Forecasting the volatility and density of returns on the Tel Aviv index: frequentist and Bayesian estimation of GARCH models

Sophia Esmeralda Dekker

(2594564)

July 2019

Bachelor Thesis Econometrics and Operations Research

Thesis commission:

Dr. Hoogerheide (supervisor)

Dr. A. Borowska (co-reader)

Abstract

The goal of this thesis is to develop a model for the returns of different indices: Tel Aviv Index 25 and Tel Aviv 100. The model should take care of the conditional variance. The model must also be able to capture some stylized facts: volatility clustering, fat tails (leptokurtosis), leverage effect and nonlinear dependence. The models for volatility that we are going to take a look at, with different densities are: GARCH, GJR-GARCH, EGARCH and APARCH. The densities are symmetric and asymmetric. To know which model performs better, in comparison to the others, the forecast ability of all models is going to be inspected. The results show that indeed (as already concluded in Alberg et al. (2008)) asymmetric GARCH models with fat-tailed densities are better to use for measuring conditional variance than the other models. The EGARCH model performs best in forcasting TASE indices. The different models are estimated using two approaches: the frequentist approach and the Bayesian approach.

Keywords: Forecasting, GARCH, day-of-the-week, TASE, Prediction Performance, Optimization,frequentist,Bayesian.

1 Introduction

Investors are interested in indices. In this world, there are many of those indices. A well known (most followed) index is the SP500. This one contains 500 of the biggest American companies. In this way, it gives an insight in the U.S. equities. Another index is the Tel Aviv Stock Exchange (TASE). It is of high importance (for e.g. portfolio analysis) to predict future returns/volatility. How are the returns produced? Is there a model behind this process? And which model captures most of the stylized facts? Is it the GARCH/GJR/EGARCH/APARCH model? And what could be the distribution of the innovations? This paper is structured as follows: in Section 2 the return characteristics are discussed. After that, in Section 3, an introduction of different volatility models follows. Then, in Section 4, an analysis of the three different densities used for the innovations is brought up. For In-Sample comparison of the models, we are going to use the Log Likelihood/AIC and BIC. For Out-Of-Sample comparison we are going to use the performance measures of Section 5. Then, finally we are going to take a look at the data of the TA25 and TA100. In Section 7, the parameters are estimated. Before doing that, the day-of-the-week effect is taken into account. Normally by looking at returns, this is not taken into account. However, since we want to follow the paper of Alberg et al. (2008), we will perform these calculations. Then we will follow up with the conclusions. After that, in Section 9, we will see how robust the conclusions are. By enhancing the Out-Of-Sample period from 30 to 500, we will still see the same results! In Section 10 the Article of Hentschel (1995) is studied. This paper takes a look at the AGARCH and TGARCH model, so we do this too! Only instead of the data of the original article, we will use the data of the TA25 and TA100. Then in Section 9.2 a short trip to another index is made: the Amsterdam Exchange index is studied. Last but not least in Section 11 we conclude with some Bayesian analysis based on the article of Hoogerheide et al. (2012). It is convenient to use the Bayesian approach because of the change of perspective. Instead of thinking of parameters being point estimates, they have a certain distribution in the latter framework! We will find a posterior interval for different volatility models. Also a comparison of the forecasting performance between the Bayesian and frequentist estimation method is made. While the model stays the same (only a prior density is added in the Bayesian approach, the method is certainly different.

2 Returns

2.1 Return Characteristics

Often we have to deal with time series which are not stationary. It is difficult to make an analysis of non stationary time series data. An example of non stationary time series are the end closing prices of the Tel Aviv Index. To deal with this kind of data, it is possible to look at the returns instead.
First, it is important to understand what returns are. Returns represent the relative change in price of a financial asset, over a given time interval. Often, returns are expressed as percentage. There exist two types of returns: simple returns and compound returns. The formulas for both types are stated below.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \qquad \text{(Simple Returns)}$$

$$Y_t = \log(P_t) - \log(P_{t-1}) \qquad \text{(Compound Returns)}$$

In a strict sense, only simple returns are correct. However, compound returns are handy for calculations. The mathematics of compound returns are easier, and the difference between R_t and Y_t is small when we talk about small returns (i.e. daily). Therefore, in this paper, we are going to use Compound Returns.

2.2 Characteristic of returns: Volatility

One of the characteristics of returns is volatility. It is the most common measure of market uncertainty and thus important for investors. There is clear evidence of cyclical patterns in volatility over time, both in the short run and in the long run. This means that volatility may be (partially) predictable. It is also handy to know that predicted volatility in a model assuming normally distributed returns represents risk if and only if the returns are normally distributed. However, keep in mind that this is often not the case! We will see in Table 5 that normality is rejected for the TA25 and TA100. If one wrongly assumes normality, the volatility model systematically underestimates risk.
The volatility can be seen as a measure of variability of asset prices over some time period. Typically, we consider the standard deviation of returns. It is of high importance in many financial decision problems: for portfolio construction, risk management and option pricing. In literature, there exist two types of volatility: historical and conditional volatility. The notation for historical (unconditional) volatility is: σ (notice there is no subindex with time). This is the volatility for a whole time period. The other one is conditional volatility: σ_t. This is volatility conditional on a given time period. In this paper, we want to model the conditional volatility of different indices. Note that, ex post, we can use squared returns $(Y_t - \mu)^2$ or just $(Y_t)^2$ (as a simple approximation of the variance at day t). However, we want ex ante a forecast of volatility: $\sigma_t^2 = E[(Y_t - \mu)^2 | \Omega_{t-1}]$.

2.3 Characteristic of returns: Fat tails

Another characteristic of returns is the presence of fat tails. This means that the random variable (returns) has more extreme outcomes than a normally distributed random variable with the same mean and variance. If the tails are fat, there is a higher probability of extreme outcomes than one would get from a normal distribution. This also implies a lower probability of non-extreme outcomes. In literature this is called leptokurtosis. If returns are normalized, by the time-varying variances, the leptokurtosis is reduced. The question of the reader might be: how to know if there are fat tails? There are two ways to find out! One is the graphical way: a QQ-plot might help. However, sometimes the statistical way is better. By performing a Jarque-Bera test or a Kolmogorov-Smirnov test for example! The advantage of the last one is that it is non-parametric and distribution free.

2.4 Characteristic of returns: Nonlinear dependence

Returns suffer also from something called: non-linear dependence. The correlation between two variables, x and y is represented as: $y = \alpha x + \epsilon$. This α represents the linear relationship. However, when there is non-linearity, the relationship between the variables depend on some (market)factor. The way to detect nonlinear dependence is by looking at exceedence correlations.

2.5 Characteristic of returns: Asymmetry

According to the article of Hentschel (1995), the effect of equity returns on the variance in the following periods is strongly asymmetric. This means that after a negative return, the volatility increases more than after an equally large positive return. In literature this is called: 'the leverage effect'. This is also important for the debt to equity ratio. This is the total liability of a company divided by the shareholders equity. According to the Article of Black and Scholes (1976): when the equity goes down, and the debt goes not equally down: the debt-to-equity ratio would increase. This means that the risk would increase. However, there is not enough evidence with only the financial leverage effect to explain the asymmetry. To model the asymmetric effect of returns, asymmetric models such as EGARCH are needed!

3 Different volatility models

3.1 Introduction to: Conditional Variance and Distribution of the Innovations

To somehow capture different aspects of the time-varying volatility we need a model for the conditional variance. The different aspects are: volatility clustering, leptokurtosis and the leverage effect (asymmetric impacts of negative vs. positive returns). In literature, many of such conditional models exist. In this, and the following sections, we are going to discuss the advantages and disadvantages of five of such models. We will take a look at: ARCH, GARCH, EGARCH, GJR-GARCH and APARCH models. First the mathematics of the model is stated, and then the model will be explained. Keep in mind that the ultimate goal is to estimate and predict volatility of the returns. In later sections, when prediction takes place, we need some parameter restrictions. To have all the model information at the same place, the parameter restrictions are already stated here. These restrictions are made to guarantee stationary time series and positive variance. This thesis also tries to verify the results found in Alberg et al. (2008). Are the parameter estimates kind of similar? We will find this out in Section 7.2.

3.2 ARCH

ARCH(q)

$y_t = E(y_t \mid I_{t-1}) + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\sigma_t^2 = \omega + \sum_{i=1}^{q}(\alpha_i \epsilon_{t-i}^2)$
$z_t \sim i.i.d(0,1)$

Parameter Restrictions ARCH(1)

$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2$
$\theta = (\mu, \omega, \alpha)$
$\omega > 0\,, \alpha \geq 0$

To model the conditional variance, Engle (1982) proposed a model that uses the lagged disturbances. So the "news" of the last period, has influence on the volatility of this period. This so called Autoregressive Conditional Heteroskedasticity (ARCH) model needs a lot of lags to capture all the dynamic behaviour of the conditional variance. This automatically implies that a lot of estimated parameters are needed. The advantage of the model is that it captures volatility clustering and leptokurtosis. The disadvantage of this model is that it does not capture the leverage effect.

3.3 GARCH

GARCH(p,q)

$\text{y}_t = E(y_t \mid I_{t-1}) + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\sigma_t^2 = \omega + \sum_{i=1}^{q}(\alpha_i\epsilon_{t-i}^2) + \sum_{j=1}^{p}(\beta_j\sigma_{t-j}^2)$
$z_t \sim i.i.d(0,1)$

Parameter Restrictions GARCH(1,1)

$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
$\theta = (\mu, \omega, \alpha, \beta)$
$\omega > 0, \alpha \geq 0, \beta \geq 0\,, \alpha + \beta < 1$

The ARCH model did only take the "news" factor into account (the α). It is also possible to take a "memory" factor into account (the β). Then we are talking about the GARCH model. This model, by Bollerslev (1986), does not need a high order of lags to capture the dynamic behaviour of conditional variance. It can be shown that a GARCH(1,1) model is the same as an ARCH(∞) model! The advantage of this model is the same as the ARCH model, it captures volatility clustering and leptokurtosis. It does not capture the leverage effect.

3.4 GJR-GARCH

GJR-GARCH(p,q)

$\text{y}_t = \mu + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\sigma_t^2 = \omega + \sum_{i=1}^{q}(\alpha_i + \gamma_i\mathbb{1}_{t-i})\epsilon_{t-i}^2 + \sum_{j=1}^{p}(\beta_j\sigma_{t-j}^2)$
$\mathbb{1}_{t-i} = \begin{cases} 1, & \text{if } y_{t-i} \leq \mu \\ 0, & \text{if } y_{t-i} > \mu \end{cases}$
$z_t \sim i.i.d(0,1)$

Parameter Restrictions GJR-GARCH(1,1)

$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma\mathbb{1}_{[\epsilon_{t-1}<0]}\epsilon_{t-1}^2$
$\theta = (\mu, \omega, \alpha, \beta, \gamma)$
$\omega > 0\,, \alpha \geq 0\,, \beta \geq 0\,, \alpha + \gamma \geq 0\,,\ \alpha + \beta + \gamma E[\mathbb{1}_{[\epsilon_{t-1}<0]}] < 1$
$E[\mathbb{1}_{[\epsilon_{t-1}<0]}]$ depends on the distribution of z_t

$$E[\mathbb{1}_{[\epsilon_{t-1}<0]}] = \begin{cases} \frac{1}{2}, & \text{if } z_t \text{ is normal or student-t distributed} \\ \frac{1}{1+\xi^2}, & \text{if } z_t \text{ is skewed student-t distributed} \end{cases}$$

The two models discussed above did not capture the leverage effect. To take this factor into account, a non-linear extension of the GARCH model is needed! Glosten et al. (1993) came up with such a non-linear model: the GJR-GARCH model. The γ can capture the asymmetric volatility clustering. This parameter models the leverage effect. The indicator function is one if the shock of last period is less than expected. So negative shocks lead to a higher volatility the next day in comparison to positive shocks (keep in mind that γ is nonnegative). The indicator function is zero if the shock of last period is more than expected. Notice that when $\gamma = 0$, the GJR-GARCH model is just a GARCH model.

3.5 EGARCH

EGARCH(p,q)

$\mathrm{y}_t = E(y_t \mid I_{t-1}) + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q}(\alpha_i(|z_{t-i}| - \mathbb{E}(|z_{t-i}|)) + \gamma_i z_{t-i}) + \sum_{j=1}^{p}(\beta_j \log(\sigma_{t-j}^2))$

$$\mathbb{E}(|z_t|) = \begin{cases} \sqrt{\frac{2}{\pi}}, & \text{if } z_t \sim N(0,1) \\ 2\sqrt{\frac{\upsilon-2}{\pi}}\frac{\Gamma(\frac{1+\upsilon}{2})}{\Gamma(\frac{\upsilon}{2})}, & \text{if } z_t \sim \text{Student-t}(0,1,\upsilon) \\ \frac{4\xi^2}{\xi+\frac{1}{\xi}}\sqrt{\frac{\upsilon-2}{\pi}}\frac{\Gamma(\frac{1+\upsilon}{2})}{\Gamma(\frac{\upsilon}{2})}, & \text{if } z_t \sim \text{Skewed Student-t}(0,1,\upsilon,\xi) \end{cases}$$

$\mathrm{z}_t \sim i.i.d(0,1)$

Parameter Restrictions EGARCH(1,1)

$\log(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - \mathbb{E}(|z_{t-1}|)) + \gamma z_{t-1} + \beta\log(\sigma_{t-1}^2)$
$\theta = (\mu, \omega, \alpha, \beta, \gamma)$
$\mid \beta \mid < 1$

The EGARCH model can capture volatility clustering, leptokurtosis and the leverage effect. It was developed by Nelson (1991). Because of the logarithmic transformation, the volatility always stays positive. Not many parameter restrictions are needed! Just one, to guarantee stationary time series.

3.6 APARCH

APARCH(p,q)

$y_t = \mu + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i((|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^\delta) + \sum_{j=1}^{p}(\beta_j \sigma_{t-j}^\delta)$
$z_t \sim i.i.d(0,1)$

Parameter Restrictions APARCH(1,1)

$\sigma_t^\delta = \omega + \alpha_1(|\epsilon_{t-1}| - \gamma\epsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$
$\theta = (\mu, \omega, \alpha, \beta, \gamma, \delta)$
$\omega > 0$,$\alpha \geq 0$,$\beta \geq 0$,$\delta > 0$, $|\ \gamma\ | < 1$, and $\alpha E[(|\ z_t\ | - \gamma z)^\delta] + \beta < 1$
E$[(|\ z_t\ | - \gamma z)^\delta]$ depends on the assumed distribution for the innovations
For the Normal and Skewed Student-t distributions it is stated below.
Student-t follows from Skewed Student-t with $\xi = 1$

$$\mathbb{E}((|\ z_t\ | - \gamma z)^\delta) = \begin{cases} \frac{1}{\sqrt{2\pi}}[(1+\gamma)^\delta + (1-\gamma)^\delta]2^{\frac{\delta-1}{2}}\Gamma(\frac{\delta+1}{2}) \\ \\ [\xi^{-(1+\delta)}(1+\gamma)^\delta + \xi^{(1+\delta)}(1-\gamma)^\delta]\frac{\Gamma(\frac{\delta+1}{2})\Gamma(\frac{\upsilon-\delta}{2})(\upsilon-2)^{\frac{1+\delta}{2}}}{(\xi+\frac{1}{\xi})\sqrt{(\upsilon-2)\pi}\Gamma(\frac{\upsilon}{2})} \end{cases}$$

The APARCH model can capture volatility clustering, leptokurtosis and the leverage effect (modelled by the γ parameter). It was developed by Ding et al. (1993). The delta parameter: δ plays an important role in measuring the serial dependence of ϵ_t. There is a correlation between different $|\epsilon_t|^\delta$. This parameter is used for a Box-Cox power transformation on σ_t.

4 Different density functions

4.1 Introduction to: Distribution of the Innovations

The returns of the indices often have very thick tails in comparison to the normal distribution. If the tails where just slightly thicker than the tails of a normal distribution then this could be caused by the heteroscedasticity. Namely: if the data follows a GARCH-normal model, then the conditional distribution of y_t (given the past: y_{t-1}, y_{t-2}, etc) is normal. But the unconditional distribution is not normal but has slightly thicker tails because σ_t^2 is different for different t. The periods with a relative high σ_t^2 then causes the thicker tails. However, in our case we have $very$ thick tails, and not slightly thicker tails. To account for this, the GARCH models stated above should have a different distribution for the innovations. For example: Student-t or Skewed Student-t. The last one is an example of an asymmetric distribution. In this section, the densities are presented. The formulas are used for the programming part. The programming part is done in OxMetrics.

4.2 Innovations: Normal Distributed

Normal Distribution

$$\mathrm{L}(\mathrm{y}_t|I_{t-1};\theta) = \frac{1}{\sqrt{2\pi}\sigma_t}\exp\left[-\frac{(y_t-\mu)^2}{2\sigma_t^2}\right]$$

$$L(\overrightarrow{y}|I_{t-1};\theta) = \prod_{t=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_t}\exp\left[-\frac{(y_t-\mu)^2}{2\sigma_t^2}\right]$$

• $z_t \sim N(0,1)$ i.i.d

$$f(z_t|\theta) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}z_t^2)$$

$$\log l(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}(\log(\sigma_t^2) + z_t^2)$$

This is the normal (Gaussian) density. It has a skewness of 0 and a kurtosis of 3. The innovations are $\epsilon_t = z_t\sigma_t$. With z_t from a standard normal distribution.

4.3 Innovations: Student-t Distributed

Student-t Distribution

$$\mathrm{L}(\mathrm{y}_t|I_{t-1};\theta) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu-2)\pi\sigma_t^2)}}(1+\frac{(y_t-\mu)^2}{(\nu-2)\sigma_t^2})^{\frac{-\nu+1}{2}}$$

$$L(\overrightarrow{y}|I_{t-1};\theta) = \left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu-2)\pi)}}\right)^{n}\prod_{t=1}^{n}\frac{1}{\sigma^t}\prod_{t=1}^{n}\left(1+\frac{(y_t-\mu)^2}{\nu\sigma_t^2}\right)^{-\frac{\nu+1}{2}}$$

•$\mathrm{z}_t \sim$ Student t (0,1) i.i.d

$$\mathrm{f}(\mathrm{z}_t|\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left(1+\frac{z_t^2}{\nu-2}\right)^{-(\nu+1)/2}$$

$$\log l(\theta) = n\left[log\Gamma(\frac{\nu+1}{2}) - \log\Gamma(\frac{\nu}{2}) - \frac{1}{2}\log(\pi(\nu-2))\right] - \frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + (1+\nu)\log(1+\frac{z_t^2}{\nu-2})\right]$$

Multiple studies have used the Student-t distribution: Bollerslev (1987), Baillie and Bollerslev (1989) and Beine et al. (2002). This distribution can capture fat tails of returns. The parameter ν represents the degrees of freedom. If it is close to infinity, this distribution is close to the normal one. If ν is very low (around 4) the tails are fat. The kurtosis is equal to $3(\nu-2)/(\nu-4)$ for $\nu > 4$ and it is ∞ for $2 < \nu \leq 4$.

4.4 Innovations: Skewed Student-t Distributed

Skewed Student-t Distribution

$$L(y_t|I_{t-1};\theta) = \frac{2}{\xi+\frac{1}{\xi}}\left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu-2)\pi\sigma_t^2}}\right)\frac{1}{\sigma_t}\Bigg[1+\frac{(y_t-\mu)^2}{(\nu-2)\sigma_t^2}$$
$$\left\{\frac{1}{\xi^2}\mathbb{1}[y_t>\mu]+\xi^2\mathbb{1}[y_t<\mu]\right\}\Bigg]^{-\frac{\nu+1}{2}}$$

$L(\overrightarrow{y}|I_{t-1};\theta) = \prod_{t=1}^{n} L(y_t|I_{t-1};\theta)$

- $z_t \sim$ Skewed Student-t$(0,1)$i.i.d

$f(z_t|\theta) = \frac{2}{\xi+\frac{1}{\xi}} s\Big[g((sz_t+m)\xi|\nu)\mathbb{1}[z_t+\frac{m}{s}<0]+g((sz_t+m)/\xi|\nu)\mathbb{1}[z_t+\frac{m}{s}\geq 0]\Big]$

- $g(.)$ is a symmetric density with an expectation of 0 and a variance of 1

$\log l(\theta) = n\Big[\log\Gamma(\frac{\nu+1}{2}) - \log\Gamma(\frac{\nu}{2}) - \frac{1}{2}\log(\pi(\nu-2)) + \log(\frac{2}{\xi+\frac{1}{\xi}}) + \log(s)\Big] - \frac{1}{2}\sum_{t=1}^{n}\Big[\log(\sigma_t^2)+(1+\nu)\log(1+\frac{(sz_t+m)^2}{\nu-2}\xi^{-2I_t})\Big]$

- $I_t = 1$ if $z_t \geq -\frac{m}{s}$ and $I_t = -1$ else
- $m = \frac{\Gamma(\frac{\nu-1}{2})\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})(\xi-\frac{1}{\xi})}$
- $s = \sqrt{(\xi^2+\frac{1}{\xi^2}-1)-m^2}$

To not only capture skewness, but also kurtosis, a Skewed Student-t comes in hand! This is studied by Lambert and Laurent (2001). This density has two extra parameters in comparison to the normal density: ν and ξ. The first one is for the degrees of freedom and the second one is the asymmetry parameter. If $\xi > 1$: the density is skewed to the right. If $\xi < 1$: the density is skewed to the left. If $\xi = 1$: there is no asymmetry.

5 Performance measurements

5.1 Introduction to: standard performance measures

In the previous sections we talked about models for the conditional variance. If we use these models to predict/forecast future conditional variances, we want to know which model performs best. In order to compare the models we use the below stated measures. It is studied by Poon and Granger (2003). The reason for using multiple performance measures lies in the fact that it is more robust. The advantage/disadvantage of the first two metrics is that they are more affected by outliers. The MedSE and AMAPE are in comparison to the MSE and MAE less affected by outliers.

5.2 Standard Performance Measures

The several metrics that we are going to use are stated below. The meaning of h is: the number of lead steps. That is, the number of out-of-sample observations. This can be for example 30. The meaning of S is: sample size. That is, the number of in-sample observations that are used for the estimation of the models. The meaning of $\hat{\sigma}^2$ is: predicted/forecasted variance. This can be for example the forecasted variance by the GARCH(1,1) model or the EGARCH(1,1) model. The meaning of σ^2 is: the 'actual' variance. In this case we take the actual squared returns, which are obviously just an approximation of the actual variance. Nevertheless, we expect that a better model will yield smaller forecast errors.

Mean Squared Error

MSE = $(\frac{1}{h+1}) \sum_{t=S}^{S+h} (\hat{\sigma_t^2} - \sigma_t^2)^2$

Median Squared Error

MedSE =$Inv(f_{med}(\mathrm{e}_t)$

$\mathrm{e}_t = (\hat{\sigma_t^2} - \sigma_t^2)^2$ and t $\in [S, S+h]$

Median Absolute Error

MAE = $(\frac{1}{h+1}) \sum_{t=S}^{S+h} \mid (\hat{\sigma_t^2} - \sigma_t^2) \mid$

Adjusted Mean Absolute Percentage Error

AMAPE = $(\frac{1}{h+1}) \sum_{t=S}^{S+h} \mid \frac{(\hat{\sigma_t^2} - \sigma_t^2)}{(\hat{\sigma_t^2} + \sigma_t^2)} \mid$

Theil's Inequality Coefficient (TIC)

TIC = $\frac{\sqrt{(\frac{1}{h+1}) \sum_{t=S}^{S+h} (\hat{\sigma_t^2} - \sigma_t^2)^2}}{\sqrt{(\frac{1}{h+1}) \sum_{t=S}^{S+h} (\hat{\sigma_t^2}^2)} + \sqrt{(\frac{1}{h+1}) \sum_{t=S}^{S+h} (\sigma_t^2)^2}}$

For all the measures we want a number that is very low. The TIC is always between zero and one. The model is good when it is close to zero. If the TIC is close to one, then the model performs the same as a random guess.

5.3 Diebold-Mariano Test

Another way of comparing forecasts with each other is via the Diebold-Mariano test. This test is not intended to compare models, it is intended to compare forecasts according to the article of Diebold (2015). This is a test that looks at whether or not there exists a significant difference in the MAE (or an other measure) between two models. For this, we need the 'actual' variance, which is of course unknown. However, we could use the squared returns. We could use those, because the expectation of the squared returns is equal to the real variance (if $E(R_t$=0)). To do this, first the loss differential has to be calculated. This is equal to the difference in absolute prediction errors of two models at a certain point in time. The two models can for example both be GARCH(1,1) but with different distributions for the innovations. Or it could be that one model is the same as the other, only with a restriction on the parameters. We are taking for time: 30 timepoints. In this way we get for example 30 loss differentials d_j. These d_j are variables out of some distribution with expectation μ_d. We need to know: is the expectation equal to zero?! Once all the loss differentials of different time periods are obtained, we can actually test if μ_d=0. For this we use a t-statistic that is obtained as the t-statistic in a regression of the d_j on a constant term, where we use the Newey-West standard error to account for possible serial correlation in the series of the d_j. Under the null hypothesis $H_0 : \mu_d = 0$ this t-statistic has approximately a N(0,1) distribution if the number of out-of-sample observations is large enough (due to the Central Limit Theorem). So we do not reject H_0 if |t-statistic| < 1.96. This means that we do not know whether the two models perform equally well or not. If the t-stat is outside the acceptance region, we know a lot more! This means that we can conclude that one model performs significantly better than the other model. If we would assume that the d_j are independent, then we would perform the Diebold-Mariano test in the following way. Note that we use the Newey-West standard error instead of $S_d/\sqrt{N}$ and that the t-statistic then asymptotically has a N(0,1) distribution under H_0. Notice that the d_j in this test could change. In Section 12 it is for example different.

Diebold-Mariano Test

$H_0 : \mu_d = 0$
$H_1 : \mu_d \neq 0$
T = $\sqrt{N}\frac{d-0}{S_d}$
$d_j =| (\hat{\sigma^2_{j1}} - \sigma^2_j) | - | (\hat{\sigma^2_j} - \sigma^2_{j2}) |$
N= number of out of sample predictions
$S^2_d = \frac{1}{N-1}\sum_{j=1}^{N}(d_j - \bar{d})^2$
T $\sim$Student-t(N-1)

5.4 Logarithmic Scoring Rule

Sometimes the difference between MAE is small for two models with the same equation for the conditional variance but with different distributions for the standardized innovations. After all, although the distribution affects the loglikelihood and therefore the MLE, the distributions do not 'directly' affect the predicted variance. For example with a GARCH(1,1) model with normal innovations in comparison to student-t innovations. Then the logarithmic scoring rule (Diks et al. (2011)) comes in hand! This metric can be seen as an out-of-sample log likelihood. First the parameters are being estimated with only the in-sample data. Then the log-likelihood of the models is calculated. The model with the highest log-likelihood 'wins', so this means that in comparison, it predicts the best. The logarithmic scoring rule already punishes over-fitting, because overfitting harms the out-of-sample performance. This is the reason that we do not need an extra penalty term for the number of parameters. It is possible for a small model to have a higher Log-Score than a bigger model, if the restrictions for the smaller model are correct. We could also test for this scoring rule by testing whether the sample mean of d_t differs from 0. In this case, a significant positive sample mean means that model 1 performs better in comparison to model 2.

Logarithmic Scoring Rule

Let $\hat{f}_t(.)$ be the chosen density of innovations
Let y_{t+1} be the observed return at time $t + 1$
$S(\hat{f}_t; y_{t+1})$=log $\hat{f}_t(y_{t+1})$
$\hat{f}_t(y_{t+1}) = \frac{1}{\sigma_{t+1}}\hat{f}_t(\hat{z_{t+1}})$
The loss differential of this: d$_t$
d$_t = S_1(\hat{f}_t; y_t) - S_2(\hat{f}_t; y_t)$

Table 2: Number of observations in and out of sample

Index	#In-Sample	#Out-of-Sample
TA25	3047	30
TA100	1917	30

6 Data

6.1 Introduction to TA25 and TA100

We have data of the Tel Aviv 25 and of the Tel Aviv 100 indices. For the TA25 we have observations from 20 October 1992 to 31 May 2005. In total we have 3077 observations. For the TA100 we have observations from 2 July 1997 to 31 May 2005. In total we have 1947 observations. The data are obtained from investing.com. Both indices contain the shares of companies with the highest market capitalization that are traded on the TASE. The indices are value weighted. Currently, in 2019, both indices no longer exist in this old form. They are now the: TA35 and TA125. We do however use the old data because we want to check the article of Alberg et al. (2008). All of the above information is summarized in Table 1. Later (in Section 9) the out-of-sample period is extended. The reason for this is discussed there.

Table 1: General information

Index	From	To	#Observations
TA25	20-10-1992	31-5-2005	3077
TA100	2-7-1997	31-5-2005	1947

6.2 Transformation to stationary time series

If we take a closer look to the indices in their raw form, one can notice that they are non-stationary time series. See Figure 1 and Figure 2. To obtain stationary time series, first a transformation is needed. If we take the returns, we have solved the problem. One can either take Simple Returns or Compound Returns. In the article of Alberg et al. (2008), the compound returns are taken. This means: $r_t = 100(\log(P_t) - \log(P_{t-1}))$. For P_t we take the closing value of the TA25/TA100 at time point t. In Table 5 the descriptive statistics for the compound returns can be found. The sample kurtosis of the TA25 is 6.463. The sample kurtosis of the TA100 is 7.9449. Both numbers are far away from 3, indicating excess kurtosis. The sample skewness of the TA25 is -0.1606. The sample skewness of the TA100 is -0.4043. This means that we also have: negative skewness. The JB statistic for the compound returns of TA25 and TA100 is: 1550.7 and 2036.7. This is far above 5.991 so we reject H_0 of normality. For more information on the Jarque-Bera Test, see the information-box on page 14.

Figure 1: Tel Aviv 25

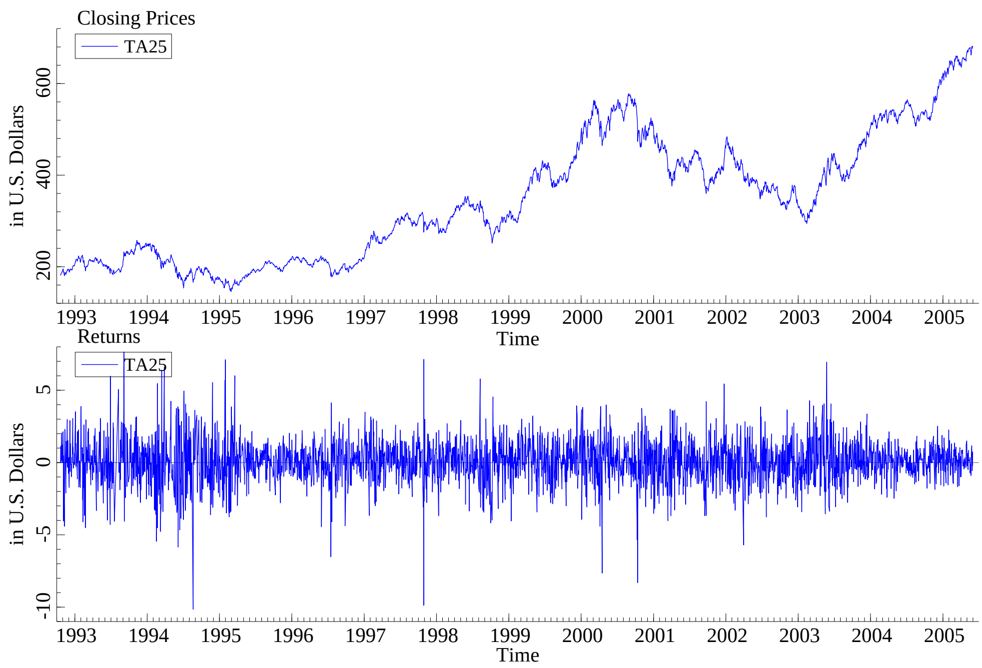
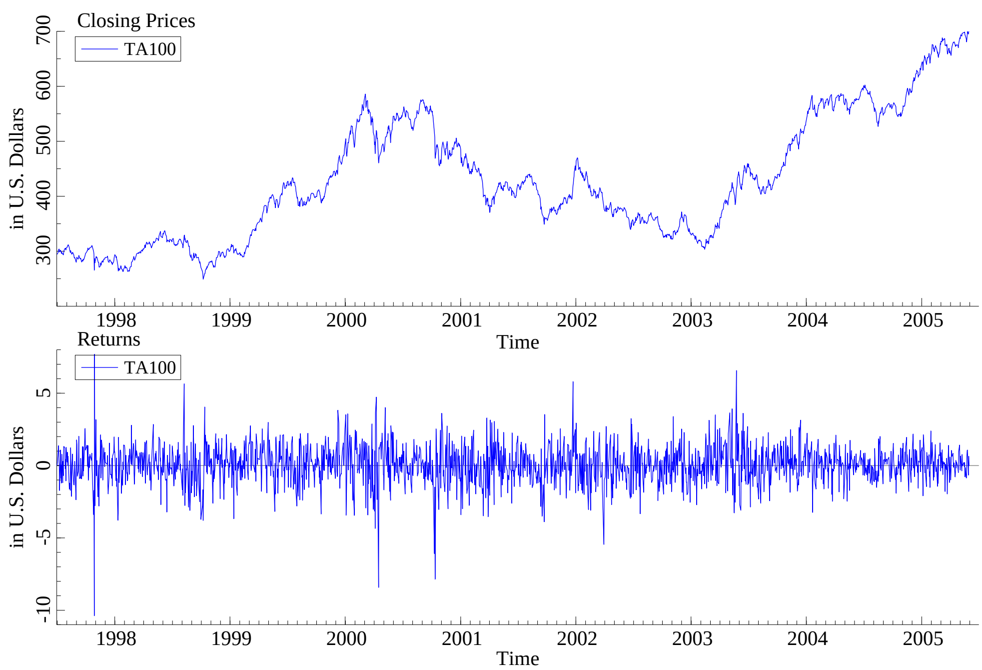


Figure 2: Tel Aviv 100



6.3 Day of the Week effect

One would normally use the returns to do further research/do model analysis. However, in the article of Alberg et al. (2008), another transformation is made. They state that the daily returns are probably correlated with the day-of-the-week effect. To avoid this potential calendar effect on further volatility analysis, the seasonalities are filtered away! This has to be done in two steps, through regression. The Israeli Trading Week is different from the European and American stock markets. It starts on Sunday and ends on Thursday instead of Monday to Friday. For this reason: there might be increasing trading activity on Sunday and Thursday. To eliminate those daily effects, the below steps are performed.

Day of the Week effect

•$r_t = \alpha_1 SUN_t + \alpha_2 MON_t + \alpha_3 TUE_t + \alpha_4 WED_t + \alpha_5 THU_t + \delta_t$
OLS regression then finds $\hat{r}_t$
•$(r_t - \hat{r}_t)^2 = \beta_1 SUN_t + \beta_2 MON_t + \beta_3 TUE_t + \beta_4 WED_t + \beta_5 THU_t + \epsilon_t$
OLS regression then finds $\widehat{(r_t - \hat{r}_t)} = \hat{\eta}_t$
$SUN_t, MON_t, TUE_t, WED_t$ and THU_t are dummy variables for Sunday, Monday, Tuesday, Wednesday and Thursday
•$y_t = (r_t - \hat{r}_t)/\sqrt{\hat{\eta}_t}$
The normalized returns are now: y_t

We would like to know if the coefficient of the Mean and Variance equation are significant. Therefore we compare the T-Stat to the standard normal quantiles. It is significant if it is outside the 95% interval: [-1.96,1.96]. The first regression gives five alphas that are not significant for the TA25. For the TA100, only the alpha for Sunday is (positive and) significant. The second regression gives five betas that are significant for the TA25 as well as the TA100. For an overview of the results, see Table 3 and Table 4. The filtered data that we now have obtained are called: y_t throughout this whole thesis. We can take a look at some of the descriptive statistics of y_t.

Table 3: Regression coefficients for day-of-the-week effect TA25

TA25	*Mean*	SE	T-stat	*Variance*	SE	T-stat
Sunday	0.115	0.062	1.856	3.303	0.215	**15.384**
Monday	0.048	0.059	0.803	1.849	0.204	**9.041**
Tuesday	0.025	0.060	0.426	2.190	0.206	**10.643**
Wednesday	-0.051	0.060	-0.845	1.850	0.207	**8.936**
Thursday	0.085	0.060	1.430	2.012	0.206	**9.770**

Table 4: Regression coefficients for day-of-the-week effect TA100

TA100	*Mean*	SE	T-stat	*Variance*	SE	T-stat
Sunday	0.194	0.066	**2.929**	2.788	0.228	**12.210**
Monday	0.076	0.066	1.159	0.971	0.226	**4.288**
Tuesday	0.037	0.067	0.552	1.617	0.229	**7.054**
Wednesday	0.115	0.067	-1.724	1.629	0.230	**7.079**
Thursday	0.101	0.067	1.513	1.580	0.230	**6.866**

6.4 Descriptive Statistics

Once the returns are ‘standardized’ by performing the steps above, the characteristics of the returns slightly change. The old mean for the TA25/TA100 was for example 0.043/0.044 and is now 0. The minimum (-10.1555) of TA25(r_t) and the minimum (-10.3816) of TA100(r_t) are both exactly the same as in the article of Alberg et al. (2008). This means that the data that we use is similar to that of the article, so that we can compare our results to the article. For a total view of the characteristics, see table 5. Notice that with r_t the returns are meant, and with y_t the standardized returns are meant. Both time series are stationary but do contain negative skewness and excess kurtosis in comparison to the normal distribution. The skewness should be close to zero and the kurtosis close to 3 when the random variable is normally distributed. As one can see from Table 5, the Jarque Bera Statistic is so high, that at any α the null hypothesis of normality is rejected. Notice that JB-stat=$\frac{n-k}{6}(S^2 + \frac{1}{4}(C-3)^2)$. The meanings of the components in the formula are stated below in the box.

6.5 Model for the data

Now that we have the compound standardized returns: TA25(r_t) and TA100(r_t) we can go further. In order to find a model that suits the data, we need to make some assumptions. We need to (1) choose a mean model, (2) choose a volatility model and (3) choose a distribution of the innovations. In Section 3 we can find the models that we are going to use. The mean equation is the whole time: $r_t = \mu + \epsilon_t$. The volatility models change from: GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1) to APARCH(1,1). The distribution of the innovations change from Normal, Student-t to Skewed Student-t.

Table 5: Descriptive Statistics for logarithm differences and 'standardized returns'

	TA25(r_t)	TA25(y_t)	TA100(r_t)	TA100(y_t)
Observations	3077	3077	1947	1947
Mean	0.0430	0.0000	0.0444	0.0000
Median	0.0140	0.0000	0.0301	0.0060
St.dev	1.4822	0.9933	1.3144	1.0003
Min	-10.1555	-6.6958	-10.3816	-8.1352
Max	7.6524	5.5931	7.6922	6.1172
Skewness	-0.1606	-0.1358	-0.4043	-0.4364
Kurtosis	6.4630	6.2135	7.9449	7.5680
JB	1550,7	1333,4	2036,7	1754,5

Jarque Bera Test

H_0 : Skewness=0 and Kurtosis=3
H_1 : Skewness $\neq$ 0 and/or Kurtosis $\neq$ 3
JB-stat=$\frac{n-k}{6}(S^2 + \frac{1}{4}(C-3)^2)$
S= sample skewness = $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2)^{\frac{3}{2}}}$
C= sample kurtosis = $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2)^2}$
n= number of observations
k= number of regressors, and is 1 outside regression context
Under H_0 : JB-stat $\sim\chi^2_2$
Critical value when $\alpha = 0.05$: 5.991

7 Results

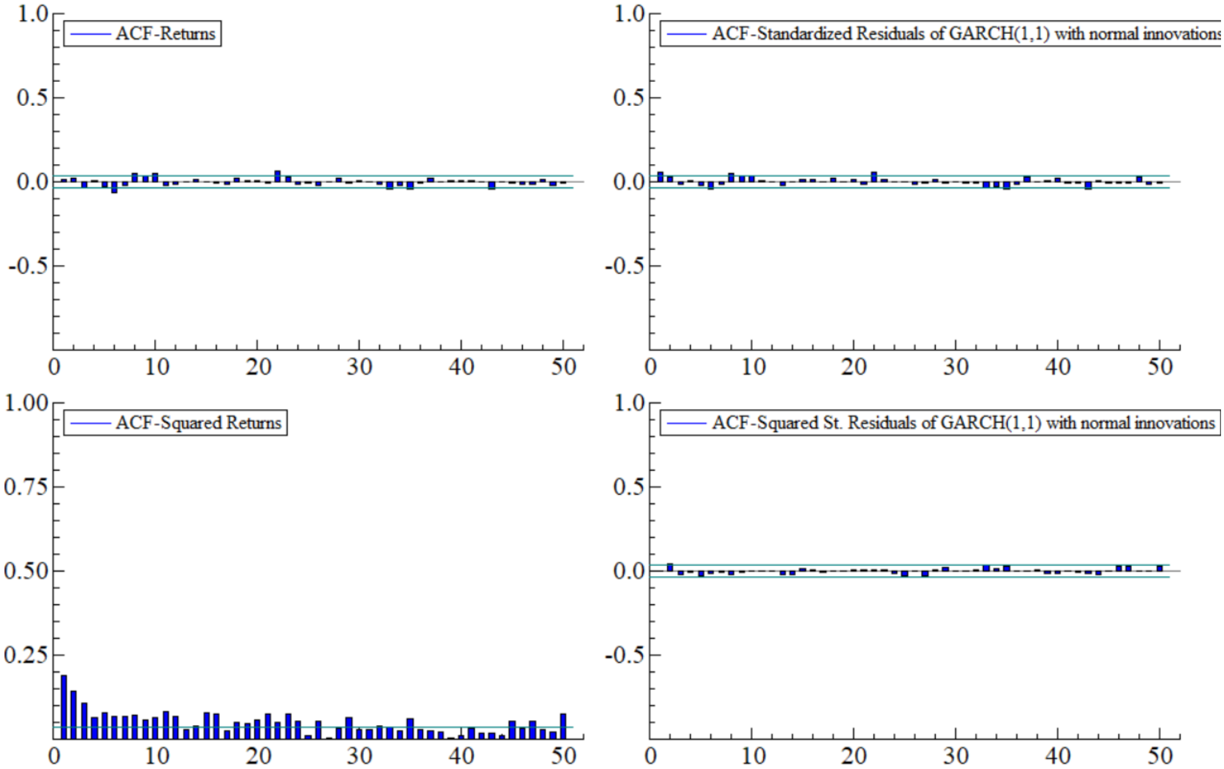
7.1 Estimation Method

Now that we have in mind the models for the data, the real estimation should take place! Just a quick recap: we have for the returns two equations. One is for the mean, and the other one for the conditional variance. For the mean, a simple model with only a constant is used to estimate μ. For the conditional variance, the models discussed in Section 3 are used. To be clear: GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1) and APARCH(1,1). The calculations are made in the Programming Language: OxMetrics. The MaxBFGS function is used. This function, developed by Broyden, Fletcher, Goldfarb and Shanno, calculates the likelihood for a given density. The likelihood-function is different for different densities of the innovations and depends on the past information. To overcome numerical problems, the Log-Likelihood is maximized instead of the likelihood. This does not have an influence on the parameter estimates, since the Log-Likelihood function has the same shape as the likelihood-function. For this algorithm, starting values are of high importance! If the wrong starting values are used, the algorithm may find a local maximum instead of a global maximum. To have the right starting values, the parameter estimation results of Alberg et al. (2008) are used. Also, we need a starting value: σ_0^2. For this we use the sample variance. The maximum likelihood estimators are generally consistent and have asymptotic standard errors that are valid under non-normality. For In-Sample comparison of the models we use the Maximum Likelihood Value, the Akaike Information Criterion(AIC) and the Bayesian Information Criterion(BIC). For Out-Of-Sample comparison, the performance measures of Section 5 are used.

7.2 Estimation Results

The parameter estimates can be found in Tables 6, 7, 8 and 9. At the left side of the tables, the parameters are stated. Below every parameter estimate, the standard error can be found. The parameters that are significant, with (at least) an alpha (significance level) of 0.05 (5 percent), have a star. Significant parameters have a t-statistic outside the [-1.96,1.96] interval. This is because if the T-stat has a normal distribution, one should use a two sided test based on the 0.025 and 0.975 quantiles. The returns, and squared returns have some characteristics. They must be captured by the models. See for example Figure 3. In this figure, the ACF of the returns and squared returns are plotted at the left side. They are plotted for 50 lags. At the right side the ACF of the standardized residuals and the squared standardized residuals are plotted. Remember the formula for the standardized residuals as: $r_t = \frac{y_t - \mu}{\sigma_t}$. If the GARCH(1,1) model with normal innovations captures all the characteristics, there should be no auto-correlation.

Figure 3: Autocorrelogram of the GARCH(1,1) model with normal innovations



We want to know which model is best for capturing the characteristics of the returns. One way to look at this problem, is see what the Log-L stat is for different models. For which model is the Log L the highest? For the TA25: the Log-Likelihood function is the highest (-4035.7) for the APARCH(1,1) model with Student-t density. For the TA100: the Log-Likelihood function is the highest (-2589.7) also for the APARCH(1,1) model with Student-t density. This model has a lot of parameters, but not the most. (The model with the most parameters is the EGARCH/APARCH model with Skewed-t density). Since the maximum of the likelihood function is prone to over-fitting, it is often the case that a model with more parameters has a higher log likelihood. Over-fitting can be a problem! To solve this problem, one could take a look at the AIC or BIC instead of the Log-L. The reason for this is: the number of parameters of a model is punished by the AIC and BIC. The difference between the AIC and BIC is the size of the punishment, the BIC prefers a smaller model (the exact formula of both measures can be found below in the box). They both take into account the number of parameters, and a punishment is given to them. For TA25: the model with the lowest AIC is again the APARCH(1,1) with Student-t (2.6536) for the TA25. However, it is very close to the EGARCH with Skewed-t (2.6589). For TA100: the model with the lowest AIC is still the APARCH(1,1) with Student-t (2.7091) for the TA25. This conclusion is the same as the log likelihood function, which can mean that there may not be that much over-fitting after all! One thing is clear: the GARCH model performs poorest in comparison to the GJR-GARCH, EGARCH and APARCH model. When talking about densities, using Student-t and Skewed Student-t instead of the Normal density will make the likelihood function higher. The AIC criteria of the GARCH goes from 2.7133 to 2.6692 for TA25. For TA100: the AIC criteria of the GARCH goes from 2.7703 to 2.7243. For GJR-GARCH, EGARCH and APARCH similar decreasing results can be found in Tables 7, 8 and 9. To summarize these conclusions, see: "Summary of In-Sample Results".

AIC and BIC

Akaike information criterion (AIC) = 2k - 2ln($\hat{L}$)
Bayesian information criterion (BIC) = ln(n)k - 2ln($\hat{L}$)
With n= number of observations
With k= number of parameters
With $\hat{L}$= the maximized value of the likelihood function
Note: *We want a high in-sample quality. We do not want parameters if they do not add up any new information to the model. The number of parameters therefore should be penalized. The penalty term is added, which makes the AIC and BIC different from the Log-Likelihood.*

Summary of In-Sample Results

In this box, the models are compared to each-other. The data that are used are only the IN-Sample data. Based on the three measures: Log-L, AIC and BIC, the APARCH(1,1) model performs best. The density of the innovations by this APARCH(1,1) model is: Student-t.

Log-Likelihood for different models with (the best) different densities

	TA25		TA100	
	Log-L	Density	Log-L	Density
GARCH	-4060.6	sk-t	-2605.2	sk-t
GJR	-4046.2	sk-t	-2595.7	sk-t
EGARCH	-4043.9	sk-t	-2596.3	sk-t
APARCH	**-4035.7**	st-t	**-2589.7**	st-t

Model with best Log-L, AIC and BIC for TA25 and TA100

Based on:	TA25	TA100	Density
Log-L	APARCH(1,1)	APARCH(1,1)	st-t
AIC	APARCH(1,1)	APARCH(1,1)	st-t
BIC	APARCH(1,1)	APARCH(1,1)	st-t

TA25	**GARCH(1,1)**			**TA100**	**GARCH(1,1)**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
μ	0.0282	0.0262	0.0960*	0.0327	0.0240	0.1203*
(SE)	(0.0152)	(0.0149)	(0.0285)	(0.0206)	(0.0199)	(0.0404)
ω	0.0350*	0.0157*	0.0164*	0.1719*	0.0543*	0.0620*
(SE)	(0.0095)	(0.0051)	(0.0052)	(0.0392)	(0.0222)	(0.0252)
α	0.1210*	0.0602*	0.0642*	0.1798*	0.0710*	0.0805*
(SE)	(0.0176)	(0.0115)	(0.0118)	(0.0292)	(0.0193)	(0.0215)
β	0.8504*	0.8995*	0.8934*	0.6566*	0.8294*	0.8066*
(SE)	(0.0226)	(0.0187)	(0.0190)	(0.0568)	(0.0499)	(0.0562)
ν		7.6530*	7.9048*		7.4637*	7.8711*
(SE)		(0.9288)	(0.9905)		(1.1226)	(1.2532)
ξ			0.9385*			0.9220*
(SE)			(0.0208)			(0.0273)
Log L	-4129.7	-4064.7	-4060.6	-2651.4	-2609.0	-2605.2
AIC	2.7133	2.6713	2.6692	2.7703	2.7272	2.7243
BIC	2.7212	2.6810	2.6811	2.7819	2.7414	2.7417

Table 6: Parameter Estimates GARCH(1,1)

TA25	**GJR(1,1)**			**TA100**	**GJR(1,1)**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
μ	0.0107*	0.0154*	0.0708*	0.0438*	0.0134	0.0728*
(SE)	(0.004)	(0.0004)	(0.0004)	(0.0002)	(0.0201)	(0.0014)
ω	0.0374*	0.0149*	0.0194*	0.1784*	0.0630*	0.0619*
(SE)	(0.0106)	(0.0048)	(0.0050)	(0.0375)	(0.0255)	(0.0229)
α	0.1111*	0.0490*	0.0598*	0.1709*	0.0738*	0.0807*
(SE)	(0.0192)	(0.0110)	(0.0124)	(0.0264)	(0.0194)	(0.0197)
β	0.7902*	0.8574*	0.8192*	0.5582*	0.7418*	0.7035*
(SE)	(0.0333)	(0.0250)	(0.0315)	(0.0606)	(0.0687)	(0.0669)
γ	0.1302*	0.1134*	0.1431*	0.1959*	0.1481*	0.2049*
(SE)	(0.0268)	(0.0242)	(0.0297)	(0.0507)	(0.0492)	(0.0519)
ν		7.9161*	8.2736*		7.7141*	8.1224*
(SE)		(0.9822)	(1.0665)		(1.1983)	(1.3253)
ξ			0.9565*			0.9546*
(SE)			(0.0110)			(0.0140)
Log L	-4116.3	-4052.9	-4046.2	-2642.9	-2603.1	-2595.7
AIC	2.7052	2.6642	2.6605	2.7626	2.7221	2.7154
BIC	2.7150	2.6761	2.6743	2.7776	2.7395	2.7357

Table 7: Parameter Estimates GJR-GARCH(1,1)

TA25	EGARCH			TA100	EGARCH	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
μ	0.0061	0.0129	0.0487	-0.0104	0.0049	0.0510*
(SE)	(0.0136)	(0.0118)	(0.0295)	(0.0206)	(0.0164)	(0.0186)
ω	0.0005	0.6667*	0.6211*	-0.0113	0.9386*	0.8444*
(SE)	(0.0040)	(0.1717)	(0.1653)	(0.0136)	(0.3323)	(0.3055)
α	0.2351*	0.1498*	0.1493*	0.3274*	0.2028*	0.2001*
(SE)	(0.0233)	(0.0226)	(0.0226)	(0.0402)	(0.0391)	(0.0388)
β	0.9513*	0.9686*	0.9702*	0.7564*	0.8798*	0.8846*
(SE)	(0.0112)	(0.0088)	(0.0087)	(0.0555)	(0.0402)	(0.0393)
γ	-0.0620*	-0.0495*	-0.0446*	-0.1683*	-0.0865*	-0.0777*
(SE)	(0.0124)	(0.0111)	(0.0114)	(0.0280)	(0.0241)	(0.0232)
ν		8.0948*	8.1703*		8.4621*	8.5050*
(SE)		(1.0136)	(1.0320)		(1.4416)	(1.4519)
ξ			0.9687*			0.9627*
(SE)			(0.0221)			(0.0180)
Log L	-4106.4	-4044.8	-4043.9	-2626.0	-2597.0	-2596.3
AIC	2.6987	2.6589	2.6589	2.7449	2.7157	2.7160
BIC	2.7085	2.6707	2.6728	2.7594	2.7331	2.7363

Table 8: Parameter Estimates EGARCH(1,1)

TA25	APARCH			TA100	APARCH	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
μ	0.0033	0.0128	0.0503	-0.0040	0.0135	0.0528
(SE)	(0.0161)	(0.0151)	(0.0295)	(0.0211)	(0.0202)	(0.0328)
ω	0.0441*	0.0222*	0.0234*	0.2513*	0.0770*	0.1050*
(SE)	(0.0109)	(0.0063)	(0.0070)	(0.0523)	(0.0223)	(0.0374)
α	0.1245*	0.0727*	0.0776*	0.1697*	0.0920*	0.1051*
(SE)	(0.0164)	(0.0107)	(0.0120)	(0.0290)	(0.0169)	(0.0220)
β	0.8559*	0.9031*	0.9015*	0.5818*	0.8202*	0.7689*
(SE)	(0.0214)	(0.0155)	(0.0170)	(0.0688)	(0.0367)	(0.0640)
γ	0.2652*	0.3718*	0.3251*	0.5181*	0.4747*	0.4077*
(SE)	(0.0544)	(0.0831)	(0.0836)	(0.1045)	(0.1287)	(0.1200)
δ	1.2480*	1.2265*	0.9900*	1.5996*	1.1997*	1.2081*
(SE)	(0.1551)	(0.1895)	(0.1886)	(0.2902)	(0.2800)	(0.3049)
ν		8.6376*	8.0506*		9.3240*	8.4845*
(SE)		(1.1506)	(1.0020)		(1.7272)	(1.4551)
ξ			0.9675*			0.9616*
(SE)			(0.0221)			(0.0243)
Log L	-4108.0	-4035.7	-4043.9	-2625.5	-2589.7	-2596.0
AIC	2.7003	2.6536	2.6596	2.7454	2.7091	2.7168
BIC	2.7122	2.6674	2.6754	2.7628	2.7294	2.7399

Table 9: Parameter Estimates APARCH(1,1)

7.3 News Impact Curve

In this subsection, the news impact curve is going to be introduced. This curve shows what happens to the predicted variance, $\hat{\sigma_t^2}$, when there is a positive or negative shock (on day $t-1$). This means that it shows the leverage effect. A shock can be seen as new information, and we want to know what happens with the variance by obtaining this new information. The news impact curve shows how asymmetric the effect of a positive/negative shock is. We are going to obtain 4 images of News Impact Curves. We take the Tel Aviv 25 as example (the same steps could also be done with the TA100). We also need a distribution for the innovations in order to estimate θ, so we take Skewed Student-t. This distribution is chosen because results have shown that it is better to use an asymmetric density for forecasting. On the y-axis σ_t^2 is going to be stated. On the x-axis the residual ϵ_{t-1} is going to be stated. We need to know what the maximum and minimum shock is, in order to know exactly what the x-axis looks like. In OxMetrics we obtain the maximum shock (in Dollars) for the APARCH model with this distribution: 5.54. This is obtained by taking the maximum of the whole vector: $\epsilon_{t-1} = y_{t-1} - \hat{\mu}$. The minimum shock is: -6.75. The steps we performed are:

- We obtain θ for the GARCH/GJR/EGARCH/APARCH model, the distributions of the innovations is Skewed Student-t.
- We need a grid for ϵ_{t-1}. We take [-10,10] with steps of 0.1.
- For the four models we get the y-coordinates. This is done by filling in the equations. GARCH: $h_t = \hat{\omega}+\hat{\alpha}\epsilon_{t-1}^2+\beta h_{t-1}$, GJR-GARCH: $h_t = \hat{\omega}+\hat{\alpha}\epsilon_{t-1}^2+\beta h_{t-1}+\hat{\gamma}\mathbb{1}_{[\epsilon_{t-1}<0]}\epsilon_{t-1}^2$, EGARCH: $h_t = \exp(\hat{\omega} + \hat{\alpha}(|z_{t-1}| - \mathbb{E}(|z_{t-1}|)) + \hat{\gamma}z_{t-1} + \hat{\beta}\log(h_{t-1}))$ and APARCH: $h_t = (\hat{\omega} + \hat{\alpha_1}(|\epsilon_{t-1}| - \gamma\epsilon_{t-1})^{\delta} + \hat{\beta}h_{t-1}^{\delta})^{2/\hat{\delta}}$.
- We assume for $h_{t-1} = 1$, so we take it constant.

Figure 4: News Impact Curve GARCH(1,1)

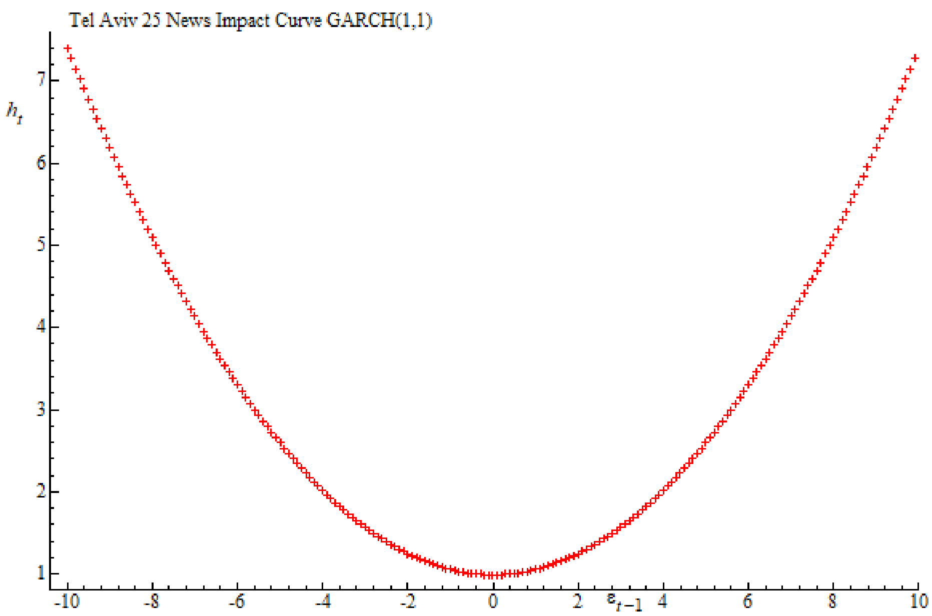


Figure 5: News Impact Curve GJR-GARCH(1,1)

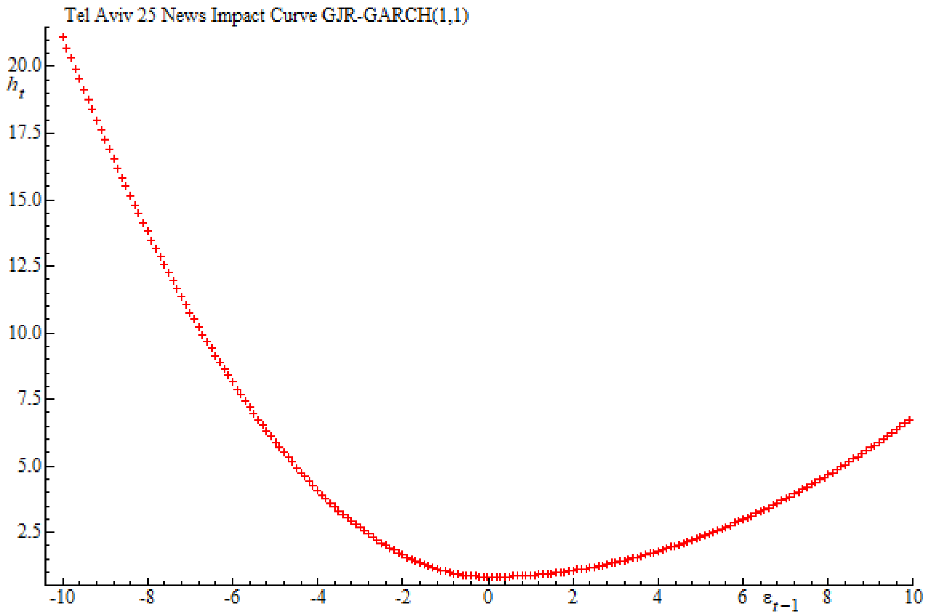


Figure 6: News Impact Curve EGARCH(1,1)

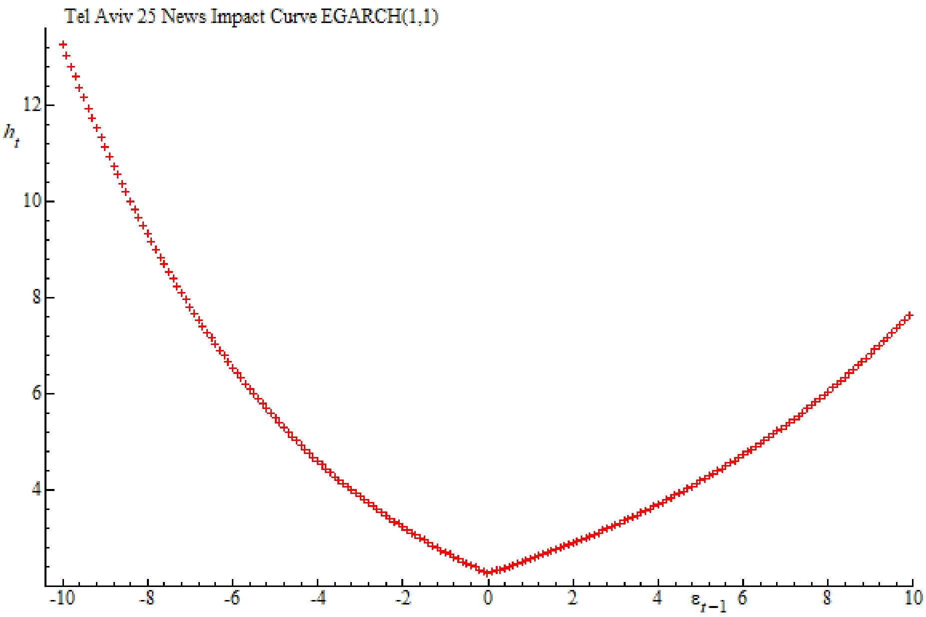
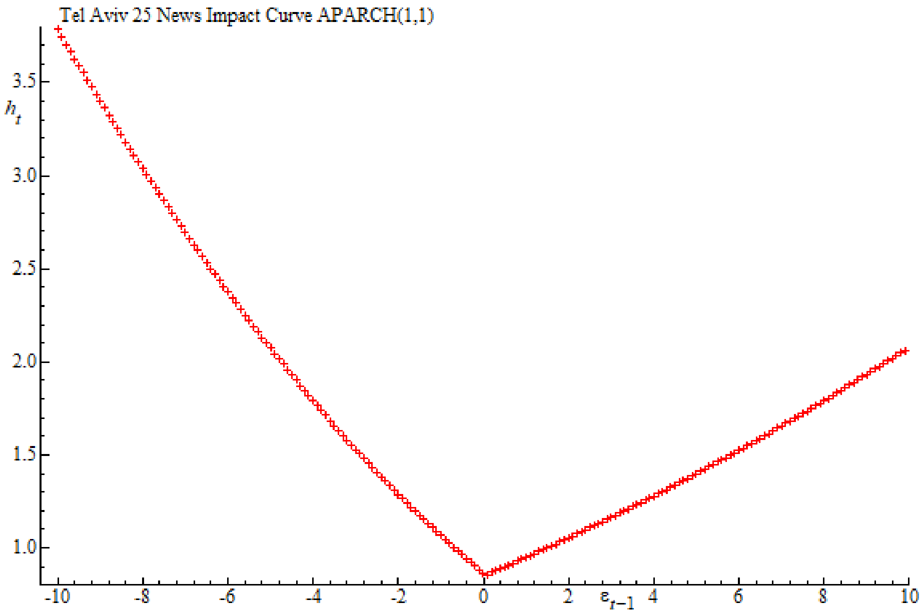


Figure 7: News Impact Curve APARCH(1,1)

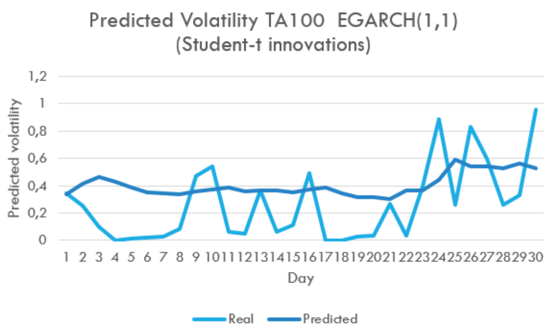


In Figure 4, the GARCH model, we see that there are no asymmetric effects, because the GARCH model does not allow for these. For the other Figures: 5, 6, 7 we do see asymmetric effects. Notice that we take also unobserved ϵ_{t-1} in the grid. For negative shocks, the predicted variance (h_t) changes more than for the same positive shocks. This is what we would have expected.

7.4 Out-Of-Sample Measures

For risk analysis and several other important aspects, GARCH models must be able to predict 'good enough'. In this section the forecast-ability of the various conditional variance models are being compared. The estimation results in the section above are based only on the in-sample data. We could for example take the parameter values of the EGARCH(1,1) model and predict 30 one-day-ahead forecasts. This is shown in Figure: 8.

Figure 8: 30 One-Day-Ahead Forecasted Volatility EGARCH(1,1)



The prediction measures discussed in Section 5 are now simply filled in with the out-of-sample data. We have 30 out-of-sample observations: r_t^2. We compare these with the 30 one-step-ahead forecasts for σ_t^2.

Notice that we want the MSE/MAE/MedSe/RMSE/TIC/AMAPE to be as small as possible. Since we are generally more interested in the variance equation, not the mean equation, we calculate the performance measures for this one. We also calculate the Log-Score. The Log-Score can be viewed as an out-of-sample log-likelihood. Notice that we want the Log-Score to be as high as possible. Ultimately, we want to compare our results with the ones of the article of Alberg et al. (2008). Do we get similar conclusions? What are the conclusions? The results can be found in Tables 10,11,12 and 13.

Besides the already discussed out-of-sample measures, we also calculate the DM-Statistic. This statistic can compare two models. We want to know if it is significantly different from zero or not. If it is significantly different from zero and positive, we prefer the model at the left side of the table (column). If it is significantly different from zero and negative we prefer the model at the upside (row) of the table. If it is not significantly different from zero, there is no evidence that the one model outperforms the other model. On the diagonal of the table, the Log-Score of that specific model for that specific density can be found. By using the DM-Statistic, it is not only possible to compare different models to each other, but also to compare the same model, with a different density.

First we are going to look at the forecasting results for the Tel Aviv 25 index. Let us look at the densities of the GARCH(1,1) model in Table 10. What we see is that the MSE, when the innovations are normally distributed, is: 0.0726. This is quite high in comparison to 0.0465 (Student-t) and 0.0453 (skewed Student-t). What we can see is that also the MAE/MedSe/RMSE/TIC/AMAPE are all declining if we go from Normal to Student-t to skewed Student-t. This is an indication that for the GARCH(1,1) model, the skewed Student-t density performs best in comparison to the other two densities. The Log-Score suggests that it is better to use a Student-t density, since -0.6115 is bigger than -0.6579 (skewed Student-t). Let us look at the densities of the GJR-GARCH(1,1) model in Table 11. The MAE of the student-t density is the lowest (0.1901). Also the MedSe,RMSE,TIC are lowest for the Student-t density. The Log-Score is the best for Student-t density. This results suggest that the performance of the Student-t density is better than the performance of the skewed Student-t density. The reason for this can be that the out-of-sample data has less skewness than the in-sample data. Another reason could be that the GARCH equation 'just' is better estimated with the Student-t density: it performs better out-of-sample in this case.

Let us look at the densities of the EGARCH(1,1) model in Table 10. It is clear that the Normal density performs worst of the three densities. The Student-t and skewed Student-t densities perform quite similar. Let us look at the densities of the APARCH(1,1) model in Table 13. Also in this case the Student-t and skewed Student-t density outperform the Normal density on predicting. The Log-Score of the student-t density is the highest of all three. Now that we looked at the Tables 10,11,12 and 13 separately we also could compare the models to each other. The EGARCH model with student-t density for the innovations performs better than the other models with student-t density for the innovations. We can conclude that for most measures the EGARCH model outperforms the APARCH model. The APARCH model outperforms the GJR and GARCH model. The GARCH model has the poorest forecast. We also can conclude that for the TA25 index the results of the MSE/MAE/MedSe/RMSE/TIC all give an indication that it is better to use the Skewed Student-t density for the innovations. However, if one would only be interested in a low Log-score, the results give an indication to use the Student-t density. In comparison to the GARCH(1,1), GJR-GARCH(1,1) and APARCH(1,1) the forecast ability of the EGARCH model. Since the MSE/MAE/MedSe/RMSE/TIC is very low and the Log-Score very high for this model (with different densities).

Now, we are going to look at the forecasting results for the Tel Aviv 100 index. The results are kind of similar to that of TA25. The EGARCH performs for most measures the best. The lowest Log-Score (-0.7809) can be found by this model with Student-t innovations. The 'worst' models for TA100 are the APARCH and GJR-GARCH. The best density for the innovations is Skewed Student-t or Student-t, depending on which measure one chooses. The Log-Score is often better for the Student-t density. If we take a look at Table 16 different models with Skewed Student-t density are compared. The DM Statistic is significant for the EGARCH(-2.0815) and APARCH(-1.9902) in comparison to the GARCH. To summarize, for most measures the (Student-t distributed) EGARCH model outperforms the other models. However it very close to the skewed Student-t EGARCH. One thing is clear, the EGARCH model gives better forecasts than for example the GARCH/GJR model. This is the same conclusion as is made in the article of Lambert and Laurent (2001). Other studies (Lambert and Laurent (2001)) also show similar results, for example that skewed Student-t is better for modelling the NASDAQ index.

Table 10: Prediction Measures GARCH(1,1)

	TA25	**GARCH**		**TA100**	**GARCH**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	0.0726	0.0465	0.0453	0.1938	0.0852	0.0843
MAE	0.2430	0.1914	0.1896	0.3929	0.2560	0.2545
MedSe	0.0742	0.0421	0.0419	0.2564	0.1126	0.1075
RMSE	0.2695	0.2156	0.2128	0.4402	0.2918	0.2903
TIC	0.3474	0.3173	0.3146	0.3699	0.3224	0.3209
AMAPE	0.5670	0.5489	0.5483	0.5609	0.5076	0.5058
Log-Score	-0.6399	-0.6115	-0.6579	-0.8537	-0.7944	-0.8523

Table 11: Prediction Measures GJR-GARCH(1,1)

	TA25	**GJR**		**TA100**	**GJR**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	0.0705	0.0448	0.0488	0.2082	0.0864	0.0805
MAE	0.2391	0.1901	0.1991	0.4068	0.2564	0.2474
MedSe	0.0854	0.0465	0.0563	0.2450	0.1144	0.1051
RMSE	0.2655	0.2116	0.2210	0.4563	0.2939	0.2837
TIC	0.3432	0.3132	0.3192	0.3719	0.3210	0.3149
AMAPE	0.5624	0.5470	0.5463	0.5665	0.5037	0.4990
Log-Score	-0.6358	-0.6058	-0.6473	-0.8617	-0.7969	-0.8128

Table 12: Prediction Measures EGARCH(1,1)

	TA25	**EGARCH**		**TA100**	**EGARCH**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	0.0558	0.0412	0.0410	0.1877	0.0801	0.0801
MAE	0.2134	0.1786	0.1783	0.3738	0.2462	0.2460
MedSe	0.0570	0.0320	0.0320	0.2173	0.0832	0.0851
RMSE	0.2363	0.2031	0.2025	0.4332	0.2831	0.2831
TIC	0.3268	0.3051	0.3044	0.3619	0.3149	0.3148
AMAPE	0.5571	0.5504	0.5500	0.5505	0.4997	0.4990
Log-Score	-0.6006	-0.5839	-0.6072	-0.8422	-0.7809	-0.8045

Table 13: Prediction Measures APARCH(1,1)

	TA25	**APARCH**		**TA100**	**APARCH**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	0.0655	0.0429	0.0417	0.2183	0.0827	0.0903
MAE	0.2285	0.1856	0.1826	0.4124	0.2506	0.2610
MedSe	0.0716	0.0397	0.0366	0.2782	0.0949	0.1094
RMSE	0.2560	0.2071	0.2043	0.4672	0.2877	0.3005
TIC	0.3379	0.3087	0.3058	0.3740	0.3175	0.3222
AMAPE	0.5584	0.5504	0.5492	0.5675	0.5013	0.5048
Log-Score	-0.6244	-0.5917	-0.6133	-0.8683	-0.7817	-0.8200

Table 14: DM Statistic of the Logarithmic Scoring Rule Innovations are Normal distributed

TA25	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	*-0.6399*	-0.2512	-1.6827	-0.8798
GJR-GARCH	0.2512	*-0.6358*	-1.5303	-0.6626
EGARCH	1.6827	1.5303	*-0.6006*	1.4281
APARCH	0.8798	0.6626	-1.4281	*-0.6244*

Table 15: DM Statistic of the Logarithmic Scoring Rule Innovations are Student-t distributed

TA25	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	*-0.6115*	-0.3497	-1.2959	-1.3174
GJR-GARCH	0.3497	*-0.6058*	-0.9522	-0.8480
EGARCH	1.2959	0.9522	*-0.5839*	0.4614
APARCH	1.3174	0.8480	-0.4614	*-0.5917*

Table 16: DM Statistic of the Logarithmic Scoring Rule Innovations are Skewed Student-t distributed

TA25	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	*-0.6579*	-0.4741	-2.0815	-1.9902
GJR-GARCH	0.4741	*-0.6473*	-1.4071	-1.3395
EGARCH	2.0815	1.4071	*-0.6072*	0.4291
APARCH	1.9902	1.3395	-0.4291	*-0.6133*

Table 17: DM Statistic of the Logarithmic Scoring Rule, Innovations are Normal distributed

TA100	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	*-0.8537*	0.4762	-0.5502	0.8749
GJR-GARCH	-0.4762	*-0.8617*	-0.8240	0.3668
EGARCH	0.5502	0.8240	*-0.8422*	1.5431
APARCH	-0.8749	-0.3668	-1.5431	*-0.8683*

Table 18: DM Statistic of the Logarithmic Scoring Rule Innovations are Student-t distributed

TA100	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	*-0.7944*	0.1732	-0.7444	-0.7761
GJR-GARCH	-0.1732	*-0.7969*	-0.9971	-1.1604
EGARCH	0.7444	0.9971	*-0.7809*	0.0738
APARCH	0.7761	1.1604	-0.0738	*-0.7817*

Table 19: DM Statistic of the Logarithmic Scoring Rule Innovations are Skewed Student-t distributed

TA100	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	*-0.8523*	-1.8077	-1.8207	-1.2077
GJR-GARCH	1.8077	*-0.8128*	-0.4609	0.4154
EGARCH	1.8207	0.4609	*-0.8045*	1.1663
APARCH	1.2077	-0.4154	-1.1663	*-0.8200*

Table 20: DM-Statistic of Logarithmic Scoring Rule, GARCH model with different densities

	TA25			**TA100**		
GARCH	Normal	Student-t	Sk St-t	Normal	Student-t	Sk St-t
Normal	0.0726	-0.9938	0.5250	0.1938	-2.0592	-0.0377
Student-t	0.9938	0.0465	1.7053	2.0592	0.0852	1.9608
Sk st-t	-0.5250	-1.7053	0.0453	0.0377	-1.9608	0.0843

Table 21: DM-Statistic of Logarithmic Scoring Rule, GJR-GARCH model with different densities

	TA25			**TA100**		
GJR	Normal	Student-t	Sk St-t	Normal	Student-t	Sk St-t
Normal	0.0705	-1.0953	0.3941	0.2082	-2.2803	-1.5404
Student-t	1.0953	0.0448	1.6727	2.2803	0.0864	0.6682
Sk st-t	-0.3941	-1.6727	0.0488	1.5404	-0.6682	0.0805

Table 22: DM-Statistic of Logarithmic Scoring Rule, EGARCH model with different densities

	TA25			**TA100**		
EGARCH	Normal	Student-t	Sk St-t	Normal	Student-t	Sk St-t
Normal	0.0558	-0.5890	0.2081	0.1877	-2.2425	-1.1930
Student-t	0.5890	0.0412	1.1339	2.2425	0.0801	1.1635
Sk st-t	-0.2081	-1.1339	0.0410	1.1930	-1.1635	0.0801

Table 23: DM-Statistic of Logarithmic Scoring Rule, APARCH model with different densities

	TA25			**TA100**		
APARCH	Normal	Student-t	Sk St-t	Normal	Student-t	Sk St-t
Normal	0.0655	-1.1734	-0.3425	0.2183	-3.0189	-1.5941
Student-t	1.1734	0.0429	1.0578	3.0189	0.0827	2.2024
Sk st-t	0.3425	-1.0578	0.0417	1.5941	-2.2024	0.0903

8 Conclusions

We finally arrive at the conclusions. We have first analyzed P_t and transformed it into R_t. After eliminating the weekly effect, we got: Y_t. We needed to know what the underlying model was for the mean and variance of the returns. We compared both in-sample measures and out-of-sample measures for different models. For the mean equation we took a simple model with only a constant. For the variance equation we took the models discussed in Section 3. For estimating the parameters, we needed to make a distributional assumption. We did not use only the Normal distribution, but also skewed Student-t and Student-t. In-sample measures (Log L, AIC, BIC) show that using a skewed Student-t distribution is best (with exception for the APARCH model where student-t performs better). The smallest AIC/BIC can be found by the APARCH model with Student-t density. This result is the same for both TA25 as TA100. Out-of-sample measures suggest to use the EGARCH model with (skewed) Student-t distribution for TA25 and TA100. However, often the results only slightly improve by using a skewed Student-t density instead of a Student-t density. One thing is clear: fat-tailed densities perform better than non fat-tailed densities. For risk implementing and forecasting, I would therefore suggest the EGARCH model with an asymmetric or fat tailed density. However, an APARCH model would also be fine.

9 Checking Robustness Of The Conclusions

9.1 Bigger number of Out-Of-Sample observations

For predicting both indices TA25, TA100 we based our conclusions on 30 out-of-sample observations. One could argue that 30 might be too small, and that if the out-of-sample period is extended, other conclusions could be reached. To check if this is the case, we extend the out-of-sample period. Instead of only taking one month, we will take two years. The quality of the model can then be compared for more out-of-sample observations. The in-sample period does not change, so the parameter estimates for θ stay the same. This means that we still use the Tables 6,7,8 9 with the parameter estimates for the four volatility models. However, the out-of-sample period ends 31-05-2007 instead of 31-05-2005. This information is summarized in Table 24. The number of observations that are out-of-sample is 519 for TA25 and 518 for TA100. This means roughly 500 out-of-sample observations instead of only 30.

Table 24: General information, extension of Out-of-Sample period

Index	From	To	#Observations	#In-Sample	#Out-of-Sample
TA25	20-10-1992	31-5-2007	3566	3047	519
TA100	2-7-1997	31-5-2007	2435	1917	518

Table 25: Prediction Measures GARCH(1,1) with an extended out-of-sample period

	TA25	**GARCH**		**TA100**	**GARCH**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	1.0557	1.0352	1.0313	1.6326	1.6002	1.5882
MAE	0.5762	0.4935	0.4914	0.7633	0.6377	0.6341
MedSe	0.1860	0.1006	0.0981	0.3806	0.1922	0.1908
RMSE	1.0275	1.0175	1.0155	1.2778	1.2650	1.2602
TIC	0.3496	0.3457	0.3444	0.3503	0.3437	0.3418
AMAPE	0.5729	0.5526	0.5516	0.5775	0.5542	0.5521
Log-Score	-0.6781	-0.6492	-0.6952	-0.8923	-0.8334	-0.8909

Table 26: Prediction Measures GJR-GARCH(1,1) with an extended out-of-sample period

	TA25	**GJR**		**TA100**	**GJR**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	1.0504	1.0465	1.0386	1.6024	1.5836	1.5710
MAE	0.5710	0.4888	0.4916	0.7427	0.6256	0.6266
MedSe	0.1779	0.0880	0.0933	0.3614	0.1751	0.1772
RMSE	1.0249	1.0230	1.0191	1.2659	1.2584	1.2534
TIC	0.3481	0.3472	0.3453	0.3461	0.3403	0.3381
AMAPE	0.5726	0.5516	0.5526	0.5729	0.5497	0.5506
Log-Score	-0.6689	-0.6498	-0.6542	-0.8973	-0.8350	-0.8585

Table 27: Prediction Measures EGARCH(1,1) with an extended out-of-sample period

	TA25	**EGARCH**		**TA100**	**EGARCH**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	1.0438	1.0256	1.0272	1.5777	1.5675	1.5689
MAE	0.5675	0.4840	0.4838	0.7427	0.6290	0.6278
MedSe	0.1660	0.0903	0.0894	0.3387	0.1770	0.1791
RMSE	1.0217	1.0127	1.0135	1.2561	1.2520	1.2526
TIC	0.3455	0.3423	0.3426	0.3424	0.3381	0.3382
AMAPE	0.5715	0.5483	0.5482	0.5709	0.5482	0.5479
Log-Score	-0.6394	-0.6208	-0.6434	-0.8784	-0.8180	-0.8410

Table 28: Prediction Measures APARCH(1,1) with an extended out-of-sample period

	TA25	**APARCH**		**TA100**	**APARCH**	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	1.0463	1.0257	1.0290	1.6168	1.5806	1.5713
MAE	0.5704	0.4867	0.4838	0.7586	0.6321	0.6317
MedSe	0.1746	0.0924	0.0905	0.3811	0.1821	0.1916
RMSE	1.0229	1.0128	1.0144	1.2715	1.2572	1.2535
TIC	0.3460	0.3417	0.3426	0.3449	0.3388	0.3382
AMAPE	0.5722	0.5495	0.5483	0.5762	0.5488	0.5489
Log-Score	-0.6621	-0.6288	-0.6506	-0.9052	-0.8185	-0.8564

The model with the highest Log-Score is again the EGARCH model with Student-t innovations. With a Log-Score of -0.6208 and -0.8180 for TA25 and TA100 respectively. However the measures for asymmetric densities of the APARCH and EGARCH are really close to each other. These results suggest to use either one of these two models. To be entirely certain about this conclusion, the DM-Statistic for the Log-Score is calculated just like we calculated it with 30 out-of-sample observations. The results can be found in Tables: 29, 30, 31 and 32. Since we already concluded that using an asymmetric density is for modelling and predicting purposes the best thing to do, the tables only have the asymmetric densities. The table with the normal distributed innovations is left out. On the diagonals of the Tables, the usual Log Score can be found. For Table 29, none of the statistics are significant. They all fall between [-1.96,1.96]. The same holds for Table 31 and 32. However for Table 30, we have some significant results. With a DM Statistic of 2.1182 and 2.0114 we can conclude that in comparison to the GARCH model with Skewed Student-t distribution, the EGARCH and APARCH model with the same distribution can better predict out-of-sample. The conclusions in Section 12, do no not change with this extension of the out-of-sample period.

Table 29: DM Statistic of the Logarithmic Scoring Rule for TA25, Innovations are Student-t distributed and Out-Of-Sample period is extended

TA25	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	-0.6115	0.0351	-1.3248	-1.3502
GJR-GARCH	-0.0351	-0.6058	-1.1747	-1.1205
EGARCH	1.3248	1.1747	-0.5839	0.4706
APARCH	1.3502	1.1205	-0.4706	-0.5917

Table 30: DM Statistic of the Logarithmic Scoring Rule for TA25 , Innovations are Skewed Student-t distributed and Out-Of-Sample period is extended

TA25	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	-0.6579	-1.6107	-2.1182	-2.0114
GJR-GARCH	1.6107	-0.6473	-0.4420	-0.1741
EGARCH	2.1182	0.4420	-0.6072	0.5094
APARCH	2.0114	0.1741	-0.5094	-0.6133

Table 31: DM Statistic of the Logarithmic Scoring Rule for TA100, Innovations are Student-t distributed and Out-Of-Sample period is extended

TA100	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	-0.7944	0.1143	-0.8464	-0.9305
GJR-GARCH	-0.1143	-0.7969	-1.0311	-1.2474
EGARCH	0.8464	1.0311	-0.7809	0.0426
APARCH	0.9305	1.2474	-0.0426	-0.7817

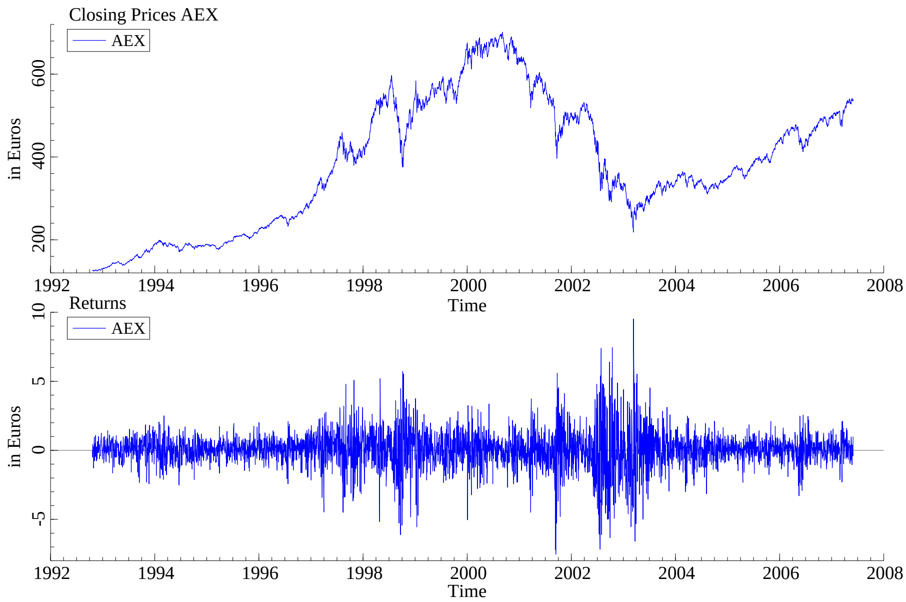
Table 32: DM Statistic of the Logarithmic Scoring Rule for TA100, Innovations are Skewed Student-t distributed and Out-Of-Sample period is extended

TA100	GARCH	GJR-GARCH	EGARCH	APARCH
GARCH	-0.8523	-1.3645	-1.9046	-1.3068
GJR-GARCH	1.3645	-0.8128	-0.9052	-0.1253
EGARCH	1.9046	0.9052	-0.8045	1.1618
APARCH	1.3068	0.1253	-1.1618	-0.8200

9.2 Another index: AEX

It is also possible to study the returns of the Amsterdam Exchange Index (AEX). It is interesting to know if the volatility model that suits best this index is still the EGARCH/APARCH with a skewed distribution. In this subsection we will take a look at this. We take as period of observations the same period as the Tel Aviv 25: 20-10-1992 until 31-05-2007. The data come from Investing.com. The reason why we take this same period is to be as close to other model as possible. If we suddenly take another time-frame, it might be possible that the economic situation during that period is too much different so that the conclusions (of which volatility model suits best) do not hold. We have in total 3736 daily observations. This is divided into in-sample observations and out-of-sample observations. The in-sample observations are until date: 19-04-2005. In total there are: 3194 observations. The out-of-sample observations are naturally until 31-05-2007. In total there are: 542 observations. Since we need stationary time series, we perform the log-returns calculations as we made before: $r_t = 100(\log(P_t) - \log(P_{t-1})$. The results of this calculation can be viewed graphically in Figure 9.2.

Figure 9: Plot made in OxMetrics to show behaviour of AEX: Closing Prices and Log Returns



To get to know the data better, let us take a look at the descriptive statistics. In Table 33, we see that the mean is 0.0389. The Skewness is quite close to 0 (-0.1502), however the kurtosis is very different in comparison to the normal distribution (8.0404). So the JB test statistic rejects the null hypothesis of the returns being normally distributed.

Table 33: Descriptive Statistics AEX

	AEX(rt)
Observations	3736
Mean	0.0389
Median	0.0853
St.dev	1.3134
Min	-7.5310
Max	9.5169
Skewness	-0.1502
Kurtosis	8.0404
JB	3968.9

The same models as in Subsection 6.5 are going to be used for this index. We again estimate the models using ML optimization in OxMetrics. The estimation results are stated below.

Table 34: Parameter Estimates GARCH(1,1) for AEX

AEX	**GARCH(1,1)**		
	Normal	Student-t	Skewed Student-t
μ	0.0619	0.0727	0.1812
(SE)	(0.0160)	(0.0146)	(0.0263)
ω	0.0191	0.0122	0.0116
(SE)	(0.0042)	(0.0034)	(0.0036)
α	0.1035	0.0840	0.0828
(SE)	(0.0112)	(0.0100)	(0.0106)
β	0.8827	0.8925	0.8936
(SE)	(0.0121)	(0.0121)	(0.0129)
ν		12.4593	14.0382
(SE)		(2.1483)	(2.9288)
ξ			0.9098
(SE)			(0.0168)
Log L	-4804.6	-4782.0	-4769.0
AIC	3.0101	2.9966	2.9890
BIC	3.0177	3.0042	3.0004

In this GARCH(1,1) model, the estimate for μ is somewhat different for a model with skewed Student-t distribution (0.1812). All the estimates for ω, which appears in the variance equation model are around the same number: 0.012. All the estimates of α and β are also around the same numbers: 0.09 and 0.90 respectively. This means that the predicted volatility of time t depends for roughly ninety percent on the volatility of the day before and 10 percent on the news-impact. What also can be viewed from Table 34: the Log Likelihood improves by moving to a skewed student-t distribution for the innovations. The same holds for the AIC and BIC which both get better using this distribution.

Table 35: Parameter Estimates GJR-GARCH(1,1) for AEX

AEX	**GJR(1,1)**		
	Normal	Student-t	Skewed Student-t
μ	0.0498	0.0623	0.1244
(SE)	(0.0162)	(0.0159)	(0.0004)
ω	0.0222	0.0151	0.0151
(SE)	(0.0043)	(0.0034)	(0.0040)
α	0.1003	0.0825	0.0853
(SE)	(0.0106)	(0.0104)	(0.0106)
β	0.8293	0.8375	0.8165
(SE)	(0.0182)	(0.2025)	(0.0219)
γ	0.1108	0.1109	0.1419
(SE)	(0.0249)	(0.0274)	(0.0286)
ν		12.9494	13.9696
(SE)		(2.4776)	(2.8367)
ξ			0.9419
(SE)			(0.1027)
Log L	-4794.2	-4773.2	-4757.2
AIC	3.0042	2.9917	2.9823
BIC	3.0137	3.0031	2.9956

By estimating the mean and volatility model by using a GJR-GARCH equation, the Log L improves. See Table 35. In the GJR-GARCH model, one extra parameter is added. This parameter, ν, shows the asymmetric effect of a shock. For the AEX index it is with all three types of densities around: 0.11. The skewness parameter is 0.9419 indicating negative skewness.

Table 36: Parameter Estimates EGARCH(1,1) for AEX

AEX	**EGARCH**		
	Normal	Student-t	Skewed Student-t
μ	0.0455	0.0566	0.1374
(SE)	(0.0162)	(0.0160)	(0.0134)
ω	0.0060	1.5512	1.4413
(SE)	(0.0025)	(0.4335)	(0.4230)
α	0.1936	0.1757	0.1778
(SE)	(0.0182)	(0.0182)	(0.0186)
β	0.9840	0.9856	0.9872
(SE)	(0.0032)	(0.0034)	(0.0033)
γ	-0.0472	-0.0494	-0.0389
(SE)	(0.0090)	(0.0095)	(0.0095)
ν		13.4032	14.1760
(SE)		(2.6626)	(2.9715)
ξ			0.9346
(SE)			(0.0124)
Log L	-4800.0	-4779.1	-4772.7
AIC	3.0078	2.9953	2.9920
BIC	3.0173	3.0067	3.0053

In Table 36 the parameter estimates of the EGARCH(1,1) model are stated. The Log L of all three densities are around a lower number than the Log L of all three densities at the GJR-GARCH model! This is different than what we saw for the TA25 and TA100 index (where the EGARCH model performed best). This could be because we have another index.

Table 37: Parameter Estimates AGARCH(1,1) for AEX

AEX	**APARCH**		
	Normal	Student-t	Skewed Student-t
μ	0.0412	0.0562	0.1425
(SE)	(0.0166)	(0.0162)	(0.0287)
ω	0.0195	0.0129	0.0134
(SE)	(0.0042)	(0.0033)	(0.0038)
α	0.0986	0.0722	0.0808
(SE)	(0.0120)	0.0107	(0.0126)
β	0.8932	0.9108	0.8963
(SE)	(0.0119)	0.0103	(0.0128)
γ	0.2043	0.3099	0.1728
(SE)	(0.0481)	(0.0676)	(0.0582)
δ	1.7547	1.6727	1.8178
(SE)	(0.2437)	(0.2612)	(0.2743)
ν		13.5146	14.5341
(SE)		(2.6731)	(3.1166)
ξ			0.9311
(SE)			(0.0182)
Log L	-4793.6	-4771.9	-4766.8
AIC	3.0044	2.9915	2.9889
BIC	3.0158	3.0048	3.0041

Lastly, the estimation results of the AGARCH(1,1) model are summarized in Table 37. What one immediately can see is that the Log L of -4766.8 is still too low in comparison to -4757.2 (GJR-GARCH model with skewed Student-t density). This results conflict with what we have found for TA25 and TA100. Still, we are going to take a look at some out-of-sample performance measures to check those results. They are summarized in Tables 38 and 39.

Table 38: Prediction Measures GARCH(1,1) and GJR-GARCH(1,1) for AEX

	AEX	GARCH		**AEX**	GJR	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	1.2202	1.2050	1.1924	1.2023	1.1933	1.1881
MAE	0.6848	0.6361	0.6339	0.6709	0.6260	0.6295
MedSe	0.2155	0.1492	0.1488	0.1916	0.1473	0.1513
RMSE	1.1046	1.0977	1.0919	1.0965	1.0924	1.0900
TIC	0.3371	0.3356	0.3338	0.3346	0.3339	0.3334
AMAPE	0.5661	0.5576	0.5557	0.5610	0.5522	0.5529
Log-Score	-1.0543	-1.0412	-1.1270	-1.0519	-1.0401	-1.0904

Table 39: Prediction Measures EGARCH(1,1) and APARCH(1,1) for AEX

	AEX	EGARCH		**AEX**	APARCH	
	Normal	Student-t	Skewed Student-t	Normal	Student-t	Skewed Student-t
MSE	1.2137	1.1897	1.1933	1.1998	1.1711	1.1743
MAE	0.6831	0.6326	0.6308	0.6856	0.6321	0.6350
MedSe	0.2045	0.1510	0.1489	0.2131	0.1469	0.1493
RMSE	1.1017	1.0907	1.0924	1.0954	1.0822	1.0836
TIC	0.3362	0.3340	0.3344	0.3344	0.3313	0.3317
AMAPE	0.5634	0.5538	0.5529	0.5638	0.5537	0.5540
Log-Score	-1.0560	-1.0456	-1.0967	-1.0695	-1.0570	-1.1099

Based on these Tables, a bigger Log-Score is often accomplished by a Student-t distribution. The biggest Log-Score can be found by the GJR-GARCH(1,1) model, again contradicting the results of the TA25 and TA100 index. We can conclude that it is not always the best way to just model the volatility with an EGARCH/APARCH model. It is useful to estimate all four models and then it depends on which index one wants to predict which model performs best in and out of sample!

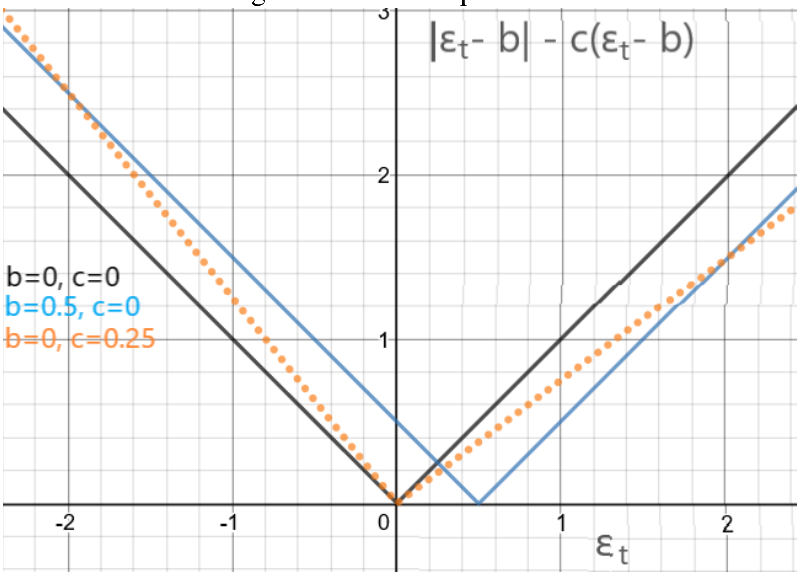
10 Two other GARCH models: AGARCH and TGARCH

In this section, some further analysis will take place. The paper of Hentschel (1995) is taken as a common thread. In this article, the US equity returns are being modelled. But for this thesis: instead of the US equity returns, the data of TA25 and TA100 are used. Just like US equity returns, our data have empirical regularities: asymmetry of returns and leptokurtosis. The meaning of these two regularities is already discussed in Section 2.
Basically the asymmetry of returns can be described as follows: when a return is positive, the volatility is going up. When the return is equal as before, but now with a negative sign, the volatility will relatively go more up. This is called the leverage effect. Another behavioural aspect are the so called fat-tails. This means that there are too many extreme events. In case the reader forgot: fat tails of the returns can be reduced by normalizing. However, it is not entirely gone by normalizing. Still, too many extreme events! This has an effect on the estimates of the different models. Suppose that there is a large shock. This happens, when an extraordinary event takes place. For a standard GARCH model (in which the shock of the last period is being squared), the variance increases a lot. This can be seen: $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. Some propose that these dramatic increases in volatility should be truncated! This can for example be found in the article of Friedman and Laibson (1989).

The article of Hentschel (1995) starts with an introduction. The paper develops a framework with models that are related to each other. These nested models are then being compared to each other. A lot of the models of the paper: EGARCH, GARCH, GJR-GARCH and APARCH, we have already seen. However, the paper also comes up with: TGARCH and AGARCH! These two models also can be interesting for our TA25 and TA100 data, so we are going to investigate this.
For these two models, a mean equation and a variance equation should be specified. In the mean equation, the data are described as a function of a constant, possible other variables and an error term. The error term has a certain variance. In the variance equation, the evolution of this conditional variance is modelled. This conditional variance model can for example contain: the lags of past conditional variance and lags of error terms.
One way to model the mean equation is: $r_t = \mu + \gamma\sigma_{t-1}^2 + \eta_t$. In literature this is called: GARCH-in-Mean (GARCH-M) model and is developed by Engle et al. (1987). This equation is used for

time-varying risk premia as well. The first component, r_t is the logarithm of the return on the market portfolio. However, first the risk free rate is subtracted. The second component, μ is a constant. The third component, $\gamma\sigma_t$ is a time-varying riskpremium. And lastly, the fourth component, η_t is a heteroskedastic error term. This last component has a conditional variance so can be written as: $\eta_t = \sigma_{t-1}\epsilon_t$. With ϵ_t having a zero mean and variance of one. This notation (of Hentschel (1995)) is illogical for this thesis, because of the naming of some of the parameters. So to keep it up in the style of Alberg et al. (2008) we are going to use mean equation: $y_t = \mu + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$. The variance equation is often modelled as a GARCH(1,1) process. This can be written in two ways. One way that we have seen before is: $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. Another way, that is totally identical is: $\sigma_t = \omega + \alpha|\epsilon_{t-1}| + \beta\sigma_{t-1}$. In this second way of writing it, the square root of the conditional variance is taken, and the squared errors are replaced by the absolute value of the error. This last way of writing it has some advantages. This is because large shocks have a smaller effect on the conditional variance, just like we wanted! The reason for this can be found in Jensen's inequality. This way of writing it, is somewhat different from what we have modelled until now. It is interesting to look further into this topic!
Let us take a deeper look to: $\sigma_t = \omega + \alpha|\epsilon_{t-1}| + \beta\sigma_{t-1}$. What one can see is that, asymmetry of the shocks is still not taken into this equation. We need to somehow change this, because empirically we know that there is asymmetry. To make an asymmetric version of this, we introduce the function $f(\epsilon_{t-1})$: $\sigma_t = \omega + \alpha f(\epsilon_{t-1}) + \beta\sigma_{t-1}$. Where: $f(\epsilon_{t-1}) = |\epsilon_{t-1} - b| - c(\epsilon_{t-1} - b)$. In literature (Hentschel (1995)), this model is labeled AGARCH(1,1) model. Let us plot $f(\epsilon_{t-1})$, to get a feeling of the function and why it can be asymmetric!

Figure 10: News impact curve



The parameters that one could change are b and c. If b is changed, for example it becomes positive, the black line is moved to the right. It turns into a blue line. The effect of a shock is now different, depending on whether it is a positive or negative shock. This is just what we wanted to have! If c is changed, for example it becomes positive, the black line is shifted into the orange line. This rotation of the news impact curve also shows the asymmetric variance response. If c is negative, the rotation would be the opposite, so counterclockwise. These two transformations of the black line into the colorful ones should be treated different from each other. The blue line is best for small shocks. When the shocks are larger, the asymmetric effect is almost enterily gone. The orange line however keeps the shock when it is zero, almost at the origin, which means the conditional variance does barely change. The asymmetric effect of small shocks is very small with the orange line. When combining the shift and rotation, according to Hentschel (1995), it is possible to get a nice model which can capture the asymmetric effect best. With the right c and b, it is possible to have an asymmetric effect for small shocks. A symmetric effect for moderate shocks, and an asymmetric effect for large shocks. Let us now state the AGARCH and TGARCH model, two models that also contain the b and c parameter! Notice that the TGARCH model is just the same as the AGARCH model, only with the restriction that b=0.

AGARCH(1,1)

$y_t = \mu + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\sigma_t = \omega + \alpha(|\epsilon_{t-1} - b| - c(\epsilon_{t-1} - b)) + \beta\sigma_{t-1}$
$z_t \sim i.i.d(0, 1)$

Parameter Restrictions AGARCH(1,1)

$\sigma_t = \omega + \alpha(|\epsilon_{t-1} - b| - c(\epsilon_{t-1} - b)) + \beta\sigma_{t-1}$
$\theta = (\mu, \omega, \alpha, \beta, c, b)$
$\omega > 0, \alpha \geq 0, \beta \geq 0, |\ c\ | < 1$

This model permits an asymmetric response. We want to know what the parameter estimates are. For this, we need to make an assumption about the innovations. In the article of Hentschel (1995), the assumption of normality is made. So this is what we also are going to do. However, as we previously have seen in other sections, our TASE25/TASE100 data does not seem to be normally distributed. To overcome numerical problems with the absolute value, the absolute value is going to be replaced! Just like in the article, we assume that: $|\epsilon_t| \approx \sqrt{a^2 + \epsilon_t^2}$. By using this hyperbola, the estimation can take place and is still quite accurate. We are now going to estimate θ for the TA25 data. Of course the model can also be estimated for the TA100 data and many, many more indices. However, since we want to see if it performs better than previous models, we are starting with these data. It is useful to know if the in-sample log-likelihood is better than before, and/or if the standard performance measures are better in comparison to the EGARCH(1,1) model. The estimation results can be found in Table 40. The b is positive: 0.1974. This is what we expected because a positive shift to the right means that when there is a negative shock, the sigma is prone to change more in comparison to the same positive shock. The c parameter is almost 0, so it means there is no rotation. The Log Likelihood is -4113.6. This number is better than the GARCH(1,1) (-4129.7) and GJR-GARCH(1,1) (-4116.3) models (with normal innovations). This number is worse than the EGARCH(1,1) (-4106.4) and APARCH(1,1) (-4108.0) models (with normal innovations). When comparing this AGARCH(1,1) model with one of the best other models: the EGARCH(1,1) with student-t innovations, one can easily notice that the EGARCH(1,1) model still performs better out of sample. All the numbers of the MSE/MAE/MedSe/RMSE/TIC and AMAPE are better for the EGARCH(1,1) model, see: Table 41. This means that we can not draw any new conclusions, so we move on to the TGARCH(1,1) model! Let us first take a look at the model and the parameter restrictions, and then after that at the estimation results.

Table 40: Estimates of variance equation parameters of AGARCH(1,1)

TA25	Estimate	(SE)
μ	0.0024	0.015872
α	0.1176	0.015574
β	0.8706	0.018571
ω	0.0354	0.007982
c	0.0000	0.055605
b	0.1974	0.031707
Log L	-4113.6	
AIC	2,703	
BIC	2,711	

Table 41: Out of sample performance measures TA25 for different models

	(Normal)	(Student-t)
TA25	AGARCH(1,1)	EGARCH(1,1)
MSE	0.0667	0.0412
MAE	0.2303	0.1786
MedSe	0.0667	0.0320
RMSE	0.2583	0.2031
TIC	0.3391	0.3051
AMAPE	0.5570	0.5504

TGARCH(1,1)

$y_t = \mu + \epsilon_t$ where $\epsilon_t = \sigma_t z_t$
$\sigma_t = \omega + \alpha(|\epsilon_{t-1}| - c(\epsilon_{t-1})) + \beta\sigma_{t-1}$
$z_t \sim i.i.d(0, 1)$

Parameter Restrictions AGARCH(1,1)

$\sigma_t = \omega + \alpha(|\epsilon_{t-1}| - c(\epsilon_{t-1})) + \beta\sigma_{t-1}$
$\theta = (\mu, \omega, \alpha, \beta, c)$
$\omega > 0, \alpha \geq 0, \beta \geq 0, | c |< 1$

Just like before, the standard deviation is a function of previous standard deviations and shocks. In this Threshold GARCH model, b is set to zero. This means that a shift is not allowed. It does however allow a rotation! The same assumption $|\epsilon_t| \approx \sqrt{a^2 + \epsilon_t^2}$ is made to overcome numerical problems in estimating. The estimation results can be found in Table 42. The Log Likelihood is -4109.2. This is better than the AGARCH(1,1) model with normal innovations(-4113.6). However, when we look at the different out-of-sample performance measures, the EGARCH(1,1) model with Student-t distribution still is better!

Table 42: Estimates of variance equation parameters of TGARCH(1,1)

TA25	Estimate	(SE)
μ	0.0062048	0.015065
α	0.11544	0.015049
β	0.86751	0.019336
ω	0.042474	0.009709
c	0.29812	0.057082
Log L	-4109.2	
AIC	2,700	
BIC	2,708	

Table 43: Out of sample performance measures TA25 for different models

	(Normal)	(Student-t)
TA25	TGARCH(1,1)	EGARCH(1,1)
MSE	0.0632	0.0412
MAE	0.2247	0.1786
MedSe	0.0685	0.0320
RMSE	0.2515	0.2031
TIC	0.3352	0.3051
AMAPE	0.5567	0.5504

11 Density Prediction: Frequentist and Bayesian Approach

Now that we have seen the frequentist approach for estimating several GARCH models, there is one thing left to do. This one thing is called: Bayesian approach! In this approach, the parameters that we are going to estimate have an underlying model and are not just point estimates! It is assumed that they are coming from some (posterior) distribution! We can simulate draws from this distribution, and come up with a posterior mean of θ. With this vector of parameters $\theta_{bay\hat{e}sian}$ we can make point predictions for future returns. It is also possible to make density predictions for future returns. The density predictions that one gets from this approach of the Tel Aviv 25 and Tel Aviv 100 index returns can then be compared with the one obtained from the frequentist approach. The quality of the density forecast can then be compared between the two. Our interest is in the comparison of those two approaches. We will follow the same procedure as is done in the article of Hoogerheide et al. (2012).

11.1 Introduction

We have already estimated different GARCH models using the Maximum Likelihood Approach. The Estimation results of this can be found in Tables 6,8,7,9. By doing ML estimation, we came across some numerical difficulties. For example: we needed equality constraints. By choosing for an alternative approach, the Bayesian Approach, we will get some extra benefits. Some of the benefits include: small sample estimation results, robust estimation, model discrimination and probabilistic statements of parameters. The purpose of this Section is to compare the performance of the GARCH models via these two approaches. For both approaches we will take a look at the relative predictive accuracy.

11.2 Bayesian Approach

The four models that we have already estimated using the ML method are: GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1) and APARCH(1,1). For these models we have come up with $\hat{\theta}$, a vector of parameter estimates. The assumption is that the vector contains point estimates. The frequentist namely assumes that the parameters are fixed but unknown. The data are however random according to them. We could also assume that each of the elements of the theta vector have a certain distribution. This means the parameters are random. The Bayesian view is that functions of the data are given after being observed. This is therefore a change of perspective! The whole parameter-vector can thus be viewed as a vector from a posterior distribution. Sadly, we do not know the exact name of this distribution. Somehow we still want draws from it. We know that this so called 'posterior distribution' of the parameter vector follows a certain formula. This formula is: $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(\theta))p(y|\theta)d\theta}$. The y in this formula are the observed log-returns. This is also known as Bayes' theorem. All the parameters in θ are stochastic. In order to simulate the draws from the posterior, we need prior ideas about the density of θ. Since we have no clue, we choose for $\mu, \omega, \alpha, \beta$ a flat uniform prior. We do take in account the area of the parameters that is allowed! So: ω, α, β non-negative. This is also known as a non-informative prior. For ν, we specify the following prior: $\lambda exp(-\lambda(\nu - 2))$ with $\lambda = 0.05$ (a translated exponential density). We want: $\nu > 2$. This is also known as a non-informative proper prior. We need a proper prior because otherwise the posterior will not be proper! This is also done in the paper of Geweke (1993). For $p(y|\theta)$, the likelihood function, we use a GARCH model with Student-t distributed innovations. To get draws from the posterior, we will adapt the Random Walk Metropolis Hastings Method (see below). This method is easier and more robust than the Independence Chain Metropolis Hasting Method. Besides a prior and likelihood function, we also need a candidate distribution. As a candidate we take a Student-t distribution with mean given by: the last accepted draw and variance equal to: the estimated co-variance matrix of the ML-estimator. In total we will get 30,000 draws but we discard the first 5,000 (burn-in). So this means we have 25,000 draws from the posterior density. Below in the boxes, the steps of the RWMH are summarized and one can find the restrictions/implementations that are performed in OxMetrics regarding this method.

Random Walk Metropolis Hasting Method

Choose feasible initial value $\theta_0 = \hat{\theta}_{ml}$
Do for draws i=1 until n_{draws}
Simulate candidate draw: $\tilde{\theta} = (\mu, \omega, \alpha, \beta, \nu)$ from symmetric density Q(.)
Compute acceptance probability α=min($\frac{P(\tilde{\theta}}{P(\theta_{i-1})}$,1) with posterior density kernel P(θ)=p(θ)p(y|θ)
simulate U from u(0,1)
if U$\leq \alpha$ accept: $\theta_i = \tilde{\theta}$
if U> α reject: $\theta_i = \theta_{i-1}$

Ox implementations RWMH

• For numerical reasons we need log prior and log posterior
•$\frac{P(\tilde{\theta})}{P(\theta_{i-1})} = \frac{exp[\ln P(\tilde{\theta}]}{\exp[\ln P(\theta_{i-1}]} = exp[\ln(p(\tilde{\tilde{\theta}}) + \ln p(y|\theta) - \ln p(\theta_{i-1}) - \ln p(y|\theta_{i-1})]$
• As starting value for the $\ln$ p(θ) one needs a σ_0^2 and this should be chosen as the sample mean variance
• If the candidate draw is at least as good as the previous draw, it should be accepted!
• The candidate distribution is Student-t, and needs a fixed value for the Degrees of Freedom parameter. In this paper, we choose: 4.

Let us take a look at the posterior mean of the TA25/TA100 index of the GARCH(1,1) model. The number of accepted draws was around 21 percent for the TA25 and around 23 percent for the TA100 index ("Indicating a reasonable performance"). The results can be found in Tables 44 and 45. The trace-plots of the parameters can be seen in Figures 11 and 12. The Figures are showing that theta moves through the sample space and that it does not get stuck at some points. The quantiles are shown for 2.5 percent and 97.5 percent. The qunatiles are (except for μ) all positive. This means that they are significant different from zero.

Table 44: Posterior Mean of GARCH(1,1) model including the left quantile and right quantile for TA25

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0253	-0.0011	0.0512
α	0.0638	0.0442	0.0873
β	0.8938	0.8562	0.9250
ω	0.0172	0.0087	0.0285
ν	8.0348	6.4484	10.0580

Table 45: Posterior Mean of GARCH(1,1) model including the left quantile and right quantile for TA100

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0243	-0.0107	0.0587
α	0.0859	0.0527	0.1259
β	0.7871	0.6797	0.8699
ω	0.0739	0.0355	0.1279
ν	7.9976	6.0562	10.5429

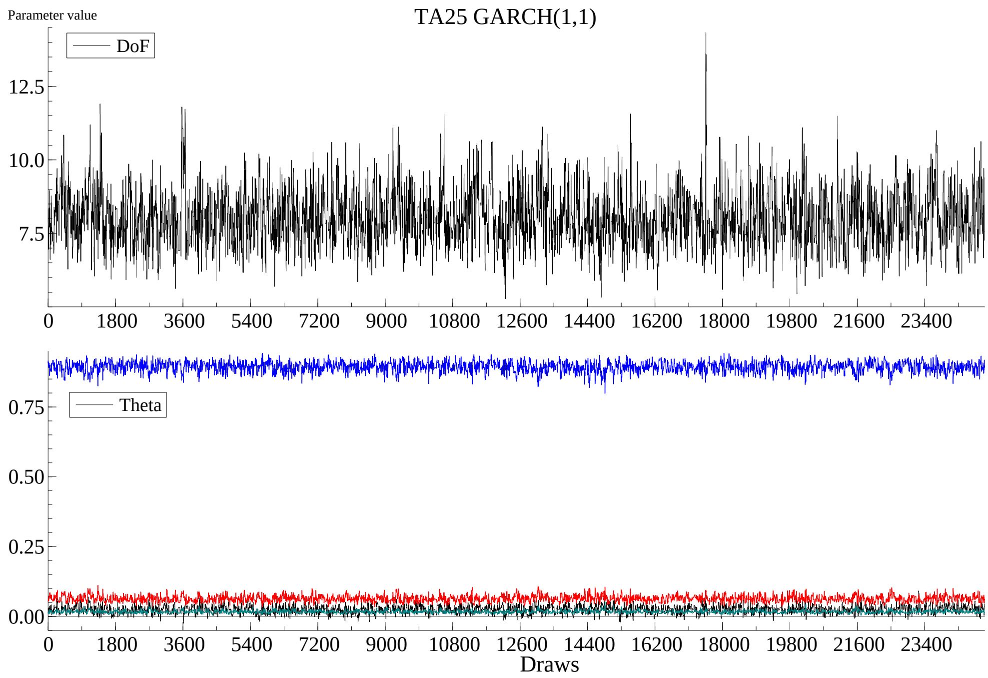


Figure 11: Traceplot of the parameters of the GARCH(1,1) model for TA25 returns

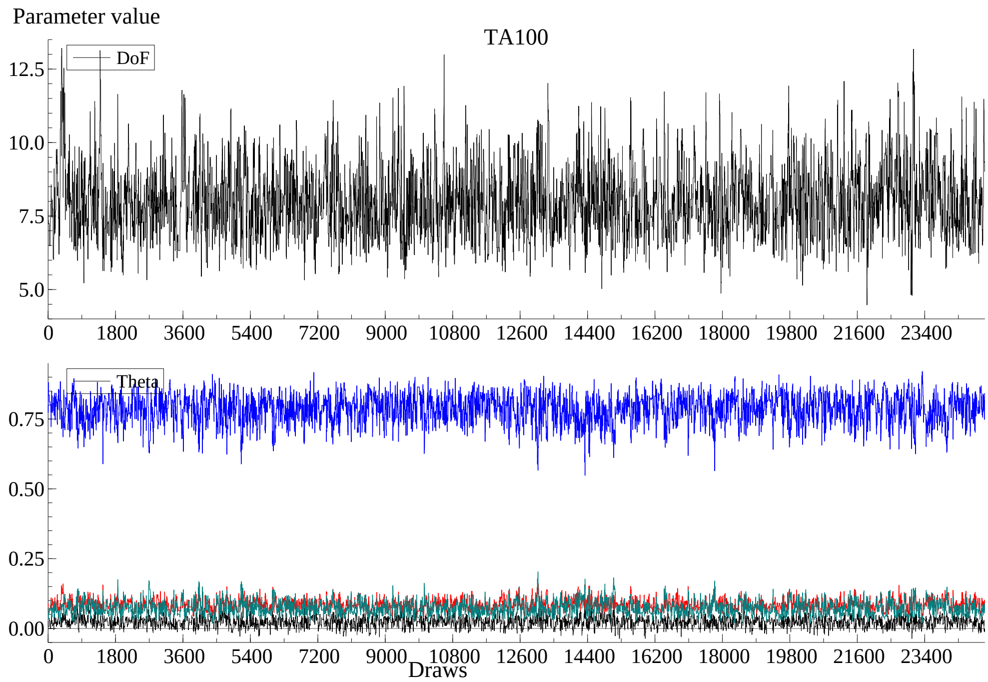


Figure 12: Traceplot of the parameters of the GARCH(1,1) model for TA100 returns

For the GJR-GARCH(1,1), EGARCH and APARCH models the Bayesian results can be found in Tables: 46,47,48,49,50,51. It includes the posterior mean and the 95 percent credible interval. For the GJR-GARCH(1,1) model the acceptance rate was around 21 percent (TA25) and 18 percent (TA100). The γ parameter in this model is the parameter for the leverage effect and is for both TA25 and TA100 significant positive. The γ parameter in the EGARCH model is significantly negative. This means a negative shock times this negative parameter makes the log variance go up. For the APARCH and EGARCH model the acceptance rate was around 15 and 10 percent respectively: so quite low.

Table 46: Posterior Mean of GJR-GARCH(1,1) model including the left quantile and right quantile for TA25

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0104	-0.0047	0.0181
α	0.0531	0.0346	0.0753
β	0.8490	0.7992	0.8918
ω	0.0167	0.0088	0.0279
γ	0.1170	0.0725	0.1636
ν	8.2650	6.6445	10.3616

Table 47: Posterior Mean of GJR-GARCH(1,1) model including the left quantile and right quantile for TA100

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0085	-0.0249	0.0491
α	0.0879	0.0563	0.1277
β	0.6822	0.5400	0.7975
ω	0.0871	0.0435	0.1473
γ	0.1750	0.0814	0.2795
ν	8.2178	6.2707	10.9677

Table 48: Posterior Mean of EGARCH(1,1) model including the left quantile and right quantile for TA25

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0075	-0.0744	0.0463
α	0.1378	-0.3068	0.2300
β	0.9584	0.8937	0.9817
ω	0.5901	-1.6279	1.0137
γ	-0.0537	-0.1261	-0.0124
ν	8.2173	5.5935	12.4756

Table 49: Posterior Mean of EGARCH(1,1) model including the left quantile and right quantile for TA100

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0041	-0.0463	0.0764
α	0.2413	0.1065	0.3632
β	0.7917	0.3899	0.9182
ω	1.1628	0.3511	1.9275
γ	-0.1222	-0.2725	-0.0579
ν	9.1181	5.5012	15.6804

Table 50: Posterior Mean of APARCH(1,1) model including the left quantile and right quantile for TA25

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0125	-0.0166	0.0392
α	0.0746	0.0564	0.0963
β	0.8971	0.8650	0.9244
ω	0.0247	0.0142	0.0384
γ	0.3761	0.2430	0.5346
δ	1.2624	0.9498	1.6103
ν	8.8869	7.0437	11.3552

Table 51: Posterior Mean of APARCH(1,1) model including the left quantile and right quantile for TA100

	$\theta_{posterior}$	Left Quantile	Right Quantile
μ	0.0132	-0.0215	0.0474
α	0.0944	0.0646	0.1257
β	0.7958	0.7198	0.8556
ω	0.0905	0.0532	0.1418
γ	0.4844	0.2809	0.7348
δ	1.3222	0.8787	1.8511
ν	10.1597	7.3067	14.9391

11.3 Forecasting Performance

It is possible to do a one-day-ahead point forecast or a one-day-ahead density forecast. We choose the latter. The ultimate goal is to find a density forecast that is almost the same as the true unobserved density for daily returns. We can do this by comparing $\hat{f}_t(y_{t+1})$ and $\hat{g}_t(y_{t+1})$. Let us first take a look at how $\hat{f}_t(y_{t+1})$ is calculated. This is the Bayesian density forecast. In the previous section we have simulated θ_i with $i = 1, 2, \ldots, 25000$. We can now fill in y_{t+1} in f($y_{t+1}|I_t, \theta_i$) for each of the i. We then get 25.000 filled in likelihood values. We take the average of this and are left with: $\hat{f}_t(y_{t+1})$. By doing it this way, the uncertainty about θ is taken into account completely. All the simulated draws of θ (after the discarded burn-in) are used! Let us now take a look at how $\hat{g}_t(y_{t+1})$ is calculated. This is the frequentist density forecast. Since we only have one θ: θ_{ML} it is much easier. This θ is simply used for g($y_{t+1}|I_t, \theta_{ML}$). By filling in y_{t+1} we are left with $\hat{g}_t(y_{t+1})$. Now, we can calculate the Kullback-Leibler information criterion (KLIC) and use it to obtain the loss differential vector. This is calculated as $d_{t+1} = \ln \hat{g}_t(y_{t+1}) - \ln \hat{f}_t(y_{t+1})$. Once this vector is obtained, a Diebold and Mariano (1995) type test statistic can be calculated. This should ofcourse be combined with (robust) Newey-West standard errors. It is not only possible to look at the forecasting performance of the whole density. We could also take a look at the forecasting performance of the left tail of the density. The measurement that we should then use is the: censored likelihood (CSL). However, this is left for the Future Research Section.

11.4 Results comparison Bayesian and frequentist approach

In this section the results of the predictive accuracy of different GARCH models are discussed. We already know from previous sections that for the TA25 and TA100 the EGARCH/APARCH model outperforms the standard GARCH model. We can also make the comparison: Frequentist GARCH versus Bayesian GARCH. The null hypothesis is: equal predictive accuracy of the frequentist and Bayesian approach. We use 30 daily forecasts as out-of-sample number of observations. The results are summarized in Table 52. When the t-stat is outside the [-1.96;1,96] interval the Bayesian approach is significantly better. This only happens one time: at the APARCH(1,1) model. However, since the calculations are based on only 30 daily density forecast, one should be careful with the interpretation of this. It may be wise to investigate this further with more out-of-sample observations, to be entirely certain the EGARCH(1,1) model estimated with the Bayesian approach outperforms the 'frequentist/ML point estimate' model. However since there is no evidence of the Bayesian approach being less accurate in comparison to the frequentist approach, one can certainly do the first. In the Article of Hoogerheide et al. (2012), there are also significant results that the Bayesian approach is more accurate for risk management applications!

Table 52: Diebold and Mariano t-statistic (using Newey-West standard error) in favour of the Bayesian approach (versus the frequentist approach). Calculations are based on the average loss differential for KLIC and based on 30 daily density forecasts.

Diebold t-stat	**TA25**	**TA100**
GARCH(1,1)	0.4492	0.9437
GJR-GARCH(1,1)	0.3620	1.0278
EGARCH(1,1)	0.7442	1.1704
APARCH(1,1)	8.7168	0.5717

12 Suggestions For Future Research

In the previous section, we have calculated the Kullback-Leibler information criterion. We did not find any significant results between the forecasting performance of the two approaches. However, it is also possible to only look at the left-tail of the predicted density. According to the article of Hoogerheide et al. (2012), this is important for risk management applications. To test for this one have to calculate the censored likelihood (CSL) scoring rule. More information about this criterion could be found in the Article of Diks et al. (2008).

Censored likelihood scoring rule

$CSL(\hat{f}_t|y_{t+1} = \mathbb{1}_{[y_{t+1}\leq r]}\hat{f}_t(y_{t+1}) + \mathbb{1}_{[y_{t+1}>r]}\ln(1-\hat{F}_t(r))$

- Where $\ln(1-\hat{F}_t(r))$ is the total non-tail probability if y_{t+1} is not a tail event
- Where $\hat{f}_t(y_{t+1})$ is the forecasted density if y_{t+1} is a tail event
- With $\hat{F}_t(r)$ is the Student-t cumulative density at time t evaluated at threshold r
- The treshold could be time varying or could be fixed, for example at -2.5 percent

Once the CSL score function is calculated, different (competing) models can be compared. This can be done in the same way as we did by the KLIC. So: comparing the difference in the score of both models and performing a Diebold and Mariano type test. The null hypothesis of equal predictive accuracy of the frequentist and Bayesian approach can then be rejected (or accepted). So for future research I would suggest performing this test on the data of TA25 and TA100. It is important to use a large out-of-sample period (so not only 30 observations but for example 500) because there should be enough tail events. It is also interesting to look at the difference in forecasting performance when different priors are used (here we used non-informative ones) or different distributions for the innovations (for example skewed Student-t).

Acknowledgements

The author acknowledges L. Hoogerheide for the usefull comments and the great supervision throughout the whole period 4, 5 and 6 of the VU academic year 2019. During this period of time Dr. Hoogerheide not only supervised this thesis project, but also several others. He managed to do this besides lecturing for Econometrics II and Data Analysis II and writing an article for the Journal of Econometrics. Although his busy schedule he still took time to answer my questions, for which I am thankful.
Also the knowledge obtained by several subjects taught by N.J Seeger (Empirical Finance), A. Fernandez (Analysis 1,2), H. Houba and R. Brink (Mathematical Economics), S.J. Koopman (Empirical Marketing), A. Ridder (Operations Research) helped a lot by understanding the material used for this thesis.

A Mistakes in the Article of Alberg et al. (2008)

This is a list with mistakes that are made:

• Page 1203: the log likelihood function of the Student-t distribution misses an $\Gamma(.)$ in the second term
• Page 1203: the log likelihood function of the skewed Student-t distribution misses an $\Gamma(.)$ in the second term
• Page 1203: for the equation of m: $\Gamma(\nu + 1)$ should be: $\Gamma(\nu - 1)$
• Page 1206: the denominator of the AMAPE should be: $|\sigma_t^2 \hat{-} \sigma_t^2|$
• Page 1206: the denominator of the TIC should contain a '+'

References

Alberg, Dima, Haim Shalit, and Rami Yosef (2008). "Estimating stock market volatility using asymmetric GARCH models". In: *Applied Financial Economics* 18, pp. 1201–1208. DOI: `10.1080/09603100701604225`.

Baillie, Richard and Tim Bollerslev (1989). "The Message in Daily Exchange Rates: A Conditional-Variance Tale". In: *Journal of Business Economic Statistics* 7.3, pp. 297–305. URL: `https://EconPapers.repec.org/RePEc:bes:jnlbes:v:7:y:1989:i:3:p:297-305`.

Beine, Michel, Sebastien Laurent, and Christelle Lecourt (2002). "Accounting for conditional leptokurtosis and closing day effects in FIGARCH models of daily exchange rates". In: *Applied Financial Economics* 12, pp. 589–600. DOI: `10.1080/09603100010014041`.

Black, Fischer and Myron Scholes (1976). "TAXES AND THE PRICING OF OPTIONS". In: *The Journal of Finance* 31.2, pp. 319–332. DOI: `10.1111/j.1540-6261.1976.tb01889.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1976.tb01889.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1976.tb01889.x`.

Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3, pp. 307–327. URL: `https://EconPapers.repec.org/RePEc:eee:econom:v:31:y:1986:i:3:p:307-327`.

— (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return". In: *The Review of Economics and Statistics* 69.3, pp. 542–47. URL: `https://EconPapers.repec.org/RePEc:tpr:restat:v:69:y:1987:i:3:p:542-47`.

Diebold, Francis X. (2015). "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of DieboldfffdfffdfffdMariano Tests". In: *Journal of Business & Economic Statistics* 33.1, pp. 1–1. DOI: `10.1080/07350015.2014.983236`. eprint: `https://doi.org/10.1080/07350015.2014.983236`. URL: `https://doi.org/10.1080/07350015.2014.983236`.

Diks, Cees, Valentyn Panchenko, and Dick van Dijk (2008). "Partial Likelihood-Based Scoring Rules for Evaluating Density Forecasts in Tails". In: *SSRN Electronic Journal*. DOI: `10.2139/ssrn.1135531`.

Ding, Zhuanxin, Clive W.J. Granger, and Robert F. Engle (1993). "A long memory property of stock market returns and a new model". In: *Journal of Empirical Finance* 1.1, pp. 83–106. ISSN: 0927-5398. DOI: `https://doi.org/10.1016/0927-5398(93)90006-D`. URL: `http://www.sciencedirect.com/science/article/pii/092753989390006D`.

Engle, Robert (1982). "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50, pp. 987–1007. DOI: 10.2307/1912773.

Engle, Robert, David M Lilien, and Russell P Robins (1987). "Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model". In: *Econometrica* 55.2, pp. 391–407. URL: https://EconPapers.repec.org/RePEc:ecm:emetrp:v:55:y:1987:i:2:p:391-407.

Friedman, Benjamin and David Laibson (1989). "Economic Implications of Extraordinary Movements in Stock Prices". In: *Brookings Papers on Economic Activity* 20, pp. 137–190. DOI: 10.2307/2534463.

Geweke, J. (1993). "Bayesian Treatment of the Independent Student-t Linear Model". In: *Journal of Applied Econometrics* 8, S19–S40. ISSN: 08837252, 10991255. URL: http://www.jstor.org/stable/2285073.

Glosten, Ravi, Lawrence Jagannathan, and David E Runkle (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". In: *Journal of Finance* 48, pp. 1779–1801. DOI: 10.1111/j.1540-6261.1993.tb05128.x.

Hentschel, Ludger (1995). "All in the family Nesting symmetric and asymmetric GARCH models". In: *Journal of Financial Economics* 39.1, pp. 71–104. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(94)00821-H. URL: http://www.sciencedirect.com/science/article/pii/0304405X9400821H.

Hoogerheide, Lennart F., David Ardia, and Nienke Corré (2012). "Density prediction of stock index returns using GARCH models: Frequentist or Bayesian estimation?" In: *Economics Letters* 116.3, pp. 322–325. ISSN: 0165-1765. DOI: https://doi.org/10.1016/j.econlet.2012.03.026. URL: http://www.sciencedirect.com/science/article/pii/S016517651200122X.

Lambert, Philippe and Laurent (2001). "Modelling skewness dynamics in series of financial data using skewed location-scale distributions". In:

Nelson, Daniel B (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach". In: *Econometrica* 59.2, pp. 347–370. URL: https://ideas.repec.org/a/ecm/emetrp/v59y1991i2p347-70.html.

Poon, Ser-Huang and Clive W.J. Granger (2003). "Forecasting Volatility in Financial Markets: A Review". In: *Journal of Economic Literature* 41.2, pp. 478–539. URL: http://www.aeaweb.org/articles?id=10.1257/002205103765762743.