VRIJE UNIVERSITEIT AMSTERDAM

BACHELOR THESIS

The impressive power of flexible parametric GARCH models applied to exchange rate returns

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION
DEPARTMENT OF ECONOMETRICS AND OPERATIONS RESEARCH

Author:
Rajni RASIAWAN

Supervisor:
Dr. L. F. Hoogerheide
Second reader:
A. Borowska MPhil

June 30, 2018

Abstract

In this research, the GARCH, EGARCH and GJR-GARCH models in combination with the student-t, the exponential generalized beta distribution of the second kind (EGB2) and generalized error distribution (GED) are being compared based on their fit to seven exchange rates series with the Canadian dollar as base currency. Exchange rate returns are characterized by leptokurtosis, skewness, high peakedness and volatility clustering. Since the traditional normal and student-t distribution do not seem to capture these properties well, the EGB2 and GED distributions are employed. To assess the validity and compare the models, the Akaike Information Criterion, Ljung-Box statistic, Goodness of Fit statistic and predicted and observed higher order moments are used. In addition, the forecasting performance of the estimated models is examined using Value at Risk backtesting methods and the logarithmic scoring rule. The results indicate that the EGARCH-EGB2 outperforms the other models. Subsequently, this model is also estimated using Bayesian estimation, which is carried out using the Metropolis-Hastings algorithm of simulation. Although little difference between the parameter estimators was found, the Bayesian density forecasts significantly outperform the frequentist density forecasts in five out of the seven analyzed cases.

Keywords: GARCH, EGARCH, GJR-GARCH, EGB2, GED, Goodness of Fit, Value at Risk, Logarithmic Scoring Rule, Diebold-Mariano, Bayesian

Contents

1 Introduction

Globalization or increased global trade and financial integration has been one of the major trends shaping the world economy over recent years. Nowadays, $5.1 trillion is traded on the foreign exchange market every day. For comparison: the GDP of the Netherlands was $771 billion in 2016, which means we transfer more than six times as much money on a daily basis than the Netherlands produces in an entire year. Analyzing and predicting exchange rates and the risk that is accompanied with them is therefore highly beneficial for international companies and for enhancing international trade in general.

For modelling the complicated exchange rate dynamics, it is of great importance to incorporate the most common properties that exchange rate time series exhibit. First of all, economic theory and econometric evidence show that the kurtosis of financial data is often larger than the three, which is called leptokurtosis. There are two likely explanations for the leptokurtosis found in exchange rate series. The first is the exchange rate overshooting hypothesis, first developed by economist Dornbusch (1976). This theory states that foreign exchange rate will temporarily overreact to changes in monetary policy to compensate for sticky prices of goods in the economy. The second explanation is speculative attacks against fixed exchange rates (Krugman 1979). A speculative attack in the foreign exchange market refers to the massive and sudden selling of a nation's currency. Both imply that extreme exchange rate returns will occur relatively often and thus cause thicker tails than the normal distribution.

However, in the absence of monetary expansion and speculative attacks, modal daily exchange rate returns are near zero (Obstfeld and Rogoff 1996) yielding high peakedness. Besides, several studies have found the 'leverage-effect' for majority of currencies indicating that negative shocks have a larger impact on the volatility than positive shocks. Wang et al. (2001) give two reasons. Firstly, permanent shocks that lead to changes in the equilibrium exchange rate may be asymmetric. Secondly, speculative attacks against a currency tend to be one-sided causing depreciation/devaluation. Skewness might therefore also be important to take into account.

Another phenomenon observed is volatility clustering. This refers to the tendency that large changes in prices of financial assets tend to cluster together, resulting in persistence in the magnitude in price changes (Mandelbrot 1963). Economically, this means that as markets respond to new information with large price movements, these high-volatility environments tend to endure for a while after that first shock. This finding is usable in predicting future volatility and of key importance, since volatility is one of the most important measures of financial risk.

In short, it is important to capture the leptokurtosis, high peakedness, asymmetry and volatility clustering. Contemporary modeling of exchange rate time series makes widespread use of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. GARCH models can not only capture the volatility clustering often found in exchange rate series, they also accommodate some of the leptokurtosis. But GARCH models with conditionally normal errors generally fail to capture sufficiently the leptokurtosis evident in asset returns (Bollerslev 1987, Baillie and Bollerslev 1989, Hsieh 1989, Baillie and DeGennaro 1990, Wang, Barrett, and Fawson 1996). For this reason, the student-t distribution has been adopted for the conditional distribution of the error (Bollerslev 1987, Bollerslev et al. 1994). However, the student-t distribution does still seem not capable enough to capture the high peakedness and fat-tail property. Therefore, Wang et al. (2001) introduce the exponential generalized beta distribution of the second kind (EGB2) that can accommodate for both leptokurtosis and skewness. The generalized error distribution (GED) can have both thicker and thinner tails and larger or smaller kurtosis than the normal distribution. However, it does not allow for asymmetry. To incorporate asymmetric news impact effect, extensions of the basic GARCH model will be studied. These extensions include the Exponential GARCH (EGARCH) model by Nelson (1991) and the GJR model by Glosten et al. (1993).

An important application of volatility modelling is Value at Risk (VaR) analysis. VaR is a measure of the risk of loss for investments. It focuses on the left tail of the return distribution and computes the worst probable loss within a certain level of confidence. Ignoring leverage may result in underestimating that risk (Engle, 2004). Therefore, the performance of the previously mentioned GARCH-models on the left tails will be compared. Using this information, firms and regulators can gauge the amount of assets needed to cover possible losses which can contribute in preventing financial disasters.

This thesis is organized as follows. First the datasets are described and their general statistics are represented in section 2. The model specifications for conditional volatility are presented in section 3, the Maximum Likelihood estimation in section 4 and the distributions for the innovations in section 5. The models in-sample fit on the data are compared according to the Ljung-Box statistic, Akaike Information Criterion (AIC), Goodness of Fit (GoF) and the higher order moments as described in section 6. Thereafter, the forecasts will be evaluated based on the relative predictive accuracy in; the entire density and the one-day-ahead Value at Risk forecasts, which is more relevant for risk management purposes, are considered in section 7. In addition, the EGARCH-EGB2 model is also estimated using Bayesian estimation, which is carried out using the Metropolis-Hastings algorithm described as in section 8. The Bayesian forecasts are also compared to the frequentist forecasts based on the logarithmic scoring rule. The results are presented in section 9. Section 10 contains the conclusions and section 11 gives suggestions for further research. Additional output can be found in the Appendix of the paper.

2 Analysis of the data

The data used are noon spot Canadian dollar exchange rates for six currencies: Deutsche mark (DM), British pound, Japanse yen, French franc (FF), Belgian franc (BF) and Italian lira (IL) from the period 1 January 1985 to 21 November 1996, as reported by the Exchange Rate Service of the Pacific Data Center at the University of British Columbia. The noon spot Canadian dollar exchange rates for the euro were sourced from *Investing* and cover the period 17 August 2006 to 9 March 2018. To obtain stationarity, we take the log-difference from each series. $r_{i,t} = \log[S_{i,t}/S_{i,t-1}] * 100$, where $r_{i,t}$ can be interpreted as the procentual change in the nominal exchange rate of currency i at period t. In this way, $r_{i,t} > 0$ implicates appreciation and $r_{i,t} < 0$ depreciation. In the table below descriptive statistics are given.

Table 1: Descriptive statistics for the log return data of the six exchange rates.

	DM	£	¥	FF	BF	IL	€
Observations	3016	3016	3016	3016	3016	3016	3016
Mean	0.02528	0.01321	0.02745	0.0219	0.02435	0.009078	-0.002948
Maximum	3.650	3.714	4.162	3.232	3.667	3.016	4.190
Minimum	-3.806	-3.118	-3.557	-3.641	-3.559	-6.914	-3.087
Skewness	-0.03711	-0.1202*	0.2863*	0.01973	0.02488	-0.6161*	-0.03356
	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
Kurtosis	5.122	5.181	6.143	5.018	5.037	8.772	5.095
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
$f_{0.75} - f_{0.25}$	1.126	1.081	1.022	1.14	1.118	1.093	1.184
$f_{0.6} - f_{0.4}$	0.4117	0.3961	0.3815	0.3926	0.4066	0.3216	0.4032
Jarque-Bera	566.7*	604.9*	1283*	511.9*	521.5*	4378*	552*
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
$Q(30)$	30.04	37.98	37.8	41.16	41.69	33.39	37.39
	[0.4638]	[0.1502]	[0.155]	[0.08436]	[0.07613]	[0.3057]	[0.1662]
$Q^2(30)$	391.4*	451.3*	234*	391.7*	370.8*	641.7*	676.4*
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]

Notes: the standard errors are given in parentheses and the p-values in brackets. JB denotes the Jarque-Bera test statistic, which is used to test whether the data is normally distributed. $Q(30)$ and $Q^2(30)$ represent the Ljung-Box test statistics for up to 30th order serial correlation for each exchange rate series. $f_{\alpha 1-\alpha 2}$ is the interpercentale range of the standardized log returns, i.e. $f_{\alpha 1-\alpha 2}$ is the distance between the chosen quantiles. The asymptotic standard errors of the skewness and kurtosis are obtained by $(6/T)^{0.5}$ and $(24/T)^{0.5}$ respectively.

Several inferences can be made based on this table. First of all, the pound, yen and lira show significant skewness, which means that the data are not symmetrically distributed. The negative skewness in the pound and lira exchange rate series probably reflect the speculative attacks on the Italian lira and British pound and the subsequent exit from the exchange rate mechanism (ERM) in autumn 1992. The positive skewness in the Japanese yen can be explained by the rapid improvements in Japanese productivity over the past fourthy years. This shows that permanent shocks, that lead to changes in the equilibrium exchange rate, may be asymmetric. Furthermore, the kurtosis of all exchange rate series are above the normal distribution's reference value of three, so there is leptokurtosis. This supports the fact that normality of all exchange rate series is rejected.

Noteworthy is the fact that the euro has been introduced in 2002 and is a merging of twelve currencies at the time, including the Deutsche mark, French franc, Belgian franc and Italian lira. One of the reasons of the introduction of the euro, was to have a stable currency with less fluctuations. This is reflected in the kurtosis; the euro has one of the lowest kurtosis-coefficients.

The interperentile range $f_{\alpha 1} - f_{\alpha 2}$ (with $\alpha 1 > \alpha 2$) illustrates the peakedness of the data, where f is the inverse of the cumulative distribution function (CDF). It gives the difference between two quantiles. As reference values we have $f_{0.75} - f_{0.25} = 1.36$ and $f_{0.75} - f_{0.25} = 0.5$ for the normal distribution. The smaller the interquartile range, the higher the peak of the data, assuming a unimodal distribution. Since interquartile values are all smaller than the reference values of the normal distribution, these datasets have higher peaks than the normal distribution.

Furthermore, based on the Ljung-Box test statistic $Q(30)$ the null hypothesis of no serial autocorrelation in the log-returns is not rejected. This is in line with the efficient market hypothesis which states that current returns can not be predicted using past returns. Tests on most of the different exchange rate series cannot reject the hypothesis of a zero conditional mean (Yang, 2006). However, the Ljung-Box test statistic $Q^2(30)$ shows there is significant autocorrelation in the squared log-returns. This implies that we might need an autoregressive structure to model the variance. For this reason we will consider GARCH-models.

All in all, it is clear that the logreturns are not normally distributed, leptokurtotic and in some cases significantly skewed. To model this properties, we will employ the EGB2 and GED in addition to the regular normal and student-t distributions. To model the autocorrelation in the squared logreturns, we will consider GARCH models. This is elaborated in the next section.

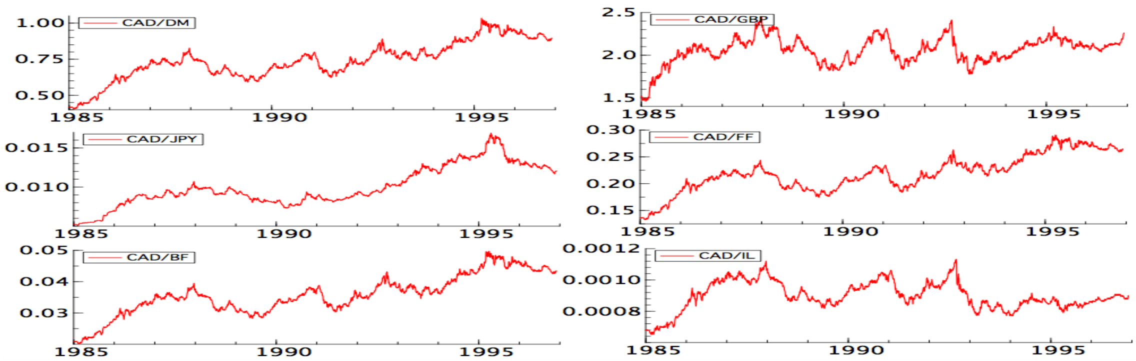

Figure 1: The six exchange rates over the period 1 January 1985 to 21 November 1996.

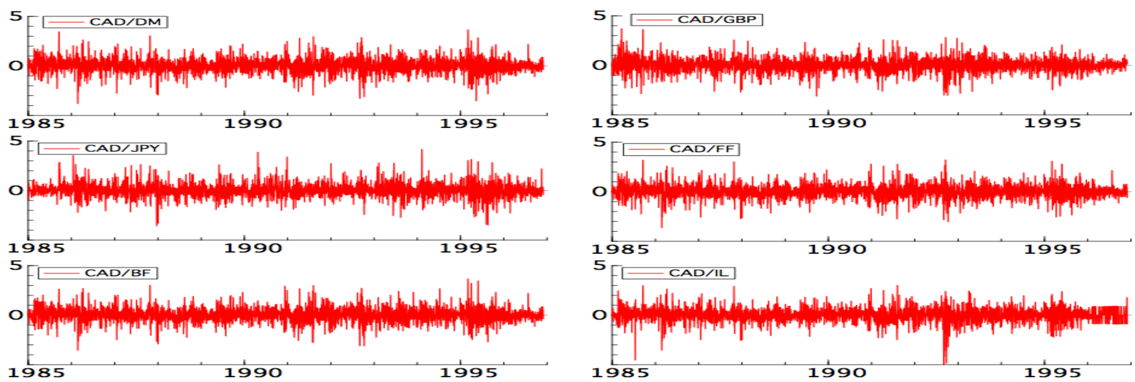

Figure 2: The log-returns of the six exchange rates over the period 1 January 1985 to 21 November 1996.

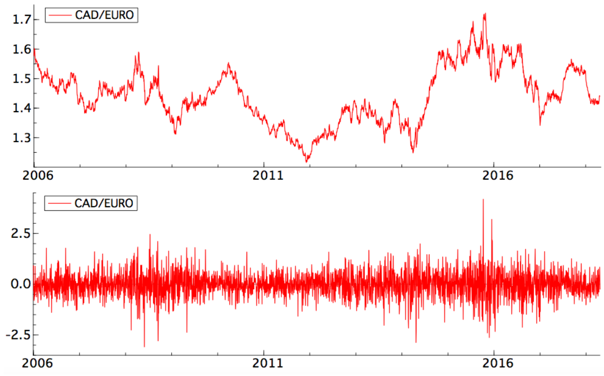

Figure 3: The exchange rates (above) and log-returns (below) of the CAD/Euro over the period 17 August 2006 to 9 March 2018.

3 Models

The mean equation of each model is specified in one of two ways. If the exchange rate series exhibits significant first order autocorrelation, the mean equation is specified as an AR(1) process:

$$r_{i,t} = \phi_0 + \phi_1 r_{i,t-1} + \epsilon_t \tag{1}$$

where r_t is the logreturn of exchange rate and ϵ_t is the error term in period t. If the coefficient ϕ_1 is insignificant, the mean equation is specified as:

$$r_{i,t} = \phi_0 + \epsilon_t \tag{2}$$

The error term can be decomposed as $\epsilon_t = \sigma_t e_t$ where σ_t is the conditional volatility at time t and e_t is an independent and identically distributed sequence of shocks with mean zero and unit variance $\{e_t\}_{t\in\mathbb{Z}}$ ~NID(0,1). In addition to the mean equation stated above, another specification is needed to describe how the conditional volatility σ_t evolves over time and a specification of the distribution for the innovations e_t. As already mentioned in the introduction, GARCH models are capable to capture heteroskedasticity and volatility clustering in financial data. Therefore the symmetric GARCH(1,1) and asymmetric EGARCH(1,1) and GJR-GARCH(1,1) are adopted for σ_t. The normal, student-t, GED and EGB2 distributions are employed for $\{e_t\}_{t\in\mathbb{Z}}$.

3.1 GARCH(1,1)

One of the corner-stones in modelling volatility is the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model of Bollerslev (1986) and Taylor (1986). Let $\sigma_t^2 = Var(\epsilon_t|F_{t-1})$ denote the conditional variance at time t. The general form of the GARCH(p,q) is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-1}^2 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 \tag{3}$$

To ensure that the conditional variance σ_t^2 is bigger than zero, it must hold that $\omega > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$. Furthermore, to have that $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a weakly stationarity white noise sequence, the GARCH(p,q) must satisfy $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. However, in most empirical applications it turns out that the simple specification $p = q = 1$ is able to reproduce the volatility dynamics of financial data. Therefore, the GARCH(1,1) will be used in this thesis:

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \epsilon_{t-1}^2 \tag{4}$$

To understand why the GARCH(1,1) model is able to capture high persistence in the variance, we can unfold the GARCH(1,1) model as an ARCH(∞) model:

$$\sigma_t^2 = \omega + \beta(\omega + \beta\sigma_{t-2}^2 + \alpha\epsilon_{t-2}^2) + \alpha\epsilon_{t-1}^2 \tag{5}$$

$$= \omega + \beta\omega + \alpha\epsilon_{t-1}^2 + \beta\alpha\epsilon_{t-2}^2 + \beta^2\sigma_{t-2}^2 \tag{6}$$

$$= \frac{\omega}{1-\beta} + \alpha\sum_{i=1}^{\infty}\beta^i\epsilon_{t-1-i}^2 \tag{7}$$

The squared observation ϵ_{t-1}^2 can be seen as an estimate of the variance at time $t-1$. When ϵ_{t-1}^2 is large, then σ_t^2 also tends to be large (for positive α). In that way, ϵ_t is more likely to be large and the return $r_{i,t}$ through equation (1) or (2) as well. Thus, the parsimonious GARCH(1,1) model with only two parameters has the impressive power to capture high volatility persistence.

3.2 EGARCH(1,1)

In order to incorporate asymmetric effects depending on the sign of the innovation, the Exponential GARCH (EGARCH) is considered. The EGARCH, introduced by Nelson (1991), is described as:

$$\log(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - E[|z_{t-1}|]) + \gamma z_{t-1} + \beta\log(\sigma_{t-1}^2) \tag{8}$$

In contrast to the most models, no restrictions need to be imposed to ensure positive conditional variance. Moreover, note that $E[|z_{t-1}|]$ depends on the density function of z_{t-1}. For the standard normal distribution $\mathbb{E}(|z_{t-1}|) = \sqrt{\frac{2}{\pi}}$ and for the Student-t distribution with ν degrees of freedom,

$$\mathbb{E}(|z_{t-1}|) = \frac{\sqrt{\nu}\Gamma(0.5(\nu-1))}{\sqrt{\pi}\Gamma(0.5\nu)}$$

and for the GED distribution,

$$\mathbb{E}(|z_{t-1}|) = \lambda 2^{1/\nu}\frac{\Gamma(2/\nu)}{\Gamma(1/\nu)}.$$

For the EGB2 distribution there is no analytical expression for $\mathbb{E}(|z_{t-1}|)$. However, $\mathbb{E}(|z_{t-1}|)$ only influences the constant term ω. It has no effect on the log-likelihood, the predictions of σ_t and the density or the other parameters, since it is a constant of which the omission is simply 'corrected' by ω. Noteworthy is the fact that ln(σ_t^2) is an AR(1) process and therefore $|\beta|$<1 is a sufficient condition for stationarity and finite kurtosis. This model allows positive and negative shocks with the same magnitude to have a different effect. A positive shock has an effect of $\alpha + \gamma$ and a negative shock has an effect of $\alpha - \gamma$.

3.3 GJR-GARCH(1,1)

The Glosten-Jagannathan-Runkle GARCH (GJR) model by Glosten, Jagannathan and Runkle (1993) is another extension of the GARCH model that takes the asymmetric effect into account. It is introduced by Glosten, Jagan and Runkle (1953):

$$\sigma_t^2 = \omega + (\alpha + \gamma\mathbb{1}_{\{\epsilon_{t-1}<0\}})\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{9}$$

The indicator variable $\mathbb{1}_{\{\bullet\}}$ takes on a value of one if the condition holds and zero otherwise. A positive shock has therefore an effect of α and a negative shock has an effect of $\alpha + \gamma$. Just as for the GARCH(1,1) model, non-negativity constraints on the parameters assure that the variance is positive. Therefore $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \gamma \geq 0$. The stationarity requirement is $\alpha + \beta + \gamma P(\epsilon_{t-1} < 0) < 1$, where $P(\epsilon_{t-1} < 0) = \frac{1}{2}$ for symmetric distributions like the normal, Student-t and GED distributions..

4 Maximum Likelihood Estimation

In order to estimate the models, the Maximum Likelihood estimation is employed using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. The Maximum Likelihood estimation aims at obtaining an estimate of the fixed, unknown parameter vector θ by maximizing the loglikelihood function, given the observations:

$$l(\epsilon;\theta) = \log(\epsilon|\theta) = \frac{1}{N}\sum_{t=1}^{N} \log f(\epsilon_t|\theta) \tag{10}$$

The idea behind this so called frequentist approach is that the data are random and the parameters are fixed. In the BFGS algorithm, the parameters are chosen in such a way that the log-likelihood function is maximized. The computed Maximum Likelihood estimators are consistent estimators for the parameters when the model is correctly specified. In the BFGS algorithm, parameter restrictions are not taken into account. For that reason, the parameters are first transformed, causing the transformed parameters to be unrestricted. After the optimization, the parameters are transformed back. In the next section the distributions for the innovations will be discussed with the corresponding log-likelihood function $l(\epsilon;\theta)$. For each model the log-likelihood function divided by the number of observations will be maximized in order to prevent numerical problems.

5 Distributions for the innovations

5.1 Normal distribution

The standard normal distribution is a commonly used assumption for the distribution of the error e_t. The standard normal Gaussian distribution has mean equal to zero and variance equal to one. The probability density is given by:

$$f(\epsilon|\theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2}\frac{\epsilon_t^2}{\sigma_t^2}\right) \tag{11}$$

The log-likelihood can subsequently be derived as:

$$l(\epsilon;\theta) = -\frac{1}{2}\Big[N\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\big(\log(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2})\Big] \tag{12}$$

The skewness of the standard normal distribution is zero and the kurtosis is three. The Gaussian GARCH models regularly fail to account adequately for the fat tails found in unconditional asset price distributions (Hsieh 1989; Wang, Barrett and Fawson 1996). Therefore, we will consider the next distributions.

5.2 Student-t distribution

The standardized student-t distrubution also has mean zero and variance one. The degrees of freedom shape parameter ν determines the thickness of the tail and thereby the kurtosis. The density of the student-t probability distribution is given by:

$$f(\epsilon_t|\theta,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)\sigma_t^2}}\left(1+\frac{\epsilon_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}} \tag{13}$$

with log-likelihood:

$$l(\epsilon;\theta) = \sum_{t=1}^{N}\Big[\log\Gamma(\frac{\nu+1}{2}) - \log(\Gamma(\frac{\nu}{2})) - \frac{1}{2}\log[\pi(\nu-2)]\Big] - \frac{1}{2}\sum_{t=1}^{N}\Big[\log(\sigma_t^2) + (\nu+1)\log(1+\frac{\epsilon_t^2}{\sigma_t^2}(\nu-2))\Big]. \tag{14}$$

This distribution has skewness 0 if $\nu > 3$, the skewness is undefined for $2 < \nu \leq 3$. The kurtosis is $3 + \frac{6}{\nu-4}$ for $\nu > 4$, the kurtosis is ∞ for $2 < \nu \leq 4$. The student-t distribution converges to the normal distribution as $\nu \to \infty$. The student-t distribution can be an accurate distribution of the daily exchange rate returns, as long as the empirical distribution of the daily returns is (almost) symmetric.

5.3 Generalized Error Distribution

The density of the Generalized Error Distribution (GED) with mean zero, variance σ_t^2 and shape parameter ν is:

$$f(\epsilon_t|\theta) = \frac{\nu\exp(-0.5|\frac{\epsilon_t}{\lambda}|^{\nu})}{2^{1+\frac{1}{\nu}}\sigma_t\Gamma(\frac{1}{\nu}+1)} \tag{15}$$

The skewness of the GED distribution is zero and the kurtosis is equal to $\frac{\Gamma(\frac{5}{\nu})\Gamma(\frac{1}{\nu})}{\Gamma^2(3/\nu)}$. The GED distribution can accommodate both fatter and thinner tails than the normal distribution. If $0 < \nu < 2$, the GED has fatter tails and larger kurtosis than the normal distribution. For $\nu = 2$ it is the normal distribution. If $\nu > 2$ it has thinner tails and smaller kurtosis than the normal distribution. Since the exchange rate series have larger kurtosis the normal distribution, we expect that ν will be larger than 2. The log-likelihood function is given by:

$$l(\epsilon;\theta) = \sum_{t=1}^{N}\Big[-\log(2^{(1+(1/\nu))}\gamma(1/\nu)\lambda - \frac{1}{2}\log(\sigma_t^2) + \log(\nu) - \frac{1}{2}\Big|\frac{\epsilon_t}{\lambda\sigma_t}\Big|^{\nu}\Big] \tag{16}$$

5.4 EGB2 distribution

The Exponential Generalized Beta Distribution of the second kind (EGB2) was introduced by McDonald (1984). The EGB2 distribution is a special kind of the Exponential Generalized Beta distribution (EGB) and is an alternative representation of the generalized logistic distribution (Johnson and Kotz 1970, Patil et al. 1984). The EGB2 is capable of modelling the peakedness and skewness. The density function is given by:

$$EGB2(z;\delta,\sigma,p,q) = \frac{e^{\frac{p(z-\delta)}{\sigma}}}{|\sigma|B(p,q)(1+e^{\frac{z-\delta}{\sigma}})^{p+q}} \tag{17}$$

where δ is a location parameter that affects the mean of the distribution, σ the scale of the density function and p and q are shape parameters that together determine the skewness and kurtosis of the distribution. The EGB2 incorporates the normal distribution approximately when $p = q$ approaches infinity. The distribution is symmetric if $p = q$ and positively skewed (negatively skewed) if $p > q$ ($p < q$) with $\sigma > 0$. However, the accommodation of skewness is restrictive. The skewnesscoefficient can take on values between -2 and 2 and the kurtosiscoefficient can take values up to 9 (McDonald 1991). For the standardized EGB2 distribution with the shape parameters p and q, the log- likelihood function of GARCH-EGB2 model is:

$$l(\epsilon;\theta) = N[\log(\sqrt{\Omega}) - \log(B(p,q)) + p\Delta] + \sum_{t=1}^{n} \Big[p(\frac{\sqrt{\Omega}\epsilon_t}{\sqrt{\sigma_t^2}}) \tag{18}$$

$$-0.5\log(\sigma_t^2) - (p+q)\log(1+\exp(\frac{\sqrt{\Omega}\epsilon_t}{\sqrt{\sigma_t^2}}) + \Delta\Big] \tag{19}$$

where

$$\Delta = \psi(p) - \psi(q)$$
$$\Omega = \psi'(p) - \psi'(q)$$

Furthermore, the skewness is equal to:

$$\frac{\psi''(p) - \psi''(q)}{(\psi'(p) + \psi'(q))^{3/2}} \tag{20}$$

and the kurtosis is equal to:

$$3 + \frac{\psi'''(p) + \psi'''(q)}{(\psi'(p) + \psi'(q))^2} \tag{21}$$

in which ψ, ψ', ψ'', ψ''' represent the polygamma function of the first, second, third and fourth order, respectively.

6 In-sample fit

6.1 Jarque-Bera Test

To test whether the data are normally distributed or not, the Jarque-Bera statistic is computed. The Jarque-Bera test is a goodness-of-fit test to test whether the sample data have the skewness and kurtosis matching a normal distribution.

$$JB = (\sqrt{\frac{N}{6}} \times skewness)^2 + (\sqrt{\frac{N}{24}} \times (kurtosis - 3))^2 \tag{22}$$

Under the null hypothesis the skewness is zero and kurtosis three, which are the reference values of the normal distribution. Asymptotically it holds under the null hypothesis of normality that $JB \sim \chi^2(2)$.

6.2 Ljung-Box

The Ljung-Box test statistic is used to test whether a group of autocorrelations of a time series are different from zero. The test statistic is as follows:

$$Q(k) = n(n+2)\sum_{j=1}^{m} \frac{\hat{\rho_j}^2}{n-j} \sim \chi^2(m) \tag{23}$$

where m the number of lags, $\hat{\rho_j}$ the sample autocorrelation for lag j and n the number of observations. Clusters of volatility may reveal themselves through autocorrelation in squared log-returns, whereby the null hypothesis of the Ljung-Box test would be rejected. If however, the variance structure is modelled correctly, the squared standardized residuals should contain no autocorrelation anymore. In that case the null hypothesis would not be rejected.

6.3 Goodness of Fit

The χ^2 Goodness of Fit (GoF) statistic compares the differences between observed frequencies of standardized residuals and the theoretically predicted frequencies based upon estimated distribution. The whole domain is divided into 40 intervals and the frequency of the values in each interval is compared with the frequency of values based on the estimated distribution in each interval. The Goodness of Fit statistic is calculated by:

$$GoF = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i} \tag{24}$$

where f_i the observed count frequency of actual residuals in the ith interval, F_i the predicted frequency derived from the estimated shape parameters of the distribution and k the number of intervals in $(-\infty, \infty)$. Under the nulhypothesis the observed and predicted distributions are the same and the GoF has asymptotically a chi-squared distribution with degrees of freedom equal to the number of intervals minus one minus the number of estimated shape parameters of the distribution. The number of shape parameters is one for the normal, student-t and GED-distribution and two for the EGB2 distribution. Furthermore, the number of intervals is chosen to be 40 as in Wang et al. (2001).

7 Forecasting

After having compared the fit of the models on the data, we want to compare the forecasts. The dataset is therefore split into an in-sample part, consisting of the first 2000 observations (from 1 January 1985 till 2 December 1992) and an out-of-sample part, consisting of the last 1015 observations (from 3 December 1992 till 21 November 1996). To assess the validity and to compare the forecasts of the different models, Value at Risk backtests and the logarithmic scoring rule will be used.

7.1 Value at Risk

The term Value at Risk (VaR) entered the financial terminology in the early 1990s. It is now one of the most common techniques used to measure and quantify the level of financial risk within a firm or investment portfolio over a specific time frame. The daily α-VaR is the minimum amount the investor stands to lose with probability α over a period of one day. Mathematically, this means that the α-VaR in percentage loss at time t is defined as the percentage value c that satisfies P($r_{i,t} \leq$ c) = α. However, VaR has the tendency to underestimate the actual loss if an extreme event actually occurs. The one-day-ahead VaR estimation is given by:

$$VaR_{T+1} = \underbrace{\mu_t}_{\text{mean equation}} + \big[CDF^{-1}(\alpha)\big]\hat{\sigma}_{T+1} \tag{25}$$

where $CDF^{-1}(\alpha)$ denotes the α-th quantile of the assumed distribution. The mean equation is either defined as equation (1) or equation (2). Moreover, $\hat{\sigma}_{T+1}$ is the one-step ahead predicted conditional volatility that also depends on the assumed distribution. Conditional on the past, it is a constant.

Define a violation as an observation for which the historical return is smaller than the estimated VaR-estimate. A valid Value at Risk model not only produces the correct number of violations but also violations that are equally spread over time. If the exceptions are clustered, this indicates that the model does not accurately capture the changes in volatility and correlations. To assess the quality of the models' VaR estimation, three tests proposed by Christoffersen (1998) will be applied: the unconditional coverage, independence and conditional coverage test.

Unconditional Coverage

The Unconditional Coverage test is used to test whether the proportion of violations of the out-of-sample differs significantly from α. Define N = $\sum_{t=1}^{T} \mathbb{1}_t$ as the number of violations over the out-of-sample period T, where

$$\mathbb{1}_t = \begin{cases} 1 & \text{if } r_{i,t} < VaR_t \\ 0 & \text{if } r_{i,t} \geq VaR_t \end{cases}$$

Ideally $\frac{N}{T}$ is equal to α, 1 minus the confidence level. In that case, a violation will occur with probability α and a 'success' (i.e. no violation) with probability $1-\alpha$. Testing whether $\mathbb{E}[\mathbb{1}_t] = \alpha$ is equivalent to testing $\mathbb{1}_t \sim Bernoulli(\alpha)$ for all t. For the LR_{uc} statistic for unconditional coverage, the null hypothesis $H_0 : \mathbb{E}[\mathbb{1}_t] = \alpha$ is tested against the althernative hypothesis $H_1 : \mathbb{E}[\mathbb{1}_t] \neq \alpha$. The likelihood under the null hypothesis looks like:

$$L(\alpha, \mathbb{1}_t, \mathbb{1}_t ... \mathbb{1}_T) = (1-\alpha)^{n_0}\alpha^{n_1} \tag{26}$$

where $n_0 = 1 - \sum_{t=1}^{T} \mathbb{1}_t$ and $n_1 = \sum_{t=1}^{T} \mathbb{1}_t$. Under the alternative hypothesis, the likelihood is:

$$L(\pi, \mathbb{1}_t, \mathbb{1}_t ... \mathbb{1}_T) = (1-\pi)^{n_0}\pi^{n_1} \tag{27}$$

The likelihood ratio statistic for unconditional coverage LR_{uc} can subsequently be composed as:

$$LR_{uc} = -2\log\left[L(\alpha; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T)/L(\hat{\pi}; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T)\right] \sim \chi^2(s-1) = \chi^2(1) \tag{28}$$

where $\hat{\pi} = \frac{n_1}{n_0+n_1}$ and $s = 2$, the number of possible outcomes of the sequence.

Independence test

One shortcoming of the unconditional coverage test is that it does not take into account whether or not the violations are clustered. Violation clustering is important as it implies repeated severe capital losses to the institution which together could result in bankruptcy. Probably the most common test for independence of the violations has been proposed by Christoffersen (1998). Define $\pi_{ij} = \mathbb{P}(\mathbb{1}_t = j|\mathbb{1}_{t-1} = i)$, so it is the probability of a transition to state j given that the current state is i. The markov transition probability matrix is then:

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

Also define n_{ij} as the number of observations with value i followed by j with $i, j \in \{0, 1\}$. The independence likelihood ratio statistic can be written as:

$$L(\Pi_1; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T) = (1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{11}. \tag{29}$$

The likelihood function from equation (27) can be optimized with the Maximum Likelihood method, which yields the estimate of $\hat{\Pi}_1$:

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \frac{n_{10}}{n_{10}+n_{11}} & \frac{n_{11}}{n_{10}+n_{11}} \end{bmatrix}$$

The transition probability matrix under the null hypothesis that the sequence $\{\mathbb{1}_t\}$ is independent corresponds to:

$$\Pi_0 = \begin{bmatrix} 1-\pi_1 & \pi_1 \\ 1-\pi_1 & \pi_1 \end{bmatrix}$$

The likelihood under the null hypothesis is equal to:

$$L(\Pi_0; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T) = (1-\pi_1)^{n_{00}+n_{10}} \times \pi_1^{n_{11}+n_{01}} \tag{30}$$

Finally, the likelihood ratio statistic for independence becomes:

$$LR_{ind} = -2log\left[L(\hat{\Pi}_0; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T)/L(\hat{\Pi}_1; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T)\right] \sim \chi^2((s-1)^2) = \chi^2(1) \tag{31}$$

where $\hat{\pi}_1 = \frac{n_1}{n_0+n_1}$ in $\hat{\Pi}_0$ and where s is again 2, because $\{\mathbb{1}_t\}$ can take on two values.

Conditional Coverage

The conditional coverage test is a joint test of unconditional coverage and independence. This test examines jointly whether the fraction of violations is significantly different from that expected under the null hypothesis and whether there is violation clustering. Since the quality of the model for forecasting VaR depends both on the correct probability of violations and on the independence property of the violations, the conditional coverage needs to be tested as well. If we combine the two statistics described above we obtain the conditional coverage likelihood ratio statistic:

$$LR_{cc} = -2\log\left[L(\alpha; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T)/L(\hat{\Pi}_1; \mathbb{1}_1, \mathbb{1}_2, ..., \mathbb{1}_T)\right] \tag{32}$$

LR_{cc} is under the null hypothesis asymptotically χ^2 distributed with two degrees of freedom, since there are two restrictions now. Conditional on the first observation, the Conditional Coverage likelihood ratio can be simplified to: $LR_{cc} = LR_{uc} + LR_{ind}$.

7.2 Logarithmic Scoring Rule

To compare the performance of the different forecast densities the logarithmic scoring rule will be used. The logarithmic scoring rule compares the forecasts of two models at the same time. The loss differential is composed as:

$$d_{t+1} = \log\hat{g}_t(y_{t+1}) - \log\hat{f}_t(y_{t+1}) \tag{33}$$

Under the null hypothesis of equal forecast accuracy the loss differential should be zero on average. The null hypothesis H_0: $\mathbb{E}(d_t) = 0$ is tested against the alternative hypothesis: H_1: $\mathbb{E}(d_t) \neq 0$, where $\bar{d} = \frac{1}{H}\sum_{h=1}^{H} d_{T+h}$. This can be tested by a Diebold and Mariano (1995) statistic:

$$DMT = \frac{\bar{d}}{\sqrt{LRV_{\bar{d}}/H}}, \tag{34}$$

where $LRV_{\bar{d}}$ represents the long-run variance estimator. The Diebold-Mariano test statistic converges under the null hypothesis to the standard normal distribution. Because $\{d_t\}_{t=h}^{H}$ can be serially correlated, the robust Newey-West standard errors are used.

8 Bayesian Estimation

The Maximum Likelihood estimation described above can be classified as classical statistical estimation. Conclusions can differ between different classical statisticians, because different significance levels and/or one-sided/two-sided tests can be used.
A Bayesian can speak about probabilities of hypotheses (probability distributions of parameters), instead of merely about rejecting the null hypothesis. Furthermore, it takes parameter uncertainty in a natural way into account in forecasts and decisions based on forecasts.

The most important theorem in Bayesian statistics is the Bayes' rule for parameters θ and data Y:

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)} \tag{35}$$

$p(\theta|Y)$ denotes the posterior density of the parameter θ, $p(\theta)$ the prior density of the parameter θ, $p(Y|\theta)$ the likelihood-function and $p(Y)$ the prior probability of the data Y. The Metropolis Hastings algorithm is a Markov Chain Monte Carlo (MCMC) method. This simulation method was introduced by Metropolis et al (1953) and eleborated by Hastings (1970). The algorithm consists of the following steps:

1. Simulate candidate draw $\tilde{\theta}$ from candidate density $Q(\cdot)$.
2. Compute acceptance probability $\alpha = min\{\frac{p(\tilde{\theta})p(Y|\tilde{\theta})}{p(\theta_{i-1})p(Y|\theta_{i-1})}, 1\}$.
3. Simulate U from uniform distribution on [0,1]. If $U \leq \alpha$, then accept draw: $\theta_i = \tilde{\theta}$. Else the acceptance does not take place and the previous draw $\theta_i = \theta_{i-1}$ is repeated.
4. Repeat steps 1-3 until a sufficiently large sample of parameter draws is collected.

Since we are interested in θ and $p(Y)$ is a scale factor to ensure $p(\theta|Y)$ integrates to 1, we could rewrite (15) as:

$$p(\theta|Y) \propto p(\theta)p(Y|\theta) \tag{36}$$

where the symbol $\propto$ means 'proportional to'. The right hand side of expression (36) is called the kernel of the posterior probability density function.

The estimated parameters obtained by the Maximum Likelihood method are used as initialisation. Furthermore, a non-informative prior distribution is used, which is uniform on the allowed region for all parameters. As candidate distribution a normal distribution is used with the mean equal to the previous draw and covariance matrix equal to the obtained estimated covariance matrix of the ML estimator. A sufficiently large number of draws, 100000 draws, are used in the random walk Metropolis-Hastings method with a discarded burn-in period of 10000 draws, as the first draws may depend too much on the initial values of the sequence.

In order to check the validity of the estimation, we look at the acceptance percentage, the first order serial correlation between the different steps in estimation for each parameter and the trace plot of each parameter. A higher acceptance percentage, lower serial correlation and trace plot with fewer horizontal lines (of consecutive rejections) are indications of validity of the candidate distribution. We make use of a non-informative prior distribution, which is uniform on the allowed region for all parameters except for the degrees of freedom parameter (for which we use a proper, non-informative prior distribution).

9 Results

9.1 On the estimated parameters

The estimated parameters are given in Appendix A. The mean equation for all exchange rates are defined as an AR(1) process, except for the Italian lira, because the AR(1) coefficient is not significantly different from zero.

The parameter estimates $\hat{p}$ and $\hat{q}$ of the GARCH-EGB2, EGARCH-EGB2 and GJR-GARCH-EGB2 determine the shape of the estimated EGB2 distribution. For all three models, $\hat{p}$ is bigger than $\hat{q}$ for the Deutsche mark, Japanese yen, French franc and Belgian franc which implies that the standardized errors are positively skewed. The British pound and Italian lira are negatively skewed. This is in line with the results found in Table 1, except for the Deutsche mark. However, it is important to notice that table 1 represents skewness in the returns, whereas the shape parameters determine the skewness in the standardized errors. The difference between $\hat{p}$ and $\hat{q}$ are furthermore small and the Deutsche mark did not show significant skewness in the returns.

The Ljung-Box statistics already showed evidence of volatility clustering in Table 1. This effect is also displayed in the parameter estimates for the conditional variance. The persistence effect for the GARCH(1,1) is $\hat{\alpha} + \hat{\beta}$, for EGARCH $\hat{\beta}$ and for GJR-GARCH $\hat{\alpha} + \hat{\beta} + \hat{\gamma}\,\mathrm{Pr}(\epsilon_t < 0)$, so that it is $\hat{\alpha} + \hat{\beta} + \frac{1}{2}\hat{\gamma}$ for the normal and Student-t distributions. First of all it can be noticed that the stationarity condition is satisfied for all models and all exchange rate series. However, in all cases it is almost equal to one. Hence, all models considered reflect strong persistence of the effect of shocks on future variances.

Furthermore, the shape parameter $\hat{\nu}$ for the GED distribution is smaller than two for all models and all currencies, as expected. The smaller the value $\hat{\nu}$ the fatter the tails and higher the kurtosis is. Noteworthy is the fact that $\hat{\nu}$ is considerably larger for the euro. We have already seen that the tails for the euro are not as thick as for the other distributions. The same conclusion can be drawn from the estimates of the degrees of freedom parameter $\hat{\nu}$ in the student-t distribution. This parameter is in the range (4.2, 6.25) for all exchange rate series, except the euro. The degrees of freedom for the euro is almost twice as big, namely 12.1.

The asymmetric effect of negative and positive shocks is reflected by $\hat{\gamma}$ in the EGARCH model and GJR-GARCH. However, a positive shock in the EGARCH has an effect of $\hat{\alpha} + \hat{\gamma}$ and a negative shock an effect of $\hat{\alpha} - \hat{\gamma}$, while in the GJR-GARCH a positive shock has an effect of $\hat{\alpha}$ and a negative shock of $\hat{\alpha} + \hat{\gamma}$. It turns out that not all of $\hat{\gamma}$ parameters are significantly different from zero. This means that a negative shock does not have a significantly different effect than a positive shock on the conditional volatility. Besides, the parameter $\hat{\gamma}$ is alternately positive or negative for the exchange rates. This is in line with the fact that the posterior distribution of γ in the Bayesian estimation contains both positive and negative values as will be seen in section 9.6.

The estimates for the euro over the period 17 August 2006 to 9 March 2018 seem to differ considerably from the estimates for the other currencies over the period 1 January 1985 to 21 November 1997. The shape parameters that determine the kurtosis, the degrees of freedom parameter $\hat{\nu}$ in the student-t distribution, $\hat{\nu}$ in the GED distribution and $\hat{p}$ and $\hat{q}$ for the EGB2 distribution, are all larger for the euro. This indicates that the estimated kurtosis coefficient is smaller than for the other currencies, which is already observed in section 9.3. Besides, the $\hat{\gamma}$-parameters in the EGARCH-models are positive, while $\hat{\gamma}$-parameters in the GJR-GARCH model are negative. This suggests that there is a leverage-effect for the euro, where negative shocks have a larger effect on the conditional volatility than positive shocks.

Observed is also that the estimated parameters of the conditional variance $\hat{\alpha}$ and $\hat{\beta}$ between the GARCH-EGB2, GARCH-GED and GARCH-T are closer to each other for the non-skewed exchange rate series: Deutsche mark, French franc and Belgian franc. The results exhibit (slightly) greater differences for currencies associated with skewed distributions. This also holds for most of the EGARCH and GJR-GARCH models.

9.2 Ljung-Box statistics

In Table 1 was observed that all exchange rate series exhibit significant autocorrelation in the squared residuals. If the GARCH part is correctly modelled , the estimated variance should capture the autoregressive structure and the squared standardized residuals should contain no autocorrelation. To test this the Ljung-Box test is performed on the squared standardized residuals.

Table 2: Ljung-Box statistics and p-values between brackets.

		DM	£	¥	FF	BF	IL	€
HN		390.6	453.5	236.9	393.6	372.1	641.5	641.5
		[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
GARCH	N	29.76	25.44	28.89	25.58	36.55	23.61	34.70
		[0.4782]	[0.7034]	[0.5234]	[0.6964]	[0.1905]	[0.7894]	[0.2536]
	T	28.56	24.77	30.48	25.22	35.10	25.64	34.62
		[0.5407]	[0.7363]	[0.4415]	[0.7144]	[0.2392]	[0.6933]	[0.2567]
	GED	29.06	24.95	29.81	25.28	35.79	24.23	34.65
		[0.5146]	[0.7274]	[0.4753]	[0.7114]	[0.2151]	[0.7616]	[0.2556]
	EGB2	30.75	24.30	29.64	26.35	36.39	24.62	34.61
		[0.4280]	[0.7582]	[0.4842]	[0.6571]	[0.1957]	[0.7433]	[0.2570]
EGARCH	T	27.99	23.06	31.76	23.58	32.29	31.30	40.34
		[0.5712]	[0.8128]	[0.3788]	[0.7909]	[0.3541]	[0.4010]	[0.09844]
	GED	29.06	24.95	29.81	25.28	35.79	24.23	40.06
		[0.5146]	[0.7274]	[0.4753]	[0.7114]	[0.2151]	[0.7616]	[0.1037]
	EGB2	27.91	23.26	31.83	23.48	32.24	27.80	40.39
		[0.5755]	[0.8043]	[0.3753]	[0.7952]	[0.3563]	[0.5809]	[0.09760]
GJR-GARCH	T	28.25	24.79	29.61	24.75	35.81	25.65	34.64
		[0.5573]	[0.7350]	[0.4858]	[0.7370]	[0.2143]	[0.6929]	[0.2561]
	GED	29.09	25.27	29.46	25.19	36.02	24.37	34.67
		[0.5131]	[0.7117]	[0.4938]	[0.7160]	[0.2073]	[0.7549]	[0.2549]
	EGB2	28.60	24.75	30.30	25.21	35.28	25.35	34.64
		[0.5385]	[0.7372]	[0.4503]	[0.7149]	[0.2328]	[0.7078]	[0.2559]

In Table 2 above the Ljung-Box statistics are given, together with the p-values between brackets. The nulhypothesis is only rejected for the homoskedastic model, which means that there is significant autocorrelation. For all other models autocorrelation is not rejected for all GARCH-models. This suggests the elimination of serial correlation in the conditional variance. Therefore, the volatility clustering can be captured by the GARCH-models.

9.3 Comparison of higher order moments

In order to check the validity of the distribution of the innovation, we compare the observed and higher order moments of the standardized residuals in Table 3. Note that the GED and student-t distributions are symmetric distributions. The sample skewness m_3 is therefore the same as the difference between the sample skewness and the estimated skewness $m_3 - \phi_3$ for these distributions.

As expected, the skewness and excess kurtosis of the standardized residuals still persist in the homosekedastic and GARCH-N model. The sample kurtosis values for the standardized residuals of all models are bigger than three. Moreover, the JB-test strongly rejects the null hypothesis of normality of the standardized residuals for all currencies, so it is valuable to consider the other distributions. The observed kurtosis of the Italian lira, which has the largest sample kurtosis (see Table 1), decreased considerably.

Furthermore, it can be seen that the models with a student-t distribution overestimate the sample kurtosis coefficient, especially for the Japanese yen. The sample kurtosis is also significantly different from the estimated kurtosis for all currencies, except for the Deutsche mark. Therefore, the student-t distribution does not seem to be an appropriate distribution for the innovations. For the GED distribution exactly the contrary effect is visible: the kurtosis coefficient is underestimated for every currency. However, this underestimation is smaller than the overestimation by the student-t distribution. Therefore, the GED distribution still seems to be less inappropriate than the student-t distribution.

For the GED and EGB2 distributions, the absolute difference between the sample skewness/kurtosis and the estimated skewness/kurtosis drops for all currencies compared to the student-t distribution. This decrease if especially visible for the Japanese yen. The difference between sample kurtosis and estimated kurtosis for the EGARCH-EGB2 is only 0.93 compared to 12.295 for the GARCH-T. The EGB2 distribution seems to model the kurtosis even better than the GED distribution, since the difference between sample kurtosis and estimated kurtosis is smaller for the EGB2 than for GED for the same GARCH-model.

Even though the GED distribution is capable of capturing the leptokurtosis, it can not model skewness. Even for the series that did not have significant skewness, the Deutsche mark, French franc and Belgian franch, the GED distribution performs worse than the EGB2 distribution for the same model based on the sample skewness and estimated skewness. The difference between sample kurtosis and estimated kurtosis is also smaller for the EGB2 distribution than for the GED distribution for the same model for all currencies, except the British pound. For the British pound prediction difference in the kurtosis is for example smaller for the GARCH-GED than for the GARCH-EGB2. This can be explained by the fact that the EGB2 distribution is able to accommodate for higher peaks and has enough probability mass in the tails of the distribution.

Remarkable is the fact that all models seem to predict the kurtosis of the euro well. Although the difference between the sample kurtosis and estimated kurtosis still decreases for the models with the GED and EGB2 distribution, this decrease is small compared to the other exchange rates.

Furthermore, there are no big differences visible between the GARCH, EGARCH and GJR-GARCH models for the same distribution. Having a distribution that can accommodate for skewness and leptokurtosis seems much more important in capturing the skewness and kurtosis than a model that can accommodate for asymmetry in the shocks.

All in all these results strongly show the impressive descriptive power of the EGB2 distribution for the third and fourth order moments of the data, in particular in case of the Japanese yen.

Table 3: Sample skewness and kurtosis of the standardized residuals and the skewness and kurtosis of the estimated distributions.

	DM	£	¥	FF	BF	IL	€
HN							
m_3	-0.03712	-0.1202	0.2863	0.019730	0.02488	-0.6162	-0.6162
m_4	5.124	5.183	6.145	5.020	5.038	8.775	8.775
GARCH-N							
m_3	0.07478	-0.1038	0.4654	0.093952	0.1076	-0.1286	0.01409
m_4	4.429	4.314	6.156	4.355	4.424	4.827	3.629
GARCH-T							
m_3	0.08369	-0.1043	0.4994	0.1036	0.1204	-0.1640	0.01306
m_4	4.451	4.321	6.312	4.394	4.473	5.210	3.631
ϕ_4	5.826	6.070	18.61	5.655	5.838	6.354	3.742
$\|m_4 - \phi_4\|$	1.374	1.749	12.30	1.261	1.365	1.144	0.1107
GARCH-GED							
m_3	0.07930	-0.1038	0.4838	0.09885	0.1137	-0.1445	0.01376
m_4	4.440	4.317	6.247	4.372	4.445	5.045	3.630
ϕ_4	4.100	4.130	4.807	4.059	4.100	4.685	3.531
$\|m_4 - \phi_4\|$	0.3402	0.1864	1.4340	0.3124	0.3445	0.3599	0.09894
GARCH-EGB2							
m_3	-0.08168	0.1140	0.5055	-0.09991	0.1160	-0.1516	0.01324
ϕ_3	-0.09720	0.09872	0.3474	-0.08783	0.07064	-0.04682	0.01096
$\|m_3 - \phi_3\|$	0.01551	0.01532	0.1581	0.01208	0.0454	0.1048	0.002271
m_4	4.473	4.368	6.367	4.414	4.477	5.097	3.631
ϕ_4	4.623	4.816	5.371	4.582	4.62	4.895	3.691
$\|m_4 - \phi_4\|$	0.1495	0.4480	0.9960	0.1678	0.1430	0.2021	0.05957
EGARCH-T							
m_3	0.0819	-0.1094	0.4548	0.1151	0.1245	-0.1628	0.01781
m_4	4.371	4.333	6.271	4.356	4.424	5.281	3.598
ϕ_4	5.735	6.238	18.31	5.569	5.722	6.959	3.704
$\|m_4 - \phi_4\|$	1.364	1.905	12.04	1.213	1.298	1.678	0.1064
EGARCH-GED							
m_3	0.02828	-0.1098	0.4031	0.05956	0.07064	-0.3194	0.01548
m_4	4.548	4.477	5.989	4.426	4.498	6.264	3.599
ϕ_4	4.111	4.140	4.828	4.075	4.110	4.733	3.514
$\|m_4 - \phi_4\|$	0.4364	0.3370	1.161	0.3519	0.3877	1.531	0.08557
EGARCH-EGB2							
m_3	0.08199	-0.1102	0.4521	0.1143	0.1234	-0.1443	0.01751
ϕ_3	0.09026	-0.07458	0.3228	0.08975	0.07198	-0.03313	0.01423
$\|m_3 - \phi_3\|$	0.00827	0.03564	0.1293	0.0246	0.0514	0.1112	0.003278
m_4	4.371	4.328	6.243	4.351	4.417	5.176	3.597
ϕ_4	4.552	4.719	5.313	4.498	4.555	5.077	3.662
$\|m_4 - \phi_4\|$	0.1817	0.3905	0.93	0.1463	0.1376	0.09938	0.0644
GJR-GARCH-T							
m_3	0.09058	-0.1041	0.5027	0.1075	0.1256	-0.1618	0.01646
m_4	4.433	4.321	6.306	4.391	4.465	5.190	3.631
ϕ_4	5.847	6.058	18.69	5.668	5.850	6.318	3.741
$\|m_4 - \phi_4\|$	1.414	1.737	12.39	1.277	1.385	1.128	0.1100
GJR-GARCH-GED							
m_3	0.07512	-0.1046	0.4796	0.09914	0.1107	-0.1416	0.01720
m_4	4.415	4.304	6.242	4.354	4.423	5.028	3.630
ϕ_4	4.097	4.128	4.807	4.058	4.099	4.684	3.531
$\|m_4 - \phi_4\|$	0.3178	0.1756	1.436	0.2957	0.3238	0.3445	0.09935
GJR-GARCH-EGB2							
m_3	0.08328	-0.1041	0.4950	0.1025	0.1190	-0.1439	0.01665
ϕ_3	0.09646	-0.07269	0.3363	0.09069	0.07368	-0.02663	0.01399
$\|m_3 - \phi_3\|$	0.01318	0.03143	0.1587	0.01183	0.04532	0.1173	0.002659
m_4	4.450	4.320	6.290	4.387	4.466	5.111	3.630
ϕ_4	4.578	4.698	5.324	4.524	4.582	5.896	3.689
$\|m_4 - \phi_4\|$	0.1282	0.378	0.9657	0.1372	0.1158	0.7857	0.05839

Notes: in this table m_3 and m_4 represent the sample skewness and sample kurtosis of the standardized residuals in the estimated model. ϕ_3 and ϕ_4 are the skewness and kurtosis of the estimated distributions. The asymptotic standard errors of the skewness and kurtosis coefficients are $(6/T)^{0.5} = 0.045$ and $(24/T)^{0.5} = 0.089$ respectively.

In Table 4 the Goodness of Fit statistics are given. The null hypothesis is rejected for all currencies for the GARCH-N model. Beside, none of the models seem to perform well for the Italian lira. As seen earlier, this currency has the highest (absolute) skewness and highest kurtosis. Furthermore, the GARCH-t model is rejected for the Japanese yen. The GoF statistic of the models with an EGB2 distribution are considerably smaller than the models with other distributions.
Overall, it can be concluded that the GoF statistic favors (E)GARCH-EGB2 in most of the cases.

Table 4: Goodness of Fit statistics with corresponding p-value.

		DM	£	¥	FF	BF	IL	€
GARCH	N	101.0	114.3	182.5	97.61	98.38	676.3	45.43
		[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.1900]
	T	37.08	34.77	65.09	41.77	41.24	597.6	33.66
		[0.9020]	[0.7539]	[0.02489]	[0.7386]	[0.8509]	[0.0000]	[0.6706]
	GED	34.24	39.89	57.66	49.52	35.54	610.2	39.76
		[0.9811]	[0.9014]	[0.07586]	[0.6024]	[0.4777]	[0.0000]	[0.3918]
	EGB2	25.88	**25.33**	**36.41**	36.89	**27.42**	626.6	35.83
		[0.9325]	[0.9427]	[0.5429]	[0.5206]	[0.8980]	[0.0000]	[0.5238]
EGARCH	T	**22.17**	27.29	51.11	35.14	37.82	578.2	33.31
		[0.9811]	[0.9014]	[0.07586]	[0.6024]	[0.4777]	[0.0000]	[0.6860]
	GED	60.95	95.30	74.45	61.06	61.88	641.1	38.76
		[0.01049]	[0.0000]	[0.0003726]	[0.01024]	[0.008512]	[0.0000]	[0.4353]
	EGB2	25.247	30.790	37.130	**33.50**	37.08	603.7	31.00
		[0.9441]	[0.7905]	[0.5095]	[0.6777]	[0.5120]	[0.0000]	[0.7454]
GJR-GARCH	T	24.30	30.29	53.38	44.97	36.53	594.7	36.00
		[0.9587]	[0.8090]	[0.05005]	[0.2032]	[0.5375]	[0.0000]	[0.5623]
	GED	48.28	33.20	46.56	37.52	51.60	**72.08**	**30.06**
		[0.1226]	[0.6909]	[0.1607]	[0.4915]	[0.06948]	[0.0006992]	[0.8175]
	EGB2	26.53	25.54	36.48	37.56	27.67	673.4	41.25
		[0.9192]	[0.9388]	[0.5400]	[0.4895]	[0.8917]	[0.0000]	[0.2900]

The AIC criterium is given by $2k - 2\log(L)$, where k is the number of parameters and $\log(L)$ the log-likelihood. The best model according to this criterium is the model with the lowest AIC, since we prefer a high log-likelihood and a low penalty term for the number of parameters ($2k$). For the British pound, Japanese yen and Italian lira the AIC favors the same model as the GoF. For the other currencies the EGARCH-EGB2 model is preferred. This can be explained due to the fact that the EGB2 distribution is able to accommodate for higher peaks and has enough probability mass in the tails of the distribution. Besides, the AIC value of the EGARCH and GJR-GARCH models are in most cases lower than the GARCH models with the same distribution. The AIC value of the EGARCH-GED and GJR-GARCH-GED are for example considerably lower than the AIC value of the GARCH-GED. This indicates that allowing asymmetric affects, depending on the sign of the innovation, on the conditional variance improves the fit on the data.

The (E)GARCH-EGB2 model performs best, except for the Italian lira. Compared to the other models, GJR-GARCH-GED seems to perform considerably better than the other models for the Italian lira based on the GoF and AIC criteria. The main difference between the EGB2, GED and student-t for the same excess kurtosis and standard deviation is in the peak. The peak is higher and more pointed for the GED distribution. The EGB2 distribution is in turn more peaked than the student-t distribution. The peaks of the EGB2 and GED distributions become closer together as the excess kurtosis increases and thereby become much higher than the peak of the student-t. Although the Italian Lira exhibits excess kurtosis, the GED distributions still seems to capture the high peakedness better for this exchange rate than the EGB2. In the next section, the peakedness will be studied in more detail.

Table 5: Modelselection based on AIC.

		DM	£	¥	FF	BF	IL	€
	HN	6700	6423	6488	6413	6652	6529	6529
GARCH	N	6472	6130	6256	6187	6431	6177	5292
	T	6335	5994	5957	6056	6297	6028	5261
	GED	6367	6018	6000	6088	6328	6023	5261
	EGB2	6332	5987	**5942**	6054	6295	6024	5263
EGARCH	T	6326	5992	5958	6048	6290	6034	**5254**
	GED	6349	6003	5985	6072	6310	6012	5254
	EGB2	**6323**	**5985**	5944	**6045**	**6287**	6025	5256
GJR-GARCH	T	6333	5995	5959	6058	6298	6033	5262
	GED	6342	5996	5978	6065	6305	**6001**	5262
	EGB2	6331	5988	5943	6055	6296	6024	5264

9.4 Peakedness

In this section we will study to which extent the student-t, EGB2 and GED distribution are able to accommodate for high peakedness. The interpercentile-range in Table 1 already showed that the Japanese yen and Italian lira showed high peakedness. For this reason we will examine these exchange rates in more detail. Figure 4, 5 and 6 present the standardized residuals of the Japanese yen (graphs on the left) and Italian lira (graphs on the right) with the estimated densities of the GARCH-T, GARCH-GED and GARCH-EGB2. The differences between the distributions are not clearly visible for the Japanese yen, although it can be seen that the peakedness is captured well. The standardized residuals for the Italian lira was found to be the most peaked distribution. In the graphs on the right side it can be seen that none of the distributions seem flexible enough in fitting the data well, not even the EGB2 distribution.

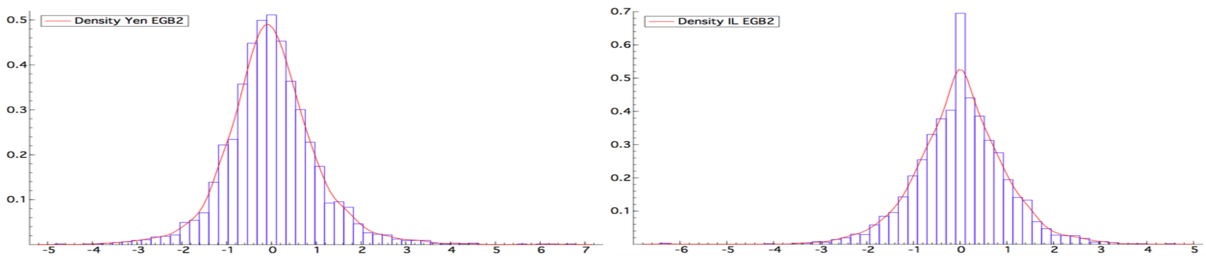

Figure 4: Standardized residuals and estimated density of the student-t distribution for the yen (left) and lira (right).

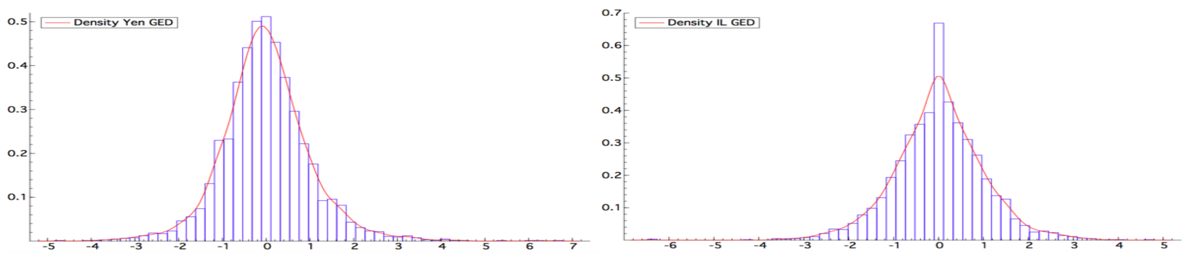

Figure 5: Standardized residuals and estimated density of the GED distribution for the yen (left) and lira (right).

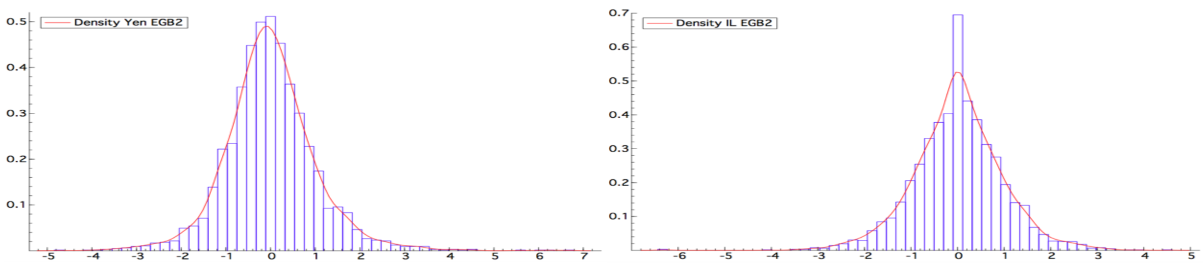

Figure 6: Standardized residuals and estimated density of the EGB2 distribution for the yen (left) and lira (right).

9.5 Forecast results

In this section the density prediction will be evaluated on the basis of the hitrates, Value at Risk backtests and logarithmic scoring rule.

The hitrates are defined as the percentage that exchange rate returns are lower than the estimated VaR. In particular, it is calculated as the number of times the loss is larger than the estimated VaR divided by the number of out-of-sample observations. Since we prefer overestimating risk over underestimating risk, we prefer too small hitrates over too large hitrates. Noteworthy is the fact that the hit rates are the smallest for the models with GED distributions. Except for the Japanese Yen and Italian Lira, all hitrates are smaller than 5%, which is an indication that the left-tail is either well-modelled or estimated too conservatively. These distributions have the highest kurtosis and the EGARCH-T model does not seem to model the left tail properly.

Table 6: 95% VaR-hitrates of the seven currencies.

		DM	£	¥	FF	BF	IL	€
GARCH	T	4.720	4.040	5.510	4.230	5.120	5.120	5.217
	GED	4.626	3.839	**5.118**	3.839	4.626	4.528	4.823
	EGB2	4.724	3.937	5.807	4.035	4.823	4.626	5.413
EGARCH	T	4.626	4.232	6.102	3.937	4.823	6.496	4.724
	GED	**4.331**	**3.051**	5.512	**3.543**	4.724	**4.331**	6.004
	EGB2	4.528	3.543	6.299	3.937	4.724	5.906	5.020
GJR-GARCH	T	4.528	4.626	5.315	4.331	4.921	5.610	5.217
	GED	4.626	3.937	**5.118**	3.740	**4.528**	4.528	**4.626**
	EGB2	4.724	4.035	5.807	4.035	4.823	4.528	5.118

Below the likelihood-ratio test statistics for the unconditional coverage (LR_{uc}), conditional coverage (LR_{cc}) and independence of violations (LR_{ind}) are given for nine models and the seven exchange rates. Surprisingly, it can be noticed that the p-values are very high overall. The null hypothesis for the conditional coverage test for the Japanese yen is rejected for the EGARCH-T, while the null hypothesis of independence is not rejected. This implies that the number of violations is not modelled correctly, while the violations are independent. The likelihood-ratio statistics of EGARCH-GED model for the British pound lead in all cases to rejections, indicating that the EGARCH-GED does not capture the left tail of the distribution well. This is in line with the GoF and AIC statistic where the EGARCH-GED also performed the worst for the British pound.

The likelihood-ratio statistics of the models with the GED distribution are lower than the corresponding model with other distributions, again indicating that the GED distribution fits the Italian lira best. The differences between the models are very small. In contrast to the statistics considered so far, the VaR does not clearly favor the EGB2 distribution over the other distributions. In the Appendix B the likelihood-ratio tests statistics for the 1%-VaR are given. The inferences drawn are similar to the inferences from the 5%-VaR.

Table 7: Backtests (UC, CC, IND) of the 5%-VaR: likelihood ratio statistics with p-values between brackets.

		DM	£	¥	FF	BF	IL	€
GARCH-T	LR_{uc}	2.01	0.496	0.322	0.496	3.25	0.496	0.104
		[0.156]	[0.481]	[0.571]	[0.481]	[0.0716]	[0.481]	[0.748]
	LR_{cc}	2.08	0.623	0.609	0.623	3.30	0.623	1.76
		[0.353]	[0.732]	[0.738]	[0.732]	[0.192]	[0.732]	[0.415]
	LR_{ind}	0.0714	0.127	0.287	0.127	0.0495	0.127	1.66
		[0.789]	[0.721]	[0.592]	[0.721]	[0.824]	[0.721]	[0.198]
GARCH-GED	LR_{uc}	2.01	0.496	0.742	0.137	2.01	0.496	0.0642
		[0.156]	[0.481]	[0.389]	[0.711]	[0.156]	[0.481]	[0.800]
	LR_{cc}	2.08	0.623	1.08	0.298	2.08	0.623	2.59
		[0.353]	[0.732]	[0.583]	[0.862]	[0.353]	[0.732]	[0.274]
	LR_{ind}	0.0714	0.127	0.337	0.161	0.0714	0.127	2.53
		[0.789]	[0.721]	[0.561]	[0.688]	[0.789]	[0.721]	[0.112]
GARCH-EGB2	LR_{uc}	2.01	2.01	2.90	0.137	3.25	0.496	0.365
		[0.156]	[0.156]	[0.0887]	[0.711]	[0.0716]	[0.481]	[0.546]
	LR_{cc}	2.08	2.08	3.41	0.298	3.30	0.623	1.66
		[0.353]	[0.353]	[0.182]	[0.862]	[0.192]	[0.732]	[0.435]
	LR_{ind}	0.0714	0.0714	0.513	0.161	0.0495	0.127	1.30
		[0.789]	[0.789]	[0.474]	[0.688]	[0.824]	[0.721]	[0.255]
EGARCH-T	LR_{uc}	0.0117	0.483	3.88	0.299	0.215	5.60	0.160
		[0.914]	[0.487]	[0.0488]	[0.585]	[0.643]	[0.0179]	[0.69]
	LR_{cc}	0.134	0.486	4.29	1.16	0.221	7.63	2.94
		[0.935]	[0.784]	[0.117]	[0.56]	[0.895]	[0.022]	[0.23]
	LR_{ind}	0.122	0.00383	0.406	0.86	0.00619	2.03	2.78
		[0.727]	[0.951]	[0.524]	[0.354]	[0.937]	[0.154]	[0.0956]
EGARCH-GED	LR_{uc}	0.712	8.35	2.05	3.10	0.0117	0.988	2.05
		[0.399]	[0.00386]	[0.152]	[0.0782]	[0.914]	[0.32]	[0.152]
	LR_{cc}	1.20	8.35	2.20	3.30	0.114	0.992	3.50
		[0.549]	[0.0154]	[0.333]	[0.192]	[0.945]	[0.609]	[0.174]
	LR_{ind}	0.486	0.00	0.145	0.202	0.102	0.00484	1.44
		[0.486]	[0.993]	[0.703]	[0.653]	[0.749]	[0.945]	[0.230]
EGARCH-EGB2	LR_{uc}	0.160	2.58	5.00	0.483	0.104	3.38	0.00129
		[0.69]	[0.108]	[0.0254]	[0.487]	[0.748]	[0.0662]	[0.971]
	LR_{cc}	0.397	2.84	5.58	1.24	0.125	6.97	2.07
		[0.82]	[0.242]	[0.0614]	[0.539]	[0.94]	[0.0307]	[0.356]
	LR_{ind}	0.237	0.260	0.583	0.755	0.0212	3.59	2.07
		[0.626]	[0.61]	[0.445]	[0.385]	[0.884]	[0.058]	[0.151]
GJR-GARCH-T	LR_{uc}	0.0117	0.0642	1.04	0.160	0.00129	1.34	0.104
		[0.914]	[0.8]	[0.307]	[0.69]	[0.971]	[0.246]	[0.748]
	LR_{cc}	0.134	0.130	1.72	1.13	0.148	2.12	1.76
		[0.935]	[0.937]	[0.424]	[0.568]	[0.929]	[0.346]	[0.415]
	LR_{ind}	0.122	0.0655	0.673	0.972	0.147	0.777	1.66
		[0.727]	[0.798]	[0.412]	[0.324]	[0.701]	[0.378]	[0.198]
GJR-GARCH-GED	LR_{uc}	0.299	2.58	0.0322	3.68	0.483	0.483	0.299
		[0.585]	[0.108]	[0.858]	[0.0551]	[0.487]	[0.487]	[0.585]
	LR_{cc}	0.608	2.84	1.51	3.83	0.486	0.486	3.34
		[0.738]	[0.242]	[0.469]	[0.147]	[0.784]	[0.784]	[0.188]
	LR_{ind}	0.310	0.260	1.48	0.150	0.00383	0.00383	3.04
		[0.578]	[0.61]	[0.223]	[0.698]	[0.951]	[0.951]	[0.0811]
GJR-GARCH-EGB2	LR_{uc}	0.00129	2.58	2.90	0.712	0.00129	0.160	0.0322
		[0.971]	[0.108]	[0.0886]	[0.399]	[0.971]	[0.69]	[0.858]
	LR_{cc}	0.0802	2.84	4.17	1.37	0.148	0.196	1.89
		[0.961]	[0.242]	[0.125]	[0.505]	[0.929]	[0.906]	[0.389]
	LR_{ind}	0.0789	0.260	1.27	0.656	0.147	0.0367	1.85
		[0.779]	[0.61]	[0.261]	[0.418]	[0.701]	[0.848]	[0.173]

Logarithmic Scoring Rule

In this section the forecasted densities based on the in-sample estimated parameters will be pairwisely compared. Based on the analysis until now, the GJR-GARCH performed the poorest compared to the GARCH and EGARCH models. For that reason we will narrow our set of considered models to the GARCH and EGARCH models with student-t, GED and EGB2 distribution.

In Appendix D the Diebold-Mariano test statistics are given for the pairwise comparison of the six models for all currencies. The model in the columns is defined as model 1 in equation (33) and the model in the rows is defined as model 2. If the Diebold-Mariano test statistic is larger than 1.96 or smaller than -1.96, the null hypothesis is rejected and there is a significant difference between the forecast accuracy of model 1 and model 2. Based on these tables a ranking is constructed in Table 8 and 9.

Table 8: Ranking of the models based on the logarithmic scoring rule.

Ranking	DM	£	¥	FF
1.	EGARCH-EGB2	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
2.	EGARCH-T	EGARCH-EGB2	EGARCH-EGB2	EGARCH-GED
3.	EGARCH-GED	GARCH-GED	GARCH-T	EGARCH-T

Table 9: Ranking of the models based on the logarithmic scoring rule.

Ranking	BF	IL	€
1.	EGARCH-EGB2	GARCH-GED	EGARCH-GED
2.	EGARCH-T	EGARCH-GED	GARCH-EGB2
3.	EGARCH-GED	EGARCH-EGB2	GARCH-GED

The ranking points out that the EGARCH-EGB2 distribution performs best or second best for all models, except for the Italian lira. For the Italian lira, the GARCH-GED and EGARCH-GED models generate the best forecasted densities similar to the results found earlier. For the Deutsche mark, Japanese yen, French franc and Belgian franc the best model corresponds with the best model based on AIC and for the Japanese yen and French franc also with the GoF-statistic.

The EGARCH models are predominantly present in the ranking for all currencies indicating that the inclusion of a leverage term significantly improves the predicted densities. Furthermore, the EGB2 and GED distribution are the most common error distributions in the ranking, once again showing the importance of including a distribution that is capable of accommodating peakedness.

9.6 Bayesian Analysis

In this section the Bayesian analysis will be discussed. Only the EGARCH-EGB2 model is estimated using the Bayesian method, since this model seems to perform better than the other models. To evaluate the difference in performance between the estimation methods, we will compare the density prediction obtained by the Maximum Likelihood estimation to the density prediction obtained by the Bayesian estimation. The dataset is therefore again split into an in-sample part, consisting of the first 2000 observations (from 1 January 1985 till 2 December 1992) and an out-of-sample part, consisting of the last 1015 observations (from 3 December 1992 till 21 November 1996). The in-sample data are used to obtain the posterior distributions for the parameters and the out-of-sample data are used to obtain the predicted density. In Table 13 the posterior means and posterior standard deviations are given. The estimates of the Maximum Likelihood estimation are given in Appendix A. All the parameters are approximately the same as the parameters obtained from the Maximum Likelihood estimation.

Table 10: Posterior means and posterior standard deviations between parentheses of the Bayesian estimation for the EGARCH-EGB2 model

	DM	£	¥	FF	BF	IL	€
ω	-0.17992	-0.11764	-0.15298	-0.16226	-0.14089	-0.18300	-0.075573
	(0.033790)	(0.021420)	(0.030442)	(0.025834)	(0.023152)	(0.029409)	(0.013174)
α	0.17947	0.13549	0.16770	0.17033	0.15379	0.19039	0.086035
	(0.032680)	(0.023059)	(0.030152)	(0.025504)	(0.023914)	(0.028264)	(0.014457)
β	0.93056	0.98033	0.96211	0.95767	0.96516	0.94432	0.99211
	(0.022518)	(0.0076807)	(0.013898)	(0.012367)	(0.010954)	(0.015677)	(0.0032871)
γ	-0.0024906	0.0010498	0.010750	0.016896	0.019535	-0.0012279	0.0070161
	(0.017401)	(0.017253)	(0.012037)	(0.012267)	(0.011532)	(0.014987)	(0.0089631)
p	0.71443	0.59997	0.44350	0.81464	0.78290	0.50953	2.4444
	(0.14948)	(0.10012)	(0.079045)	(0.13851)	(0.13533)	(0.11002)	(1.1489)
q	0.69366	0.64363	0.36494	0.76125	0.74182	0.52154	2.3845
	(0.14700)	(0.10968)	(0.062541)	(0.12670)	(0.12448)	(0.11392)	(1.0950)

Below the posterior densities for the six parameters of the EGARCH-EGB2 model (ω, α, β, γ and shape parameters p and q) are shown for the Deutsche mark. The posterior densities for the other exchange rates are similar and can be found in Appendix C. Noteworthy is the fact that the domain of the posterior distribution for γ contains both negative and positive values. Recall that γ allows for the asymmetric effect of shocks. This observation indicates that, depending on the draw from the posterior density, a negative shock could either have a larger or smaller effect than a positive shock. Furthermore, the posterior densities of p and q are very similar. The skewness of the EGB2 distribution is positive (negative) if $p > q (p < q)$. Since there is no significant skewness found in the returns for the Deutsche mark, we already expected that the distributions would be approximately the same for p and q.

One major difference between the Deutsche mark and the euro are the distributions for p and q. Although, the posterior distribution for p is similar for q as well, the peakedness is much bigger than for the Deutsche mark. Besides, the posterior mean for p and q are considerably higher for the euro than for the Deutsche mark. The kurtosis of the distribution decreases as p and q increases. This is in line with results earlier found, namely that the euro has a lower kurtosis than the Deutsche mark.

Another difference is that the posterior distribution of γ of the euro is more centered to the right. Besides, the domain of the posterior distribution of γ is narrower for the euro. This indicates that a positive shock tends to have a higher effect on the conditional volatility than a negative shock.

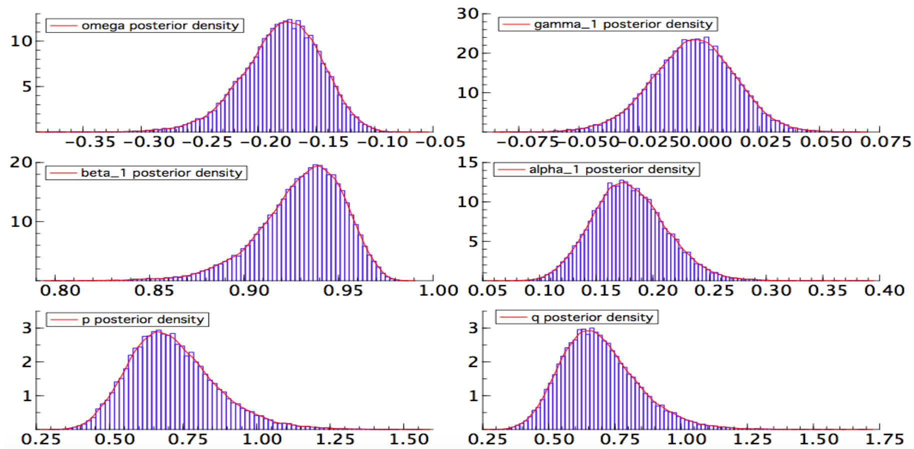

Figure 7: Posterior densities for the Deutsche mark estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

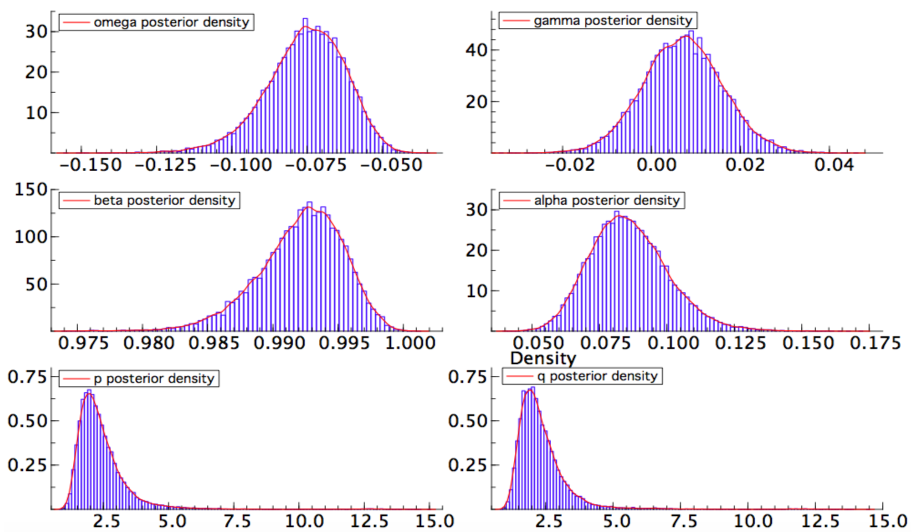

Figure 8: Posterior densities for the euro estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

Acceptance percentage, trace plot and serial correlation

As described in section 8, the validity of the estimation will be checked by looking at the acceptance percentage, the first order serial correlation between the steps in estimation for each parameter and the trace plot of each parameter.

The acceptance percentages the exchange rate series are shown in Table 11 below. All acceptance rates are above 15% which is considered reasonable. The acceptance rates might be improved by using another candidate density, for example the student-t distribution or a mixture of student-t distribution as described by Hoogerheide, Opschoor and Van Dijk (2012). These can provide a more accurate approximation to a wide variety of target densities, with substantial skewness and high kurtosis which suits the characteristics of exchange rates.

Table 11: Acceptance percentage of the candidate draws during Bayesian estimation with a simulation of 100000 draws of all exchange rate series.

	DM	£	¥	FF	BF	IL	€
Acceptance	28.79%	26.85%	28.53%	22.04%	18.64%	22.96%	31.10 %

The second way to access the reliability and quality of the estimation is to look at the trace plots. In Figure 12 and 13 the traceplot for the Deutsche mark and euro are displayed. The trace plots of the estimators indicate that many different steps were taken during the simulation, moving back and forth over space. This makes the estimation seem reliable. Furthermore, there are no horizontal lines in the figure, indicating convergence of the Markov Chain. Horizontal lines would suggest that there were few steps taken in the simulation. The trace plots of the other exchange rates can be found in Appendix C.

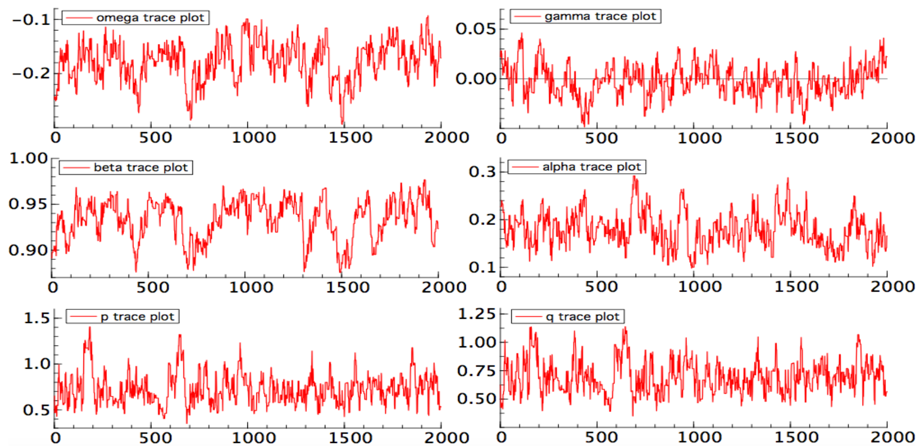

Figure 9: Trace plots for Deutsche mark estimated using Bayesian estimation with a simulation of 100000 draws.

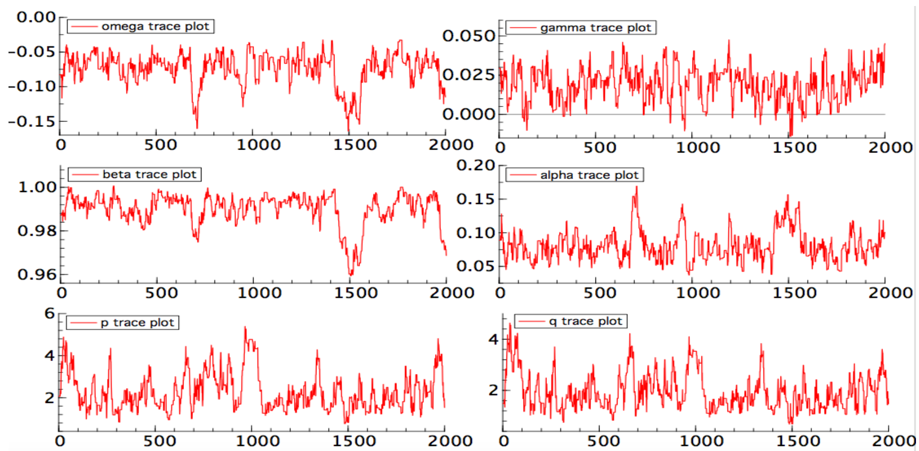

Figure 10: Trace plots for euro estimated using Bayesian estimation with a simulation of 100000 draws.

The acceptance percentages and the trace plots of the estimation indicated a valid estimation. However, the outcome of the computation of the first order serial correlation of the draws suggests the opposite. The first order serial correlations are very high in all cases indicating that the candidate steps are (possibly too) small. It indicates that the inferenceres made based on the Bayesian estimation have to be done carefully. More draws or a different prior distribution may be required to obtain accurate estimates of the posterior mean and posterior standard deviation.

Table 12: First order serial correlation if the Bayesian estimation of the EGARCH-EGB2 model for all seven exchange rate series.

	DM	£	¥	FF	BF	IL	€
ω	0.9331	0.9190	0.9213	0.9200	0.9226	0.9150	0.9107
α	0.9173	0.9140	0.9133	0.9187	0.9240	0.9126	0.9058
β	0.9706	0.9337	0.9377	0.902	0.8999	0.9113	0.9186
γ	0.9321	0.9048	0.9094	0.8983	0.9083	0.9132	0.8942
p	0.8986	0.9039	0.9004	0.9436	0.9497	0.9459	0.9655
q	0.9081	0.9048	0.9000	0.9399	0.9429	0.9450	0.9650

Comparison between frequentist and Bayesian forecasted densities

To compare the frequentist forecasts with the Bayesian forecast, the logarithmic scorig rule will be used again. In this case the loss differential is defined as:

$d_t =$ logarithmic scoring rule of frequentist method at t - logarithmic scoring rule of Bayesian method at t

The null hypothesis H_0: $\mathbb{E}(d_t) = 0$ of no difference between the forecast accuracy of frequentist and Bayesian estimation will be tested against the alternative hypothesis of H_1: $\mathbb{E}(d_t) \neq 0$ of significant difference in the estimation methods. The results are presented in Table 9.6 below.

Table 13: Diebold-Mariano test statistic of the logarithmic scoring rule of the frequentist versus Bayesian estimation.

	DM	£	¥	FF	BF	IL	€
DMT	2.141	5.121	-3.209	-5.473	-4.953	-2.993	-2.798

First of all it can be noted that the null hypothesis of equal forecasting accuracy is rejected for all exchange rate series because $|DMT| > 1.96$. It turns out that the forecasts obtained by Bayesian estimation are in most cases significantly better. However, the DMT statistic is significantly positive for the Deutsche mark and British pound which means that the forecasts obtained by frequentist estimation are more accurate. This may seem odd, but if we take a closer look at the differences between the sample variances in the in-sample and out-of-sample periods, it turns out to be explainable.

It is highly likely that there is a structural break between the in-sample and out-of-sample period for the Deutsche mark and British pound. The sample variance of the out-of-sample period is (considerably) lower than the in-sample period: $\mathbb{V}ar(r_{DM,in-sample}) = 0.55341$ versus $\mathbb{V}ar(r_{DM,out-of-sample}) = 0.50530$ and $\mathbb{V}ar(r_{pound,in-sample}) = 0.54585$ versus $\mathbb{V}ar(r_{pound,out-of-sample}) = 0.37743$. The frequentist method ignores the parameter uncertainty meaning that the estimated distribution underestimates the involved risk. If the out-of-sample period by coincidence turns out to have a lower variance than the in-sample period, the predictive distribution with the thinner tails from the frequentist method can accidentally fit the out-of-sample data better than the distribution with thicker tails from the Bayesian method.

In Figure 14 a graphical exposition of the one-step-ahead density forecasts of the frequentist and Bayesian estimation are displayed for the Deutsche mark and euro. Since the difference between the two forecasts is not clearly visible, the loss differentials d_t are plotted as well.

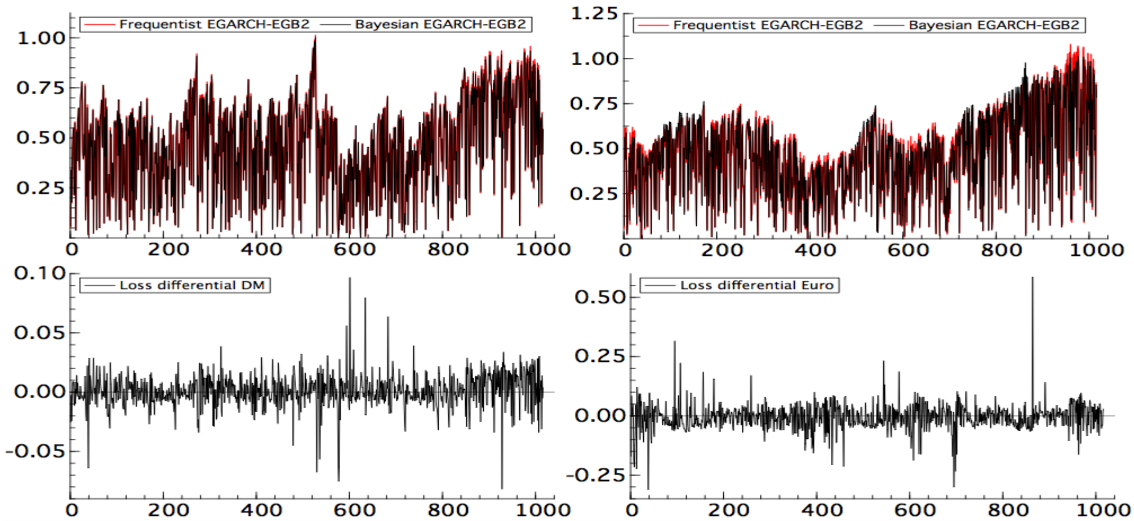

Figure 11: The predicted densities of the Deutsche mark (left) and euro (right) are shown above. The loss differential score for the EGARCH-EGB2 model using frequentist and Bayesian approach for the Deutsche mark and euro are shown below. Positive (negative) values indicate a better forecast performance of the frequentist (Bayesian) framework.

10 Conclusion

Modelling and forecasting the volatility of exchange rate returns has become an important field of empirical research in finance. This is because volatility is considered to be an important concept in many economic and financial applications like risk management. Most common characteristics found in exchange rate returns are leptokurtosis, high peakedness, asymmetry and volatility clustering. Although the widely used GARCH model with student-t distribution seems to capture the volatility clustering and leptokurtosis, it cannot accommodate high peakedness and skewness. The proposed flexible EGB2 distribution emphasizes the importance of specifying an appropriate distribution which is able to capture the underlying statistical properties. Based on both the fit on the data as on the predictive performance, the EGARCH-EGB2 outperformed the competing models. Since the models with EGB2 distribution seem to perform better than the GED distribution, it is furthermore shown that not only accommodating high peakedness in the error distribution is of importance, but also accommodating skewness. However, the EGB2 distribution also showed its limitations for exchange rate returns with extreme leptokurtosis like the Italian lira. The performance of the EGB2 distribution detoriated as the kurtosis reached the boundary.

The EGARCH and GJR-GARCH improved the models fit even more compared to the GARCH models. The GJR-GARCH models performed in most cases worse compared to the EGARCH models. Therefore, the EGARCH specification of conditional volatility seemed to incorporate the asymmetry in shocks better than the GJR-GARCH specification. Comparisons based on the models' Value at Risk forecasting accuracy did not clearly favor one particular model over the other models. However, the EGARCH-EGB2 performed best or second best for all currencies, except for the Italian lira, according to the logarithmic scoring rule. Besides, the EGARCH models were predominantly present in the ranking of the best predictive models based on the logarithmic scoring rule. Inclusion of a leverage term therefore significantly improved the predicted densities.

In order to check whether the estimation method could be improved, the EGARCH-EGB2 model is also estimated using Bayesian estimation. While the acceptance percentage and the trace plots showed no sign of invalidity, the high first order serial correlations imply that a large number of draws is required in order to obtain precise estimates. Small differences between the Bayesian and Maximum Likelihood estimates were found. However, the Bayesian approach still appeared to produce significantly better forecasted densities compared to the frequentist approach for five of the seven exchange rates. The fact that the frequentist method performed better in two cases can probably be attributed to a structural break in the data, where the sample variance was much smaller in the out-of-sample period.

11 Suggestion for further research

The models used in this thesis were not able to capture the extreme peak of the Italian lira, not even the EGB2 distribution. The EGB2 can accommodate coefficient of kurtosis values up to 9 (McDonald 1991). To capture high kurtosis better, other distributions could be considered in further research. For example the skewed-t distribution or the Feasible EGB2 (FEGB2) introduced by Fischer (2000). The latter is an extension of the EGB2, where the shortcoming of modelling kurtosis and skewness only in a restricted region is removed by introducing an additional parameter. Unfortunately, the FEGB2 has no closed form for the pdf and needs to be numerically approximated.

Besides, a rolling window estimation could be used instead of a fixed estimation. This means that the parameter estimates are updated regularly (daily, weekly or even quarterly). However, Hoogerheide and Ardia et al. found that it seems much worse to choose a simpler model (with daily updated parameter estimates) than to infrequently update the parameter estimates of a more advanced model. The latter substantially reduces the computational efforts without seriously harming the models' performance.

In this thesis Value at Risk is used as risk measure. However, it does not fully describe the tail behavior of the loss function; it is the quantile of the loss function and hence not informative about the magnitude of the losses. Consequently, VaR may underestimate the potential loss if the actual loss exceeds the VaR. The Expected Shortfall (ES) is closely related to the VaR and measures the expected loss given that the loss is bigger than the corresponding VaR. Therefore, the ES might give a better assessment of the potential loss.

Furthermore, the Bayesian analysis could be extended and improved. As already mentioned, other candidate densities could be used like mixtures of the student-t distribution, to lower the first order serial correlation and improve the quality of the estimation. Since the predicted densities turned out to be significantly better in five out of the seven cases in case of the Bayesian estimation, the posterior densities could be used to obtain the Value at Risk forecasts and even Expected Shortfall forecasts.

References

Abdalla, S. (2012). Modelling Exchange Rate Volatility using GARCH Models: Empirical Evidence from Arab Countries. *International Journal of Economics and Finance, 4*(3), 216-228.

Ali, G. (2013). EGARCH, GJR-GARCH, TGARCH, AVGARCH, NGARCH, IGARCH and APARCH Models for Pathogens at Marine Recreational Sites. *Journal of Statistical and Econometric Methods 2*(3), 57-73.

Ardia, D. and L. Hoogerheide (2009). Bayesian estimation of the GARCH(1,1) model with Student-t innovations. *The R Journal 2*(2), 41-47.

Bollerslev T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.

Caivano, M. and A. Harvey (2014). Time series models with an EGB2 conditional distribution. Bank of Italy and Faculty of Economics, Cambridge University.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review, 39*(4), 841–862.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253-265.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 50*(4), 987–1007.

Fischer, M. (2000). The Folded EGB2 distribution and its application to financial return data. Discussion Papers from Friedrich-Alexander-University Erlangen-Nuremberg, Chair of Statistics and Econometrics.

Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Appli- cations. *Biometrika, 57*(1), 97–109.

Hoogerheide, L., D. Ardia, and N. Corré (2011). Stock Index Returns' Density Prediction using GARCH Models: Frequentist or Bayesian Estimation? *Tinbergen Institute Discussion Paper*, (2011-020/4).

Jarque, C. and A. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residu- als. *Economics Letters 6*(3), 255–259.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives, 3*(2), 73–84.

McDonald and Xu (1995). A generalization of the beta distribution with applications. *Journal of econometrics 66*, 133–152.

McNeil, A.J., Frey, R. (1987). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7*, 271-300.

Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica, 59*(2), 347–370.

Omari, C., P. Nyamuhanga Mwita, A. Gichuhi Waititu. Modeling USD/KES Exchange Rate Volatility using GARCH Models. *IOSR Journal of Economics and Finance (IOSR-JEF), 8*(1), 5-26.

Wang, K., Fawson, C., Barrett, C.B. and McDonald, J.B. (2001). A Flexible Parametric GARCH Model with an Application to Exchange Rates. *Journal of Applied Econometrics, 16*, 521-536.

12 On the original article

Based on the results found in this paper, some notes can be made on the differences between our findings and those from the original paper of Wang et al. (2001).

First of all can be noted that the data the authors put on their website is not the same as the data they used in their paper. For the Deutsche mark, British pound and French franc the data on the website of the authors is the inverse of the exchange rate. They have given the local currency/$ instead of $/local currency. Furthermore, although in the original article is stated that the US dollar was the base currency, the Canadian dollar has been used as base currency.

The GARCH-EGB2 log-likelihood function in the article misses the parentheses in exponent of the logarithm. It should be $\log\left[1+\exp(\frac{\sqrt{\Omega}\epsilon_t}{\sqrt{\sigma_t}}+\Delta)\right]$ instead of $\log\left[1+\exp\frac{\sqrt{\Omega}\epsilon_t}{\sqrt{\sigma_t}}+\Delta\right]$. In the latter case, one would interpret Δ outside the exponent operator, whereas it should be within the operator.

Also, in the article of Wang et al. (2001), the kurtosis for the French franc is incorrectly reported as 6 whereas it should be 5.02. Furthermore, the inter-percentile ranges for the Italian lira are different than the values we computed. Besides, the formula for the predicted kurtosis is incorrect. The right formula is given by equation (20).

The calculated χ^2 GoF statistics in Table 4 are different from the statistics in the article. It is likely that the authors tried to fit a non-standardized distribution on the standardized residuals. If one wants to test whether the standardized residuals have a student-t distribution with ν degrees of freedom for example, the standardized student-t distribution should be used and not the student-t distribution with ν degrees of freedom. In the latter case the assumed distribution has a variance $\nu/(\nu-2)$ whereas the data have variance one. The standardized residuals should either be multiplied by $\sqrt{\nu/(\nu-2)}$ or the standardized student-t distribution should be used. However, the inferences drawn are equivalent to the inferences we drew.

A Appendix - Parameter Estimates

Table 14: Parameter estimates for the GARCH(1,1) model.

	DM	£	¥	FF	BF	IL	€
				Student-t			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	0.03472	0.02709	0.02678	0.03246	0.03341	0.03626	0.002393
	(0.008992)	(0.007745)	(0.009747)	(0.008041)	(0.008480)	(0.008762)	(0.001044)
$\hat{\alpha}$	0.07685	0.07339	0.09627	0.08727	0.08304	0.08795	0.0349
	(0.01354)	(0.01372)	(0.02211)	(0.01455)	(0.01396)	(0.01601)	(0.006112)
$\hat{\beta}$	0.8634	0.8797	0.8558	0.8534	0.8584	0.8443	0.9587
	(0.02416)	(0.02312)	(0.03610)	(0.02422)	(0.02406)	(0.02712)	(0.007837)
$\hat{\nu}$	5.835	5.994	4.194	5.839	6.116	5.155	12.09
	(0.62787)	(0.6554)	(0.3402)	(0.6278)	(0.6944)	(0.4919)	(1.960)
LL	-2137	-2126	-1877	-2054	-2113	-2021	-2627
				GED			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	0.03309	0.02465	0.02578	0.03166	0.03254	0.03923	0.001921
	(0.008326)	(0.007308)	(0.009347)	(0.007474)	(0.007904)	(0.008817)	(0.0009687)
$\hat{\alpha}$	0.07734	0.06594	0.08237	0.0888	0.08344	0.1065	0.0278
	(0.01335)	(0.01247)	(0.01858)	(0.01444)	(0.01369)	(0.01777)	(0.005911)
$\hat{\beta}$	0.8630	0.8887	0.8614	0.8502	0.8568	0.8180	0.9667
	(0.02344)	(0.02238)	(0.03512)	(0.02356)	(0.02324)	(0.02859)	(0.007567)
$\hat{\nu}$	1.348	1.371	1.16	1.36	1.373	1.269	1.574
	(0.04686)	(0.04684)	(0.03767)	(0.04682)	(0.04794)	(0.04327)	(0.05862)
LL	-2138	-2129	-1882	-2057	-2114	-2025	-1617
				EGB2			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	0.0337	0.02633	0.02505	0.03178	0.03278	0.0379	0.002047
	(0.008735)	(0.0076623)	(0.009221)	(0.007832)	(0.008301)	(0.008844)	(0.001025)
$\hat{\alpha}$	0.07667	0.07065	0.08336	0.08711	0.08289	0.09654	0.02921
	(0.01349)	(0.01330)	(0.01902)	(0.01449)	(0.01392)	(0.01681)	(0.006083)
$\hat{\beta}$	0.8632	0.882	0.862	0.8523	0.8579	0.8293	0.965
	(0.02421)	(0.02325)	(0.03547)	(0.02431)	(0.02405)	(0.02832)	(0.007734)
$\hat{p}$	0.6743	0.6647	0.3957	0.686	0.6994	0.5483	1.54
	(0.1084)	(0.1066)	(0.07434)	(0.1093)	(0.1148)	(0.08517)	(0.3374)
$\hat{q}$	0.6510	0.7135	0.3220	0.6622	0.6943	0.5778	1.374
	(0.1056)	(0.1180)	(0.05767)	(0.1051)	(0.1150)	(0.09036)	(0.2965)
LL	-2135	-2124	-1869	-2053	-2111	-2021	-1617

Table 15: Parameter estimates for the EGARCH(1,1) model.

	DM	£	¥	FF	BF	IL	€
				Student-t			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	-0.006259	-0.003829	0.005134	-0.01117	-0.008935	-0.004538	-0.002157
	(0.011074	(0.010251	(0.018332	(0.012661	(0.011838	(0.013712	(0.003669)
$\hat{\alpha}$	0.166	0.1604	0.21	0.1905	0.1811	0.1929	0.06341
	(0.025349)	(0.024501)	(0.037403)	(0.026261)	(0.026339)	(0.028493)	(0.01301)
$\hat{\gamma}$	-0.002969	-0.02269	0.01299	-0.01398	-0.006187	-0.01436	0.01854
	(0.013969)	(0.013789)	(0.015054)	(0.014754)	(0.014400)	(0.015043)	(0.008427)
$\hat{\beta}$	0.9432	0.951	0.9354	0.9376	0.9398	0.9367	0.9935
	(0.015049)	(0.013730)	(0.021832)	(0.015116)	(0.015552)	(0.016227)	(0.003038)
$\hat{\nu}$	5.837	5.892	4.198	5.878	6.077	5.097	10.99
	(0.63066)	(0.63503)	(0.34070)	(0.63699)	(0.69011)	(0.48218)	(2.0490)
LL	-2138	-2125	-1876	-2054	-2115	-2023	-1615
				GED			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	-0.14988	-0.12673	-0.19651	-0.17923	-0.15521	-0.1987	-0.058434
	(0.02505)	(0.02360)	(0.04305)	(0.02700)	(0.02624)	(0.03125)	(0.01315)
$\hat{\alpha}$	0.105	0.090834	0.16341	0.14991	0.08378	0.14946	0.057469
	(0.05275)	(0.04593)	(0.06025)	(0.05623)	(0.05585)	(0.04558)	(0.02490)
$\hat{\beta}$	0.93047	0.94651	0.9266	0.93228	0.92155	0.91654	0.99153
	(0.01894)	(0.01624)	(0.02219)	(0.01801)	(0.01877)	(0.02003)	(0.005970)
$\hat{\gamma}$	0.034166	0.030299	0.018803	0.022419	0.052776	0.034335	0.0071849
	(0.02658)	(0.02337)	(0.02510)	(0.02868)	(0.02842)	(0.01908)	(0.01676)
$\hat{\nu}$	1.3491	1.3714	1.1598	1.3631	1.3726	1.264	1.5764
	(0.04700)	(0.04686)	(0.03762)	(0.04708)	(0.04797)	(0.04297)	(0.05887)
LL	-2139	-2129.7	-1884.8	-2058.5	-2114.1	-2030.5	-1616.3
				EGB2			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	-0.16083	-0.14789	-0.20386	-0.18847	-0.17638	-0.20611	-0.056253
	(0.02497)	(0.02383)	(0.04109)	(0.02641)	(0.02614)	(0.03024)	(0.01186)
$\hat{\alpha}$	0.16624	0.15501	0.19755	0.19026	0.18155	0.20889	0.06346
	(0.02529)	(0.02411)	(0.03476)	(0.02607)	(0.02621)	(0.01291)	(0.01291)
$\hat{\beta}$	0.9446	0.9531	0.93303	0.93858	0.94087	0.9323	0.9937
	(0.01478)	(0.01349)	(0.02217)	(0.01485)	(0.01522)	(0.01679)	(0.002963)
$\hat{\gamma}$	-0.001035	-0.01877	0.006797	-0.01164	-0.004376	-0.01453	0.01877
	(0.01363)	(0.01332)	(0.01444)	(0.01439)	(0.01416)	(0.01536)	(0.008463)
$\hat{p}$	0.67337	0.65399	0.39357	0.70791	0.69101	0.54081	1.595
	(0.1088)	(0.1038)	(0.07386)	(0.1146)	(0.1137)	(0.08420)	(0.3550)
$\hat{q}$	0.6472	0.69912	0.3213	0.67847	0.68198	0.5652	1.409
	(0.1052)	(0.1141)	(0.05762)	(0.1089)	(0.1130)	(0.08848)	(0.3075)
LL	-2135.5	-2123.2	-1868.6	-2052.2	-2112.4	-2023.3	-2621.9

Table 16: Parameter estimates for the GJR-GARCH(1,1) model.

	DM	£	¥	FF	BF	IL	€
				Student-t			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.023271	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	0.01337	0.00448	0.01025	0.01447	0.01345	0.02139	0.002257
	(0.004865)	(0.002467)	(0.004303)	(0.005373)	(0.005204)	(0.006778)	(0.001063)
$\hat{\alpha}$	0.07559	0.05104	0.06468	0.07711	0.07202	0.07188	0.03832
	(0.01434)	(0.01179)	(0.01520)	(0.01558)	(0.01480)	(0.01580)	(0.008674)
$\hat{\beta}$	0.9153	0.9434	0.9241	0.9038	0.9132	0.8854	0.9594
	(0.01722)	(0.01379)	(0.01861)	(0.02058)	(0.01867)	(0.02394)	(0.007801)
$\hat{\gamma}$	-0.02691	-0.003407	-0.008711	-0.0155	-0.01588	0.003149	-0.007357
	0.01480	0.01214	0.01371	(0.01626)	(0.01530)	(0.01881)	(0.01168)
$\hat{\nu}$	6.108	5.962	4.382	6.249	6.106	5.808	12.1
	(0.6729)	(0.6612)	(0.3683)	(0.7047)	(0.6793)	(0.6429)	(2.445)
LL	-2097	-1985	-1973	-2006	-2086	-1997	-1742
				GED			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.02327	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01821)	(0.0008787)
$\hat{\omega}$	0.03371	0.02626	0.0255	0.03306	0.03334	0.0428	0.001734
	(0.008537)	(0.007569)	(0.009286)	(0.007791)	(0.008175)	(0.009564)	(0.0008787)
$\hat{\alpha}$	0.07328	0.05688	0.07961	0.07699	0.07826	0.08702	0.0359
	(0.01550)	(0.01326)	(0.01977)	(0.01623)	(0.01586)	(0.01934)	(0.00781)
$\hat{\beta}$	0.8615	0.8843	0.8619	0.8468	0.8548	0.8088	0.9687
	(0.02383)	(0.02270)	(0.03489)	(0.02414)	(0.02378)	(0.03004)	(0.01041)
$\hat{\gamma}$	0.009036	0.02083	0.006364	0.02538	0.01155	0.04325	-0.01923
	(0.01917)	(0.01721)	(0.01906)	(0.02095)	(0.01999)	(0.02636)	(0.01041)
$\hat{\nu}$	1.348	1.369	1.16	1.36	1.373	1.271	1.577
	(0.04684)	(0.04674)	(0.03767)	(0.04677)	(0.04791)	(0.04340)	(0.05878)
LL	-2138	-2128	-1882	-2056	-2114	-2024	-1616
				EGB2			
$\hat{\phi}_0$	0.02398	0.01222	0.02637	0.02058	0.02327	0.009078	-0.003213
	(0.01336)	(0.01274)	(0.01290)	(0.01274)	(0.01326)	(0.01300)	(0.01119)
$\hat{\phi}_1$	0.04929	0.07398	0.04399	0.05771	0.04003	0.04003	0.01474
	(0.01820)	(0.01817)	(0.01820)	(0.01819)	(0.01820)	(0.01820)	(0.01821)
$\hat{\omega}$	0.03453	0.02813	0.02464	0.03350	0.03374	0.04104	0.001837
	(0.008989)	(0.01551)	(0.02472)	(0.01981)	(0.1087)	(0.1058)	(0.0009274)
$\hat{\alpha}$	0.07129	0.05935	0.07919	0.07399	0.07648	0.08069	0.03781
	(0.01551)	(0.01407)	(0.02346)	(0.01857)	(0.1061)	(0.1165)	(0.008014)
$\hat{\beta}$	0.8611	0.87712	0.86269	0.84756	0.85541	0.821	0.96731
	(0.02472)	(0.01983)	(0.03508)	(0.01911)	(0.07458)	(0.05780)	(0.007034)
$\hat{\gamma}$	0.01216	0.02551	0.009685	0.02980	0.01448	0.03579	-0.02062
	(0.01981)	(0.01609)	(0.02511)	(0.02161)	(0.1096)	(0.1048)	(0.01071)
$\hat{p}$	0.6759	0.6630	0.3952	0.6884	0.7008	0.5521	1.565
	(0.1087)	(0.01600)	(0.02468)	(0.02062)	(0.1151)	(0.1151)	(0.3458)
$\hat{q}$	0.6517	0.7078	0.3213	0.6621	0.6943	0.5790	1.383
	(0.1058)	(0.01841)	(0.02984)	(0.02480)	(0.08612)	(0.09080)	(0.3000)
LL	-2134.9	-2123	-1869.1	-2051.9	-2111	-2019.8	-1615.7

B Appendix - Results Value at Risk

Table 17: Backtests (UC, CC, IND) of the 1%-VaR: likelihood ratio statistics with p-values between brackets.

		DM	£	¥	FF	BF	IL	€
GARCH-T	LR_{uc}	2.01	0.496	0.322	0.496	3.25	0.496	0.137
		[0.156]	[0.481]	[0.571]	[0.481]	[0.0716]	[0.481]	[0.711]
	LR_{cc}	2.08	0.623	0.609	0.623	3.3	0.623	3.55
		[0.353]	[0.732]	[0.738]	[0.732]	[0.192]	[0.732]	[0.169]
	LR_{ind}	0.0714	0.127	0.287	0.127	0.0495	0.127	3.41
		[0.789]	[0.721]	[0.592]	[0.721]	[0.824]	[0.721]	[0.0647]
GARCH-GED	LR_{uc}	2.01	0.496	0.742	0.137	2.01	0.496	0.137
		[0.156]	[0.481]	[0.389]	[0.711]	[0.156]	[0.481]	[0.711]
	LR_{cc}	2.08	0.623	1.08	0.298	2.08	0.623	3.55
		[0.353]	[0.732]	[0.583]	[0.862]	[0.353]	[0.732]	[0.169]
	LR_{ind}	0.0714	0.127	0.337	0.161	0.0714	0.127	3.41
		[0.789]	[0.721]	[0.561]	[0.688]	[0.789]	[0.721]	[0.0647]
GARCH-EGB2	LR_{uc}	2.01	2.01	2.90	0.137	3.25	0.496	0.07
		[0.156]	[0.156]	[0.0887]	[0.711]	[0.0716]	[0.481]	[0.791]
	LR_{cc}	2.08	2.08	3.41	0.298	3.3	0.623	2.71
		[0.353]	[0.353]	[0.182]	[0.862]	[0.192]	[0.732]	[0.258]
	LR_{ind}	0.0714	0.0714	0.513	0.161	0.0495	0.127	2.64
		[0.789]	[0.789]	[0.474]	[0.688]	[0.824]	[0.721]	[0.104]
EGARCH-T	LR_{uc}	2.01	0.496	0.742	0.496	3.25	2.01	0.137
		[0.156]	[0.481]	[0.389]	[0.481]	[0.0716]	[0.156]	[0.711]
	LR_{cc}	2.08	0.623	1.08	0.623	3.3	2.08	3.55
		[0.353]	[0.732]	[0.583]	[0.732]	[0.192]	[0.353]	[0.169]
	LR_{ind}	0.0714	0.127	0.337	0.127	0.0495	0.0714	3.41
		[0.789]	[0.721]	[0.561]	[0.721]	[0.824]	[0.789]	[0.0647]
EGARCH-GED	LR_{uc}	0.322	1.32	2.04	0.00225	2.04	0.07	2.04
		[0.571]	[0.251]	[0.153]	[0.962]	[0.153]	[0.791]	[0.153]
	LR_{cc}	2.64	3.09	2.49	0.201	3.58	0.311	3.58
		[0.267]	[0.213]	[0.288]	[0.904]	[0.167]	[0.856]	[0.167]
	LR_{ind}	2.32	1.77	0.45	0.199	1.54	0.241	1.54
		[0.128]	[0.183]	[0.502]	[0.656]	[0.215]	[0.623]	[0.215]
EGARCH-EGB2	LR_{uc}	0.16	2.58	5.00	0.483	0.104	3.38	0.00129
		[0.69]	[0.108]	[0.0254]	[0.487]	[0.748]	[0.0662]	[0.971]
	LR_{cc}	0.397	2.84	5.58	1.24	0.125	6.97	2.07
		[0.82]	[0.242]	[0.0614]	[0.539]	[0.94]	[0.0307]	[0.356]
	LR_{ind}	0.237	0.26	0.583	0.755	0.0212	3.59	2.07
		[0.626]	[0.61]	[0.445]	[0.385]	[0.884]	[0.058]	[0.151]
GJR-GARCH-T	LR_{uc}	0.0117	0.0642	1.04	0.160	0.00129	1.34	0.104
		[0.914]	[0.800]	[0.307]	[0.69]	[0.971]	[0.246]	[0.748]
	LR_{cc}	0.134	0.13	1.72	1.13	0.148	2.12	1.76
		[0.935]	[0.937]	[0.424]	[0.568]	[0.929]	[0.346]	[0.415]
	LR_{ind}	0.122	0.0655	0.673	0.972	0.147	0.777	1.66
		[0.727]	[0.798]	[0.412]	[0.324]	[0.701]	[0.378]	[0.198]
GJR-GARCH-GED	LR_{uc}	0.299	2.58	0.0322	3.68	0.483	0.483	0.299
		[0.585]	[0.108]	[0.858]	[0.0551]	[0.487]	[0.487]	[0.585]
	LR_{cc}	0.608	2.84	1.51	3.83	0.486	0.486	3.34
		[0.738]	[0.242]	[0.469]	[0.147]	[0.784]	[0.784]	[0.188]
	LR_{ind}	0.31	0.26	1.48	0.15	0.00383	0.00383	3.04
		[0.578]	[0.61]	[0.223]	[0.698]	[0.951]	[0.951]	[0.0811]
GJR-GARCH-EGB2	LR_{uc}	0.00129	2.58	2.90	0.712	0.00129	0.16	0.0322
		[0.971]	[0.108]	[0.0886]	[0.399]	[0.971]	[0.69]	[0.858]
	LR_{cc}	0.0802	2.84	4.17	1.37	0.148	0.196	1.89
		[0.961]	[0.242]	[0.125]	[0.505]	[0.929]	[0.906]	[0.389]
	LR_{ind}	0.0789	0.26	1.27	0.656	0.147	0.0367	1.85
		[0.779]	[0.61]	[0.261]	[0.418]	[0.701]	[0.848]	[0.173]

C Appendix - Posterior Distributions and Trace Plots

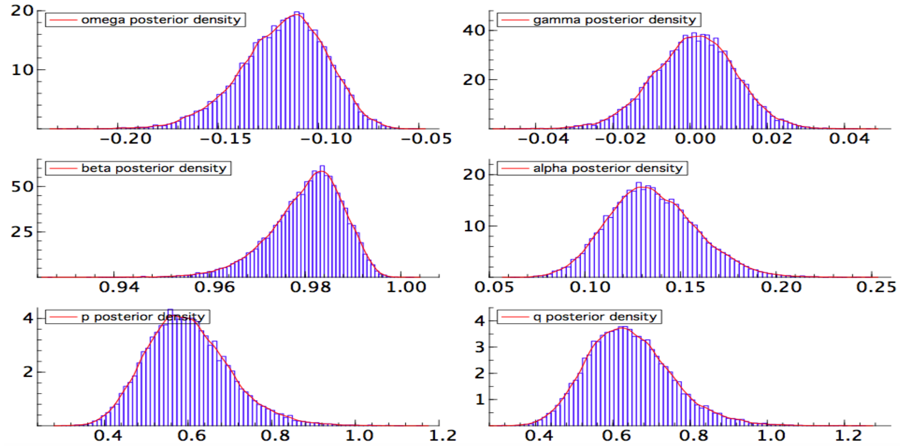

Figure 12: Posterior densities for the British pound estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

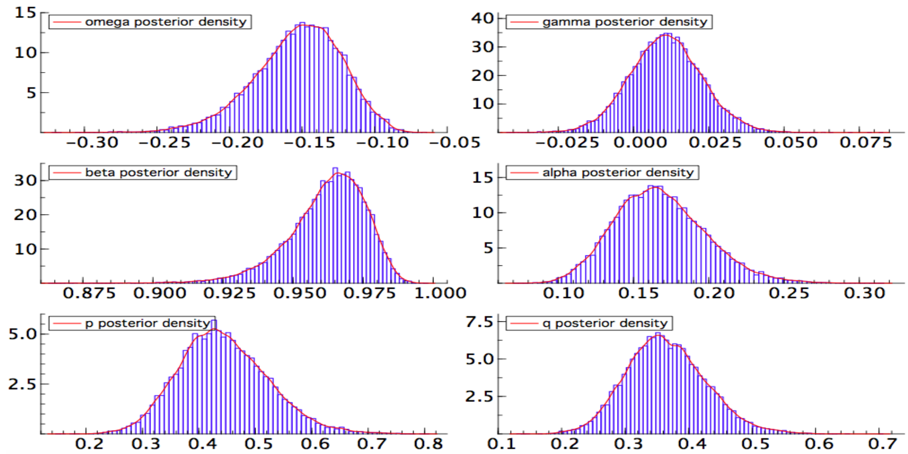

Figure 13: Posterior densities for the Japanese yen estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

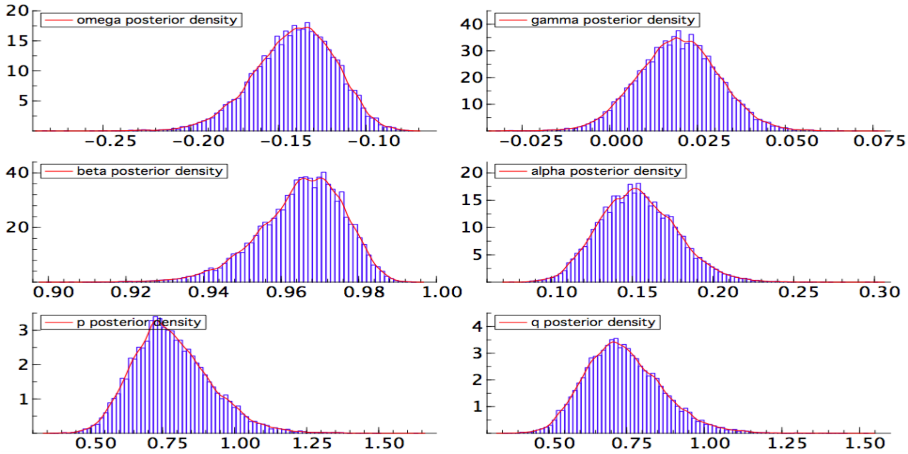

Figure 14: Posterior densities for the Belgian franc estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

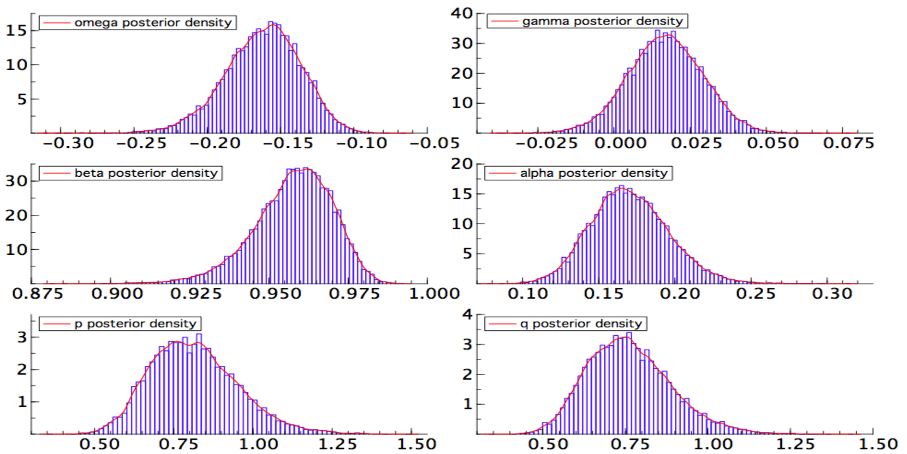

Figure 15: Posterior densities for the French franc estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

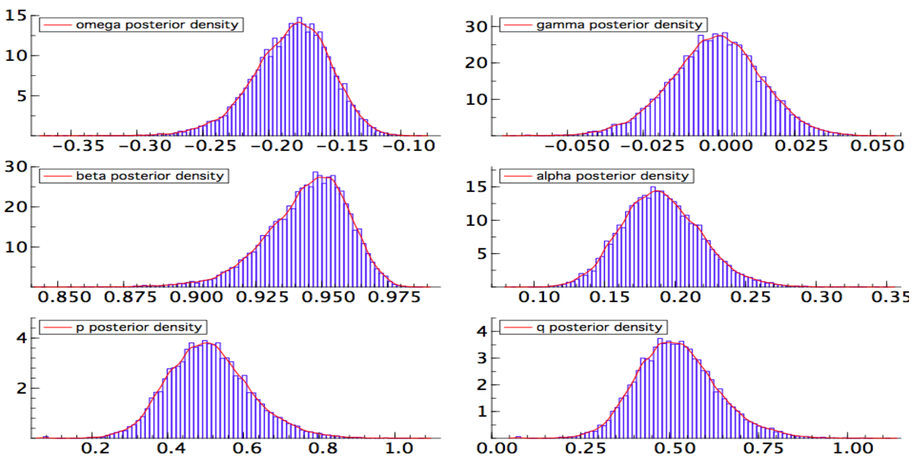

Figure 16: Posterior densities for the Italian lira estimated using Bayesian estimation with a simulation of 100000 draws and 10000 burn-in draws.

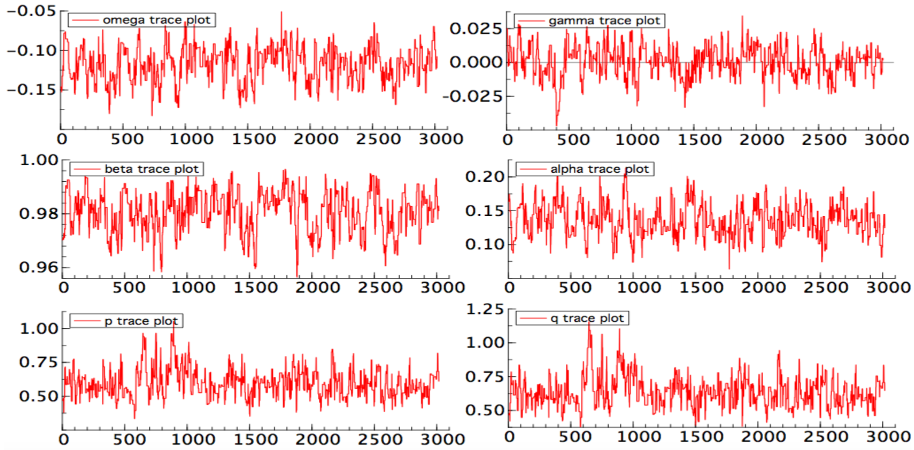

Figure 17: Trace plots for British pound estimated using Bayesian estimation with a simulation of 100000 draws.

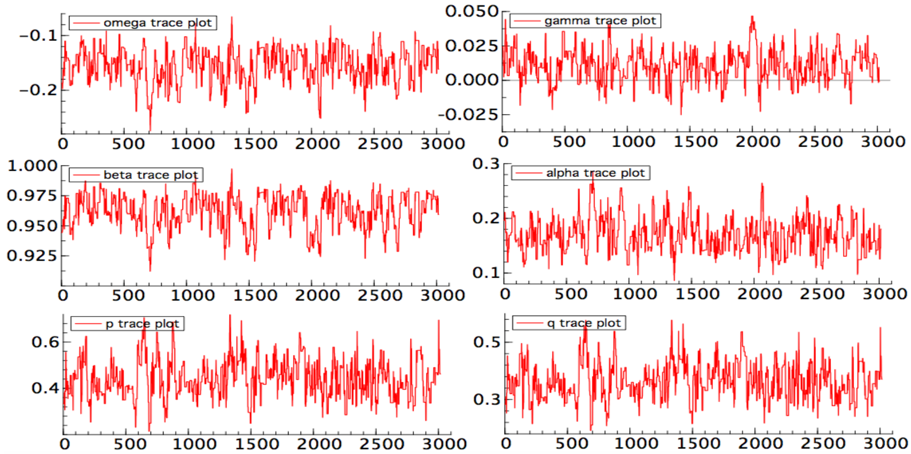

Figure 18: Trace plots for Japanese yen estimated using Bayesian estimation with a simulation of 100000 draws.

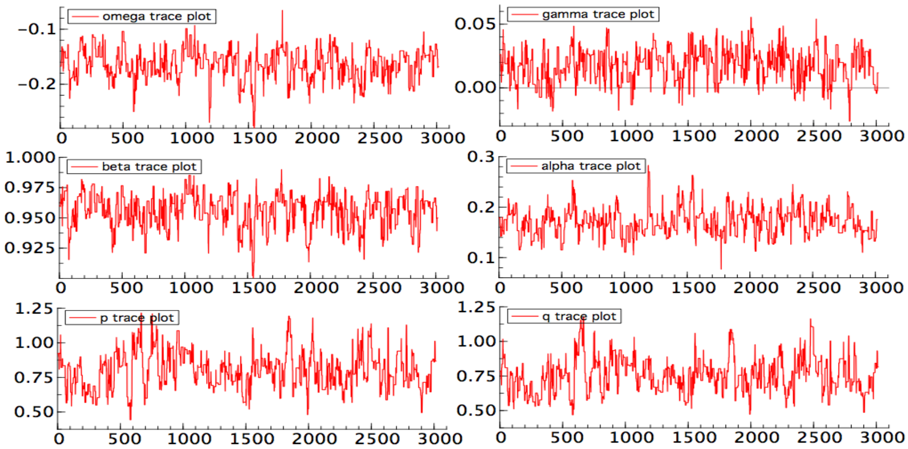

Figure 19: Trace plots for French franc estimated using Bayesian estimation with a simulation of 100000 draws.

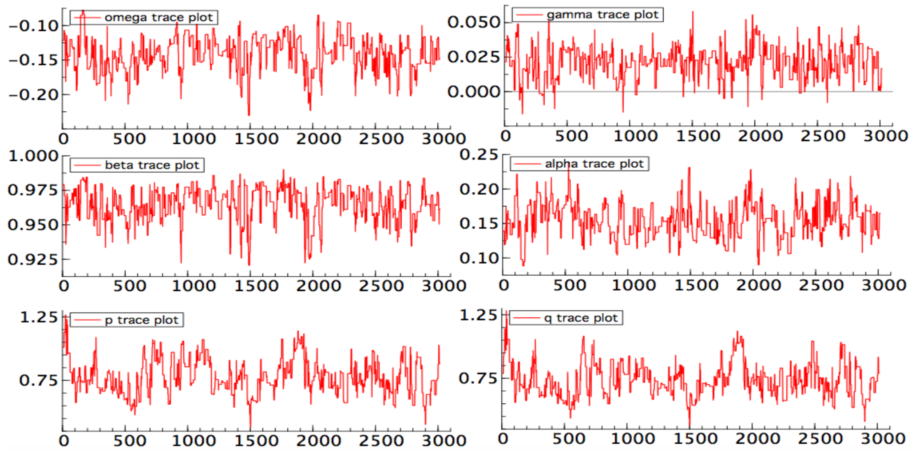

Figure 20: Trace plots for Belgian franc estimated using Bayesian estimation with a simulation of 100000 draws.

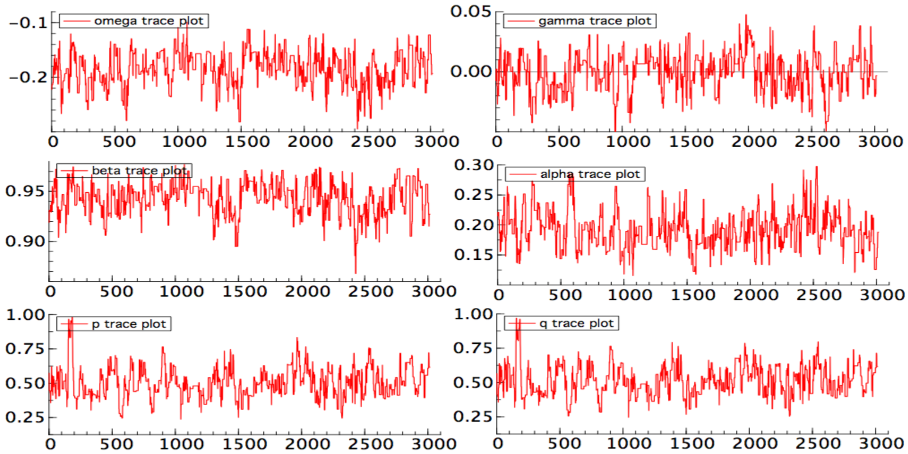

Figure 21: Trace plots for Italian lira estimated using Bayesian estimation with a simulation of 100000 draws.

D Appendix - Results Logarithmic Scoring Rule

Table 18: Diebold-Mariano statistics of the logarithmic scoring rule for the Deutsche mark. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	2.0676	-				
GARCH-GED	-0.85280	-2.1013	-			
EGARCH-GED	-0.26672	-1.0441	0.12820	-		
GARCH-EGB2	0.58623	-1.6782	1.4967	0.40079	-	
EGARCH-EGB2	2.1338	0.81247	2.7481	1.2552	2.2248	-

Table 19: Diebold-Mariano statistics of the logarithmic scoring rule for the British pound. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	2.2195	-				
GARCH-GED	0.87909	-1.0952	-			
EGARCH-GED	1.7651	0.77384	1.3998	-		
GARCH-EGB2	1.9402	-1.2750	0.095998	-1.3401	-	
EGARCH-EGB2	3.0004	2.3204	2.1348	-0.22651	2.4882	-

Table 20: Diebold-Mariano statistics of the logarithmic scoring rule for the Japanese yen. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	-0.094349	-				
GARCH-GED	-0.83335	-0.64048	-			
EGARCH-GED	-2.3142	-2.4123	-2.2829	-		
GARCH-EGB2	0.49868	0.45832	1.3471	2.6947	-	
EGARCH-EGB2	0.15655	0.22063	0.86214	2.7013	-0.57856	-

Table 21: Diebold-Mariano statistics of the logarithmic scoring rule for the French franc. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	1.7244	-				
GARCH-GED	-0.39195	-1.5266	-			
EGARCH-GED	1.7746	0.93670	2.8240	-		
GARCH-EGB2	0.58592	-1.3417	0.84350	-1.8010	-	
EGARCH-EGB2	1.7939	0.75450	2.1197	0.80420	1.8399	-

Table 22: Diebold-Mariano statistics of the logarithmic scoring rule for the Belgian franc. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	2.3274	-				
GARCH-GED	-0.84560	-2.2338	-			
EGARCH-GED	-1.6383	-2.0361	-1.4391	-		
GARCH-EGB2	0.18364	-2.2498	1.2299	1.6852	-	
EGARCH-EGB2	1.9414	0.17879	2.6217	2.0823	2.3211	-

Table 23: Diebold-Mariano statistics of the logarithmic scoring rule for the Italian lira. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	0.66340	-				
GARCH-GED	3.6902	4.5551	-			
EGARCH-GED	3.1963	2.8813	0.47166	-		
GARCH-EGB2	3.8063	1.0862	-3.1611	-2.6627	-	
EGARCH-EGB2	-0.43111	-0.48490	-0.86009	-0.85991	-0.55872	-

Table 24: Diebold-Mariano statistics of the logarithmic scoring rule for the euro. A negative (positive) value means that the model in the row (column) is better.

Model	GARCH-T	EGARCH-T	GARCH-GED	EGARCH-GED	GARCH-EGB2	EGARCH-EGB2
GARCH-T	-					
EGARCH-T	0.23740	-				
GARCH-GED	-0.39645	-0.41932	-			
EGARCH-GED	-1.1800	-1.3566	-1.1425	-		
GARCH-EGB2	-2.0290	-0.95761	-0.76841	0.76727	-	
EGARCH-EGB2	-0.55688	-2.0199	-0.27170	0.89094	0.19322	-