Running Head: TIME SERIES VOLATILITY FORECASTING

Time Series Volatility Forecasting Using Linear Regression and GARCH

Mario Situm

February 2010

Abstract

Returns and volatilities of prices are very important parameters in finance. In many cases it is interesting to predict future developments of volatilities. Theory and practice provide different methods and techniques for this problem. This paper uses a linear regression and a GARCH(1,1) model to predict volatility for data of the Dow Jones Industrial Average for the period between 6th April 2004 and 4th December 2009. The investigations aim to analyze, how the volatility changed before and after financial crisis and if the mentioned models are able to be used under circumstances of higher volatility.

Introduction

The stock markets had a turbulent time due to the so called financial crisis. This one had remarkable impacts on many economies and several surprising news about facts none ever thought they would occur. The CNN (2008) provides an interesting article about the time line of the financial crisis. Within this article the first indicator for financial crisis was the 6th April 2007, where HSBC announced losses linked to U.S. subprime mortgages. From many other articles it is known that the crisis had its origin in earlier occurrences, which had not been taken for serious. Several well established banks and other companies were able to hide faulty business relations and the majority of the people did not know anything about it. This is more than interesting as this would totally disagree with the assumptions of strong efficient markets. Even if the market participants were able to hide information, the markets should have – based on this theory – reflect them immediately in the market prices of stocks.

In such situations we always seek for answers and it is once again time for empirical research. This is also the starting point for this paper. The focus is driven on volatility, where several definitions can be found. They will be highlighted shortly in this paper and some of them will be used for the empirical research concerning the Dow Jones Industrial Average. The literature review contains some chosen papers using different methods to test and explain occurrences of volatility. After that a short theoretical description of the GARCH model is given, as this technique is used for the empirical part of this paper.

Based on obtained data from 6th April 2004 till 4th December 2009 for the Dow Jones Industrial Average some statistical test concerning the assumption of normal distribution and significance of certain variables are made. After that a discriminant analysis is applied in order to test, whether the indicated start of the financial crisis (6th April 2007) is well chosen. The core section compares two methods of time series volatility forecasts concerning their ability to predict future volatility. As last section the paper closes with a summary and a short conclusion.

Definitions of volatility

First of all it must be defined, how volatility is measured and which types of volatility have to be considered. The most known measure of volatility is the standard deviation or variance computed from a set of observations of returns. The obtained parameter is also called the historical volatility (Abken/Nandi 1996, 23). If we assume an asset with daily prices, we can compute the daily return from one day to another. The simplest way of calculation is shown in equation 1 (Abken/Nandi 1996, 23).

$$r_t = \frac{(P_t - P_{t-1})}{P_{t-1}} \tag{1}$$

The return of the asset is the difference of the price today minus the price of the previous day divided by the price of the previous day. Another possible formula for obtaining the return assumes the link between normally distributed continuously compounded returns and the lognormality of assets (McDonald 2006, 593). The continuously compounded returns from t-1 to t are defined in equation 2.

$$r_t = \ln(P_t / P_{t-1}) \tag{2}$$

We can derive the actual price of the asset, if we rearrange equation 2.

$$\begin{aligned} \exp(r_t) &= P_t / P_{t-1} \\ P_t &= P_{t-1} \exp(r_t) \end{aligned} \tag{3}$$

The last expression in equation 3 defines, that if continuously compounded returns are normally distributed, the asset prices are lognormally distributed (McDonald 2006, 593). Knowing how to derive the returns from prices of assets, we can next compute the mean or average return for a certain period.

$$\tilde{r} = \frac{\sum_{t=1}^{T} r_t}{T} \tag{4}$$

T denotes the number of observations for the returns. The standard deviation is the square root of the variance computed with equation 5.

$$\sigma^2 = \frac{\sum_{t=1}^{T}(r_t - \widetilde{r})^2}{T-1}$$
$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{t=1}^{T}(r_t - \widetilde{r})^2}{T-1}} \tag{5}$$

Abkin and Nandi (1996) explain that historical volatility is an inaccurate estimate for the future volatility. Future volatility is unobservable and can differ significantly from historical volatility. This depends on the fact that returns are unpredictable. From observations one can derive a large number of extreme values (so called fat tails). Additionally extreme and quiet periods are clustered in time, so that we are faced with volatility clustering (Engle 2004, 407).

This leads to the discussion of conditional or predictive volatility. Based on Engle (2004) volatility is changing over time. Historical volatility is the simplest approach of reporting volatilities. The change of volatility over time is uncertain. This effect is called heteroskedasticity. Researches assume a persistence in variance. Past volatilities are able to explain current volatilities. In order to recognize the effect, researchers compute the degree to which conditional variance is persistent (Lamoureux/Lastrapes 1990, 225). The conditional variance is a linear dependence on past behaviors of squared errors.

Following Lamoureux/Lastrapes (1990) the conditional variance is the sum of the unconditional variance and several volatility shocks shown in equation 6. Commonly known models are ARCH or GARCH.

$$h_t = \sigma^2 + \alpha[\mathrm{v}_{t\text{-}1} + \lambda v_{t-2} + \ldots] \tag{6}$$

The last type of volatility to be explained here is the implied volatility. Implied volatility is directly observable form market data without any modeling assumptions on the processes involved. This is the main difference to the conditional variance of return, which is not directly observable and has to be filtered out from price data with conditional volatility models (Cond/da Fonseca 2002, 47). In many cases the implied volatility is computed with the Black-Scholes formula or with an implied tree approach (Dumas/Fleming/Whaley 1998, 2061 -2066). This aspect is highlighted later within this paper.

Models of volatility forecasting

There are many methods used to compute and predict volatilities. Poon and Granger (2003) wrote an article, where they summarize 93 papers on forecasting methods for volatilities. This summary names the authors, the assets, the data period, the data frequency and the forecasting method. Based on these information the literature distinguishes between following models in volatility forecasting (Poon/Granger 2003, pp. 482 – 486):

Times series volatility forecasting models:

The prediction of future volatility is based on past standard deviations. Such models assume that standard deviation in t-1 can be used as a forecast for the standard deviation in t. Well known models are the historical average method, the simple moving average method, the exponential smoothing method or the exponentially weighted moving average method. Beside of this simple regression methods express volatility as a function of its past volatility and an error term. When past volatility errors are considered models like ARMA, ARIMA or ARFIMA are used.

Another group of time series models are ARCH class conditional volatility models. ARCH is the abbreviation for autoregressive conditional heteroskedastic and formulates a conditional variance with a maximum likelihood procedure. Bollerslev, Engle and Nelson (1994) provide an impressive overview about different ARCH models. The big advantage of the ARCH model is the distinction between the conditional and the unconditional second order moments. The unconditional covariance matrix is time invariant, whereas the conditional variances and covariances mostly depend on the past states of the world. It must be remarked that ARCH models can be subsumed under the so called deterministic volatility models. The simplest idea assumes volatility to depend on its past in such a way that future volatility can be perfectly predicted from its history and possibly other observable information. As shown in equation 7 the future volatility depends on a constant and a constant proportion of the last period's volatility (Abken/Nandi 1996, 24 – 25).

$$\sigma_{t+1}^2 = \theta + \kappa\sigma_t^2 \tag{7}$$

The last group of times series are stochastic volatility models. Such models assume that a derivative asset with a price is depending on the price of the underlying asset and its instantaneous variance. The price of the underlying asset obeys a stochastic process, which also includes a Wiener process. In such a model the instantaneous volatility can change randomly (Hull/White 1987, 282 and McDonald 2006, 744). This also means that volatility is driven by a random source, which is different from the random source driving the asset returns process. Most stochastic volatility processes assume a mean reversion in volatility (Abken/Nandi 1996, 29 – 30). Two well know deterministic volatility models are those of Heston and Hull-White. Both models are characterized by its mean-reversion rate, its volatility and its correlation with the underlying asset (Balland 2002, 31).

$$\mathrm{dv_t} = \left(\alpha_t - \kappa v_t\right)\mathrm{dt} + \xi\sqrt{\mathrm{v_t}}\,\mathrm{dB_t} \qquad \text{(Heston)} \tag{8}$$

$$\frac{\mathrm{dv_t}}{\mathrm{v_t}} = \alpha_t\,\mathrm{dt} + \xi\,\mathrm{dB_t} \qquad \text{(Hull - White)} \tag{9}$$

The Process B is a Brownian motion with correlation defined under risk-neutral measure. All of the other parameters are implied by calibration to the initial smile surface. One can calibrate them from historical data, whereas these data are typically much lower. In many cases implied volatilities are obtained with the Black-Scholes model (Balland 2002, 31 – 33).

Options based volatility forecasts:

The assumptions behind this model are closely related to the Black-Scholes formula. The option price at a certain time t is a function of the price of the underlying asset, the strike price, the risk-free interest rate, the time to maturity and the volatility of the underlying asset over the period from t till maturity. With a backwardation technique one can compute the standard deviation, which is used from the market as input for the pricing. We obtain the implied volatility. Due to different strike prices and maturities we derive different implied volatilities. These volatilities can be plotted and we receive volatility smiles, smirks and other nonlinear shaped curves. There are several assumptions behind the Black-Scholes formula, which are discussed in a another part of this paper.

Review of literature in the field of volatility

There are many papers examining the effects of price movements on volatility. Within this paper some articles with interesting results are summarized. Engle (2004) examined the S&P 500 composite index for daily levels from 1963 till November 2003. On time series plots the crash of October 1987 left a remarkable shape in the movement of the daily prices. At that time extreme returns appeared. Prices were falling and the volatility increased. The analysis of the kurtosis over the full observation period shows a dramatically high value. This indicates that extremes were more substantial, than the normal distribution would expect from the random variable. Engle applied a GARCH(1,1) using weights of unconditional variance, previous forecasts and news measured as square of yesterday's return. The result was that the bulk of information comes from the previous day forecast. The long-run average variance has a very small effect, whereas new information produces a small change.

Macbeth and Merville (1980) analyzed the daily closing prices of all call options traded on the Chicago Board of Trade Options Exchange for six selected companies from 31st December 1975 to 31st December 1976. They use a stochastic differential equation to measure the difference between the model price and the market price. For this they use a Cox-model and a Black-Scholes model. They recognize a superiority of the Cox-model, as the Black-Scholes model underprices in the money options and overvalues out of the money options.

Additionally the Black-Scholes model does not always overprice out of the money options, whereas the Cox-model underprices such options more often. Therefore the Black-Scholes model predicts a greater implied volatility as the true value is. Summarized it can be said that the Cox-model values option closer to market prices.

Wei (1997) analyzed a European call option on a 15-year bond, which pays an annual coupon of 10 % on a face value of US$ 100. The exercise price of the option is US$ 100 and matures in 5 years. He computes two sets of option prices at different interest rate levels. One way is to use Jamshidian´s approach (Jamshidian 1989). This one delivers a solution for European options on default free bonds assuming a term structure completely determined by the value of the instantaneous interest rate, which is following a mean reverting Gaussian (normal) process as in Vasicek. The second set used a simple approach based on a process followed by a discount bond. In the next step the author computes the accurate and approximate option prices, interest rate deltas, price deltas and gammas with the Vasicek model and the Cox-Ingersoll-Ross (CIR) model. With the simple approach the result shows an overpricing of the option when the bond value is above the exercise price. An underpricing occurs when the bond value is below the exercise price. For the purpose of this paper the result concerning pricing errors versus interest rate and volatility are interesting. For volatilities of about 0.01 at the Vasicek model the pricing error for out-of-the money options is within 2 %. For in-the-money options the error is about 2 %. Higher volatilities for out-of the money options cause a price error of 2 % and for in the-money options a lower value of 1 %. In the CIR model the pricing error for out-of the money options lies at 2 % and for in-the money options it is close to 1 %. The general conclusion is that higher volatility delivers a smaller percentage pricing error. These results show that there is a connection between estimating volatilities and the pricing of bond prices affecting the pricing error.

Duan (1995) introduces a GARCH option pricing model. The idea behind this model results form the incorrect assumption of the homoskedasticity for volatilities under the Black-

Scholes model. As explained before the GARCH model uses the heteroskedastic approach, which means a change of the volatility with time (Engle 2001, 157). Therefore a GARCH option pricing model seems to be closer to reality than the Black-Scholes model. The author uses S&P 100 index calls of European style from 2nd January 1986 to 15th December 1989. The GARCH model is fitted to the observed daily data of the index. It delivers interesting aspects for delta hedging. In times of low variance delta hedging calls for smaller option positions relative to the Black-Scholes model. For high-variance states it demands larger positions. Using the GARCH model can explain some systematic biases of the Black-Scholes model. It allows asymmetrical or skewed lower ends, which is consistent with the empirical evidence of implied volatility smiles.

Szakmary et al. (2003) are using data from 35 futures options markets from eight different exchanges to test the implied volatility as predictor for the realized volatility. For this a regression test and a GARCH model are applied. Using the regression analysis they conclude that implied volatility is stationary for most of the observed series. Even if implied volatility is a biased forecast for the realized volatility, it is a better predictor than historical volatility. Despite of the expectation the GARCH forecasts are generally not superior as a supplemental predictor of realized volatility. The implied volatility is therefore an interesting alternative to many other models and can also be used as predictor for future volatility.

Fleming (1998) tested the quality of market volatility forecasts of implied volatilities by using S&P 100 index option prices. For deriving the implied volatilities he uses the Black-Scholes model. The research hypotheses states that implied volatility should represent an unbiased forecast, whose error is orthogonal to the market's information set. The idea behind this hypotheses believes on the fact that implied volatility should incorporate the information conveyed by the historical volatility and any additional relevant information. From the empirical results he recognizes that implied volatilities for S&P100 call and put options are biased forecasts. The deviations could be interpreted as market inefficiencies, but can also stem from misspecifications in the option valuation model. Nevertheless the implied volatility can be useful as an index or market sentiment, as a link to ex ante market volatility and contains a significant relation to expected returns.

Schwert (1990) examines the stock volatility and the crash 1987. October 1987 was the largest percentage change in market value for a observation period of 29,000 days. The aim of the paper is to test several models and their behavior concerning extreme values of returns. It is remarkable that stock market volatility returned to low pre-crash levels quickly. This conclusion is supported by evidence from options and futures markets. Negative returns lead to larger increases in volatility than for positive returns. Interestingly many reversals are observable. When large drops in stock prices occur, then large increases in stock prices are following.

Dumas, Fleming and Whaley (1998) used implied volatility functions for empirical tests for the S&P 500 index options (European style expiring on the third Friday of the contract month) traded on the CBOE over the period June 19888 through December 1993. They also remark the problems of the Black-Scholes formula, when implying volatilities from reported option prices. Estimates vary systematically across exercise prices and times to expiration. Nevertheless they argue that hedge ratios determined by the Black-Scholes model are more reliable than those obtained from deterministic volatility functions option valuation models.

In sum the research about volatility uses many approaches, where weaknesses and advantages occur. The importance of volatility as a parameter in finance is increasing and it will be necessary to improve the possibilities of forecasting.

Implied volatility and the Black-Scholes Equation

From the previous sections it is known that implied volatilities can be obtained from option prices using the Black-Scholes formula or another option pricing model. This section is used to give a short overview and discuss the advantages and drawbacks of the method. Implied volatilities will not be computed within this work as we are interested in time series models. McDonald (2006) explains the process of computing implied volatilities. First the market prices for an option must be observed. Here we must distinguish between a call and put option. After that an option pricing model has to be chosen, which is able to infer volatility. Volatility is a parameter of the Black-Scholes formula, which determines the price of the option for a certain underlying asset. Computing implied volatility explains the observed prices.

Based on Lee (2001) the Black-Scholes formula for a European-style call option with a strike price K and a date of expiration T can be written as in equation 10.

$$\begin{aligned} C^{BS}(X,t,K,T,\sigma) &:= \mathrm{X}\,\mathrm{N}(\mathrm{d}_1) - \mathrm{K}\,\mathrm{N}(\mathrm{d}_2) \\ \mathrm{d}_{1,2} &:= \frac{\log\left(\frac{X}{K}\right)}{\sigma\sqrt{T-t}} \pm \frac{\sigma\sqrt{T-t}}{2} \end{aligned} \tag{10}$$

The option price C(K,T) is observable from market data. Instead of the constant volatility the implied volatility is defined as I(K,T). Now the actual price of the option is compared with the modified Black-Scholes formula containing the implied volatility.

$$C(K,T) = C^{BS}(X,t,K,T,I(K,T)) \tag{11}$$

This aspect contains a main assumption about the Black-Scholes formula. Equation 11 can not be solved directly for the implied volatility I(K,T). It is necessary to use an iterative procedure to solve the equation. It must be emphasized once again that other pricing models could be used to compute the implied volatility. The Black-Scholes implied volatilities are frequently used as benchmarks (McDonald 2006, 400 and Soczo 2003, 205).

The estimation of implied volatility has its weaknesses. The changes in the implied volatilities are to a certain degree results of pricing errors in the options. Using estimates of the actual variance therefore requires a large amount of data (Hull/White 1987, 292 – 293). Another aspect is the assumption of a non-flat instantaneous profile of the implied volatility surface with the Black-Scholes model. A implied volatility surface contains the call and put prices of a certain date for different maturities and moneyness. The shape of the surfaces is changing over time (Cont/da Fonseca 2002, 45 – 47). A main problem using the backwardation of the Black-Scholes models can be explained by two errors. The specification error explains the difference between the market price of an option and the valuation using a model. Bid-ask spreads and infrequent trading among index stock cause differences between the observed and the theoretical value – this is called the estimation error (Fleming 1998, 321).

There is another thing, which must be emphasized. The Black-Scholes model has assumptions about the economic environment like a known constant risk-free rate, no transaction costs and taxes, the possibility to short-sell costless and to borrow at the risk-free rate (McDonald 2006, 379). This explains why trading strategies based on these assumption can produce positive and significant abnormal profits (Amin/Morton 1994, 171).

The great advantage of implied volatility is the possibility to observe it directly from market data. The local or conditional volatility is not directly observable. Using market data increases the representation of the research as they also include opinions of market participants (Cont/da Fonseca 2002, 47). Additionally implied volatility contains incremental information beyond that in past volatility. This is an important fact and also an important reason, why such models are superior to result in comparison to past volatility (Christensen/Prabhala 1998,133).

Based on Jiang and Tian (2005) it is not absolutely sure, if implied volatility is the real parameter for estimation of future volatility. The implied volatility shows no correlations with future realized volatility and does not incorporate information contained in historical volatility. In general some studies state that implied volatility is a biased forecast of future volatility. If the Black-Scholes model is correct and the market is efficient, than the implied volatility should contain all relevant information. This is due to the special assumption not the case, but many studies receive remarkable good results.

Theory of GARCH

As a GARCH(1,1) model is used in this paper a short theoretical overview about the concept is given. Bollerslev, Engle and Nelson (1994) show that GARCH is a variant of ARCH allowing infinite lags to be estimated from a small number of parameters. There is clear distinction between conditional and unconditional second order moments. Based on McDonald (2006) the general formula for the GARCH(p,q) model is shown in equation 12.

$$q_t = a_0 + \sum_{i=1}^{m} a_i \cdot \varepsilon_{t-i}^2 + \sum_{i=1}^{n} b_i . q_{t-1} \tag{12}$$

The constants of the formula have to be estimated with delivered data. Epsilon denotes the continuously compounded returns, which are computed daily. q is the conditional value of the return variance. We concentrate on the GARCH(1,1) model and therefore equation 12 can be reduced to equation 13.

$$q_t = a_0 + a_1 \cdot \varepsilon_{t-1}^2 + b_1 \cdot q_{t-1} \tag{13}$$

The notation GARCH(1,1) has to be explained in detail. The first number refers to how many autoregressive lags appear in the equation, while the second number refers to how many moving average lags are used (Engle 2001,160). In order to get the constants a maximum likelihood method is used. The basic for the computations is the probability density for the continuously compounded returns conditional on the variance (McDonald 2006, 752).

$$f(\varepsilon_t ; q_t) = \frac{1}{\sqrt{2\pi q_t}} \cdot e^{-0.5\varepsilon_t^2 / q_t} \tag{14}$$

With the maximum likelihood method the set of parameters is computed, so that they maximize the probability of observing the returns actually observed. This is one of the widest used methods for computing GARCH. The function to be maximized is shown in equation 15.

$$\sum_{i=1}^{n} \left[-0.5 \ln(q_i) - 0.5\varepsilon_i^2 / q_i \right] \tag{15}$$

With the constants of GARCH model the unconditional volatility can be computed.

$$\overline{\sigma}^2 = a_0 \cdot \sum_{i=0}^{\infty} (a_1 + b_1)^i = \frac{a_0}{1 - a_1 - b_1} \qquad (16)$$

The unconditional volatility is the level of volatility to which the volatility process reverts and is assumed to be time-invariant. The conditional variance is the sum of the unconditional variance plus the weighted difference between the squared innovation of the compounded continuous returns and the unconditional variance as shown in equation 17. The model therefore considers high volatility expectations after large price changes and low volatilities for low price changes.

$$\sigma^2 = \cdot\overline{\sigma}^2 + a_1\left(\varepsilon_{t-1}^2 - \overline{\sigma}^2\right) \qquad (17)$$

To compare the results with the historical annual volatility we must compute the unconditional annual standard deviation based on equation 18.

$$\overline{\sigma} = \sqrt{\frac{a_0}{1 - a_1 - b_1} \times 252} \qquad (18)$$

Data and observation period

For the analysis the Dow Jones Industrial Average index for the period from 6th April 2004 till 4th December 2009 is taken. The data consist of the closing prices, the highest daily prices, the lowest daily prices, the opening prices and the daily volumes traded. For the statistical analysis the daily returns are computed using equation 2. The continuously compounded version for interest rates is more suitable than the simple interest rate computation, because discrete-time involves too much variability (Garcia/Renault 1997, 3). Fama (1965) names three reasons for using changes in log prices. First the change in log price is the yield from holing a certain security. Second the variability of simple price changes for a given stock is an increasing function of the price level of the stock. The last reason is that for change less than plus/minus 15 per cent the change in log price is very close to the percentage price change. It is often convenient to look at data in term of percentage price changes.

In order to avoid the mentioned problem and to benefit form the advantages the continuous model is chosen. Additionally the daily volatility of continuously compounded returns is calculated using equation 19 (continuously compounded intra day return), 20 (compounded intra day returns) and 21 (volatility of compounded intra day returns).

$$
\begin{aligned}
r_1^{\text{intraday}} &= \ln\left(\frac{\text{Closing Price}}{\text{Opening Price}}\right) \\
r_2^{\text{intraday}} &= \ln\left(\frac{\text{Highest Daily Price}}{\text{Opening Price}}\right) \\
r_3^{\text{intraday}} &= \ln\left(\frac{\text{Lowest Daily Price}}{\text{Opening Price}}\right) \\
r_4^{\text{intraday}} &= \ln\left(\frac{\text{Closing Price}}{\text{Highest Daily Price}}\right) \\
r_5^{\text{intraday}} &= \ln\left(\frac{\text{Closing Price}}{\text{Lowest Daily Price}}\right)
\end{aligned}
\tag{19}
$$

$$
\mu_{daily} = \frac{\sum_{i=1}^{5} r_{\text{i}}^{\text{intraday}}}{5} \tag{20}
$$

$$
\sigma_{\text{daily}} = \sqrt{\frac{\sum_{i=1}^{3}\left(\mu_{\text{daily}} - r_i^{\text{intraday}}\right)^2}{5-1}} \tag{21}
$$

The data assume a person investing in the opening price can exit at the lowest price, the highest price and the closing price. Additionally the person can invest at the stage of the highest price and exit at the closing price and invest in the stage of the lowest price and exit at the closing price. Due to lack of other intraday prices that approach is chosen in this paper. A similar approach for the volatility has been applied by Schwert (1990), who calculated the sample standard deviation within each month with the formula for the standard deviation.

The banking crisis had a remarkable impact for all economies. In such cases it is always difficult to define a starting point. CNN (2008) provides a timeline for the banking crisis, where the key moments of the process are summarized. The first remarkable information in this timeline is the announcement of the HSBC concerning losses linked to U.S. subprime mortgages on 7th of February 2007. In newspapers and magazines many articles about the reasons for the banking crisis had been published. They name several explanations indicating causes for the financial crisis. Interestingly these information did not appear in the prices of markets. Here one could discuss about the efficiency of the markets, which is not the purpose of this paper. Nevertheless looking on the developments of the prices for the chosen indexes (see appendix 1) one can see that the after the 7th of February 2007 there was a increase in the prices. After a certain period the prices began to decrease. For the following analyses the 7th of February 2007 will be the cutting point between the two observed periods. This date divides the data into two groups – group one contains variables before 7th of February 2007 and group two consists of data after that time.

Statistical analysis and empirical results

Descriptive statistics and tests of normal distribution:

Appendix 1 shows the prices and daily returns for the index from 6th April 2007 till 4th December 2009. The index shows a significant drop in prices at the beginning of October 2007. In March 2009 the prices began to rise again. Looking at the daily returns one can see from July 2007 till July 2009 that they are higher than for the previous period. It must be remarked that in the period from July 2008 till July 2009 the daily returns have been at the highest level for the period of observation. The daily volatilities of prices and returns are having higher movements in the period after financial crisis, which is in relation with the movements of figure 1. This is also in congruence with the analysis of Schwert (1990) for the crash of 1987, where fall in stock prices were followed by large increases of volatility.

In appendix 2 (figure 3) the data for the descriptive statistics and the frequency distributions for the daily prices and daily returns are included. The theoretical normal curve is shown in every plot. For the daily returns it can be seen that data after financial crisis do not fulfill the requirements of the normal distribution. We are having extreme values in the tails, which are not covered with the theoretical normal distribution. This is the already mentioned problem of fat tails (Engle 2004, 407). The data concerning skewness and kurtosis also indicate deviations from the assumptions of normal distribution. In case of skewness the computed values are all different from zero, but for the period before financial crisis we find values closer to zero than for the period after financial crisis. This indicates an asymmetry of the distributions. For the positive value we have right skewed and for the negative values we have left skewed distribution. All of the frequency distributions are leptokurtic as the value of kurtosis is positive – the kurtosis of the normal distribution is zero.

The normal probability plots in appendix 3 (figure 5) show a deviation of the observed data from the expected data. For the period after financial crisis the observed data are constructing an s-curve indicating a very large deviation from normal distribution. Interestingly the data before financial crisis are much closer to the expected data, so that the deviations from normal distribution are much lower. The last test using a Kolmogorov-Smirnov (KS) test confirms the already derived results that the assumption of normal distribution must be rejected for the period after financial crisis. The most extreme difference absolute for this period is higher than the value computed from the KS table for N = 714 (0.051). Additionally the asymptotic significance (2-tailed) is lower than the value of 0.05. The most extreme difference absolute for the period before financial crisis is below 0.051 and the significance is higher than 0.05. For the data before financial crisis we can assume a normal distribution.

Figure 4 in appendix 2 shows the frequencies for the daily volatilities. Visually the data do not fit into the displayed normal curve. The skewness and the kurtosis are far away from the values for a normal distribution. Figure 6 in appendix 3 displays the probability plots for the daily volatilities. In both cases we can recognize strong deviations for the expected values, so that we can not assume a normal distribution of daily volatilities. Applying the Kolmogorov-Smirnov test shows most extreme differences absolute of 0.157 (after financial crisis) and 0.067 (before financial crisis). In both cases the values are higher than 0.051. The significance is for both periods lower than 0.05. Daily volatilities of the observed periods are not normally distributed.

One-Way ANOVA:

With the ANOVA one can test whether the means for the variables of the groups are differing significantly or by change. The null hypothesis is that the means of the variables are equal across groups. The results in appendix 4 show that the variables are different between the two groups – we must reject the null hypotheses of equal means. The variables continuously compounded daily returns and lowest daily price are having values of significance higher than 0.05. In these both cases we can not reject the null hypothesis and assume the groups are having similar means for these two variables. This implies differences between the two groups and we can for the first thought indicate that the financial crisis is the cause for this differences.

Discriminant analysis and correlations:

The discriminant analysis is a technique to derive a linear function, which is able to distinguish between the groups. A detailed description of the technique is shown in Klecka (1980) or Subhash (1996). This analysis is made in order to prove, whether the chosen date defined as the beginning of the financial crisis is valid. The obtained function contains numerical discriminants consisting of variables that are significantly different between the groups. In order to get the best possible function the Fisher stepwise method is used for the data. This also means that there is no other function, which is able to guarantee a better classification result. The discriminants are the lowest daily price, the volume, the difference between the highest and lowest daily price and the volatility of the daily price. Table 7 shows the coefficients of the canonical discriminant function for the named variable. The discriminant function can be obtained by a linear combination of the data and is shown in equation 22.

$$\begin{aligned}
&\text{Discriminant function} = 0.00052 \cdot x_1 + 1.1 \cdot 10^{-9} \cdot x_2 - 0.0029 \cdot x_3 + 0.0037 \cdot x_4 - 9.0524 \\
&x_1 = \text{Lowest Daily Price} \\
&x_2 = \text{Volume} \\
&x_3 = \text{Difference between highest and lowest daily price} \\
&x_4 = \text{Volatility of Daily Prices}
\end{aligned} \tag{22}$$

The interesting thing at this equation is the variable x_3. Based on Schwert (1990) the spread between the highest and lowest price is a five time more efficient variance estimator than the daily stock return. The appearance of this variable within the discriminant function allows assuming that the spread between lowest and highest daily price could have an important role for the further analyses. Looking at the bivariate correlations for some chosen variables between the different groups we find several significant relations marked in grey (appendix 6 – table 9). It is highly significant that the continuously compounded daily returns for both groups are uncorrelated with the other variables. In general the correlations are between + 0.05 and – 0.067. For some variables of group two there are correlations with variables of group one. As an example the lowest daily price of group two is highly negative correlated with the lowest daily price of group one (- 0.621).

We derive good results for the classification shown in appendix 5 (table 8). The computed function classifies 94.3 per cent of the original grouped cases correctly. Based on statistical theory the type I error (prediction of cases from group one into group two) is about 6.4 per cent. The type II error is also relatively small with 5 per cent. This is a remarkable result as normally one of both errors is always significantly higher. The obtained function is balancing both types of error in a suitable way. Errors at classifications are due to extreme values, as the chosen frequency distributions for group two are not normally distributed. Figure 7 shows the separate group graphs. We can not find a line, which can perfectly divide both groups without having classification errors. Based on these good results we will leave the 7th February 2007 as the cutting point between both groups.

Linear regression:

In the first step linear regressions for both groups (before and after financial crisis) have been computed using the available variables. For achieving the best regression the stepwise method is used. The results of the regressions are shown in appendix 7. Figure 8 shows the computed volatilities of daily returns and the estimated values using the regression in equation 23 for the second group.

$$\begin{aligned}&\text{Daily volatilty of returns}_{\text{group 2}} = -3.625 \cdot 10^{-13} \cdot x_1 - 1.583 \cdot 10^{-6} x_2 + 5.702 \cdot 10^{-5} \cdot x_3 + 0.017\\ &x_1 = \text{Volume}\\ &x_2 = \text{Opening price}\\ &x_3 = \text{Difference between highest and lowest daily price}\end{aligned} \tag{23}$$

As it can be seen from table 9 the residuals are almost high. This means that regression is not suitable to estimate the volatility of the daily returns, which is also visible in the figure. Table 10 provided some casewise diagnostics, where the residuals are extremely high. For group one (before financial crisis) the linear regression contains four variables as displayed in equation 24.

$$\begin{aligned}&\text{Daily volatilty of returns}_{\text{group 1}} = -8.871 \cdot 10^{-7} \cdot x_1 + 4.820 \cdot 10^{-6} \cdot x_2 - 7.243 \cdot 10^{-7} \cdot x_3 + 4.562 \cdot 10^{-5} \cdot x_4 + 0.008\\ &x_1 = \text{Differnece between closing and lowest daily price}\\ &x_2 = \text{Volatility of daily price}\\ &x_3 = \text{Opening price}\\ &x_4 = \text{Difference between highest and lowest daily price}\end{aligned} \tag{24}$$

The difference between real volatility and computed from regression is smaller than for group two. Looking on table 11 at the casewise results there are also several examples, where the residual is extremely high. Therefore this regression function is also not able to predict volatility properly. It is interesting that in both regression functions the variables "opening price" and "difference between highest and lowest daily price" are appearing. The variable "difference between highest and lowest daily price" also has an important role within the computed discriminant function.

The purpose of the planned regression is to use the linear function of equation 24 and apply the formula on the data for group two. Is the derived regression from data of group one applicable on data of group two in order to estimate the volatility of daily returns? The results are surprising in that way that the residuals for the periods of lower volatility are lower than applying equation 23 on group two. In all periods of high volatility the residuals are much higher.

This allows two conclusions. First the regression analysis is not a suitable instrument to predict future volatilities, when the prices are decreasing strongly. In these cases the volatility is high, so that the regression function can not be properly computed to integrate this deviations. Second the regression analysis seems to be an interesting alternative, when volatilities are fluctuation within a small interval. If volatility is expected to remain constant and no extreme impacts are assumed, then the regression function can be applied to a certain degree. The goodness of the function can be valued by the appearing residuals. Table 12 summarizes the main parameters for evaluating the quality of regression. The best regression appears for group one, where the minimum, maximum and the sum of the residual are the smallest. Trying to apply the regression formula obtained from data of group one on data of group two shows significant differences. If we really believe in the random walk hypothesis this result is not surprising, as under this theory the past cannot predict the future in any meaningful way. Otherwise the literature states that the random walk hypothesis is not a complete accurate description of reality. This indicates that occurrences in t+1 are not totally independent of occurrences in t. Under this premises the attempt of linear regression seems to be suitable (Fama 1965, 34). Maybe the results of prediction are not that perfect, but we obtain relevant conclusions for further research.

GARCH model:

In figure 11 the estimated GARCH volatilities for the second group – data after financial crisis – are displayed. Similar like in figure 2 we derive higher volatilities in time between September and December 2008. The formula obtained by GARCH for this section is shown in equation 25.

$$\mathrm{q}_{\mathrm{t}} = 0.000004857 + 0.040969441 \cdot \varepsilon_{t-1}^{2} + 0.922998431 \cdot \mathrm{q}_{\mathrm{t-1}} \tag{25}$$

The unconditional annual standard deviation is obtained by the unconditional daily variance estimate using equation 18. The value is 18.43 per cent – the annual historical volatility is 28.03 per cent. For the first group a separate GARCH model is computed, which is shown in equation 26. In figure 12 it can be seen that the volatility does not had such extreme movements like for the data of group two. The unconditional annual standard deviation is 9.97 per cent and the annual historical volatility is 10.08 per cent.

$$\mathrm{q}_{\mathrm{t}} = 0.000001422 + 0.040969441 \cdot \varepsilon_{t-1}^{2} + 0.922998431 \cdot \mathrm{q}_{\mathrm{t-1}} \tag{26}$$

The interesting point is that both equations are very similar. They are only differing in the first part of the formula. a_0 for group one is 3,435.E – 06 smaller than for group two. This is the purpose to use equation 26 and apply it on the data of group two. The result is shown in figure 13, where the estimated curve and the curve of the GARCH(1,1) model for group two are shown. Due to the very small difference in the models there curves are in congruence. The obtained GARCH(1,1) model can be used for GARCH volatilities of group two, even if there have been the impacts of the financial crisis. The big difference is the value of the unconditional volatility.

The GARCH(1,1) model delivers remarkable results for the prediction of volatility. Even if there have been the increased volatility due to financial crisis, the model remains stable and considers the mean reversion of the volatility in times where the price changes normalized again. In opposite to the linear regression the GARCH model delivers an unconditional volatility, which can be used to compute the conditional volatility based on equation 17. Looking on the actual volatility allows computing the volatility for the next period. The prediction of market volatility is to a certain degree possible and confirms the statement from actual literature (Poon/Granger 2005, 54). Impressive is the ability of the model to consider the distinction between conditional and unconditional variance with the effect of mean reversion.

Conclusion and summary

Within the descriptive statistics we can see that assumptions concerning normal distribution are not fulfilled for all variables. The distribution of the daily continuously compounded returns for group one (data before financial crisis) are in general normally distributed. Interestingly the daily volatilities of the continuously compounded daily returns are beyond normality. These aspects are problematic as from theoretical viewpoints the condition of normality is an essential requirement for the related models. In case of this study the lack of normality can be explained by extreme values not considered in the theoretical normal distribution and a too small sample not covering the central limit theorem.

The means of the continuously compounded daily returns and their volatilities do not significantly differ between the groups. This means that even if we can observe higher volatilities for a certain period after financial crisis, the returns and volatilities reverted to a "normal" level. This is the same result as in the paper of Schwert (1990). Periods with higher price changes result in higher volatilities. After that the price changes come back to the previous levels before financial crisis, which is also followed by lower volatilities. This is having an impact on the obtained GARCH model as applying the function derived from data of group one on data of group two delivers remarkably good estimates. There are relationships between variables and between both groups. These relationships are very important for the discriminant analysis and the linear regression model. Remarkable is the lack of correlation of the continuously compounded returns with the other variables.

The volatility of continuously compounded daily returns seems to be problematic. The data have been computed with “only” five intraday prices due to lack of additional information. We derive daily volatilities of the daily returns, which in sum (for both groups) are not normally distributed. This aspect induces enough problems for the accuracy and validity of statistical analysis. The difference between the highest and lowest daily price can be seen as a parameter of daily volatility, which is not constant over time. Due to the appearance of this variable within the discriminant analysis and the regression analysis the weaknesses of the volatilities of daily returns will not be discussed in detail.

With the results of the discriminant function we obtain a good result for the chosen date. In practice it is always difficult to define the right time of a special occurrence and its point of impact. The used date of 7th February 2007 is well chosen. In the introduction of the paper it was emphasized that there is some doubt about the strong efficiency of markets, as in the history there had been many situations, where the bad news should had been reflected in the prices. Looking at the movements of the prices and the returns such bad information can not be confirmed. This lead to the assumption of lagging information and therefore the chosen cutting point is suitable as “beginning” of the financial crisis. Clear cutting points can not be obtained due to blurred information, which are causing biases in the obtained prices. Nevertheless it is important to compute a discriminant function to get an empirical argument of the reliability. Bad results in the classifications are signals of a unsuitable cutting point.

The obtained discriminants can be best explained by the bivariate correlations. There is an interesting connection between the lowest daily price and the volume of both groups. Lowest daily prices of group one and group two are significantly correlated and the volumes and lowest daily prices of both groups are having also significant correlations. The volatility of daily prices of both groups is correlated with the volume of the other groups. One interesting point is the difference between highest and lowest daily price – a significant correlation exists between this variable of group two and the lowest daily price of group one. Additionally this variable for group one has a correlation with the volume of group two. The discriminant analysis is a method using such correlations to obtain the best linear combination, which divides both groups the best. The most remarkable thing is the variable difference between highest and lowest daily price as it also appears in both obtained linear regression formulas.

The linear regression delivers interesting insights into the behavior of intraday volatility (based on the computations of this paper). From both groups two separate regressions have been computed and in both functions two characteristic variables can be found. Due to lack of a second order moment in the formula – in opposite to the GARCH model – the power of prediction is not that good. The application of the regression formula obtained from group one for the data of group two can not be used in practice. There are only partial results, which are predicting the right intra day volatility. The problem of intraday volatility was mentioned before and the weak results can in one way surely be explained by the lack of normal distribution and second in the small sample of intraday data. A comparison of both models is here not possible as they are measuring different aspects of volatility.

As last we have the remarkable results of the GARCH(1,1) model. The estimated constants of the formula are useful predictors. The amazing thing is the stability of the formula, when estimating the GARCH volatilities for the data of the second group. A main explanations for this effect is the reversion of volatility to a normal level after a certain period of higher volatility. This result is comparable to other papers examining financial crashes. Volatility has a mean reversion, which can be anticipated very well with a GARCH model. A difference between both groups must be emphasized – the unconditional annual volatility is lower for group one. This can be explained by the higher volatility for the period after financial crisis. Such aspects are then considered in the formula using the conditional variance.

The quality of data is a very important aspect when examining volatility. It is important to have sufficient data – long time series – in order to fulfill the requirements of normality for the data. Only then the central limit theorem can be (partially) fulfilled. Nevertheless problems remain with volatility clustering and extreme values. In many papers – not mentioned in this article – authors conclude the assumption of skewed distributions with abilities to better consider these aspects. Correlations leave some doubt about the concept of the random walk hypothesis. In its total concept the occurrences of the previous period are not having impact on actual situations. This can not be supported by this paper – movements of prices and returns are not predictable, but they are not totally independent of previous moments. Such aspects are considered with the GARCH model, so that it seems to be a suitable and reliable instrument for volatility forecasting. Can therefore models using stochastic volatility with aspects of Brownian motion be reliable instruments of volatility forecasting?

Situations like financial crises are perfect for research. In this case we derive similar results like in previous studies and can state a mean reversion of volatility. This could be one reason, why investors are once again bearing risks and why investment banks are focusing their businesses again. When computing VaR-models the market participants are deriving lower values and it seems that the situation has improved significantly. We have to ask, how reliable and serious such computations are? The strong efficiency of markets is not reality and due to lack of information the prices are not reflecting and revealing the truth. How can then such models be used for decisions and reports? This is surely a field, which needs additional research.

References

Abken, P. A., & Nandi, S. (1996). Options and Volatility. Economic Review, Federal Reserve Bank of Atlanta, December 1996, 21 - 35

Amin, K. I., & Morton, A. J. (1994). Implied Volatility Functions in Arbitrage-Free Term Structure Models. Journal of Financial Economics, 35, 141 - 180

Balland, P. (2002). Deterministic Implied Volatility Models. Quantitative Finance, 2, 31 - 44

Bollerslev, T., Engle, R. F., & Nelson, D. (1994). ARCH Models. In: Engle, R. F., & McFadden, D. L. (1994). Handbook of Econometrics, Volume IV, Elsevier Science B. V., 2959 - 3038

Christensen, B. J., & Prabhala, N. R. (1998). The Relation Between Implied and Realized Volatility. Journal of Financial Economics, 50, 125 - 150

CNN.com (2008). Timeline: Banking Crisis. Web site: http://edition.cnn.com/2008/BUSINESS/09/30/us.bailout.timeline/index.html

Cont, R., & da Fonseca, J. (2002). Dynamics of Implied Volatility Surfaces. Quantitative Finance, 2, 45 - 60

Duan, J.-C. (1995). The GARCH Option Pricing Model. Mathematical Finance. 5 (1), 13 - 22

Dumas, B., Fleming, J., & Whaley, R. E. (1998). Implied Volatility Function: Empirical Tests. The Journal of Finance, 53 (6), 2059 - 2106

Engle, R. (2001). GARCH 101: The Use of ARCH/GARCH Model in Applied Econometrics. The Journal of Economic Perspectives, 15 (4), 157 - 168

Engle, R. (2004). Risk and Volatility: Econometric Models and Financial Practice. The American Economic Review, June 2004, 405 - 420

Fama, E. F. (1965). The Behavior of Stock-Market Prices. The Journal of Business, 38 (1), 34 - 105

Fleming, J. (1998). The Quality of Market Volatility Forecasts Implied by S & P 100 Index Option Prices. Journal of Empirical Finance, 5, 317 - 345

Garcia, R., & Renault, E. (1997). A Note on Hedging in ARCH and Stochastic Volatility Option Pricing Models. Scientific Series, Centre Interuniversitaire de Recherche en Analyse des Organisations, 97s-13, Montreal

Hull, J., & White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. The Journal of Finance, 42 (2), 281 - 300

Jamshidian, F. (1989). An Exact Bond Option Formula. Journal of Finance, 44 (1), 205 - 209

Jiang, G. J., & Tian, Y. S. (2005). The Model-Free Implied Volatility and Its Information Content. The Review of Financial Studies, 4, 1305 - 1342

Klecka, W. R. (1980). Discriminant Analysis. Quantitative Applications in the Social Sciences Nr. 07 - 019, Newbury Park – London – New Delhi: Sage Publications 1980

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in Variance, Structural Change, and the GARCH Model. Journal of Business & Economic Statistics, 8 (2), 225 - 234

Lee, R. W. (2001). Implied and Local Volatilities Under Stochastic Volatility. International Journal of Theoretical and Applied Finance. 4 (1), 45 – 89

Macbeth, J. D., & Merville, L. J. (1980). Test of the Black-Scholes and Cox Call Option Valuation Models. Journal of Finance, 35 (2), 285 - 301

McDonald, R. L. (2006). Derivatives Markets. 2nd Edition, Boston: Pearson Education, Inc.

Poon, S.-H., & Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. Journal of Economic Literature, 16, 478 – 539

Poon. S.-H., & Granger, C. (2005). Practical Issues in Forecasting Volatility. Financial Analysts Journal, 61 (1), 45 - 56

Schwert, G. W. (1990). Stock Volatility and the Crash of ´87. The Review of Financial Studies, 3 (1), 77 - 102

Soczo, C. (2003). Estimation of Future Volatility. Periodica Polytechnica Ser. Soc. Man. Sci., 11 (2), 201 – 214

Subhash, S. (1996). Applied Multivariate Techniques, New York: John Wiley & Sons, Inc.

Szakmary, A., Ors, E., Kim, J. K., & Davidson, W. N. (2003). The Predictive Power of Implied Volatility: Evidence from 35 Future Markets. Journal of Banking & Finance, 27, 2151 – 2175

Wei, J. Z. (1997). A Simple Approach to Bond Option Pricing. The Journal of Futures Markets, 17 (2), 131 - 160

Appendix 1 – Daily Prices, Daily Returns and Daily Volatilities of the Index

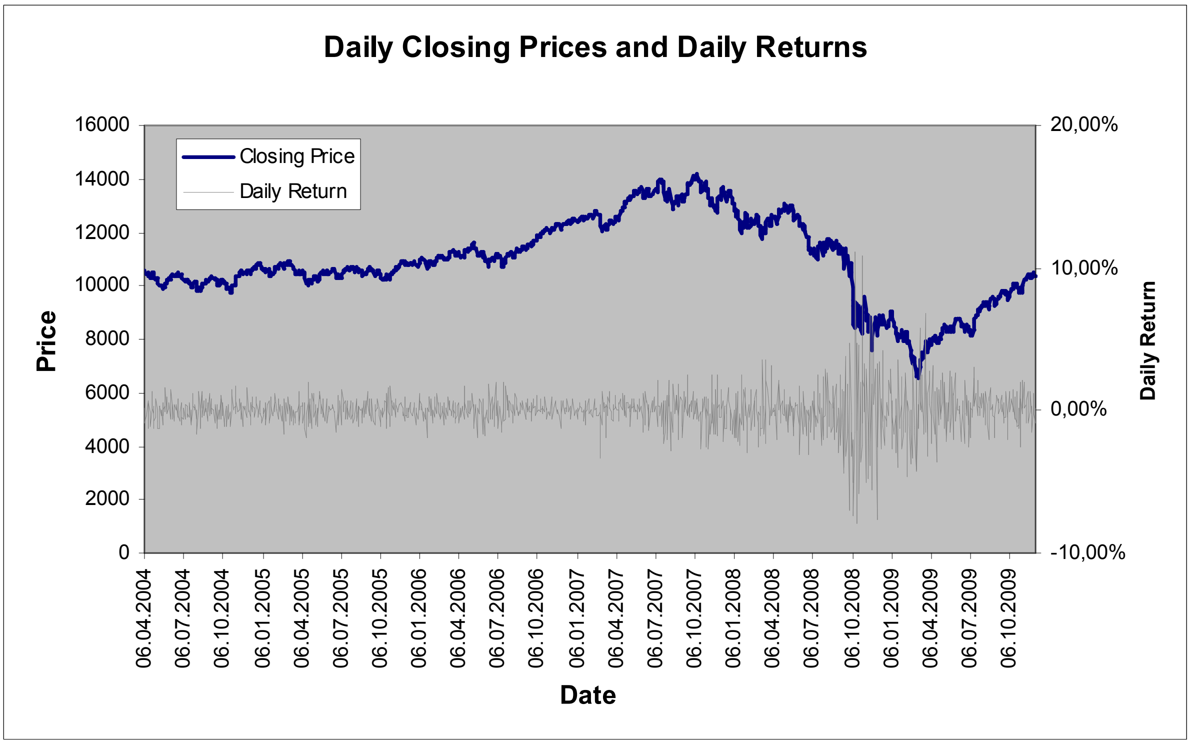


Figure 1: Daily Closing Prices and Daily Returns - Dow Jones Industrial Average

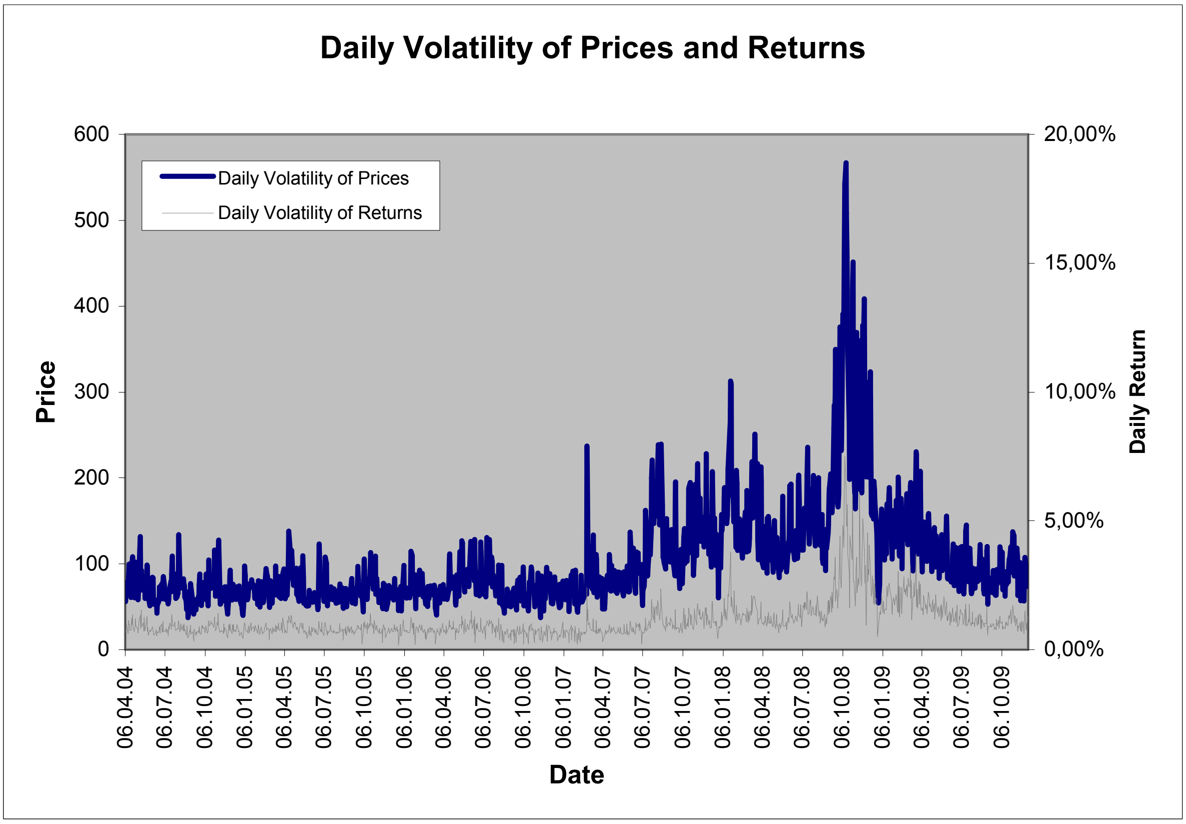


Figure 2: Daily Volatility of Prices and Returns - Dow Jones Industrial Average

Appendix 2 – Frequencies and Histogram Plots for the Index with Normal Curves

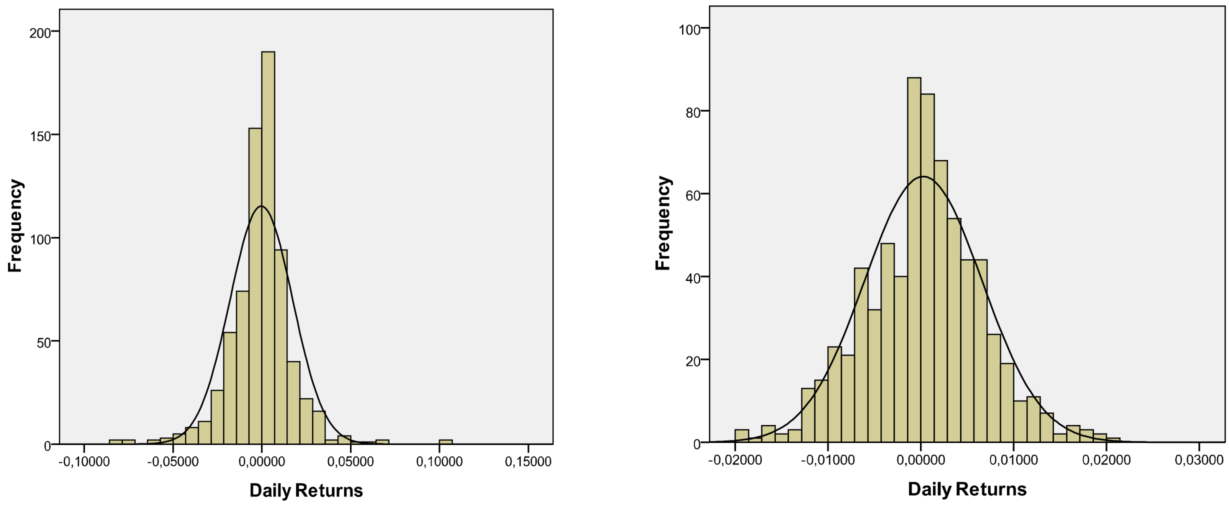
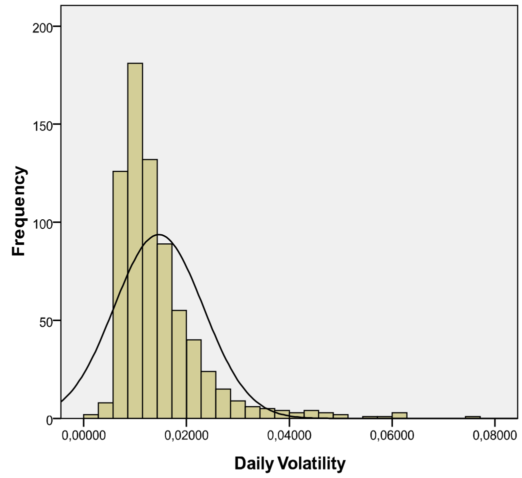


Figure 3: Frequencies of Continuously Compounded Daily Returns– Dow Jones Industrial Average
Left Side – 7th February 2007 – 4th December 2009 (Group 2)
Right Side – 6th April 2004 – 6th February 2007 (Group 1)

	7th February 2007 - 4th December 2009	6th February 2004 - 6th February 2007
N	714	714
Mean	-,0002776	,0002549
Std. Deviation	,01764490	,00634457
Variance	,00031134	,00004025
Skewness	**,08207978**	**-,03970985**
Kurtosis	**6,00537793**	**,43162627**

Table 1: Descriptive Statistics for Daily Returns – Dow Jones Industrial Average



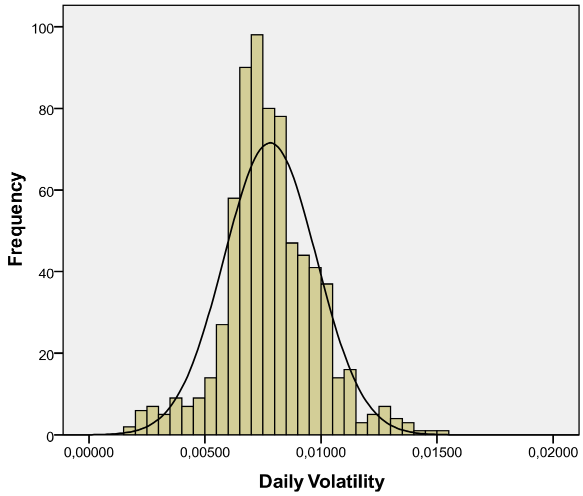


Figure 4: Frequencies of Daily Volatilities – Dow Jones Industrial Average
Left Side – 7th February 2007 – 4th December 2009 (Group 2)
Right Side – 6th April 2004 – 6th February 2007 (Group 1)

	7th February 2007 - 4th December 2009	6th February 2004 - 6th February 2007
N	714	714
Mean	,01459834	,00781601
Std. Deviation	,00868023	,00199234
Variance	,00007535	,00000397
Skewness	**2,58565454**	**,14221232**
Kurtosis	**9,59306621**	**1,21759663**

Table 2: Descriptive Statistics for Daily Volatilities – Dow Jones Industrial Average

Appendix 3 – Normal Probability Plots and Test for Normal Distribution

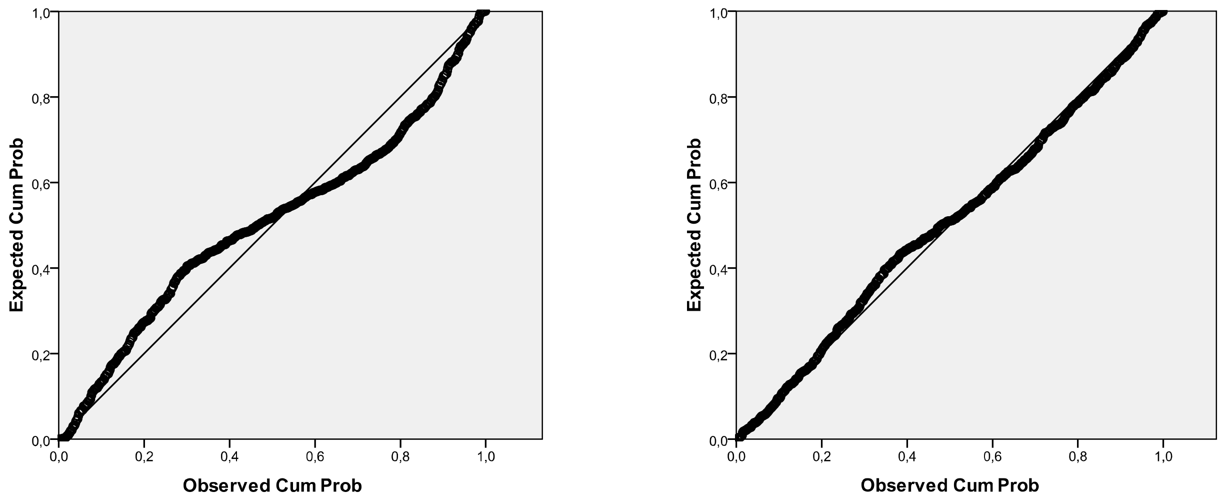


Figure 5: Normal Probability Plots of Continuously Compounded Daily Returns – Dow Jones Ind. Aver.
Left Side – 7th February 2007 – 4th December 2009 (Group 2)
Right Side – 6th April 2004 – 6th February 2007 (Group 1)

		7th February 2007 – 4th December 2009	6th April 2004 – 6th February 2007
N		714	714
Most Extreme Differences	Absolute	,105	,050
	Positive	,092	,026
	Negative	-,105	-,050
Kolmogorov-Smirnov Z		2,793	1,330
Asymp. Sig. (2-tailed)		,000	**,058**

Table 3: One Sample Kolmogorov-Smirnov Test for Daily Returns – Dow Jones Industrial Average

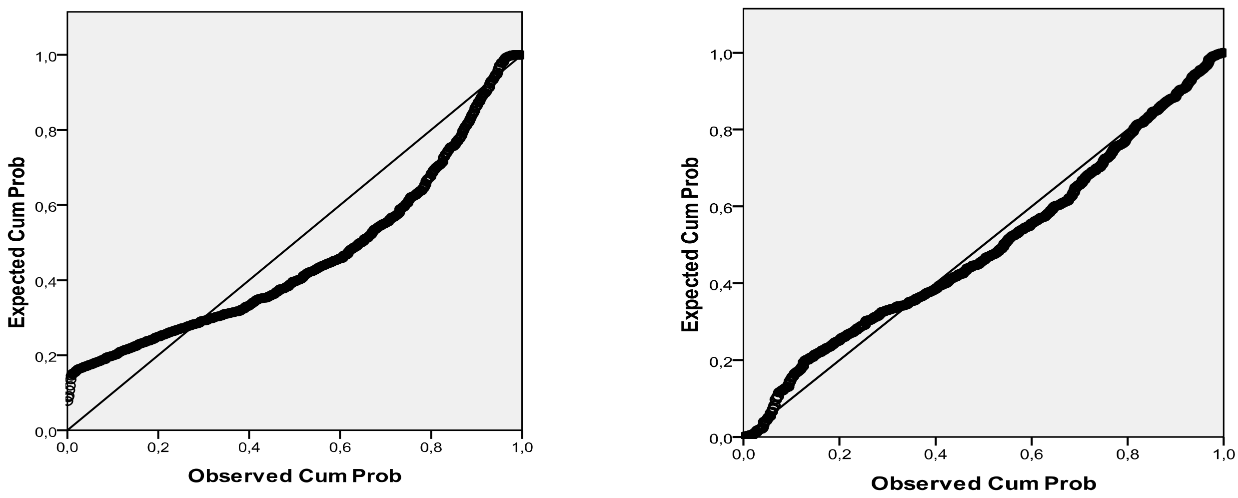


Figure 6: Normal Probability Plots of Daily Volatilities – Dow Jones Ind. Aver.
Left Side – 7th February 2007 – 4th December 2009 (Group 2)
Right Side – 6th April 2004 – 6th February 2007 (Group 1)

		7th February 2007 – 4th December 2009	6th April 2004 – 6th February 2007
N		714	714
Most Extreme Differences	Absolute	,152	,071
	Positive	,152	,062
	Negative	-,141	-,071
Kolmogorov-Smirnov Z		4,063	1,911
Asymp. Sig. (2-tailed)		,000	,001

Table 3: One Sample Kolmogorov-Smirnov Test for Daily Volatilities – Dow Jones Industrial Average

Appendix 4 – One-Way ANOVA

			Sum of Squares	df	Mean Square	F	Sig.
Opening Price	Between Groups	(Combined)	1,931E+07	1	1,931E+07	7,840	0,005
	Within Groups		3,512E+09	1426	2,463E+06		
	Total		3,531E+09	1427			
Highest Daily Price	Between Groups	(Combined)	2,987E+07	1	2,987E+07	12,239	0,000
	Within Groups		3,480E+09	1426	2,441E+06		
	Total		3,510E+09	1427			
Lowest Daily Price	Between Groups	(Combined)	9,312E+06	1	9,312E+06	3,746	**0,053**
	Within Groups		3,545E+09	1426	2,486E+06		
	Total		3,554E+09	1427			
Closing Price	Between Groups	(Combined)	1,836E+07	1	1,836E+07	7,453	0,006
	Within Groups		3,513E+09	1426	2,463E+06		
	Total		3,531E+09	1427			
Volume	Between Groups	(Combined)	2,637E+21	1	2,637E+21	1917,923	0,000
	Within Groups		1,960E+21	1426	1,375E+18		
	Total		4,597E+21	1427			
Continuously Compounded Daily Return	Between Groups	(Combined)	,000	1	,000	,576	**0,448**
	Within Groups		,251	1426	,000		
	Total		,251	1427			
Difference between Highest and Lowest Daily Price	Between Groups	(Combined)	5,827E+06	1	5,827E+06	567,922	0,000
	Within Groups		1,463E+07	1426	10259,916		
	Total		2,046E+07	1427			
Difference between Highest Daily and Closing Price	Between Groups	(Combined)	1,391E+06	1	1,391E+06	182,413	0,000
	Within Groups		1,087E+07	1426	7622,992		
	Total		1,226E+07	1427			
Difference between Closing and Lowest Daily Price	Between Groups	(Combined)	1,525E+06	1	1,525E+06	212,317	0,000
	Within Groups		1,024E+07	1426	7181,767		
	Total		1,177E+07	1427			
Volatility of Daily Price	Between Groups	(Combined)	1,489E+06	1	1,489E+06	583,769	0,000
	Within Groups		3,637E+06	1426	2550,450		
	Total		5,126E+06	1427			
Volatility of Continuously Compounded Daily Returns	Between Groups	(Combined)	,016	1	,016	414,092	0,000
	Within Groups		,057	1426	,000		
	Total		,073	1427			

Table 4: One-Way ANOVA

Appendix 5 – Discriminant Analysis

Variables Entered/Removed[a,b,c,d]

Step	Entered	Wilks' Lambda							
						Exact F			
		Statistic	df1	df2	df3	Statistic	df1	df2	Sig.
1	Volume	,426	1	1	1426,000	1917,923	1	1426,000	,000
2	Lowest Daily Price	,311	2	1	1426,000	1579,237	2	1425,000	,000
3	Difference Between Highest and Lowest Daily Price	,308	3	1	1426,000	1064,659	3	1424,000	,000
4	Volatility of Daily Price	,307	4	1	1426,000	803,233	4	1423,000	,000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 22.

b. Minimum partial F to enter is 3.84.

c. Maximum partial F to remove is 2.71.

d. F level, tolerance, or VIN insufficient for further computation.

Table 5: Fischer Stepwise Method for the Discriminant Analysis

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2,258[a]	100,0	100,0	,832

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,307	1681,844	4	,000

Table 6: Eigenvalues and Wilks´ Lambda for the Discriminant Analysis

Canonical Discriminant Function Coefficients

	Function
	1
Lowest Daily Price	,0005145795
Volume	,0000000011
Difference Between Highest and Lowest Daily Price	-,0028908549
Volatility of Daily Price	,0036888000
(Constant)	-9,0523697472

Unstandardized coefficients

Table 7: Coefficients of the Canonical Discriminant Function

Classification Results[a]

		Group	Predicted Group Membership		Total
			1	2	
Original	Count	1	668	**46**	714
		2	**36**	678	714
	%	1	93,6	**6,4**	100,0
		2	**5,0**	95,0	100,0

a. **94,3%** of original grouped cases correctly classified.

Table 8: Classification Results for the Discriminant Analysis

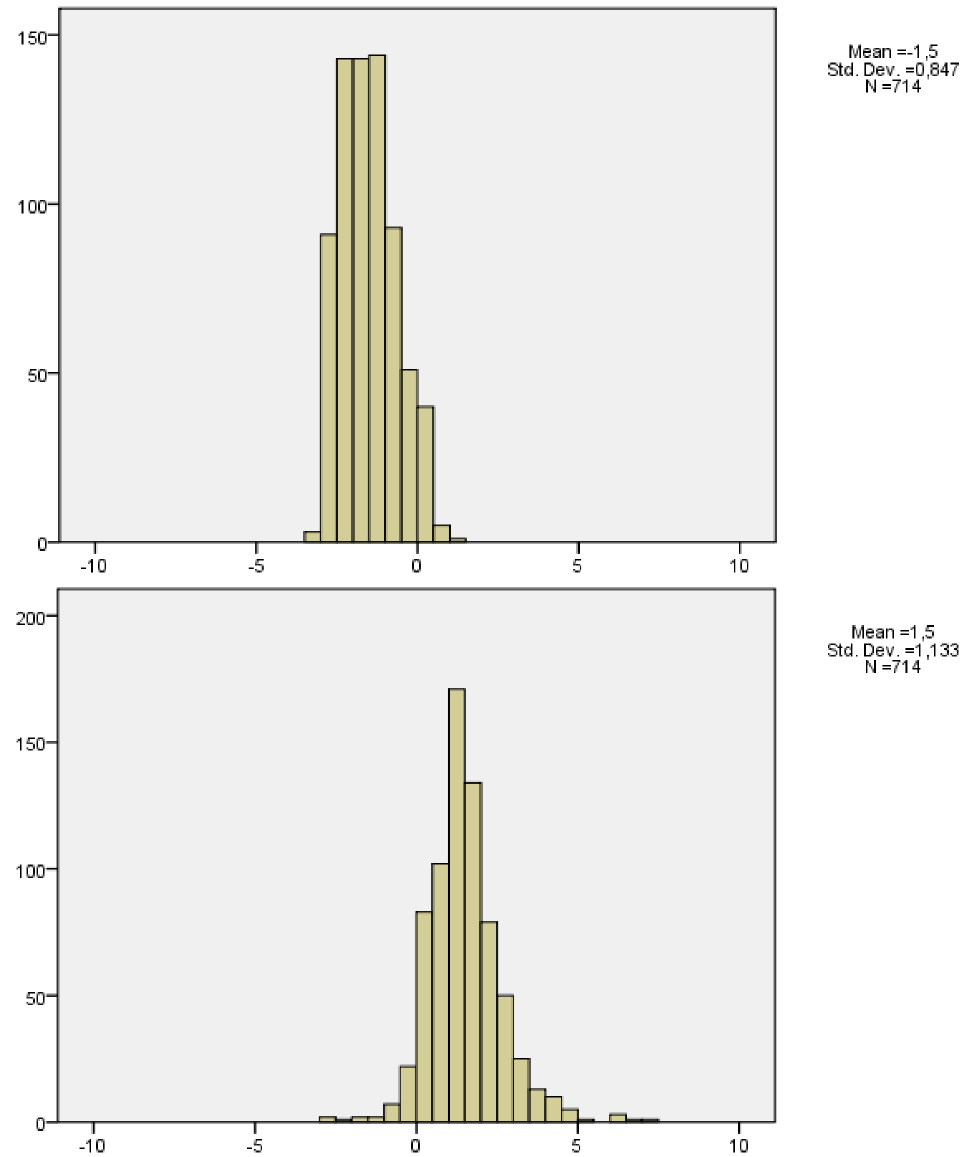


Figure 7: Separate-Group Graphs – Dow Jones Ind. Aver.
Above – Classification for Group 1 - 6th April 2004 – 6th February 2007
Below – Classification for Group 2 - 7th February 2007 – 4th December 2009

Appendix 6 – Bivariate Correlation between Groups and Variables

Bivariate Correlations

		Lowest Daily Price (Group 2)	Lowest Daily Price (Group 1)	Volume (Group 2)	Volume (Group 1)	Continuously Compounded Daily Return (Group 2)	Continuously Compounded Daily Return (Group 1)	Volatility of Daily Prices (Group 2)	Volatility of Daily Prices (Group 1)	Volatility of Continuously Compounded Daily Returns (Group 2)	Volatility of Continuously Compounded Daily Returns (Group 1)
Lowest Daily Price (Group 2)	Pearson Correlation	1	-,621**	-,730**	-,697**	,001	-,022	-,215**	-,107**	-,502**	,068
	Sig. (2-tailed)		,000	,000	,000	,977	,559	,000	,004	,000	,068
	N	714	714	714	714	714	714	714	714	714	714
Lowest Daily Price (Group 1)	Pearson Correlation	-,621**	1	,389**	,634**	,024	,050	-,132**	-,128**	,060	-,338**
	Sig. (2-tailed)	,000		,000	,000	,522	,185	,000	,001	,111	,000
	N	714	714	714	714	714	714	714	714	714	714
Volume (Group 2)	Pearson Correlation	-,730**	,389**	1	,536**	-,016	-,013	,529**	,087*	,702**	-,033
	Sig. (2-tailed)	,000	,000		,000	,668	,729	,000	,020	,000	,377
	N	714	714	714	714	714	714	714	714	714	714
Volume (Group 1)	Pearson Correlation	-,697**	,634**	,536**	1	,040	,043	,082*	,301**	,250**	,135**
	Sig. (2-tailed)	,000	,000	,000		,281	,253	,029	,000	,000	,000
	N	714	714	714	714	714	714	714	714	714	714
Continuously Compounded Daily Return (Group 2)	Pearson Correlation	,001	,024	-,016	,040	1	,031	-,050	,008	-,050	,029
	Sig. (2-tailed)	,977	,522	,668	,281		,402	,185	,824	,179	,443
	N	714	714	714	714	714	714	714	714	714	714
Continuously Compounded Daily Return (Group 1)	Pearson Correlation	-,022	,050	-,013	,043	,031	1	-,055	-,025	-,032	-,067
	Sig. (2-tailed)	,559	,185	,729	,253	,402		,144	,510	,393	,075
	N	714	714	714	714	714	714	714	714	714	714
Volatility of Daily Prices (Group 2)	Pearson Correlation	-,215**	-,132**	,529**	,082*	-,050	-,055	1	-,026	,866**	,018
	Sig. (2-tailed)	,000	,000	,000	,029	,185	,144		,485	,000	,622
	N	714	714	714	714	714	714	714	714	714	714
Volatility of Daily Prices (Group 1)	Pearson Correlation	-,107**	-,128**	,087*	,301**	,008	-,025	-,026	1	,019	,679**
	Sig. (2-tailed)	,004	,001	,020	,000	,824	,510	,485		,611	,000
	N	714	714	714	714	714	714	714	714	714	714
Volatility of Continuously Compounded Daily Returns (Group 2)	Pearson Correlation	-,502**	,060	,702**	,250**	-,050	-,032	,866**	,019	1	,009
	Sig. (2-tailed)	,000	,111	,000	,000	,179	,393	,000	,611		,813
	N	714	714	714	714	714	714	714	714	714	714
Volatility of Continuously Compounded Daily Returns (Group 1)	Pearson Correlation	,068	-,338**	-,033	,135**	,029	-,067	,018	,679**	,009	1
	Sig. (2-tailed)	,068	,000	,377	,000	,443	,075	,622	,000	,813	
	N	714	714	714	714	714	714	714	714	714	714

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 9: Bivariate Correlations

Appendix 7 – Linear Regression

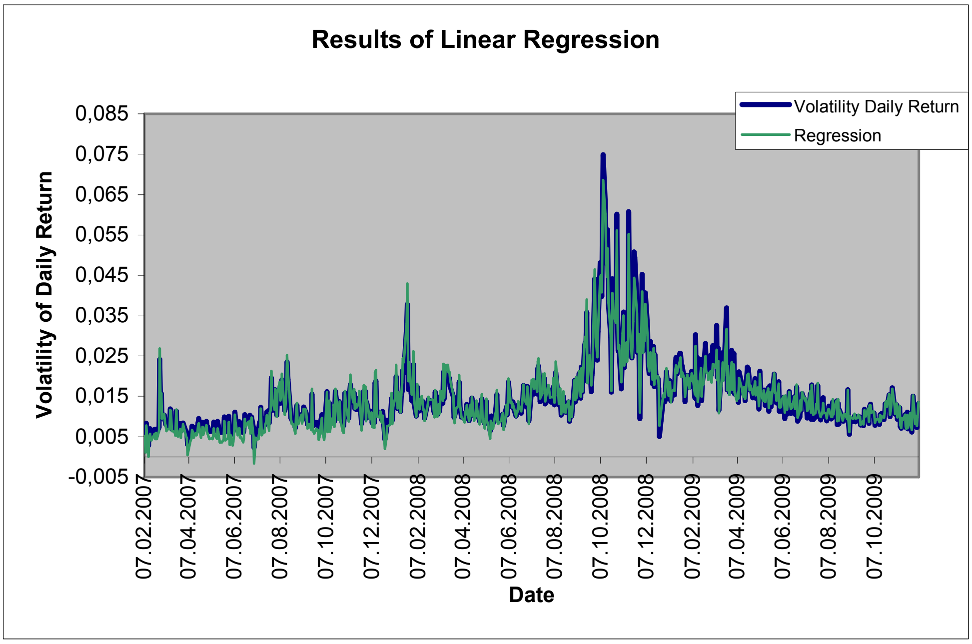


Figure 8: Linear Regression for Volatility of Daily Return - 7th February 2007 – 4th December 2009 (Group 2)

Casewise Diagnostics[a]

Case Number	Std. Residual	Volatility of Daily Return	Predicted Value	Residual
180	3,435	,03694	,0320301	,00491387
189	4,017	,03256	,0268177	,00574694
192	3,078	,02763	,0232248	,00440301
206	3,829	,02461	,0191301	,00547762
261	4,235	,04950	,0434395	,00605905
262	4,296	,05080	,0446501	,00614575
267	3,636	,06074	,0555405	,00520185
281	3,722	,04040	,0350770	,00532497
287	3,026	,05627	,0519440	,00432906
291	4,211	,07485	,0688297	,00602424
473	-3,793	,03785	,0432771	-,00542705

a. Dependent Variable: LogDailyVolatilityDJIndAver2

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-,0013330	,0688297	,0145983	,00856202	714
Residual	-,00542705	,00614575	,00000000	,00142763	714
Std. Predicted Value	-1,861	6,334	,000	1,000	714
Std. Residual	-3,793	4,296	,000	,998	714

a. Dependent Variable: Volatility of Daily Return

Table 10: Casewise Diagnostics and Residual Statistics – Linear Regression for Group 2

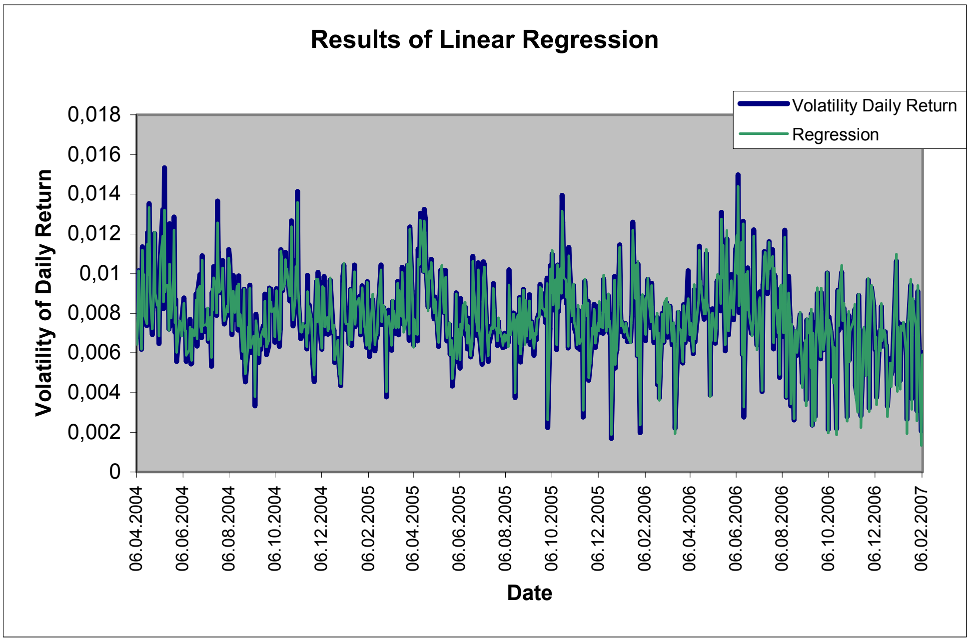


Figure 9: Linear Regression for Volatility of Daily Return - 6th April 2004 – 6th February 2007 (Group 1)

Casewise Diagnostics[a]

Case Number	Std. Residual	Volatility of Daily Return	Predicted Value	Residual
213	4,511	,01015	,0087345	,00141461
420	4,123	,00875	,0074535	,00129278
642	3,567	,01365	,0125362	,00111863
684	6,480	,01250	,0104723	,00203185
689	6,828	,01533	,0131937	,00214108
691	4,316	,01323	,0118727	,00135347

a. Dependent Variable: LogDailyVolatilityDJIndAver1

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	,0013502	,0143806	,0077813	,00192188	610
Residual	-,00052654	,00214108	-,00000499	,00030509	610
Std. Predicted Value	-3,033	3,130	,009	,909	610
Std. Residual	-1,679	6,828	-,016	,973	610

a. Dependent Variable: Volatility of Daily Return

Table 11: Casewise Diagnostics and Residual Statistics – Linear Regression for Group 1

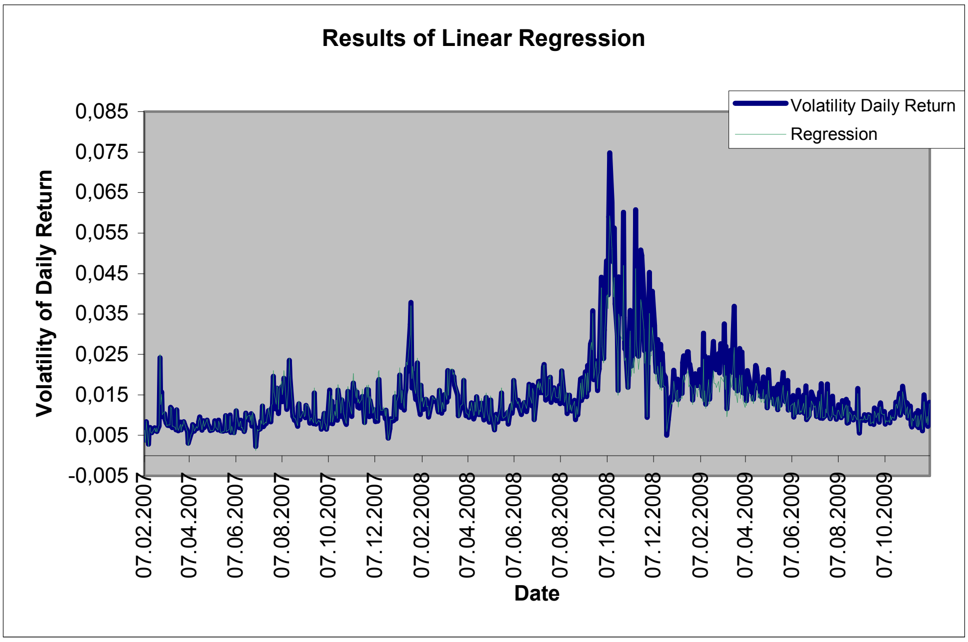


Figure10: Linear Regression for Volatility of Daily Return – Forecasting for group 2

	Minimum	Maximum	Sum of Residuals
Linear Regression Group 1	1.513E – 06	0.00214	0.1629
Linear Regression Group 2	6.222E – 08	0.00642	0.7746
Linear Regression Prediction Group 2	1.845E - 06	0.0156	1.0056

Table 12: Diagnostics of Regression Analysis

Appendix 8 – GARCH(1,1) Model

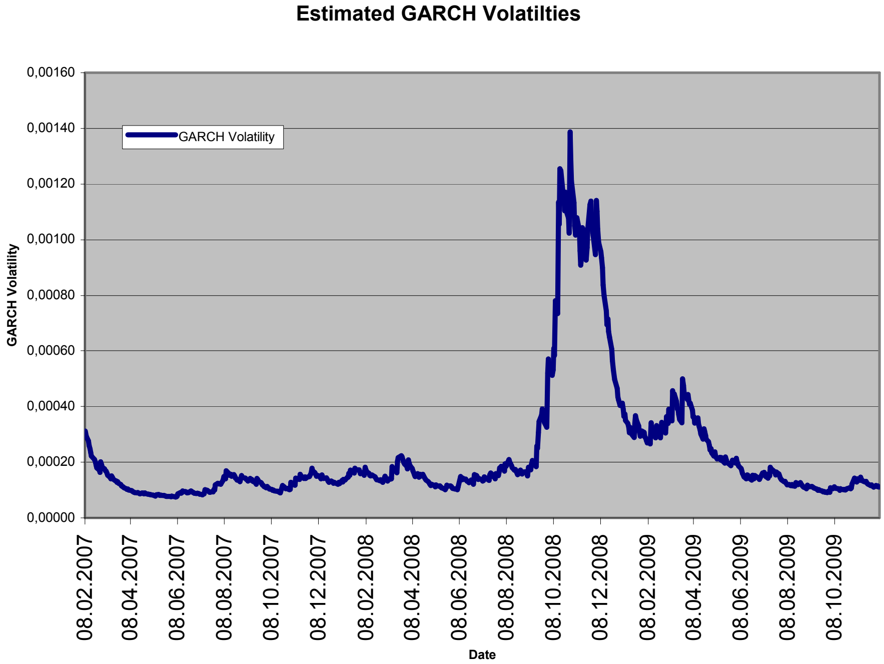


Figure11: Estimated GARCH Volatilities - 7th February 2007 – 4th December 2009 (Group 2)

	Values
a_0	0.000004857
a_1	0.040969441
b_1	0.922998431
Log Likelihood Maximum	2674.535959244
Unconditional Annual Standard Deviation	0.1843134
Historical Annual Standard Deviation	0.2803006

Table 13: Diagnostics of GARCH(1,1) Model for Group 2

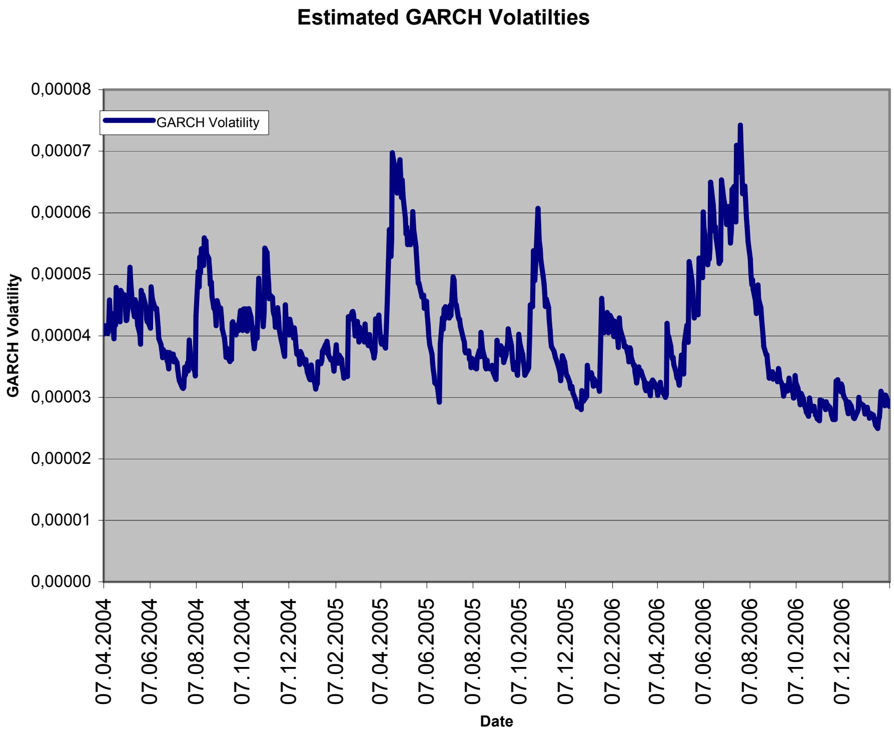


Figure12: Estimated GARCH Volatilities - 6th April 2004 – 6th February 2007 (Group 1)

	Values
a_0	0.000001422
a_1	0.040969441
b_1	0.922998431
Log Likelihood Maximum	3250.260241546
Unconditional Annual Standard Deviation	0.099707792
Historical Annual Standard Deviation	0.100786109

Table 14: Diagnostics of GARCH(1,1) Model for Group 1

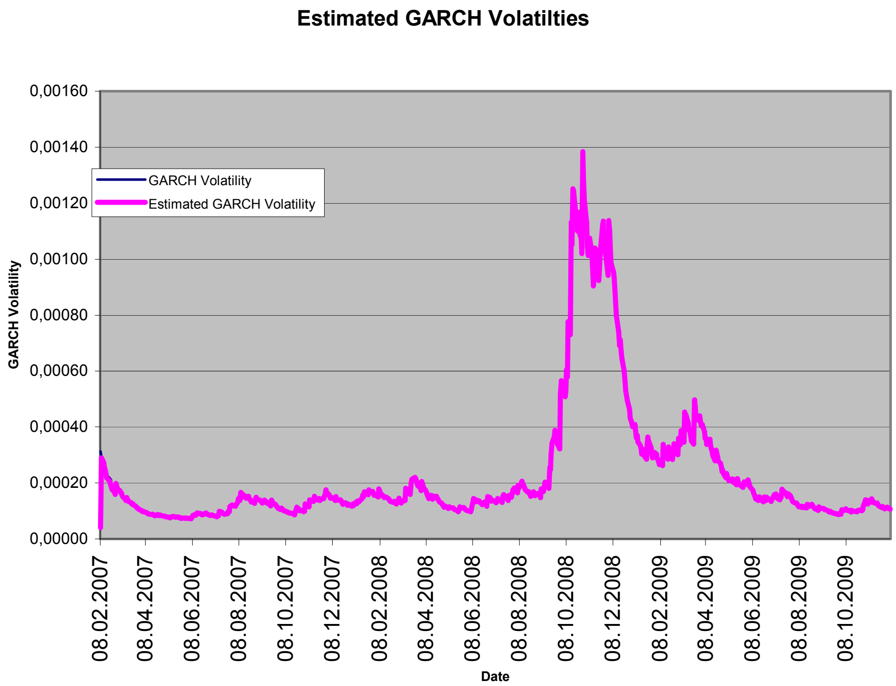


Figure13: Estimated GARCH Volatilities – Forecasting Group 2