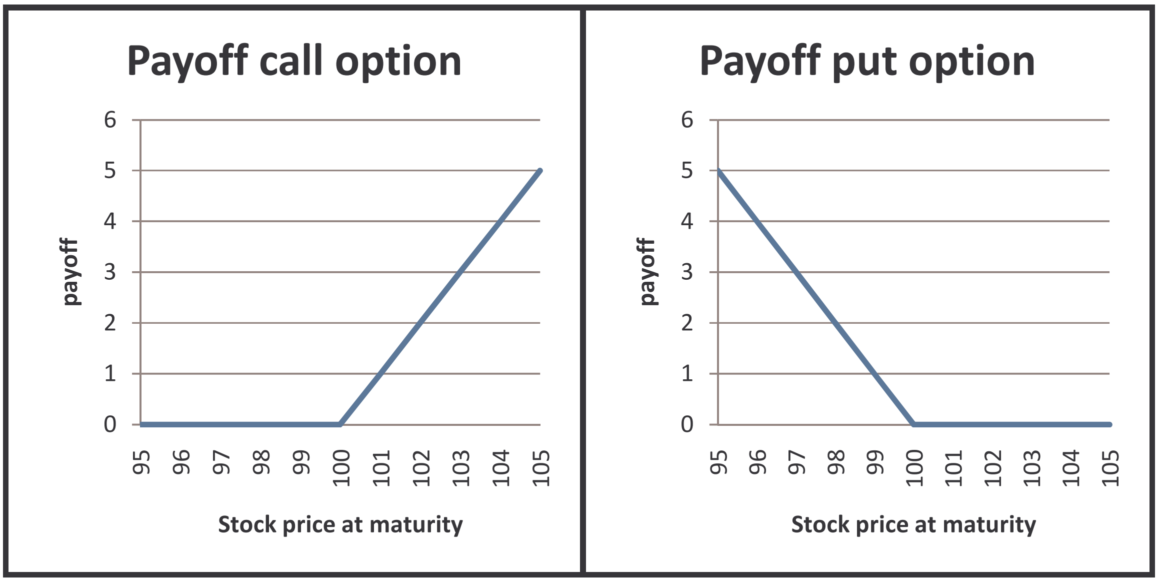


Figure 13 – Option payoff patterns.

Figure 13 clearly show that because of the right of exercise, there are no negative payoffs for the options as all payoffs are greater than zero. We can write the payoff of a call option as $max(S_{\bar{t}} - K, 0)$ and the payoff of a put option as $max(K - S_{\bar{t}}, 0)$. As the option contracts can result in non-negative payoffs only, there is a cost to acquiring an option. We will call this cost the option price.

4.2. Monte Carlo option pricing

Options are priced assuming a risk-neutral world. In a risk-neutral world, the expected return on all assets is the risk free interest rate as we make the assumption that all investors are risk-neutral. Due to that assumption the present value of all cash flows can be obtained by discounting its expected value at the risk free rate. This property of the risk-neutral world makes risk-neutral valuation convenient as we do know the discount rate for all expected payoffs in the risk free world, where we do not know the appropriate discount rate in the real-world. The resulting option prices from risk-neutral valuation give the correct price, not only in a risk neutral world but in the real world as well. Risk-neutral valuation also ensures that there are no possible arbitrage opportunities that make use of a replicating portfolio strategy.

The option price depends on the price of the underlying at maturity, but the stock price at maturity is unknown at *t*=0. To be able to estimate the option price it is common (and often necessary due to the complexity of the models) to use a Monte Carlo approach. First of all, the option price can be seen as the discounted expectation of the payoff at maturity, which depends on the price at maturity.

It is important to distinguish between an actual price path and its expected path at time 0. The price of an underlying will follow a certain path as uncertainty resolves over time, this is the actual price observed at time *t*. However, we are not trying to predict the actual path of the underlying, because this is simply impossible due to the many unknown and unobservable factors that drive the price process.

What we can do is make expectations about the future at time 0. The expectation about the future can be seen as a probability density of the price at maturity. This can be represented as

$$option\ price = D * \int_a^b g(S_{\bar{t}})\, dF(S_{\bar{t}}) \quad (4)$$

Where *D* is the discount factor (as a positive payoff in the future is worth less than the same payoff right now, given positive interest rates, due to the time value of money). $g(.)$ is the payoff function, which determines the output given a certain price. The payoff function of the call option is $\max(S_{\bar{t}} - K)$ and for the put option, $g(S_{\bar{t}}) = \max(K - S_{\bar{t}})$. $dF(S_{\bar{t}})$ is the probability density of $S_{\bar{t}}$, $\int_a^b dF(S_{\bar{t}}) = 1$.

This integral can be estimated by $E[g(S_{\bar{t}})] = \sum_{i=1}^{k} \frac{g(S_{\bar{t}}^{(i)})}{k}$ as $k \to \infty$. Where $S_{\bar{t}}^{(i)}$ is the price at maturity of the i'th $(1,2, \dots, i, \dots, k-1, k)$ price path, where all the k price paths are resolutions of a stochastic price model.

4.3. Parameter estimation from option prices

We can construct an expectation about the future by simulating a lot of price paths with our models. The resulting prices at maturity form a simulated probability density for the price at maturity. With those simulated prices the expected payoff at maturity is estimated. Of course, how realistic these simulated density and consequently the estimated option price is, depends on how good or realistic the model (including its parameter specifications) we use to simulate the prices of the underlying is.

Again, it is very important that the simulated price paths are risk neutral. The set of simulated price paths is risk neutral when the expected return on the set, the mean return, is equal to the risk free return. In our case the risk free return is the return on risk free bonds. We use the term structure to create a set of price paths that corresponds to the term structure. First we simulate a set of price paths with one of the volatility models. Then we demean this set of price paths to make sure the mean of the whole set is equal to zero. At last we add the risk free return according to the term structure to the price paths with the result that the price paths are risk neutral. Figure 14 shows the construction of risk-neutral price paths.

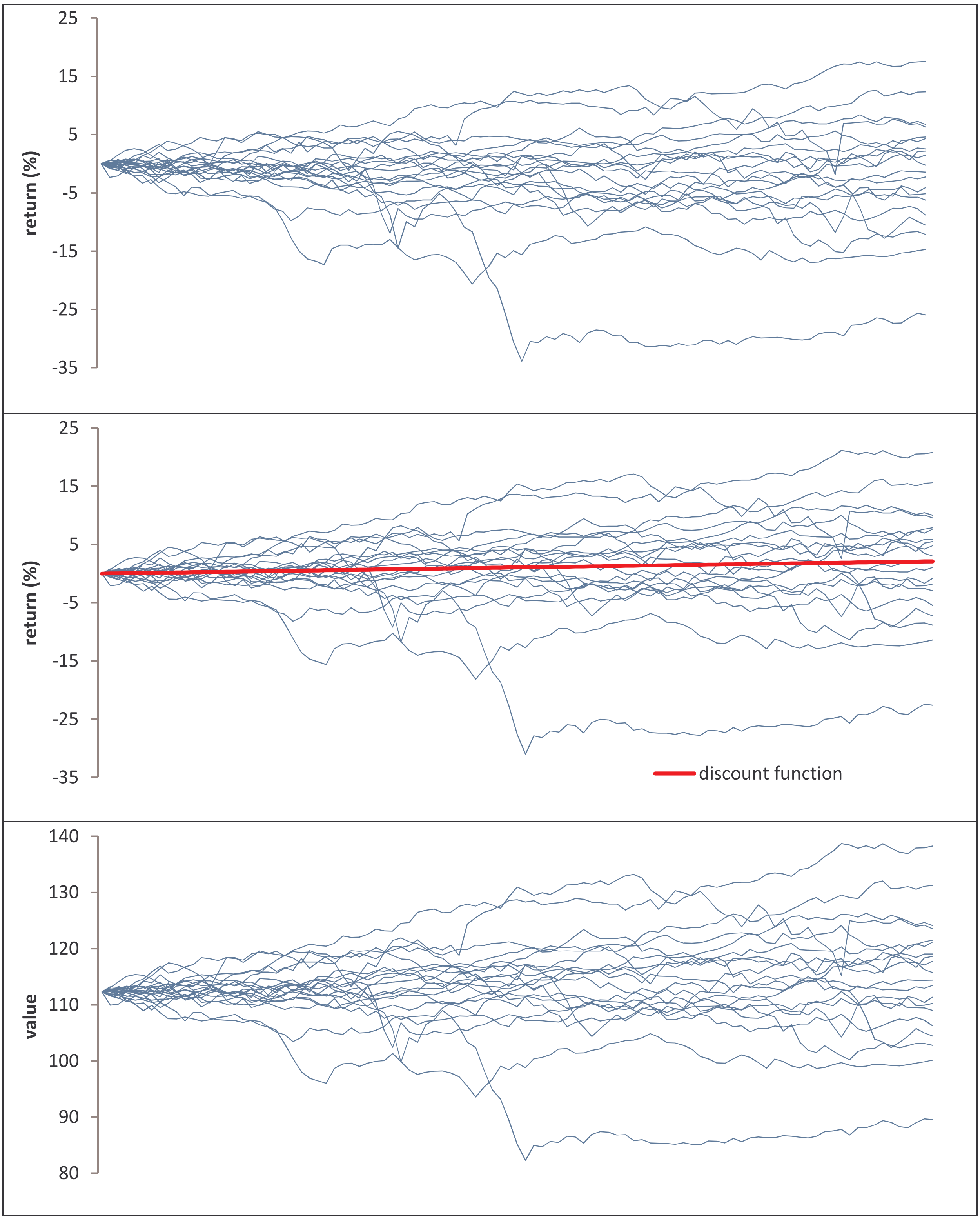


Figure 14 – Illustration on risk neutral price paths.

The first graph shows 20 cumulative return paths simulated with a GARCH model with parameters estimated on option contracts traded on June 3th 2006. The second graph shows the cumulative returns from the first graph, adjusted to risk neutral cumulative returns. The average of the 20 cumulative return paths in the second graph is now equal to the discount function (the discount function is the yield from the term structure multiplied with the corresponding maturity, for every maturity). Finally, the third graph shows the risk neutral price paths, which are the result of the risk neutral cumulative return paths from the second graph.

Now in order to examine how realistic our simulated price density is, we have to compare it with actual option prices, which are the result of the market expectation of the volatility of the underlying. With the simulated paths, the strike price, time to maturity of the observed option contract and the appropriate yield curve we can estimate what the option price would be according to the model.

The aim is to estimate parameters for the volatility models for every single day in the dataset. Therefore we group the option contracts in our dataset by the day the contract was traded. This way, the parameters of the models are estimated on a lot of option contracts.

By changing the parameters of the model the price paths will change as well, resulting in different option prices. So we are left with an optimization problem. We have to find out what parameter setting will lead to the smallest error with the actual data. The error term we minimize is the squared relative pricing error, see equation (5).

$$SRPE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{p}_i - p_i}{p_i}\right)^2 \qquad (5)$$

Where $\hat{p}_i$ is the estimated and p_i is the observed or actual option price for option i and N is the number of option contracts on the specific day. Because there is a huge variety in the value of the options in the dataset, a relative error term is more suitable than an absolute error term, because an absolute error term would give the options with a high value a bigger weight in the error term. The SRPE is the function we minimize. The SRPE is a function of the parameters of our models and the lower the SRPE the lower the error of our simulated price path with the real expected price paths.

Estimating the model parameters using observed option prices has a huge advantage over estimating the parameters from the series of the underlying as the estimated parameters are the risk-neutral parameters. For estimating from historical return series of the underlying this is not the case and the estimated parameters have to be transformed before we can use them to valuate options. For the constant volatility model this transformation is pretty straightforward as only the drift component changes. However, for the more complex volatility models this transformation is not so simple.

In this thesis all estimations are made based on simulations with 10,000 paths. Furthermore, we will make use of a control variate[23] technique in order to reduce the variance of our estimations.

4.4. Estimation of the current state of volatility

Besides the parameters that determine the price paths in a simulation it is also crucial to know where the volatility is at right now at the starting point of simulation, at *t*=0. As taking a volatility of 0.5 instead

[23] See Appendix E.

of 1.0 as a starting point will result in different volatility paths and consequently in different price paths. The disadvantage of estimating model parameters from option prices is that we don't have a history that we can use. We estimate the parameters on day *x* with the observed option contracts on day *x* and it is not possible to use the option contracts on previous days, the history, to estimate the current volatility (where this is a possibility when we estimate parameters from total return series of for example a single stock or an index).

For the constant volatility model adding the current state of the volatility is straightforward, as this model assumes constant volatility, the estimated parameter σ is the current state of volatility. For the GARCH model the estimation of one extra parameter is necessary. We call this parameter $currentVol$. This takes the total number of parameters for the GARCH model to five. Besides the four original parameters we now also optimize the fifth parameter $currentVol$ in order to minimize the SRPE.

Estimating the current volatility state of the MSM model takes some extra work. This is due to the typical structure of the MSM model. The current volatility state of the MSM model is not one single value or state, but a probability density of multiple states. This distinguishes the MSM model from the constant volatility and the GARCH model, where the current volatility is a single value. We have to come up with a smart way to estimate the probability density that defines the current volatility state of the MSM model.

Every single one of the $\bar{k}$ volatility components in the MSM model takes on either value m_0 or value m_1, where $m_1 = 2 - m_0$ at the start, *t*=0. Now define $prob_k$ to be the chance that the *k*'th component is in state m_0. With those *k* probabilities it is possible to calculate the probability for every possible 2^k different volatility states to be the state we are in at *t*=0. Recall, the volatility is driven by k components with values $M_{k,t}$. The product of the values of the different components is the volatility state at time t, $Vstate_t = \prod_{i=1}^{\bar{k}} M_{k,t}$. We have

$$P(Vstate_0 = current\ state) = \prod_{i=1}^{\bar{k}} \left(I\left(M_{k,0} = m_0\right) \cdot prob_k + I\left(M_{k,0} = m_1\right) \cdot (1 - prob_k) \right)$$

for the probability that a particular volatility state is the starting state. Where $I(.)$ is an indicator function that returns a 1 when true and a 0 when false. As $M_{k,t}$ is either m_0 or in m_1, two possibilities, we end up with $2^{\bar{k}}$ different volatility states. The probabilities that those different states are the current volatility states combined are the probability density for the current state.

Similar to the extra parameter for the GARCH model, we have to optimize the $\bar{k}$ extra parameters for the MSM model in order to minimize the SRPE. A huge drawback is that adding an extra $\bar{k}$ parameters to our solve function increases the computational effort enormously when $\bar{k}$ gets larger. Not only because of the extra parameters that have to be estimated, because the number of states grows with a factor 2 and the number of elements of the transition matrix with a factor 4 as we add one extra component to $\bar{k}$. Therefore, we stick to the MSM model with 8 volatility components in this thesis.

We are going to evaluate the volatility models on an empirical dataset. First we estimate the parameters on some part of the dataset. With those parameters we estimate the option prices of the other part of the dataset. Finally we compare the estimated and the observed prices of the option contracts.

4.5. The Dataset

The dataset consists of transaction data of European options on the Dow Jones Industrial Average Index traded on the Chicago Board Options Exchange[24]. The dataset consist of date, closing price index, experience date of option, call/put flag, strike price of option, best bid and best offer price, the implied volatility and the identification number of the option. The sample starts on July 3th 2006 and ends on December 29th 2006. This period contains 126 trading days. The sample comprises 67,656 traded option contracts. Not all this observations are in the final sample from which the parameters are estimated. Observations with an extreme implied volatility (yearly implied volatility>50%) and observations with a price really close to zero (price <0.25) are removed from the sample. The observations with extreme volatilities are removed because such extreme observations can have a huge impact on the estimation of the models. The observations with low prices are removed because pricing errors due to rounding effects can be substantial for low prices. This leaves us with a final sample of 57,476 option contracts ,30,732 call options and 26,744 put options. So we have 57,476 option contracts in a period of 126 days. This gives us an average of 456 option contracts a day. But of course, the number of option contracts traded on a day is not uniformly distributed among the days. We notice an increase in option contracts over the sample period. On the day with the least activity, August 18th, there are 328 observations. On December 22th, the most active day, we witness almost 638 trades, almost twice as many. These numbers of available option contracts for the single days seems more than sufficient in order to make reliable parameter estimations for our models on a day to day basis.

	Min.	10%	25%	50%	75%	90%	Max.
DJIA	107.39	111.04	113.45	119.86	122.78	124.08	125.11
Option price	0.25	0.80	2.20	6.15	13.50	24.70	50.20
Strike price	74	92	104	113	122	130	140
Time to maturity	1	26	55	157	274	401	499

Table 8 – Descriptive statistics of the dataset.

The time to maturity is given in trading days. There are roughly 250 trading days in one calendar year. Table gives the descriptive statistics of the dataset. The minimum, the 10th ,20th , 50th ,75th and 90th percentile and the maximum are given for several variables in the sample. During the sample period, DJIA rose from 112.82 on July 3th to 124.63 on December 29th.

[24] Source: OptionMetrics.

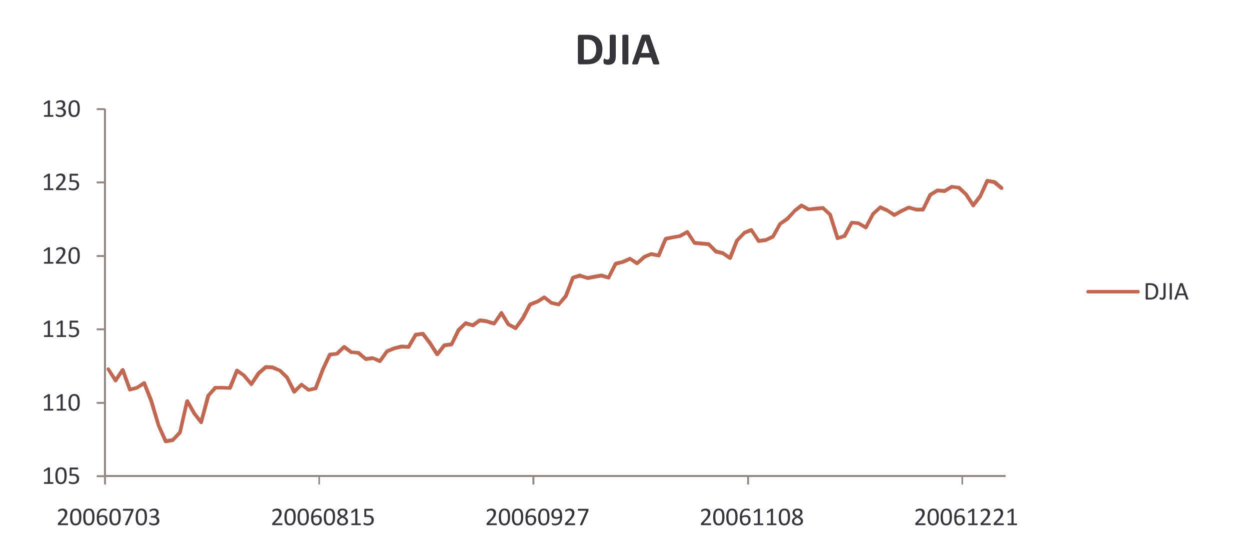


Figure 15 – The Dow Jones Industrial Average from July 3th up until December 29th 2006.

The strike prices in that period range from 74 to 140. There is a great variety in time to maturity for the different contracts. The shortest contract expires after one day, whereas the longest contract has a maturity of about two years. The choice for the Dow Jones Index is convenient as there is a lot of trading activity on the index and on the index' options, providing liquidity. Another advantage is that the index ignores returns from dividends, which means that there is no need for including the dividend yield in our option pricing setup.

Besides the dataset containing the option contracts, we make use of a dataset containing daily observations of U.S. Treasury yield curve as well[25].

[25] Gurkaynak, R. S., Sack, B. and Wright, J. H. 2006.

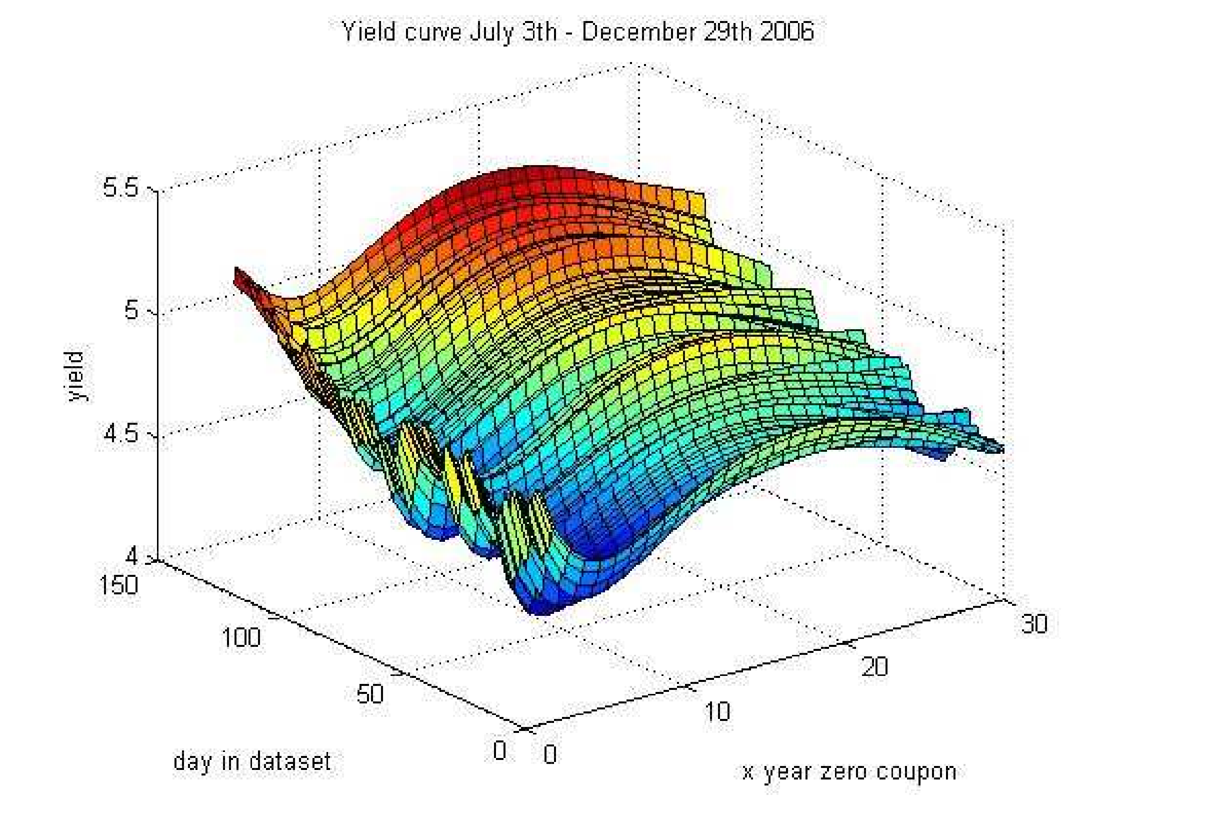


Figure 16 – The term structure from July 3th up until December 29th 2006.

Next to U.S. Treasury yield curve data, the set also contains estimated Svensson[26] parameters for every single trading day. The Svensson model is an extension to the Nelson Siegel model. These parameters make it possible to estimate the risk free rate for every maturity of any option contract in our sample by solving the Svensson model with that given maturity and the Svensson parameters for that day.

We will estimate the parameters for the constant volatility model, the GARCH and the MSM model for every day in our half-year sample of observed option contracts. 80% of these option contracts will be used for estimating the parameters (the in sample observations). The resulting 20% will be used in an out of sample test for the models. Furthermore, we will look at value at risk forecasts for the different models on multiple time scales.

4.6. Estimated parameters

As describer earlier, the parameters for each of the models are estimated on 126 consecutive trading days by minimizing the SRPE (equation $SRPE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{p}_i - p_i}{p_i}\right)^2$ (5)). Table 9, Table 10 and Table 11 report the mean parameter values and the corresponding standard deviations for the various models.

[26] See Appendix F.

Constant volatility model	σ
Mean	0.71
Standard deviation	0.04

Table 9 – Constant volatility model parameter estimates.

For the parameters of the constant volatility model the daily volatility is reported.

GARCH model	ω	α_1	γ	β_1	$currentVol$
Mean	0.08	0.16	0.50	0.49	0.57
Standard deviation	0.07	0.12	0.19	0.22	0.39

Table 10 – GARCH parameter estimates.

Table 10 displays the four parameters of the GARCH model plus the fifth model that indicates the current volatility.

MSM model	m_0	$\gamma_{\bar{k}}$	b	$\bar{\sigma}$	p_1	p_2
Mean	1.57	0.39	5.05	1.17	0.44	0.39
Standard deviation	0.17	0.21	1.34	0.36	0.15	0.16
MSM model	p_3	p_4	p_5	p_6	p_7	p_8
Mean	0.32	0.43	0.45	0.42	0.47	0.43
Standard deviation	0.11	0.14	0.19	0.15	0.14	0.19

Table 11 – MSM parameter estimates.

Table 11 shows the four parameters of the MSM model plus the eight $prob_k$ parameters, the probability that the k'th component has value m_0 at t=0.

Notice that the estimated parameters, except for the estimated σ for the constant volatility model, have a rather large standard deviation. This indicates that believes of market participants in the DJX fluctuate over the period that is covered by our dataset.

4.7. Out of sample pricing performance

The first part of the evaluation is based on the out of sample performance of the models. We divide the observed option contracts in an in sample part and an out of sample part. The in sample part (80% of the observed options on a single day) is used to estimate the parameters for that day. Subsequently, we use those estimated parameters to estimate the option prices for the option contracts in our out of sample set. The goodness of fit of the out of sample estimates can be evaluated by comparing them with the observed prices for the out of sample option contracts. We measure the deviation of the estimated option prices with the observed option prices in terms of the relative pricing error (RPE) and the absolute relative pricing error (ARPE).

$$RPE = \frac{\hat{p}_i - p_i}{p_i}$$

$$ARPE = \left|\frac{\hat{p}_i - p_i}{p_i}\right|$$

The RPE is a measure of the bias of the pricing error. A non zero RPE indicates that prices are structurally under or overpriced by the pricing model. The ARPE measures, besides the pricing bias, also the efficiency of the prices produced by the models.

The out of sample set contains 11,444 option contracts, 6,087 call option and 5,357 put option contracts. Table 12 displays the results over the whole out of sample set.

RPE	Mean(%)	Median(%)	Standard deviation(%)
Constant volatility	-0.16	0.00	0.41
GARCH	-0.07	0.01	0.29
MSM	-0.09	0.00	0.39

Table 12 – Relative pricing error results.

We see that all the models report a negative mean RPE. All models systematically underprice the option contracts. The median of the RPE over all option contracts lies close to zero for all models. This means that all the models approximately overprice just as many contracts as they underprice. So the magnitude of the RPE of the underpriced contracts is a lot larger than the magnitude of the RPE of the overpriced contracts. Furthermore we see that the Standard deviation of the RPE is large for all models. The GARCH model seems to perform the best of the different models.

ARPE	Mean(%)	Median(%)	Standard deviation(%)
Constant volatility	0.28	0.10	0.33
GARCH	0.21	0.12	0.22
MSM	0.25	0.11	0.31

Table 13 – Absolute relative pricing error results.

Table 13 shows the results for the ARPE over the whole out of sample set. The average relative pricing error is quite large. From 28% with the constant volatility model to 21% with the GARCH model which has, just as with the RPE, the lowest pricing error. The values for the median are once more a lot smaller than the values for the mean, this suggests that there are some contracts that have huge pricing errors and pull the mean away from the median.

To investigate this further, we will split up the sample in buckets, in an attempt to find out whether there are some option contract characteristics that make an option harder to price, what leads to larger pricing errors. We construct buckets by two option contract characteristics, time to maturity and the moneyness of the option.

The moneyness is a measure of the degree to which an option is likely to have a positive payoff at maturity. The moneyness of an option is calculated as $\frac{Value\ underlying - Strike\ price}{Strike\ price}$ for call options and $\frac{Strike\ price - Value\ underlying}{Strike\ price}$ for put options. A negative moneyness indicates that the option is out-of-the-money, a positive moneyness that the option is in-the-money and a moneyness around zero indicates that the option is at-the-money.

We divide the option contracts in 48 buckets. The buckets and the number of option contracts per bucket are shown in Table 14.

Moneyness	Maturity						all
	<50	50≤…<100	100≤…<150	150≤…<200	150≤…<250	≥250	maturities
<-0.2	0	1	62	116	118	382	679
-0.20≤...<-0.13	1	50	53	51	84	232	471
-0.13≤…<-0.07	54	169	155	166	183	499	1226
-0.07≤…<-0.00	578	437	289	285	265	585	2439
0.00≤…<0.07	832	450	287	275	259	606	2709
0.07≤…<0.13	484	214	184	178	194	528	1782
0.13≤…<0.2	220	94	66	92	78	195	745
≥0.2	374	231	147	124	120	397	1393
all moneyness	2543	1646	1243	1287	1301	3424	11444

Table 14 – Buckets and number of contracts per bucket.

We won’t look at the buckets with less than 50 contracts. This eliminates three buckets.

The results for the different subsets are given in Table 15 . The reported values are the mean RPE over the option contracts in the bucket.

Moneyness	Model	Maturity <50	50≤…<100	100≤…<150	150≤…<200	150≤…<250	≥250	all maturities
<-0.2	BS			-0.99	-0.98	-0.98	-0.94	-0.96
	GARCH			-0.53	-0.47	-0.57	-0.55	-0.53
	MSM			-0.47	-0.52	-0.57	-0.54	-0.54
-0.20≤...<-0.13	BS		-0.97	-0.94	-0.75	-0.44	-0.60	-0.67
	GARCH		-0.33	-0.42	-0.30	-0.17	-0.34	-0.31
	MSM		-0.45	-0.45	-0.30	0.04	-0.22	-0.23
-0.13≤…<-0.07	BS	-0.87	-0.68	-0.19	-0.08	-0.16	-0.24	-0.29
	GARCH	-0.38	-0.17	-0.01	0.00	-0.01	-0.06	-0.07
	MSM	-0.36	-0.21	0.19	0.25	0.10	-0.09	-0.01
-0.07≤…<-0.00	BS	-0.01	0.02	-0.16	-0.17	-0.21	-0.22	-0.11
	GARCH	-0.01	0.04	-0.08	-0.08	-0.11	-0.10	-0.05
	MSM	-0.06	0.00	-0.18	-0.20	-0.20	-0.19	-0.13
0.00≤…<0.07	BS	0.01	-0.05	-0.09	-0.12	-0.15	-0.18	-0.08
	GARCH	0.00	-0.03	-0.06	-0.06	-0.09	-0.10	-0.05
	MSM	-0.02	-0.09	-0.12	-0.14	-0.16	-0.18	-0.10
0.07≤…<0.13	BS	0.00	0.00	-0.02	-0.05	-0.06	-0.10	-0.04
	GARCH	0.01	0.02	-0.01	-0.02	-0.04	-0.05	-0.02
	MSM	0.01	0.02	0.01	-0.02	-0.03	-0.07	-0.02
0.13≤…<0.2	BS	0.01	0.01	0.01	0.02	0.01	0.01	0.01
	GARCH	0.01	0.02	0.02	0.03	0.02	0.03	0.02
	MSM	0.01	0.02	0.03	0.04	0.04	0.04	0.03
≥0.2	BS	0.01	0.02	0.04	0.04	0.05	0.08	0.04
	GARCH	0.01	0.03	0.04	0.04	0.06	0.08	0.04
	MSM	0.01	0.03	0.04	0.05	0.06	0.09	0.05
all moneyness	BS	-0.02	-0.10	-0.17	-0.19	-0.22	-0.25	-0.16
	GARCH	-0.01	-0.02	-0.08	-0.08	-0.10	-0.12	-0.07
	MSM	-0.02	-0.05	-0.08	-0.10	-0.11	-0.15	-0.09

Table 15 – Mean relative pricing error per bucket.

When we examine this table, we see that overall the GARCH model has the smallest RPE. In most of the subsamples, the GARCH model dominates the BS and the MSM model. The BS model performs the worst, but this was expected as the model assumes constant volatility, an assumption that empirically does not hold. There is not that big of a difference in the average RPE of the GARCH and the MSM model in the different buckets. We see that option contracts with a negative moneyness are a lot more difficult to price than option contracts with a positive moneyness. Furthermore we see that with longer maturities come larger pricing errors. This is also something we expected, as a longer time to maturity brings along more uncertainty. Those observations can be made for all the three models. Table 16 reports the mean ARPE for the subsets of option contracts.

Moneyness	Model	Maturity <50	50≤…<100	100≤…<150	150≤…<200	150≤…<250	≥250	all maturities
<-0.2	BS			0.99	0.98	0.98	0.96	0.97
	GARCH			0.53	0.48	0.57	0.57	0.55
	MSM			0.48	0.53	0.57	0.59	0.56
-0.20≤...<-0.13	BS		0.97	0.94	0.91	0.96	0.79	0.87
	GARCH		0.38	0.42	0.51	0.64	0.54	0.52
	MSM		0.52	0.45	0.61	0.92	0.70	0.68
-0.13≤…<-0.07	BS	0.87	0.84	0.72	0.71	0.50	0.45	0.60
	GARCH	0.43	0.36	0.43	0.46	0.43	0.38	0.40
	MSM	0.40	0.47	0.72	0.80	0.57	0.43	0.54
-0.07≤…<-0.00	BS	0.42	0.39	0.32	0.28	0.29	0.31	0.35
	GARCH	0.30	0.26	0.24	0.21	0.25	0.27	0.26
	MSM	0.35	0.39	0.31	0.25	0.27	0.28	0.31
0.00≤…<0.07	BS	0.05	0.08	0.12	0.15	0.18	0.24	0.13
	GARCH	0.06	0.09	0.13	0.13	0.16	0.21	0.12
	MSM	0.06	0.11	0.16	0.18	0.20	0.26	0.15
0.07≤…<0.13	BS	0.02	0.03	0.06	0.08	0.10	0.17	0.08
	GARCH	0.02	0.05	0.07	0.09	0.10	0.15	0.08
	MSM	0.02	0.04	0.06	0.08	0.10	0.17	0.09
0.13≤…<0.2	BS	0.01	0.03	0.04	0.05	0.07	0.11	0.05
	GARCH	0.02	0.04	0.05	0.06	0.08	0.11	0.06
	MSM	0.02	0.03	0.05	0.06	0.08	0.12	0.06
≥0.2	BS	0.01	0.02	0.04	0.05	0.06	0.09	0.05
	GARCH	0.01	0.03	0.04	0.05	0.06	0.09	0.05
	MSM	0.01	0.03	0.04	0.05	0.06	0.10	0.05
all moneyness	BS	0.14	0.25	0.30	0.33	0.34	0.36	0.28
	GARCH	0.11	0.15	0.20	0.22	0.26	0.28	0.21
	MSM	0.11	0.21	0.26	0.29	0.31	0.32	0.25

Table 16 – Mean absolute relative pricing errors per bucket.

From Table 16 we can conclude that the ARPE and with that the efficiency for the three models does not differ that much when we look at option contracts with a positive moneyness (in-the-money options). The GARCH model performs better pricing out-of-the-money options, even though the pricing errors of the GARCH model are very large for those option contracts. Again we find that the smaller the maturity the smaller the pricing errors over all models.

4.8. VaR on options

The final application we cover in this thesis is estimating and evaluating a Value at Risk on options. The VaR on options requires a different setup than making VaR forecasts on an index, as we did in Section

3.3. Instead of estimating the minimum level of the Dow Jones that could occur over a given holding period with a specified confidence level, we will now estimate the minimum option value that could occur given a holding period with a specified confidence level. We look at a holding period of 1 day, 10 and 50 days. In order to evaluate the *x* day VaR it is necessary to find option contracts that cover the same option, traded at day *t* and day *t+x.* In other words, we have to look for every option in our dataset whether there exists an option of the same type (call/put), with the same strike price, with a maturity that is *x* days shorter and that is recorded *x* days later in time. We collect all the options that meet these requirements.

In our sample of 57,476 option contracts we happen to find a lot of those *x* days in between option pairs. Table 17 displays the number of option pairs for the different holding periods.

Holding period (*x*)	Number of option pairs
1 day	30,549
10 days	26,312
50 days	4,462

Table 17 – Number of option pairs for various holding periods.

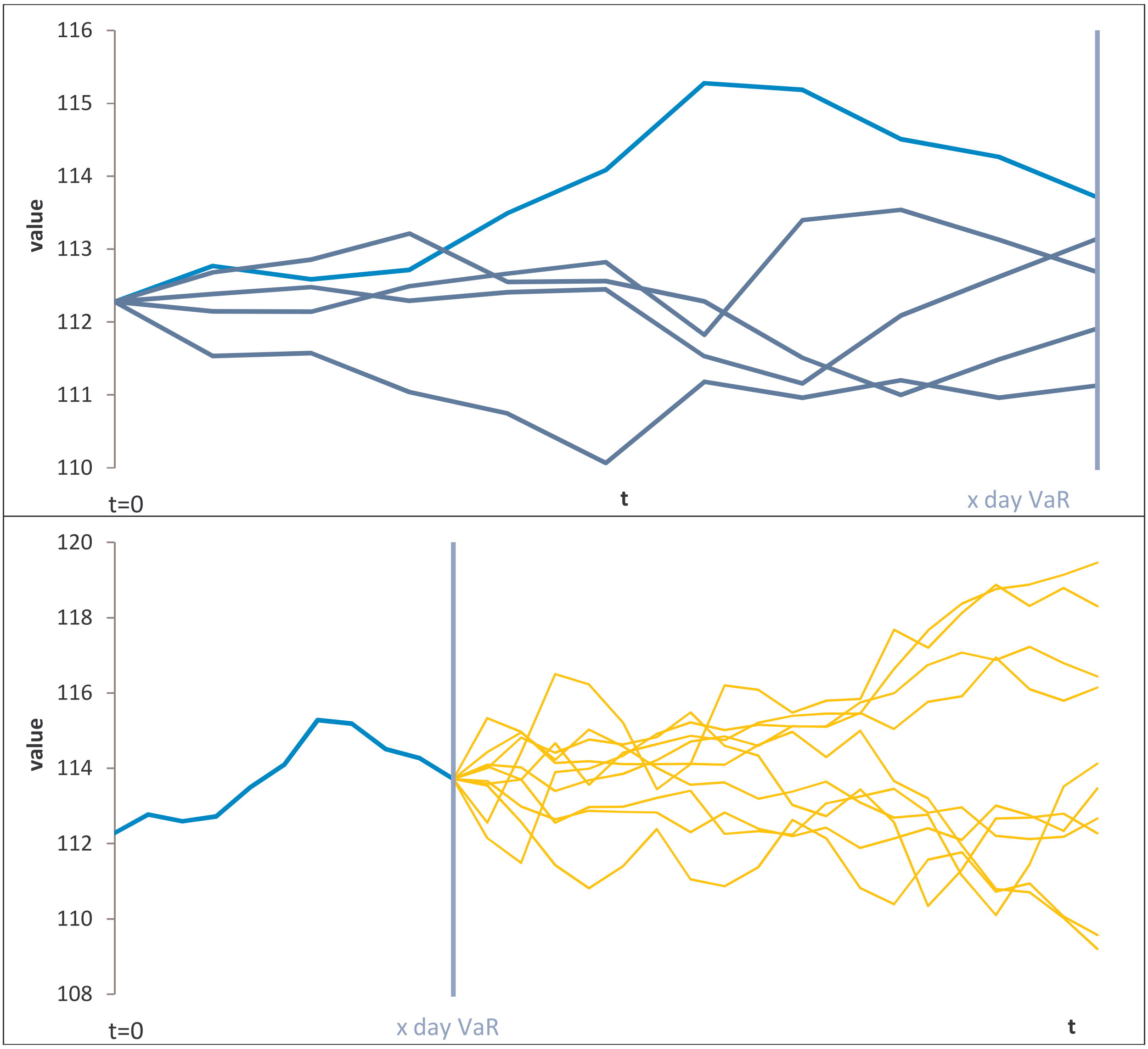
This number seems more than sufficient for the VaR evaluations. We even choose to reduce the number of option pairs we use for the 1 day and the 10 day VaR estimations in order to reduce computational time. For the 1 day VaR we draw 5,092 option pairs at random from the total number of option pairs available. For the 10 day VaR we do the same for 4,386 option pairs.

For the VaR forecast on an option, we start out with estimating the parameters for the first option of the option pair. As we already have the parameter estimates for all the days in our dataset for the different models we continue with the simulation of price paths starting at the day of the first option up until *x* days from now, the day the second option from the option pair was traded. Then the option price for the second option is estimated with the parameter estimates of the date of the first option. The parameter estimates include the parameter estimates for the Svensson model. So every price path from the first simulation now has an accompanying option price. Those option prices are all estimations of the value of the original option, *x* days later. Finally, the Value at Risk is estimated over the those estimated option prices. As the dataset also contains the value of both the options in the option pairs, it is easy to evaluate whether the observed option price of the second option lies below the VaR estimate of its value.

For every option pair, there are 2,500 price path simulations and for each of the resulting 2,500 starting prices, the corresponding option price is calculated with 1,000 simulations. This, of course, is not the case for the constant volatility model. As this model has an analytical solution to the option price the (nested) 1,000 simulations are not necessary.

Note that we assume the same term structure for both of the options in the option pairs, even though there are *x* days between them. The risk free rate is kept constant over time. This is a deliberate choice, another possibility is to use a stochastic model for the term structure but this will increase the number of necessary simulations in order to get a converged option price estimate.

Thus, a VaR forecast on an option requires a nested simulation procedure. Figure 17 shows a graphical representation of the procedure. For each of the models, the constant volatility model, the GARCH model and the MSM model we start out simulating price paths that have a length of *x* days. This is shown in the first graph.



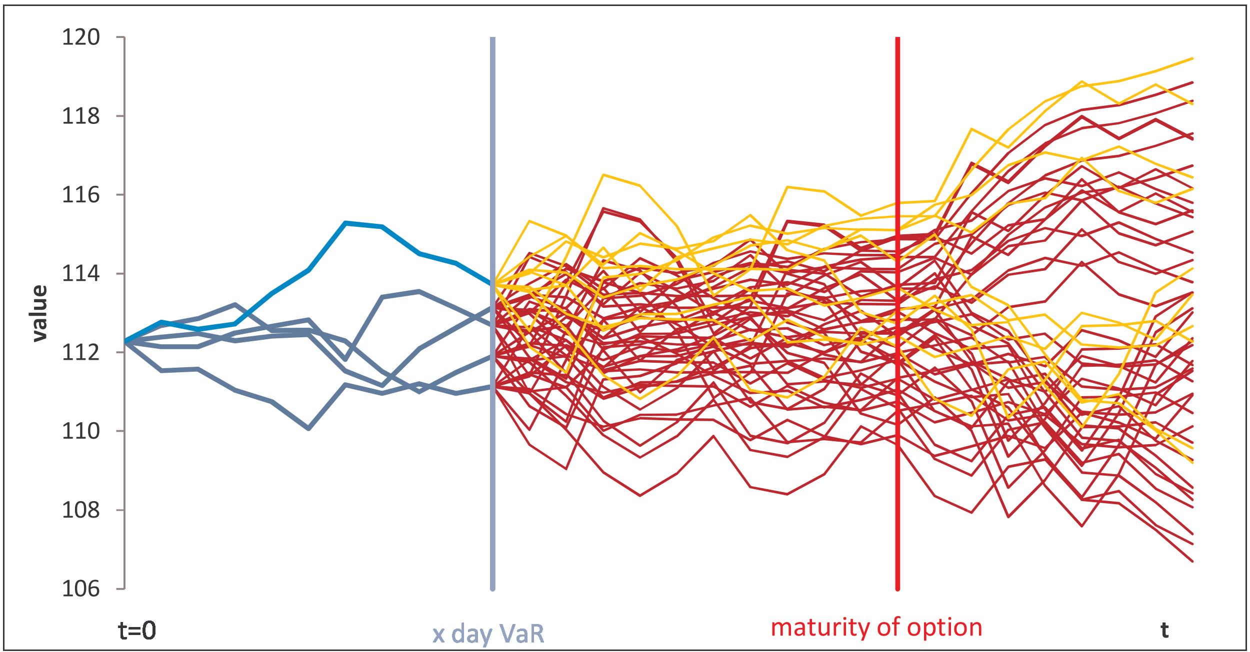


Figure 17 – Illustration on nested VaR estimation procedure on options.

The second graph shows one of the price paths from the first graph. With the simulated value of the underlying as a starting price, risk neutral price paths are simulated that are used to price the option under the assumption of the simulated starting price. The second graph results in one option price for the option *x* days from now. The third graph shows the same we see in the second graph, but now for all original price paths from the first graph. This results in option prices for all of the paths from the first graph. All what's left now is calculating the proper percentile in order to obtain the VaR forecast.

Table 18 displays the results for the VaR forecast on option prices.

x day VaR	Model	90% VaR	95% VaR	99% VaR
1	Constant volatility	0.203	0.152	0.087
	GARCH	0.315	0.272	0.216
	MSM	0.179	0.147	0.086
10	Constant volatility	0.017	0.005	0.001
	GARCH	0.063	0.031	0.008
	MSM	0.034	0.016	0.002
50	Constant volatility	0.003	0.000	0.000
	GARCH	0.017	0.007	0.004
	MSM	0.014	0.002	0.000

Table 18 – VaR forecast results.

The VaR forecasts are far off for all models for all confidence intervals. For the 1 day VaR we observe far too many violations of the VaR forecasts and for the 10 and 50 day VaR we observe far too few. It seems impossible to make good VaR estimates for option prices with the models covered in this thesis. We did not expect that the constant volatility model would perform well as the assumption on constant volatility just does not hold. A possible explanation for the failure of the GARCH and MSM model is that in Table and Table we see that the GARCH and the MSM parameters have a rather high standard deviation, which indicates that the parameters change considerably over time. Changing parameters makes forecasting very difficult. Finally, we can conclude that the GARCH and the MSM model used in this thesis are not capable of producing accurate VaR forecasts for options and that there is still enough room for improvement in their option pricing performance.

5. Discussion

We can conclude that the models in the form described in this thesis are not the 'final volatility models' the financial world is searching for. These models are not the final answers to all who are in desperate need of good volatility forecasts. Both the GARCH and the MSM model were unable to produce reasonable long term volatility forecast and the question arises whether it is even possible to forecast volatility over a long period.

But we also made some promising, positive observations to hold on to. The MSM models has a ACF that comes really close to the ACF of the time series we try to replicate. The proper reproduction of the ACF of the original series is thought to be one of the most important steps in forecasting volatility over longer periods. As the MSM model seems to cover this property better than the GARCH model does, it suggests that the MSM model could be superior in long term forecasting. So this ACF would be a great starting point for future research. From there on, the MSM model needs improvements in replicating the other phenomena we witness in financial time series.

The properties the MSM model badly failed on replicating from the original time series were the fat tails and the skewness of the returns. The simulated returns series from the MSM model were not skewed and not fat tailed. Even though researchers claim that the MSM model is able to produce fat tails and negative skewness even with a Gaussian distribution for the tails[27], this claim definitely does not hold for all time series we try to model. We believe that improving the MSM model would start here, by introducing another distributions for the return, where we leave the volatility component of the returns of the MSM model as it is right now because this volatility component seems promising. Instead of using a Gaussian distribution we could look at a t distribution with fat tails, or even a skewed t distribution.

The results for GARCH models with skewed distributions look very promising[28]. The skewed t-distribution could be used in simulating returns with the MSM model in a similar way, or even in the estimation of the MSM model. Estimating the MSM model and combining the resulting volatility paths with skewed-t draws for the returns, or estimating the MSM model with a skewed-t probability distribution (where we have to estimate the probability function as well) seems a very interesting direction of future research. Especially when we focus on estimating the model from historical volatility.

For the GARCH model there already exist a lot of extensions and additions that have proven to produce better volatility forecasts. For example the Markov-Switching GARCH model. But the number of parameters of those extensions to GARCH increase and the estimation procedure becomes more difficult.

[27] Calvet, L. E. and Fisher , A.J. 2008.
[28] Hu, W. 2005.

6. Conclusion

In this thesis we compared the popular GARCH model with the alternative MSM model, which are based on different assumptions. We started out with looking at some of the properties of the models and at their ability to replicate the stylized facts we witness in financial time series. This part was based on estimations on historical volatility.

The autocorrelation function of the squared and absolute return of the GARCH model with leverage term did not follow the same pattern as the autocorrelation function of the squared and absolute observed return does. Furthermore, it showed that the MSM model by far does not produce returns that are fat tailed and skewed enough. The GARCH model, just like the MSM model with a Gaussian distribution for the returns, also does not produce returns that are fat tailed and skewed enough but this model was less far off. These properties have of course large influences on the forecasting performances of the models.

When we look at the performances of the models in forecasting the volatility of a stock index, we observe that the GARCH model performs way better than the MSM model. Again, this is probably a direct result from the lack of fat tails and skewness in the return series produced by the MSM model.
We continued with looking at the ability of the models to forecast option prices. To get there we started out with making parameter estimations from observed option prices. With the resulting parameters we made volatility forecasts and constructed return series.

The GARCH model performed slightly better in the out of sample test on the valuation of option prices, where both the GARCH and the MSM model dominate the constant volatility model. All the models have a lot of difficulties with valuating options that are far out-of-the-money. They do a better job in valuating in-the-money options.

For the evaluation of the volatility forecast for expected volatility, trying to point out the model that performed best and worse seemed useless as all the models were way too far off to be ever useful to anybody. As the expected volatility changes a lot over time it seems very difficult, even if we would have access to the 'true volatility model', to make a volatility forecasts with a long or even medium horizon. Consequently, the VaR performance of the models for options was very poor.

In final conclusion we can state that the performance of the models we covered in this thesis is not sufficient. Especially the MSM model had a lot of difficulties with forecasting volatility. The GARCH model showed having difficulties with longer forecasting horizons. In forecasting option prices both the volatility models failed to produce reasonable forecasts but we can attribute this to the instable parameters of the models for the expected volatility. On the other hand we see a lot of room for improvement of the MSM model. The modeling of the ACF of transformations of the observed return series comes real close to reality and this should be a great starting point for further research.

References

Baxter, M. and Rennie, A. 1996. Financial Calculus; An introduction to derivative pricing. *Cambridge University Press.*

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307-27.

Bollerslev, T. 2008. Glossary to ARCH (GARCH). *CREATES Research Papers* 2008-49.

Calvet, L. E. and Fisher , A.J. 2002. Regime Switching and the Estimation of the Multifractal Processes.

Calvet, L. E. and Fisher , A.J. 2008. Multifractal Volatility: Theory, Forecasting, and Pricing. *Academic Press Advanced Finance Series.*

Fama, E. 1965. The behavior of stock market prices. *Journal of Business* 38: 34-105.

Galluccio, S., Calderelli, G., Marsili, M. and Zhang, Y.C. 1997. Scaling in currency exchange. *Physica A* 245: 423-36.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *American Finance Association* 48(5):1779-1801.

Gourieroux, C. 1997. ARCH Models and Financial Applications. *Springer-Verlag*.

Gurkaynak, R. S., Sack, B. and Wright, J. H. 2006. The U.S. Treasury yield curve: 1961 to the present.

Hu, W. 2005. Calibration of Multivariate Generalized Hyperbolic Distributions Using the EM Algorithm, with Applications in Risk Management, Portfolio Optimization and Portfolio Credit Risk.

Hull, J. 2006. Options, Futures and Other Derivatives.

Lehar, A., Scheicher, M. and Schittenkopf, C. 2002. GARCH vs Stochastic Volatility: Option Pricing and Risk Management. *Journal of Banking and Finance.*

Malmini, R. 2007. Scale invariance in financial time series. *Economics Bulletin* 24: 1-7.

Mandelbrot, B. 1963. The variation of certain speculative prices. *Journal of Business* 36: 393-413.

Mandelbrot, B. B., Fisher, A. J. and Calvet, L. E. 1997. A Multifractal Model of Asset Returns. *Cowles Foundation Discussion Paper No. 1164. Sauder School of Business Working Paper.*

McMillan, D. G. and Speight, A. E. H. 2004. Daily volatility forecasts: reassessing the performance of GARCH models. *Journal of Forecasting* 23: 449-460.

Richards, G. 2000. The fractal structure of exchange rates: measurement and forecasting. *Journal of International Financial Markets, Institutions and Money* 10: 163-80.

Ross, S. M. 2002. Simulation, Third Edition. *Academic Press Inc.*

Yamasaki, K., Wang, F., Havlin, S. and Stanley, H. E. 2006. Scaling and memory in volatility return intervals in financial markets. *Physical Review E* 73.

Appendix

A. The Black Scholes option pricing formula

'The Black-Scholes stock model assumes that the stock drift and stock volatility are constant'[29]. The stock price under the risk neutral measure follows a geometric Brownian motion with the risk free interest rate as drift and volatility σ.

$$dS_t = S_t(r_t d_t + \sigma W_t)$$

With $r_t$the risk free interest rate for a zero coupon bond with maturity t and where W_t is a Wiener process . The price of a European option under black-scholes depends besides the risk free rate and the volatility on the value of the underlying S_0, on the strike price of the option K and on the time to maturity t.

The price of an European call option is given by

$$C_{BS} = S_0 N(d_1) - Ke^{r_t t} N(d_2).$$

where $$d_{1=} \frac{\log\left(\frac{S_0}{K}\right) + \left(r_t + \frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}}$$ and

$$d_2 = d_1 - \sigma\sqrt{t}.$$

And the price of an European put option is given by

$$P_{BS} = C_{BS} + Ke^{-r_t t} - S_0.$$

B. The ARCH model

The simplest and most basic model in the GARCH family is the ARCH(q) model (Engle, 1982):

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t = z_t\sqrt{h_t}$$
$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2$$

Where h_t is the conditional volatility , z_t~N(0,1) ε_t is the (random) innovation term. All $\alpha's$ should be non-negative to ensure that $h_t \geq 0$ for all t. q stands for the number of lags that are included. When all $\alpha's$ are equal to 0, the process is homoskedastic. Remark, ARCH models cannot capture empirical ACF properties of the squared returns as the ACF decays to quick. Another drawback is that q often needs to be large.

[29] For the equations on the Black Scholes option pricing formula and an extensive derivation we refer to; 'Baxter, M. and Rennie, A. 1996.'

C. Bayes' Rule

Bayes rule states that we can write the conditional probability of A given B as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where $P(A|B)$ is the conditional probability of A given B, also called the posterior. $P(A)$ is the marginal or prior probability. $P(B)$ is the marginal probability of B and $P(B|A)$ is the conditional probability of B given A.

In this thesis $P(A)$ is our prior believe on the volatility prior to consult the data (at time t). B is the data and $P(A|B)$ is our new (posterior) believe on the volatility, after we combined our prior believes with the observed data.

D. Power law

A power law is described as $y = ax^k$. The most important property of power laws is that they are scale invariant. Given the relation $f(x) = ax^k$, scaling the argument x by a constant factor c causes only a proportionate scaling of the function itself, $f(cx) = a(cx)^k = c^k f(x) \propto f(x)$, where the symbol $\propto$ means 'is proportional to'. To round up, scaling by a constant multiplies the original power-law relation by the constant c^k.

E. A Control Variate

Suppose we want to estimate $\theta = E[X]$, where *X* is the output of a simulation. Now suppose there is a second variable Y for which the expected value is known, say $E[Y] = \mu_y$. The for any constant c the quantity $X + c(Y - \mu_y)$ is also an unbiased estimator of θ[30].

The optimal value for *c* is c^*, that minimizes the variance of $X + c(Y - \mu_y)$. This happens when

$$c^* = \frac{Cov(X,Y)}{Var(Y)}$$

For this value of c, the variance of the estimator is

$$Var\left(X + c^*(Y - \mu_y)\right) = Var(X) - \frac{[Cov(X,Y)]^2}{Var(Y)}$$

The variable Y is called the control variate. When *X* and *Y* are positively correlated, c^* is negative, and vice versa. This procedure reduces the variance because when we get a large value of *Y* out of the simulation, *Y* larger than its mean μ_y, than it is natural to assume that *X* might also be larger than its

[30] Ross, S. M. 2002.

mean θ (when we assume a positive relationship). Therefore we correct for this by lowering the value of *X,* since c^* has the opposite sign of the relationship between *X* and *Y*.

In this thesis we used the control variate technique to lower the variance of the estimated option prices. From the simulation we obtained different GARCH (and MSM) price paths from which we used to calculate the option prices according to those price paths. So the GARCH (MSM) option price is our *X* variable.

In order to reduce variance, we used the same random numbers used to construct the GARCH (and MSM) price paths to construct Black-Scholes price paths. For the Black-Scholes price paths we also calculate an option price. But for the Black-Scholes option price, there is also an analytical solution available, which is the known expectation of the option price from the simulated Black-Scholes price paths. So we have a perfect candidate for the control variable *Y*.
We estimate the GARCH (MSM) option prices with reduced variance as

$$p_{GARCHsim} + c^*(p_{BSsim} - p_{BSanalytical})$$

With $p_{GARCHsim}$ and p_{BSsim} the option prices from the simulated GARCH and Black-Scholes price paths, $p_{BSanalytical}$ as the analytical Black-Scholes price and $c^* = \frac{Cov(p_{GARCHsim}, p_{BSsim})}{Var(p_{BSsim})}$.

F. The Svensson model

The Svensson model (Svensson 1994) is a popular model to estimate the term structure. The model is an extension to the Nelson-Siegel model (Nelson and Siegel, 1987). The great benefit of Nelson and Siegel type models is that they can be used to estimate the yield for maturities for which aren't any observations available. Those estimates proof to be very accurate as well. In comparison with the Nelson-Siegel model The Svensson model contains an extra hump-shape factor with a separate decay parameter (so the model has six variables instead of the usual four). The result is a model with a better fit that has the ability to take on more shapes than the Nelson-Siegel model.

In the Svensson model, the zero-coupon yield for maturity *t* is estimated by

$$y(t) = \beta_0 + \beta_1 \left(\frac{1 - exp\left(\frac{-t}{\tau_1}\right)}{\frac{t}{\tau_1}} \right) + \beta_2 \left(\frac{1 - exp\left(\frac{-t}{\tau_1}\right)}{\frac{t}{\tau_1}} - exp\left(\frac{-t}{\tau_1}\right) \right) + \beta_3 \left(\frac{1 - exp\left(\frac{-t}{\tau_2}\right)}{\frac{t}{\tau_2}} - exp\left(\frac{-t}{\tau_2}\right) \right).$$

Where β_0, β_1, β_2, β_3, τ_1 and τ_2 are the Svensson-parameters.