Comparison of the GARCH, EGARCH, GJR-GARCH and TGARCH model in times of crisis for the S&P500, NASDAQ and Dow-Jones

Misha Dol

482752

Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis Financial Economics

Supervisor: Abhinav Bhuyan

Second assessor: Rogier Quaedvlieg

July 31, 2021

1 Author Note

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

2 Abstract

This paper investigates the relative performances in volatility forecasting of the standard GARCH, EGARCH, GJR-GARCH and TGARCH models in times of crisis. The data analyzed are the daily price indices of the S&P500, NASDAQ and Dow-Jones. The testing periods consist of the dot-com bubble, the financial crisis and the COVID-19 crisis. The forecasts are done over a five-year rolling window and are tested against each other using two evaluation measures and Diebold-Mariano tests. For all four models, forecasts assuming a normal-, t-, and skewed t-distribution are tested against each other. The results show that the t-distribution performs best for all but one combination of model and data set. To test between the models, the forecasts assuming a t-distribution are compared across the four models. The results show that none of the asymmetric models outperforms the GARCH model. The GARCH model even significantly outperforms the GJR-GARCH and TGARCH models on all data sets.

Contents

3 Introduction

When one takes a quick look at the CBOE Volatility index (VIX) it immediately becomes apparent that crises cause massive spikes in volatility. In March 2020 the world experienced a massive shock as stock prices plunged, initiated by the COVID-19 crisis. Even though this shock was set off by an unprecedented public health crisis, a good volatility prediction model should be able to detect these spikes early. Being able to predict big negative shocks is one of the main objectives of risk managers around the globe. Volatility is used as a measure to indicate the frequency and magnitude of shocks. It has numerous uses in the financial world ranging from risk management to pricing options to calculating Sharpe ratios. As it is not directly observable estimating it poses one of the biggest challenges in finance. Because of this, volatility is among the most widely researched topics in the financial literature, as well as in private businesses and financial institutions.

Before 1982 the traditional econometric models generally assumed a constant one step ahead volatility. The models assumed that there was no autocorrelation between the variance and its lags. To relax this assumption of homoskedasticity Engle (1982) introduced the AutoRegressive Conditional Heteroskedastic (ARCH) model. In this model, the conditional variance is dependent on past innovations and therefore adjusts itself over time. After the release of the paper, innumerable extensions have been made that build on the ARCH model. One of the most influential extensions is the Generalized AutoRegressive Conditional Heteroskedastic (GARCH) model introduced by Bollerslev (1986). This extension has grown out to be the most popular used model in empirical volatility modelling according to Li, Ling, and McAleer (2002). The GARCH model adds an autoregressive component to the ARCH process by letting the conditional variance be dependent on both past innovations and its lags. Briefly after the introduction of the GARCH model Akgiray (1989) showed that both the ARCH and GARCH model outperformed traditional homoscedastic models. This research proved that the assumption of homoskedasticity of traditional models needed to be relaxed and more research needed to be done to provide accurate volatility forecasts.

Like the ARCH model, the GARCH model has many extensions developed in later research. One of the problems of GARCH volatility forecasting for stock returns is the existence of excess kurtosis. The original ARCH and GARCH processes assume a Gaussian distribution. Part of this problem was alleviated by changing the underlying distribution assumption. Baillie and DeGennaro (1990) found that a GARCH model assuming a student's t-distribution (t-distribution) outperformed the GARCH model assuming a Gaussian distribution.

Another problem in GARCH volatility forecasting for stock returns is symmetry. The GARCH model requires symmetry in its estimates. Many papers have documented that stock returns display a so-called 'leverage effect' (Bouchaud, Matacz, & Potters, 2001; Poon & Granger, 2003). This effect was first introduced by Black (1976) and is widely accepted as a stylized fact of stock returns. The leverage effect implies that stock returns generally have negatively skewed distributions. As the GARCH model is unable to produce asymmetric estimates this would mean the predictions would be biased. One possible solution to the issue is changing the assumed distribution to a skewed student's t-distribution (skewed t-distribution). Other possible solutions are new models that were specifically designed to accommodate asymmetry. The three asymmetric models used in this paper are the Exponential GARCH (EGARCH) model, proposed by Nelson (1991), the GJR-GARCH model introduced by Glosten, Jagannathan, and Runkle (1993) and the Threshold GARCH (TGARCH) model by Zakoian (1994).The EGARCH model uses a conditional variance equation in logarithmic form, making it able to relax the parameter restrictions on GARCH models. On top of that, it introduces a new component that allows the model to react more or less to negative shocks. The GJR-GARCH and TGARCH model are very similar to each other. They both use an indicator function to allow the model to react more to negative shocks.

These three models are among the most widely used volatility prediction models and numerous papers have shown that they produce good predictions for stock returns (Alberg, Shalit, & Yosef, 2008; Brailsford & Faff, 1996; Hansen & Lunde, 2005; Pagan & Schwert, 1990; Poon & Granger, 2003). Many of these papers also suggest that the asymmetric models outperform the standard GARCH model for stock returns.

Most of the relevant highly cited papers stem from the 1990s and 2000s, meaning much of the relevant research in the literature is relatively old. We have since then experienced three major crises in a relatively short period of time: the dot-com bubble around 2000, the financial crisis of 2008/2009 and the ongoing COVID-19 crisis. All crashes are characterized by a sharp spike in volatility and a relatively fast recovery compared to previous crises. The financial markets have become increasingly computerized and countless new volatility forecasting models have been created. Despite these new models, ARCH family models are still used by financial institutions and taught in universities. This paper uses the new data available on the last three major stock market crashes to investigate and evaluate the performance of four models of the ARCH family. The standard GARCH model will be tested against three asymmetric GARCH models: the EGARCH model by Nelson (1991), the GJR-GARCH model by Glosten et al. (1993) and the TGARCH model by Zakoian (1994). Each model is used to forecast over the time periods of the three crises. Subsequently, the forecasts are compared to each other with two evaluation measures. Afterwards, the difference in evaluating measures is tested using a Diebold-Mariano test.

This paper updates the current literature on the relative performances of the four models. On top of that, it researches the assumed distributions and tests which of these distributions provide the best forecasts for all four models. According to most of the literature, the asymmetric models are expected to outperform the GARCH model. This should especially hold in times of crisis, as these time periods are generally characterised by big negative shocks. The EGARCH, GJR-GARCH and TGARCH models should be able to adjust to these shocks better than the standard GARCH model. Additionally, as the literature has documented the fat tails and the leverage effect in stock returns the t and skewed t-distribution are expected to outperform the normal distribution. The skewed t-distribution is expected to perform best.

The paper is split up into sections. First, in section 4 a brief elaboration on the data sets used is given. In section 5 the empirical modelling and forecasting scheme are specified. Section 6 will display the results found from the forecasts and tests. The final section will conclude and discuss points for future research.

4 Data

The data used in this paper consists of four data sets. The first three data sets contain daily price indices of the S&P500 (SPX), NASDAQ (IXIC) and Dow Jones (DJI). The time period spans from January 1, 1990, to May 11, 2021. The data is collected from the Thomas Reuters Eikon Datastream database. The last data set used contains the daily indices of the CBOE Volatility Index (VIX) from January 2, 1992, until May 11, 2021. The data is taken from Yahoo finance.

The return series of the S&P500, NASDAQ and Dow-Jones are presented in Figure 1, Figure 2 and Figure 3. The dot-com bubble, financial crisis and COVID-19 are easily recognizable in all three index series by their big spikes in returns.

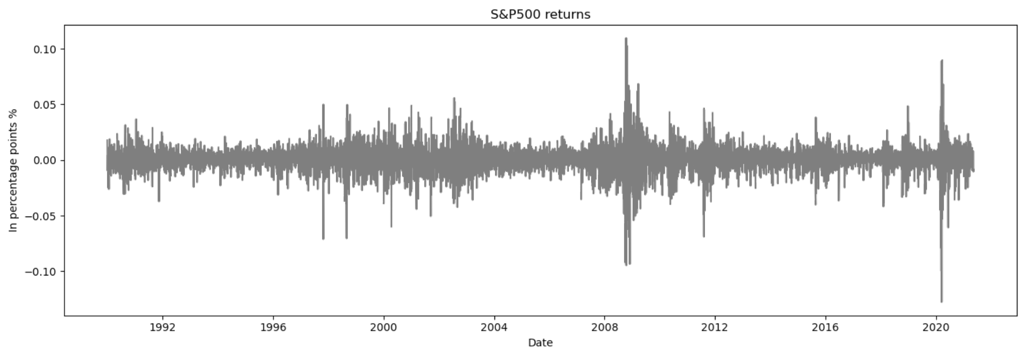

Figure 1

S&P500 returns over 1990-2021

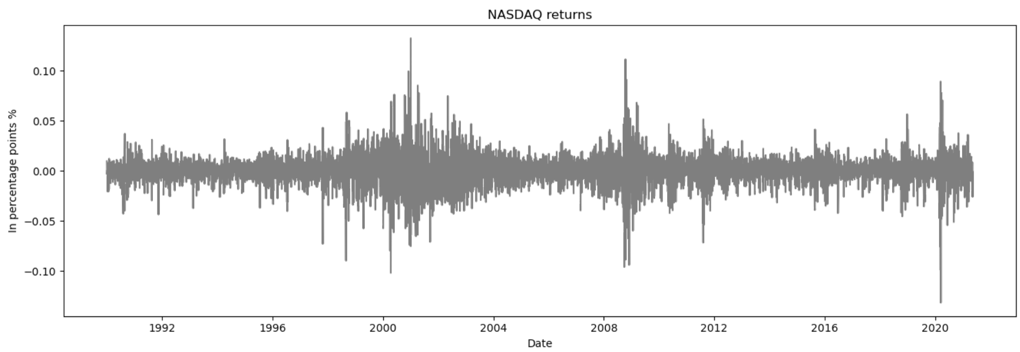

Figure 2

NASDAQ returns over 1990-2021

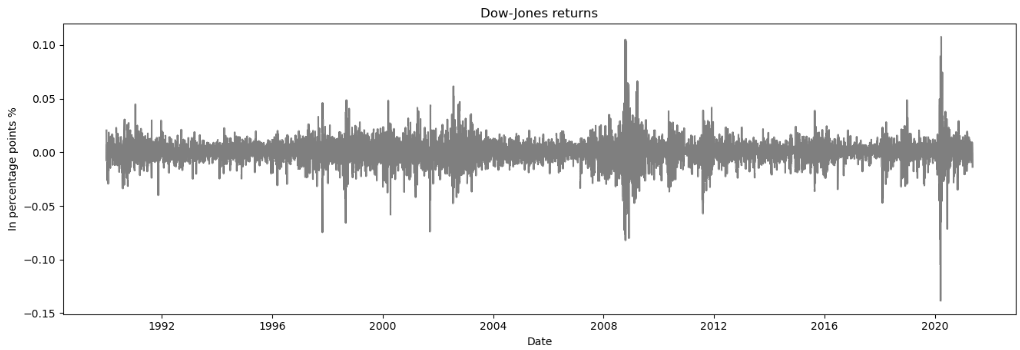

Figure 3

Dow-Jones returns over 1990-2021

Table 1

Descriptive statistics of the S&P500, NASDAQ and Dow-Jones returns

	Dataset		
Statistics	S&P500	NASDAQ	Dow-Jones
Mean	0.00030	0.00041	0.00031
Variance	0.00013	0.00021	0.00012
Skewness	-0.41536	-0.21356	-0.40820
Kurtosis	11.84366	7.49756	13.54714

The descriptive statistics of the S&P500, NASDAQ and Dow-Jones returns are reported in Table 1. The data shows that all three indices have high excess kurtosis and have negatively skewed distributions. This supports the previous literature studies that argue that stock returns have heavy tails and a leverage effect. Simply looking at Table 1 would suggest that the models would perform best if a skewed t-distribution is assumed.

5 Methodology

In this section the models are specified and the forecasting scheme is defined. The forecasting scheme is set up as follows. First, the testing data and rolling window length are defined. Afterwards, the evaluating measures are described and finally, the test statistic used for testing the relative performances is specified.

Model specification

In this paper for each data set the price indices are used to calculate the daily financial returns. Log returns are used, which are computed as follows:

$$y_t = 100 * (ln(I_{t+1}) - ln(I_t)) \tag{1}$$

y_t is defined as the daily return, It is defined as the stock index on time t and I_{t+1} is defined as the stock index at time $t + 1$.

The first forecasting model used is the linear GARCH(1,1) model. The GARCH(1,1) model is specified below:

$$y_t = \mu_t + u_t \tag{2}$$

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3}$$

$$u_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1) \tag{4}$$

In these equations y_t is the daily return. ω and α and β are constant parameters. Equation 2 describes the conditional mean equation. The μ_t in this model can be a constant, an ARMA process or any other type of mean calculation process. In this paper, a constant mean is assumed. The choice of μ_t does not affect the volatility estimates. As this paper strictly focuses on the volatility estimation of the GARCH models, any choice of μ_t suffices. u_t contains the residuals of the conditional mean equation.

Equation 3 describes the conditional variance equation with an autoregressive and a moving average component. The autoregressive component is captured by σ_{t-1}^2 and the moving average component is captured by u_{t-1}^2. Some additional restrictions are added to the model to prevent negative variances. The first restriction is that the parameters α, β and ω must be nonnegative. The second restriction is as follows: $\alpha + \beta < 1$ (Bollerslev, 1986).

As seen in equation 4, the distribution of the residuals u_t is dependent on σ_t^2, which contains the conditional variance, and ϵ_t, the innovation at time t. In this model specification, the innovations are assumed to follow a standard normal distribution.

u_t and y_t are symmetric time series, which is why a standard linear GARCH model is not able to capture the asymmetric properties of the return series. The three asymmetric GARCH models that will be used to try to capture the asymmetric properties are the GJR-GARCH(1,1), TGARCH(1,1) and the EGARCH(1,1) model.

The GJR-GARCH(1,1) is similar to the GARCH(1,1) model. The conditional mean equation is equal to that of the GARCH(1,1) model, equation 2. The GJR-GARCH model differs in its conditional variance equation. The addition of an indicator function to the conditional variance equation allows for asymmetric properties to be modelled. The conditional variance equation is specified as follows:

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma u_{t-1}^2 \mathbb{1}_{[u_{t-1}<0]} \tag{5}$$

σ_t^2 denotes the conditional variance and u_{t-1}^2 denotes the lagged squared residuals . ω α, β and γ are constant parameters. The indicator function $\mathbb{1}_{[u_{t-1}<0]}$ is equal to 1 if the past residual is negative and equal to 0 if the past residual is nonnegative. If the γ coefficient is equal to zero, the GJR-GARCH model is identical to a GARCH(1,1) model. If gamma is positive, negative shocks will have a bigger impact on the conditional volatility than positive shocks. If gamma is negative the inverse is true. The indicator function allows the model to fit itself to returns with asymmetric properties.

The TGARCH(1,1) model is very much alike the GJR-GARCH model but uses absolute residuals instead of squared residuals. The conditional variance equation is specified below:

$$\sigma_t^2 = \omega + \alpha |u_{t-1}| + \beta \sigma_{t-1} + \gamma |u_{t-1}| \mathbb{1}_{[u_{t-1}<0]} \tag{6}$$

A more general model can be used to describe both models in one equation:

$$\sigma_t^\kappa = \omega + (\alpha + \gamma \mathbb{1}_{[u_{t-1}<0]})|u_{t-1}|^\kappa + \beta \sigma_{t-1}^\kappa \tag{7}$$

κ is the power of the model. The conditional variance is $(\sigma_t^\kappa)^{\frac{2}{\kappa}}$. This more general model is very similar to the A-PARCH model, introduced by Ding, Granger, and Engle (1993). If κ equals 2 the model is equal to the GJR-GARCH model. If κ is equal to 1 the model is equal to a TGARCH model. In this paper, only the GJR-GARCH and TGARCH models are used to forecast.

The last model used to forecast is the nonlinear EGARCH(1,1) model. Once again the conditional mean equation is equal to those of the previous models. The model is however fundamentally different from the three previous models as it does not put restrictions on its parameters to ensure nonnegative variances. The conditional variance equation is specified as follows:

$$ln(\sigma_t) = \omega + \alpha \frac{u_{t-1}}{\sqrt{\sigma_{t-1}}} + \beta(\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}}} - \sqrt{\frac{2}{\pi}}) \tag{8}$$

The natural logarithm on the left side of the equation ensures that the variance is nonnegative as the exponential function is strictly positive. This allows the model to drop the restrictions on its parameters. ω, α, and β are constant parameters. The α coefficient allows the model to capture asymmetry. If $\alpha = 0$, positive shocks have the same impact as negative shocks; if $\alpha > 0$ positive shocks increase the conditional variance; if $\alpha < 0$ positive shocks reduce the conditional variance. The coefficient determines whether positive or negative shocks have a greater impact on the conditional variance.

Underlying distributions

In all four models the innovations are assumed to be normally distributed. As Cont and Bouchaud (2000) show that stock returns generally have fat tails and asymmetric properties this assumption is relaxed. All four models are tested separately assuming a normal-, t- and skewed t-distribution.

Forecasting scheme

The S&P500, the NASDAQ and Dow-Jones data sets are all analysed separately using the following forecasting scheme: All four models are used to forecast the one step ahead volatility for each day in the testing period. For each day in the testing period, the model is trained on a five-year rolling window and used to predict the volatility for the following day.

Time periods

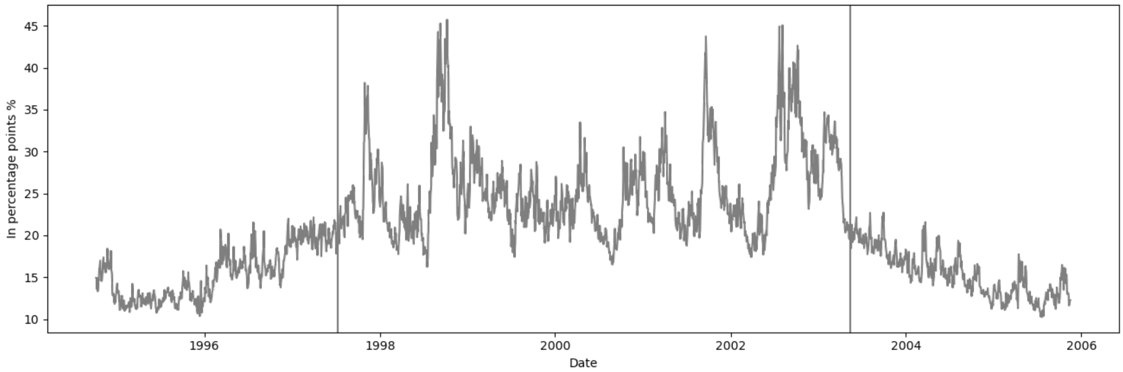

Figure 4

VIX volatility in the dot-com bubble

For the dot-com bubble analysis, the testing period consists of the data on the interval between July 8, 1997, and May 12, 2003. These dates were selected by visually inspecting the VIX. The VIX index around this time period is shown in Figure 4, the starting- and ending points are indicated by the vertical lines. When looking at the VIX, July 1997 is one of the last months before a time period of continuous high volatility up until May 2003. In this paper, this time frame will therefore be considered as the dot-com bubble.

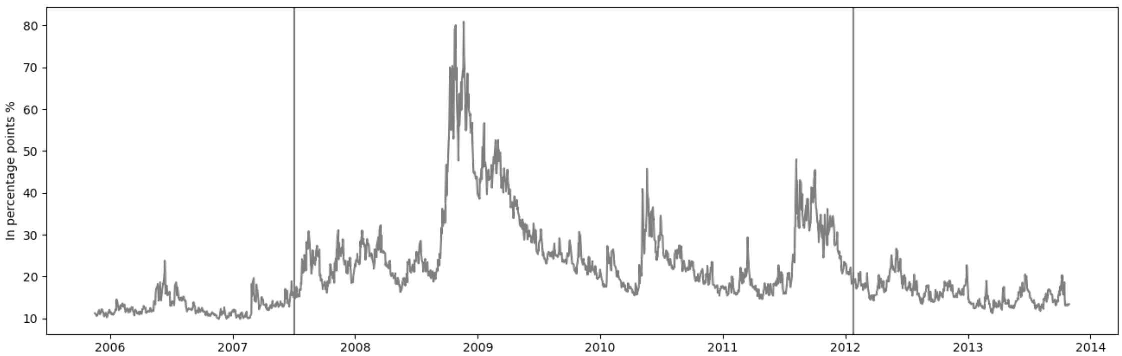

Figure 5

VIX volatility in the financial crisis

For the financial crisis analysis, the testing period consists of the data in-between July 1, 2007, and January 23, 2012. Once again the testing period is selected by inspecting the VIX. The VIX index around this time period is shown in Figure 5. July 2007 is one of the last months before the high volatility period that lasted until 2012.

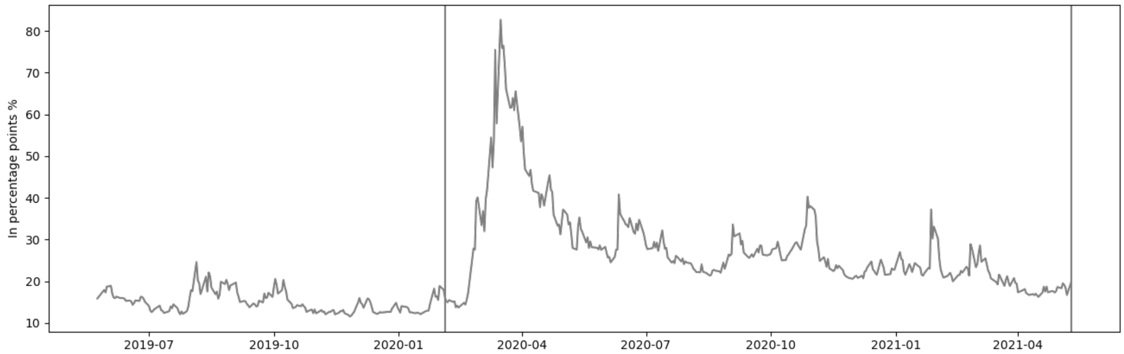

Figure 6

VIX volatility in the COVID-19 crisis

For the COVID-19 crisis the analysis, the testing period consists of the data from February 1, 2020, until May 11, 2021. The VIX index around this time period is shown in Figure 6. The big spike in volatility in February is used as the starting point of the crisis.

Evaluating measures

As volatility cannot be observed it must be estimated as well. As Poon & Granger (2003) state: “In finance, volatility is often used to refer to the standard deviation”. The benchmark used for volatility estimation in this paper is the moving average of the last trading month, which generally contains 21 days. As a robustness check, the predictions are also compared to the moving average of the last trading week, which generally consists of five days. The five-day moving average and 21 day moving averages are calculated as follows:

$$\hat{\sigma}_t^2 = \frac{1}{5}\sum_{k=1}^{5}(y_{t-k} - \bar{y_t})^2 \tag{9}$$

$$\hat{\sigma}_t^2 = \frac{1}{21}\sum_{k=1}^{21}(y_{t-k} - \bar{y_t})^2 \tag{10}$$

Where $\hat{\sigma}_t^2$ denotes the estimated variance, y_{t-k} denotes the lagged returns and $\bar{y_t}$ denotes the mean of the observation of the rolling window. The daily volatilities are then multiplied by a factor of $\sqrt{252}$ to annualize the results.

To test the accuracy of the prediction series the Mean Squared Error (MSE) is used. The MSE is calculated as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t - \hat{\sigma}_t)^2 \tag{11}$$

Where σ_t denotes the benchmark volatility value at time t and $\hat{\sigma}_t$ denotes the predicted value at time t.

The Mean Squared Error (MSE) is chosen as the accuracy measure because it can be rewritten as the sum of the variance and the bias of the estimator (Wackerly, Mendenhall, & Scheaffer, 2008):

$$MSE(\hat{\theta}) = Var_\theta(\hat{\theta}) + Bias(\hat{\theta}, \theta)^2 \tag{12}$$

The θ denotes the true population and the $\hat{\theta}$ denotes the estimator. In this paper, the benchmark volatility σ_t will be set as the true population θ and the predicted series $\hat{\sigma}_t$ will be set as the estimator $\hat{\theta}$. It will therefore test both the variance and the bias of the predicted series compared to the benchmark volatility.

As the mean squared error punishes large forecasting errors very heavily, a second evaluating measure is added. The Mean Absolute Error (MAE) is very much like the MSE, but uses absolute errors instead of squared errors. The MAE is calculated as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\sigma_t - \hat{\sigma}_t| \tag{13}$$

To compare the MSE of two different combinations of model and distribution a two-sided Diebold-Mariano (DM) test is performed, first introduced by Diebold and Mariano (1995). The hypotheses are described below:

$$H_0 : MSE_1 - MSE_0 = 0 \, versus \, H_1 : MSE_1 - MSE_0 \neq 0 \tag{14}$$

Where H_0 denotes the null hypothesis and H_1 denotes the alternative hypothesis.

The MSE_0 is the MSE of the benchmark model and the MSE_1 is the MSE of the model being compared. The Diebold-Mariano (DM) test statistic is calculated as follows:

$$DM = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\hat{S}^2}} \sum_{t=T_1}^{T_2} ((\sigma_t - \hat{\sigma}_{1,t})^2 - (\sigma_t - \hat{\sigma}_{2,t})^2) \sim N(0,1) \tag{15}$$

Where n denotes the number of observations in the sample period. T_1 and T_2 denote the starting point and the ending point of the predicted series, respectively. $\hat{S}^2$ denotes the heteroscedasticity and autocorrelation consistent (HAC) robust covariance matrix. This matrix is estimated using the Newey-West procedure (Newey & West, 1987). σ_t denotes the benchmark volatility on time t. $\hat{\sigma}_{1,t}$ denotes the volatility predictions by the model being compared. $\hat{\sigma}_{2,t}$ denotes the benchmark model. For each day in the sample, the DM-statistic compares the squared errors of the competing models.

Comparing prediction series

At first, all mean squared errors and mean absolute errors are calculated and summarized. Afterwards, Diebold-Mariano tests are done to test the significance of the differences found in MSE. All tests were performed using the 21-day moving average as the volatility benchmark. The tests were repeated using the five-day moving average volatility benchmark as a robustness check. The results were nearly identical.

Each model will compare between its assumed distributions. A DM-test is performed between the predictions made by the normally distributed model and the t-distributed model and between the normally distributed model and the skewed t-distributed model. Lastly, the t-distributed model is compared to the skewed t-distributed model. These tests are repeated for all four models on all three data sets. After analysing the results of the test for each model, the best performing assumed distribution is used to compare the different models. The remaining distributions are put aside. Once again a DM-test is performed to compare the performance of the models. All four models are tested against each other.

6 Results

Table 2

MSE and MAE of the forecasts of all combinations and distributions

	S&P500		NASDAQ		Dow-Jones	
Model - Distribution	MSE	MAE	MSE	MAE	MSE	MAE
GARCH Normal	0.01087	0.08352	0.01722	0.10922	0.01015	0.07958
GARCH t	**0.00989**	0.08061	**0.01598**	**0.10625**	**0.00939**	0.07734
GARCH Skewed t	0.00995	0.08098	0.01616	0.10706	0.00944	0.07774
EGARCH Normal	0.01194	0.08476	0.01892	0.11189	0.01129	0.08090
EGARCH t	0.01003	**0.07947**	0.01708	0.10763	0.00959	**0.07646**
EGARCH Skewed t	0.01028	0.08052	0.01724	0.10830	0.00982	0.07743
GJR-GARCH Normal	0.01168	0.08521	0.01868	0.11221	0.01101	0.08095
GJR-GARCH t	0.01086	0.08257	0.01819	0.11088	0.01041	0.07963
GJR-GARCH Skewed t	0.01093	0.08298	0.01787	0.11046	0.01050	0.08012
TGARCH Normal	0.01220	0.08697	0.01957	0.11465	0.01153	0.08254
TGARCH t	0.01164	0.08541	0.01845	0.11212	0.01100	0.08153
TGARCH Skewed t	0.01171	0.08575	0.01850	0.11242	0.01110	0.08192

Notes: MAE and MSE of the forecasts of all combinations of models and distributions. The best-performing forecasts are denoted in bold.

Table 2 shows the MSE and MAE of all combinations of models and their assumed distributions for all three data sets. The combination with the lowest MSE or MAE is denoted in bold for each data set. When the MSE is used as evaluating measure the GARCH model with a t-distribution delivers the best forecasts. When using the MAE the results differ across data sets. For the NASDAQ data set the GARCH model with a t-distribution performs best, but for the S&P500 and Dow-Jones data sets the model is outperformed by the EGARCH model with a t-distribution. Both the t- and the skewed t-distribution outperform the normal distribution for all models on all data sets. For the GARCH, EGARCH, and TGARCH models the t-distribution slightly outperforms the skewed t-distribution. Only the GJR-GARCH model on the NASDAQ data set provides better forecasts assuming a skewed t-distribution over a t-distribution. Overall, Table 2 shows that the GARCH and

EGARCH models with t-distribution provide the best forecasts.

As many of these results differ only slightly, Diebold-Mariano tests are done to test the significance of these differences.

Testing between distributions

Table 3

GARCH model distributions compared on the S&P500 data set

		Base	
Test	Normal	Student's t	Skewed t
Normal	X	1.000 (-12.54)	1.000 (-11.70)
Student's t	0.000*** (12.54)	X	0.000*** (-15.08)
Skewed t	0.000*** (11.70)	1.000 (15.08)	X

Notes: Diebold-Mariano tests p-values and test statistics. Test statistics are denoted in brackets. Base indicates the benchmark distribution. Test indicates the distribution that is being tested against the benchmark distribution. * indicates the significance level. '*' is significant on a 90% confidence level, '**' is significant on a 95% confidence level,'***' is significant on a 99% confidence level.

Table 4

GJR-GARCH model distributions compared on the NASDAQ data set

		Base	
Test	Normal	Student's t	Skewed t
Normal	X	1.000 (-4.02)	1.000 (-6.62)
Student's t	0.000*** (4.02)	X	1.000 (-11.88)
Skewed t	0.000*** (6.62)	0.000*** (-11.88)	X

Notes: Diebold-Mariano tests p-values and test statistics. Test statistics are denoted in brackets. Base indicates the benchmark distribution. Test indicates the distribution that is being tested against the benchmark distribution. * indicates the significance level. '*' is significant on a 90% confidence level, '**' is significant on a 95% confidence level,'***' is significant on a 99% confidence level.

To test which distribution assumption delivers the best predictions the normal, t- and skewed t-distributions are tested against each other. The distributions are tested separately on all four models on all three data sets. In Table 3 the results of the GARCH model on the S&P500 data are presented. As seen in the table both the t and the skewed t-distribution outperform the normal distribution on a 99% confidence level. Surprisingly, however, the t-distribution significantly outperforms the skewed student's t distribution.

The tests from the EGARCH, GJR-GARCH, and TGARCH models all show very similar results. The t- and skewed t-distribution outperform the normal distribution and the t-distribution outperforms the skewed t-distribution, all on a 99% confidence level. The same holds when the tests are repeated over the NASDAQ and Dow-Jones data sets. The only outlier is the GJR-GARCH model on the NASDAQ data set. These results are reported in Table 4.

As for nearly all combinations of models and data sets the student's t assumed distribution performs best, the comparison between models will be done assuming a t-distribution.

Model comparison

Table 5

The GARCH, EGARCH, GJR-GARCH and TGARCH models compared on the S&P500 data set

	Base			
Test	GARCH	EGARCH	GJR-GARCH	TGARCH
GARCH	X	0.113 (1.59)	0.000*** (6.58)	0.000*** (10.87)
EGARCH	0.887 (-1.59)	X	0.000*** (5.35)	0.000*** (9.84)
GJR-GARCH	1.000 (-6.58)	1.000 (-5.35)	X	0.000*** (8.72)
TGARCH	1.000 (-10.87)	1.000 (-9.84)	1.000 (-8.72)	X

Notes: Diebold-Mariano tests p-values and test statistics. Test statistics are denoted in brackets. Base indicates the benchmark model. Test indicates the model that is being tested against the benchmark model. * indicates the significance level. '*' is significant on a 90% confidence level, '**' is significant on a 95% confidence level,'***' is significant on a 99% confidence level.

Table 6

The GARCH, EGARCH, GJR-GARCH and TGARCH models compared on the NASDAQ data set

	Base			
Test	GARCH	EGARCH	GJR-GARCH	TGARCH
GARCH	X	0.000*** (9.16)	0.000*** (12.65)	0.000*** (12.95)
EGARCH	0.887 (-9.16)	X	0.000*** (6.36)	0.000*** (9.39)
GJR-GARCH	1.000 (-12.65)	1.000 (-5.35)	X	0.057* (1.91)
TGARCH	1.000 (-12.95)	1.000 (-9.84)	0.943 (-1.91)	X

Notes: Diebold-Mariano tests p-values and test statistics. Test statistics are denoted in brackets. Base indicates the benchmark model. Test indicates the model that is being tested against the benchmark model. * indicates the significance level. '*' is significant on a 90% confidence level, '**' is significant on a 95% confidence level,'***' is significant on a 99% confidence level.

Table 7

The GARCH, EGARCH, GJR-GARCH and TGARCH models compared on the Dow-Jones data set

	Base			
Test	GARCH	EGARCH	GJR-GARCH	TGARCH
GARCH	X	0.030** (2.18)	0.000*** (7.15)	0.000*** (9.79)
EGARCH	0.970 (-2.18)	X	0.000*** (5.67)	0.000*** (9.26)
GJR-GARCH	1.000 (-7.15)	1.000 (-5.67)	X	0.057* (7.37)
TGARCH	1.000 (-9.79)	1.000 (-9.26)	0.943 (-7.37)	X

Notes: Diebold-Mariano tests p-values and test statistics. Test statistics are denoted in brackets. Base indicates the benchmark model. Test indicates the model that is being tested against the benchmark model. * indicates the significance level. '*' is significant on a 90% confidence level, '**' is significant on a 95% confidence level,'***' is significant on a 99% confidence level.

The results of the Diebold-Mariano tests between models are presented in Table 5, Table 6, and Table 7. Table 5 presents the model comparison of the S&P500 data set. Table 6 and Table 7 present the model comparison of the NASDAQ and Dow-Jones data sets, respectively. Somewhat surprisingly the GARCH model seems to produce the best predictions on all data sets. It outperforms the GJR-GARCH and TGARCH models on all three data sets on a very significant level. On the S&P500 data set the GARCH model fails to significantly outperform the EGARCH model, but it does on the NASDAQ and Dow-Jones data sets. It outperforms the EGARCH model on a 99% confidence level on the NASDAQ data set and outperforms it on a 95% confidence level on the Dow-Jones data set. The GJR-GARCH model outperforms the TGARCH model on a 90% or higher confidence interval on all data sets. Both the GARCH and the EGARCH model outperform the GJR-GARCH, and therefore the TGARCH, model on a 99% confidence interval on all data sets. Overall based on MSE, the GARCH model seems to perform best, the EGARCH model second, and the GJR-GARCH and TGARCH model third and fourth, respectively.

7 Conclusion

To conclude, the results will be set out against the outstanding literature. The fat tails and leverage effect that were documented in previous literature are present in all three data sets. All data sets exhibit a negative skewness and high excess kurtosis. The results confirm the findings of previous literature in the fact that an assumed t-distribution fits the data better than a normal distribution. Somewhat surprisingly, however, the skewed t-distribution fails to outperform the t-distribution on all but the GJR-GARCH model on the NASDAQ data set. Studying the literature reviewed in this paper and looking at the descriptive statistics of the data sets would lead one to assume the skewed t-distribution to perform best.

Additionally, according to the majority of the literature stemming from the 1990s and 2000s, the EGARCH, GJR-GARCH, and TGARCH models are expected to outperform the GARCH model. Many researchers attribute this superiority over the standard GARCH model to the ability of these models to capture the asymmetric properties of stock returns. As seen in the descriptive statistics the negative skewness of the data implies the presence of asymmetry in all data sets. Especially in times of crisis, the asymmetric models are expected to be able to forecast more accurately as they should be able to model the big negative shocks that are typically associated with stock market crashes. Once again the results conflict with the expectations. None of the three asymmetric models manage to outperform the standard GARCH model. The GARCH model even manages to significantly outperform the GJR-GARCH and TGARCH model on all three data sets.

The results are conflicting with the outstanding literature. As much of the research on the four models was done in the 1990s and 2000s, it could mean the literature is dated. All models, however, are still frequently used, as they are easy to apply and understand. Managers deciding on what model to use could base their decisions on dated research, which would lead them to choose one of the asymmetric models. As the accuracy of volatility forecasts is the most important in times of crisis, using one of the asymmetric models over the standard GARCH model could result in poor performances. The results of this paper show that, out of the four models, the most accurate forecasts in times of crisis are provided by the GARCH model assuming a t-distribution, contrary to prior belief.

The data analyzed in this paper strictly focuses on times of crisis. Financial managers deciding on what models to use will most likely want to base their decisions on the performance of a model on a data set broader than solely in times of crisis. The relative performance could differ when looking at data including time periods with stable market conditions. Research analyzing data from stable periods could find opposing results. The findings do, however, allow managers to prepare for financial crises by selecting models that perform well in them. Managers could choose to switch between models in stable periods and times of financial turmoil or risk-averse managers could choose to always use crisis robust models. More research should be done on the subject to accurately uncover the relative performances of the models and distributions. The research could be repeated on different indices, specific stocks, or industries.

Testing periods could be expanded with older crises, switched to stable periods, or combine the two. The models in this paper are all estimated on a five-year rolling window. Changing the rolling window length could also affect the relative performance of the models and distributions. Finally, as there are constantly new models being developed, adding different models could find better-performing models. Strictly focusing on these four models in times of financial crises, however, the results show that the asymmetric models do not significantly outperform the GARCH model. Providing motivation for future research into the relative performances of these popular volatility forecasting models.

8 References

Akgiray, V. (1989). Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *The Journal of Business*, *62*(1), 55–80.

Alberg, D., Shalit, H., & Yosef, R. (2008). Estimating stock market volatility using asymmetric garch models. *Applied Financial Economics*, *18*(15), 1201-1208.

Baillie, R. T., & DeGennaro, R. P. (1990). Stock returns and volatility. *Journal of financial and Quantitative Analysis*, *25*(2), 203-214.

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Bouchaud, J. P., Matacz, A., & Potters, M. (2001). Leverage effect in financial markets: The retarded volatility model. *Physical review letters*, *87*(22), 228701.

Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, *20*(3), 419-438.

Cont, R., & Bouchaud, J. P. (2000). Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic dynamics*, *4*(2), 170-196.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134-144.

Ding, Z., Granger, C. W., & Engle, R. F. . (1993). A long memory property of stock market returns and a new model. *Econometrica: Journal of the econometric society*, *1*(1), 83-106.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987-1007.

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

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, *20*(7), 873-889.

Li, W. K., Ling, S., & McAleer, M. (2002). Recent theoretical results for time series models with garch errors. *Journal of Economic Surveys*, *16*(3), 245-269.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, *59*(2), 347-370.

Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(2), 703-708.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of econometrics*, *45*(1-2), 267-290.

Poon, S. H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, *41*(2), 478-539.

Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications*. Belmont, California: Cengage Learning.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, *18*(5), 931-955.