LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

Forecasting Exchange Rate Value-at-Risk and Expected Shortfall: A GARCH-EVT Approach

Master Thesis in Finance, August 2022

School of Economics and Management, Lund University

Author: Christoffer Titov

Supervisor: Birger Nilsson

Abstract

This thesis aims to investigate the accuracy of Value-at-Risk and Expected Shortfall forecasts of various GARCH-type models based on five currency exchange rate pairs. The GARCH models are employed under different conditional distributional assumptions, and extended using the two-stage Extreme Value Theory (EVT) approach of McNeil and Frey (2000). The forecasts are evaluated through simulation using the backtesting methodologies of Christoffersen (1998) and Acerbi & Szekely (2014). We find that forecasts of models assuming a skewed t-distribution are rejected the least number of times. Furthermore, the usefulness of the EVT approach of McNeil and Frey (2000) appears to be dependent on the distributional assumption as well as the choice of quintile. No conditional volatility model is consistently found to be superior to the others.

Keywords: GARCH, Extreme Value Theory, Value-at-Risk, Expected Shortfall, Exchange Rate Volatility

Acknowledgements

I would like to take this opportunity to thank my supervisor Birger Nilsson for his support. His comments provided valuable guidance in completing this thesis.

Contents

1. Introduction

This chapter aims to provide a background to the research topic of this thesis. The first section defines the concepts of risk management and risk measures, and provides a short description of Value-at-Risk and Expected Shortfall. This section also outlines the current and upcoming regulatory standards for market risk exposures. The following section highlights a few stylized facts of financial time series data, and the subsequent section presents the purpose of the thesis. The last section mentions financial literature of relevance and shortly discusses previous findings. This section also mentions the contribution of this paper to the existing literature.

1.1 Regulation, Risk Management and Risk Measures

Events such as the global financial crisis of 2007–08 and, more recently, the COVID-19 pandemic have shed a light on the need for effective and robust risk management practices. In essence, risk management is the process of identifying and measuring risks in order to ensure resilience to uncertain future events (McNeil et al., 2015, p. 7). There are various types of risks that financial institutions have to manage, including operational risk, credit risk, and market risk. The regulatory agreements regarding the latter are issued by the Basel Committee on Banking Supervision (BCBS). As a response to the flaws in the prior market risk framework that came to light during the global financial crisis, BCBS issued a consultative document, The Fundamental Review of the Trading Book (FRTB), in which the international regulatory standards for banking institutions were revised and new capital requirements for market risk exposures were proposed (BIS, 2013). As such, one of the key revisions of the document is that Value-at-Risk (VaR), which has been widely used in the last decades and currently is the required risk measure according to the Basel framework, is to be replaced with Expected Shortfall (ES). This reform is expected to be implemented in January 2023 under the Basel Accord (BIS, 2020).

Risk measures, in broad terms, determine the "riskiness" of a financial position by linking it to a quantifiable potential loss (McNeil et al., 2015, p. 61). They are used for a number of purposes, such as determining the capital and margin requirements for financial institutions and investors to buffer against unexpected losses and limit the amount of risk. Both Value-at-Risk and Expected Shortfall are distributional risk measures, i.e., they are statistical quantities that are derived from a loss distribution. Value-at-Risk corresponds to a given quantile of the loss distribution. It demonstrates the maximum loss that is expected given a pre-determined confidence level. For example, if an asset has a daily VaR(0.95) of 10%, then there is a 95% probability that the loss will not exceed 10% in one day. Although this measure has some intuitively appealing properties, such as its straightforward interpretation and robust backtesting capabilities, it does have some potential drawbacks. Besides that it lacks the desired property of subadditivity[1], which Artzner et al. (1999) were among the first to point out, it also is unable to capture "tail risk", as pointed out in FRTB (BIS, 2013). That is, Value-at-Risk does not say anything about the magnitude of the loss when the given quantile is exceeded. Expected Shortfall, on the other hand, represents the expected loss beyond a given quantile of the loss distribution. It thus provides information on both the probability of a large loss occurring and the expected magnitude of the loss when it occurs. In addition, this risk measure fulfills the property of subadditivity, thereby circumventing the main shortcomings of Value-at-Risk. There are, however, some potential disadvantages to Expected Shortfall as well, mainly in regard to its backtesting capabilities. Backtesting Expected Shortfall is more difficult than Value-at-Risk and requires a larger sample size to attain similar precision, see e.g., Yamai & Yoshiba (2005).

[1] Subadditivity is a risk aggregation property that satisfies $R(L_1 + L_2) \leq R(L_1) + R(L_2)$ for a risk measure R. The rationale behind this property is that the risk of a merged portfolio cannot exceed the risk of the two individual portfolios due to diversification effects (Artzner et al., 1999).

A variety of methods utilize historical data to predict VaR and ES. In particular, these methods are mostly centered around modelling the conditional variance, which in turn can be used to obtain estimates of VaR and ES. In the field of financial statistics, one of the most popular volatility forecasting models is the generalized autoregressive conditional heteroskedasticity (GARCH) model, introduced by Bollerslev (1986). The popularity of the model can largely be ascribed to its ability to capture the volatility clustering phenomenon that is one of the so-called "stylized facts" of financial time series. A number of extensions of the original GARCH model has since been introduced in order to, for example, make use of high frequency data and to incorporate additional stylized facts of financial data.

1.2 Stylized facts of financial time series

Empirical observations from a wide range of price series, across different assets, markets and time periods, suggest that they all have similar properties from a statistical point of view; they exhibit so-called stylized facts of financial time series. As previously mentioned, one observed phenomenon of financial time series is that volatility tends to cluster, i.e., large price changes, regardless of sign, tend to be followed by additional large price changes, and vice versa (Cont, 2001). This means that there usually are calm periods of low volatility, which then are followed by more turbulent periods, and so on. Time series data that exhibits these properties are known to be conditionally heteroskedastic, i.e., the conditional variance[2] varies over time. Another characteristic of financial time series is that returns tend to exhibit heavy tails. This has raised questions whether it is appropriate to model return series using the otherwise popular normal distribution, as it may lead to underestimation of risks. However, assessing the exact form of the tails of financial returns is often a difficult task, see Cont (2001). A third stylized fact of financial time series is the phenomenon called the leverage effect, first noted by Black (1976). This effect refers to the observation that past negative shocks tend to affect current volatility to a greater extent than equally large positive shocks do. This means that more turbulent periods can generally be expected in the aftermath of losses in comparison to gains of similar magnitude. A final empirical finding to mention is the "gain/loss asymmetry" in returns, as Cont (2001, p. 224) describes it. Returns, particularly from aggregated stock markets, tend to exhibit negative skewness. That is, there is an increased probability of negative returns than what is implied by a symmetric distribution. There are additional stylized facts of financial data, however, the ones highlighted above will be of focus as they form the basis of the modelling choices of this paper.

1.3 Purpose

The purpose of this thesis is to investigate the forecasting performance of different variations of GARCH models. The models will be applied to produce one day ahead predictions of Value-at-Risk and Expected Shortfall for five major foreign exchange rate pairs. The forecasting accuracy of each model will then be evaluated through backtesting. Different GARCH-type models will be employed in order to assess if more complex extensions of the original GARCH model, i.e., models that incorporate additional stylized facts and make use of high frequency data, yield more precise predictions. The conditional variance models that will be utilized in this study are

1. GARCH (1,1)
2. IGARCH (1,1)
3. GJR-GARCH (1,1)
4. EGARCH (1,1)
5. Realized GARCH (1,1)

Models 1 and 2 have a similar structure that enables them to capture the volatility clustering phenomenon. These models are symmetrical in the sense that both positive and negative shocks are

[2] That is, conditional on past information.

assumed to have the same effect on the volatility process. Models 3 and 4 extend the previous models by also allowing for an asymmetric response in volatility to shocks, thereby incorporating the leverage effect. Model 5 goes one step further by utilizing a realized measure of volatility derived from intraday data in its structure, while also accounting for the previous effects.

Moreover, as empirical observations suggest that the distribution of financial returns exhibits some specific characteristics, the models will be employed under different conditional distributional assumptions. In particular, the distribution of the standardized residuals will be modelled using the normal distribution, the student's t-distribution and the skewed student's t-distribution; the former to examine whether it indeed leads to an underestimation of risk, and the latter to examine if distributions that feature heavy tails and skewness yield more accurate estimates. Furthermore, given the uncertain nature of the tail properties, each variation of the models will be combined with Extreme Value Theory (EVT), following the two-stage approach of McNeil & Frey (2000). This approach provides an alternative way of obtaining estimates of VaR and ES under heavy tails by utilizing a parametric method for the tail of a distribution.

1.4 Previous research

Concepts related to risk measures and EVT are covered extensively in the financial literature. In particular, we find the book of McNeil et al. (2015) to be very useful for both theoretical and practical purposes. This book covers many of the concepts related to time series modelling and forecasting, and provides a useful introduction to GARCH models and their application in finance. It also presents useful diagnostic tools for model checking, many of which have been applied in this thesis. To navigate through the universe of GARCH related models, we recommend the paper of Bollerslev (2008). This paper lists most GARCH-type models used in the literature.

Several studies in the field of forecasting suggest that models implementing EVT yield more accurate estimates of VaR than stand-alone GARCH models, which is the main motivation for its implementation in this thesis (see e.g., Gençay et al. (2003); Ho et al. (2000)). McNeil & Frey (2000) proposed a two-stage method for which the EVT approach could be applied within the GARCH modelling framework. Applying EVT under a GARCH structure is intuitively appealing as one fundamental notion of EVT is that the observations (returns) are independent and identically distributed (i.i.d.). It is widely recognized that returns often exhibit higher order dependency, i.e., that they are not i.i.d., which may adversely affect the accuracy of quintile estimates, see e.g., Wagner & Marsh (2005). The approach of McNeil & Frey (2000) provides a remedy to this issue by first fitting a GARCH model to the return series, clearing the series of higher order dependency, and then applying EVT to the standardized residuals of the GARCH model, which should be approximately i.i.d. A number of follow-up studies suggest that the two-stage approach of McNeil & Frey (2000) yield more accurate forecasts of VaR than other conventional models do, see e.g., Byström (2004); Fernandez (2005).

It should be mentioned that numerous studies have been conducted to investigate the forecasting performance of various volatility models. We will not go into any detail of these studies as this field is too large to cover. Instead, we refer the reader to the paper of Poon & Granger (2003). This extensive survey reviews the findings of several papers related to this topic. Their review covers a wide range of time series models, including historical volatility models, GARCH-type models and stochastic volatility models.

This thesis contributes to the existing literature by providing information on the forecasting capabilities of specific GARCH models for exchange rates. To the best of my knowledge, no previous study has investigated the accuracy of *both* VaR and ES predictions of the more novel Realized GARCH model in regard to this asset class. Similarly, the existing literature on the accuracy of both VaR and ES predictions of GARCH-EVT models is quite limited, particularly for exchanges rates.

2. Data

This chapter briefly outlines some facts about the foreign exchange market and presents the data sets that are implemented in this study. It is organized as follows. The first section provides a short description of the foreign exchange rate market and the currency pairs of interest. The following section explains the data gathering process and the procedure of transforming the raw data into return series. The last part presents summary statistics and visualizations of these series to highlight some stylized facts of financial data. Note that the full data sets are analyzed in this section, however, the specifications of the models are determined beforehand, thus avoiding any potential look-ahead bias. Also note that this chapter does not cover any of the theory – that is done in the following chapter.

2.1 The foreign exchange market

According to the most recent triennial survey of Bank of International Settlements (BIS, 2019), the average trading volume on the FOREX market amounts to $6.6 trillion per day, making it the most liquid market in the world. FOREX derivatives trading account for the majority of the daily turnover, while spot trades make up approximately thirty percent ($2 trillion) of the volume. The spot market is heavily dominated by financial institutions, accounting for almost 95% of all over-the counter (OTC) transactions.

Although the FOREX market consists of numerous currencies, the survey finds that only a few of the leading exchange rate pairs comprise the majority of OTC daily turnover on a global scale. Depicted in table 1 below, the five most traded currency pairs account for roughly 57 percent of the OTC turnover, where the USD/EUR exchange rate is, by a considerable margin, the most traded currency pair with an average daily volume of $1,584 billion. In table 1 it is also evident that the US Dollar is part of each of the most traded currency pairs. In fact, the survey finds that 88 percent of worldwide FOREX transactions feature the USD as one of the currencies, demonstrating the key influence of the currency in international FOREX trading.

Currency pairs	Amount (in millions of USD)	Proportion of total turnover (%)
USD / EUR	1,584,000	24.0
USD / JPY	871,000	13.2
USD / GBP	630,000	9.6
USD / AUD	358,000	5.4
USD / CAD	287,000	4.4

Table 1: Most traded currency pairs according to the Bank of International Settlements (BIS) Triennial Central Bank Survey 2019. The amount corresponds to the average daily OTC turnover.

The currency pairs depicted in table 1 comprise the data sets that will be employed in this paper. The main rationale behind this choice is that they cover the majority of the total turnover which in turn ensures excellent liquidity. Utilizing highly liquid assets is often beneficial in the study of volatility as it tends to mitigate market microstructure impacts such as the bid-ask bounce (Demsetz, 1968). Moreover, the frequent use of these currency pairs makes them attractive to analyze from a relevance point of view.

2.2 Data description

The raw FOREX (FX) data for the currency pairs depicted in table 1 is retrieved from Histdata[3]. Histdata is a website that provides financial data at high frequencies for several asset classes. Exchange rate data from Histdata has been used in a number of studies, see e.g., Islam & Hossain (2021); Yong et al. (2018); Gbatu et al. (2017). The frequency of the FX data that was obtained from this website is of the highest resolution available, i.e., tick data consisting of bid and ask quotes. This

[3] Available at https://www.histdata.com/download-free-forex-data/

frequency allows us to utilize as much information as possible for the realized kernel estimator, which is the realized measure of volatility that will be utilized for the Realized GARCH model of Hansen et al. (2011). A more detailed explanation of the realized kernel estimator and its implementation is presented in section 3.4.

The return series of each currency pair is created using the last observed mid-quote price of each day. The timelines of the data sets cover the beginning of the century[4] until the end of 2021. Such lengthy timelines allow for a rigorous examination of the capabilities of each model as periods of different volatility regimes are covered. The data sets encompass times of global financial turbulence such as the dot-com bubble, the global financial crisis, the covid-19 pandemic but also calmer periods in between these events. Another benefit of using such extensive data sets is that it makes backtesting possible for extreme quintiles of the loss series distribution.

2.3 Summary statistics and data visualization

The return series, r_t, are defined by

$$r_t = \log\left(\frac{FX_t}{FX_{t-1}}\right) = \log(FX_t) - \log(FX_{t-1}) \quad (1)$$

where FX_t is the last observed mid-quote price of the exchange rate series for day t. The return series of the full data sets are plotted in figure 1, and the summary statistics are presented in table 2.

FX Pair	Length	Start date	End date	Mean	St.dev	Min.	Max.	Skewness	Kurtosis	JB-statistic
EUR/USD	6626	2000-05-31	2021-12-31	0.003%	0.544%	-3.346%	3.977%	0.011	5.936	2383
USD/JPY	6633	2000-05-31	2021-12-31	0.001%	0.538%	-3.736%	4.361%	0.046	8.096	7188
GBP/USD	6617	2000-05-31	2021-12-31	-0.002%	0.540%	-9.577%	3.509%	-1.161	22.997	111826
AUD/USD	6377	2001-04-27	2021-12-31	0.004%	0.715%	-7.726%	6.783%	-0.263	13.138	27408
USD/CAD	6450	2001-01-03	2021-12-31	-0.003%	0.501%	-6.406%	3.344%	-0.143	10.233	14098

Table 2: Summary statistics of the return series. JB denotes the Jarque-Bera test statistic.

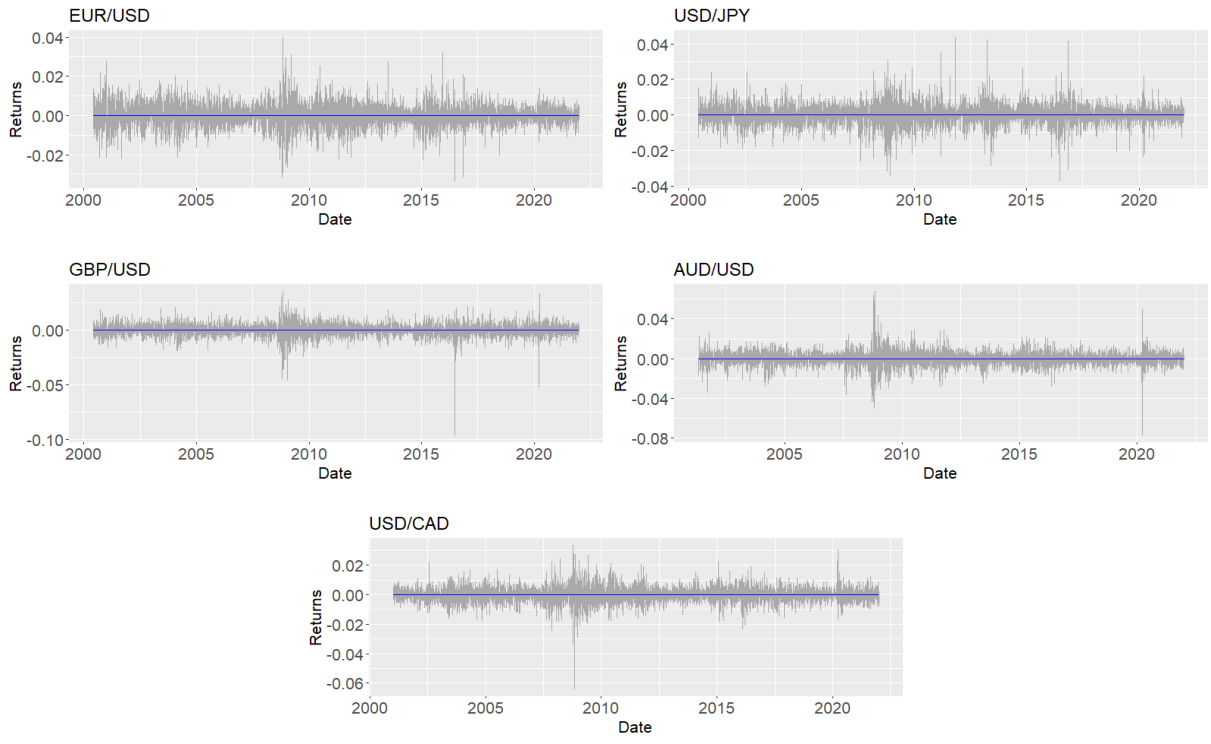

Figure 1: A visual representation of the return series. The blue horizontal line is fixed at 0.

[4] The starting point of the data sets differ slightly due to data availability.

There are a few things to note from the summary statistics presented in table 2. First, we can see that the mean return is approximately zero for all FX pairs. The empirical distribution of most FX pairs exhibits negative skewness, demonstrating a frequently observed phenomenon of asset returns. Moreover, by examining the kurtosis[5] of the FX pairs, it is evident that all distributions exhibit heavier tails than what would be implied by a normal distribution, demonstrating another key characteristic of financial returns. The null hypothesis of the Jarque-Bera test is rejected at a 0.01 significance level for all FX pairs, thus confirming that none of the empirical distributions are normally distributed. The specifications of the Jarque-Bera test can be found in section 6.4 in Appendix.

Another briefly mentioned stylized fact of financial returns is that volatility tends to cluster. By inspecting the plots of figure 1, there appears to be long periods in which volatility of returns tends to be high and other periods in which the opposite is true, suggesting that returns are dependent on past observations. During periods of financial turbulence, we observe sequences in which extreme returns are followed by additional extreme returns, which is particularly evident during the financial crisis of 08. We also observe periods in which small returns are followed by additional small returns. To get a clearer representation of this phenomenon, we will look at the realized kernel estimator in figure 2. Again, we will not present the theory behind this measure here - that is done in the subsequent chapter - for now it is sufficient to know that it represents a realized measure of volatility.

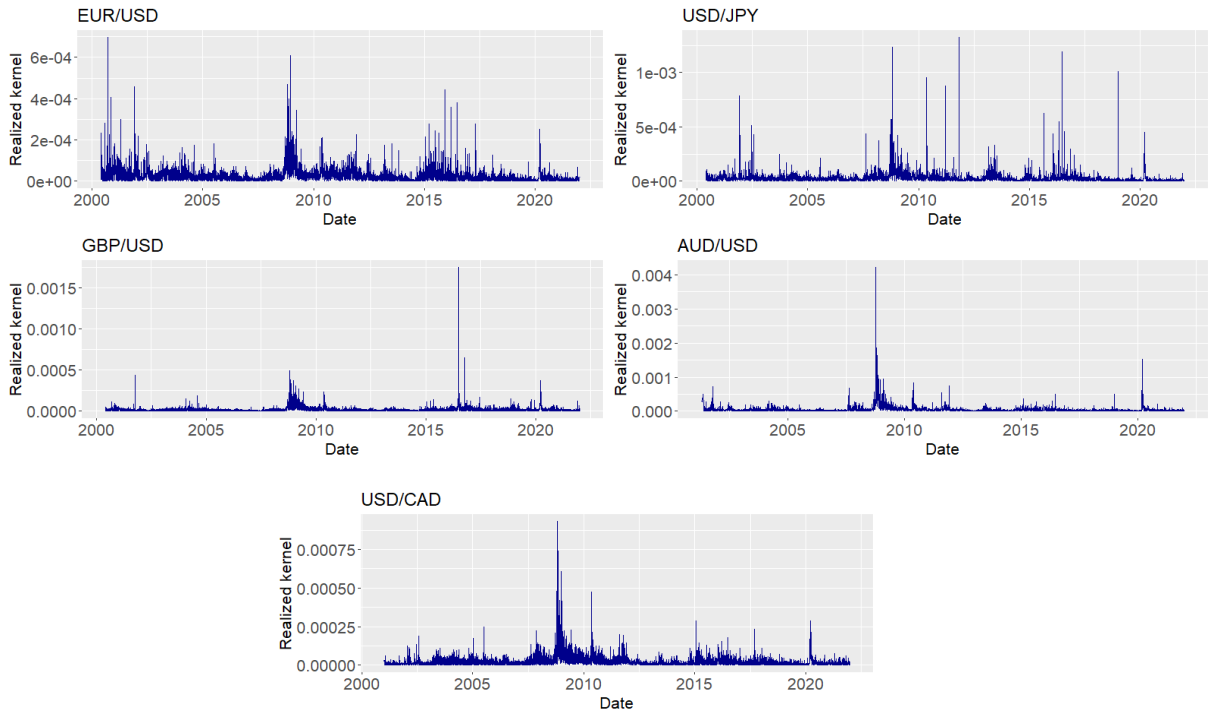

Figure 2: Realized kernel estimator for each FX pair.

For the realized kernel estimator, depicted in figure 2, the clustering phenomenon of volatility is even more prevalent, indicating that the GARCH framework is appropriate for our modelling purposes. In the face of economic crises, such as the dot-com bubble, the financial crisis and the outbreak of the COVID-19 pandemic, all currency pairs underwent significant turmoil for extended periods. It is also evident that there have been calmer periods of low volatility for extended periods of time, e.g., prior to the financial crisis and the COVID-19 pandemic. Furthermore, we also observe occasional sharp spikes in volatility that are more isolated, for instance GBP/USD during Brexit in 2016.

[5] Kurtosis is a measure that relates to the heaviness or lightness in the tails of a distribution. A normal distribution has a kurtosis of three. A kurtosis in excess of three implies a leptokurtic distribution, i.e., it has heaver tails than that of a normal distribution.

To further quantify the dependence of the return series we will examine the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the return series and their squared counterparts. Plotting these functions in so-called correlograms can be useful to detect if the observations are independent of each other or if there is any serial correlation and dependence in the series. See section 6.3 in Appendix for a formal definition of the autocorrelation function and section 3.3 for a brief explanation of processes that can be identified using correlograms. The ACF and PACF of the returns and squared returns are plotted in figure 3 and figure 4, respectively.

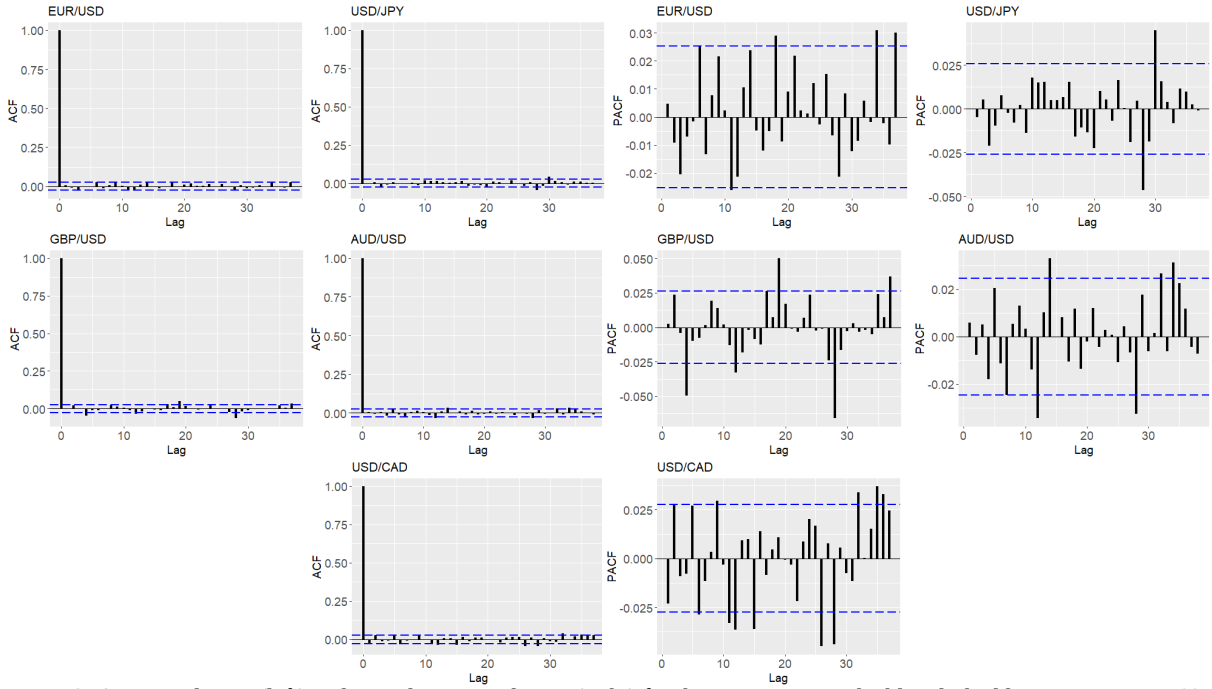

Figure 3: Autocorrelations(left) and partial autocorrelations(right) for the return series. The blue dashed line represents a 5% confidence level.

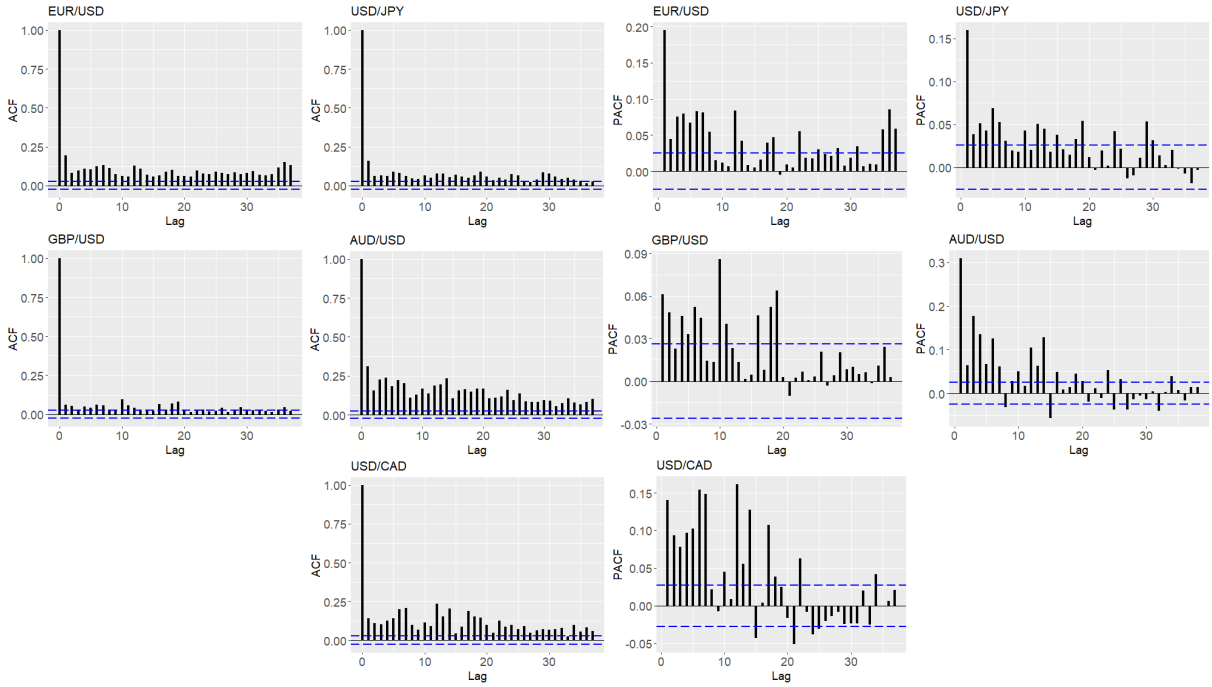

Figure 4: Autocorrelations(left) and partial autocorrelations(right) for the squared return series. The blue dashed line represents a 5% confidence level.

Based on an ocular inspection of figure 3, we observe that most lags of the ACF:s tend to stay within the given bandwidth. Assuming that distant lags are potentially spurious, the ACF:s suggest a nonexistent MA order for most return series. The return series of GBP/USD do however exhibit a significant serial correlation at lag four, potentially indicating a low moving-average (MA) order. Likewise, the plots of the PACF:s indicate low autoregressive (AR) order for GBP/USD, and perhaps also for USD/CAD, AUD/USD and EUR/USD, again assuming that distant lags are potentially spurious. In general, the dependencies tend to be small for all data sets despite that we observe occasional minor lags of significance. However, for the squared returns, depicted in figure 4, the dependence of the returns is more evident. The ACF and the PACF are significant at most lags for all return series, suggesting a higher order dependency. This is the dependence that volatility models are designed to capture, again confirming that the GARCH framework is appropriate for our modelling purposes.

Next, we examine the empirical distributions of the daily returns by analyzing Q-Q plots of different theoretical distributions. A Q-Q plot is a graphical tool that can be used to analyze the relationship between the empirical quantiles of the data and the theoretical quantiles of a probability distribution. If the two distributions are equivalent we would expect perfect linearity between the quintiles. In figure 5, the empirical distributions of the FX pairs are compared to the best fitted normal distribution, represented by the black line. For a formal definition of this probability distribution, we refer the reader to section 6.1 in Appendix.

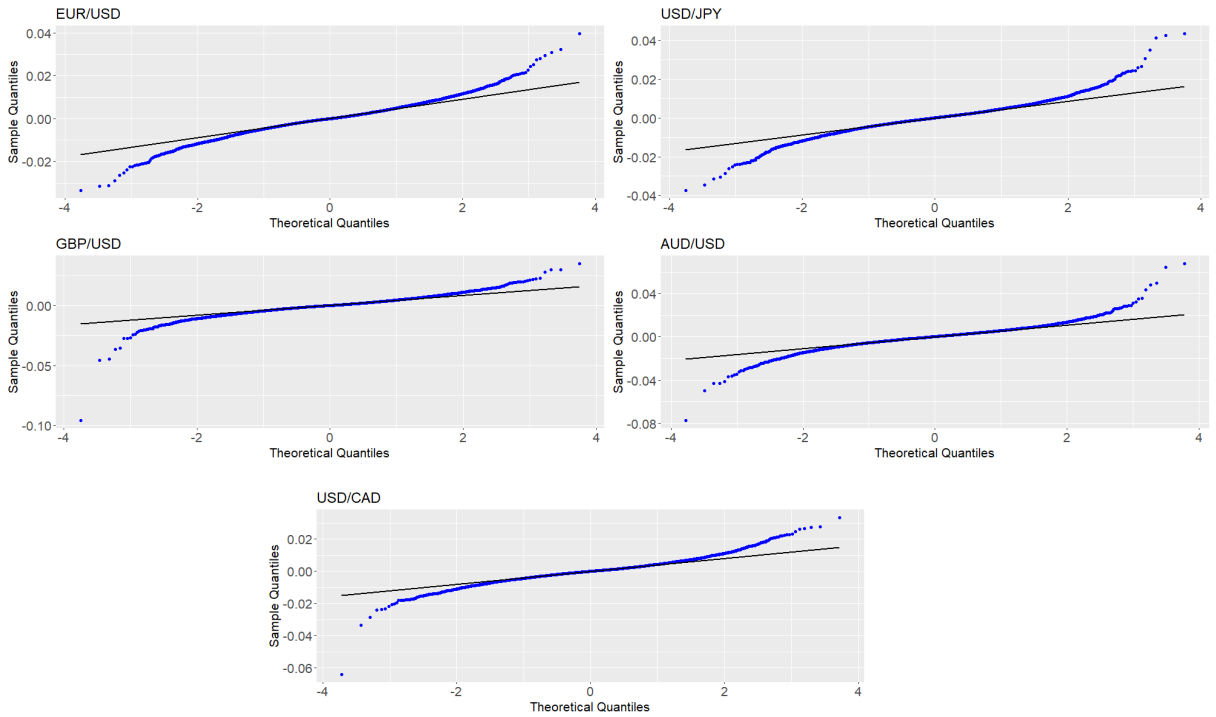

Figure 5: Q-Q plots of the empirical distribution of the returns. The black line represents the best fitted normal distribution.

The lack of linearity of the Q-Q plots depicted in figure 5 demonstrates that the normal distribution does not provide a good fit for any of the currency pairs. This is expected as this was implied by the kurtosis and Jarque-Bera test statistic presented in table 2, suggesting that the empirical distributions have heaver tails than that of a normal distribution. To get a different representation of the fit of a normal distribution, we present histograms of the daily return series with a superimposed theoretical normal distribution in figure 6.

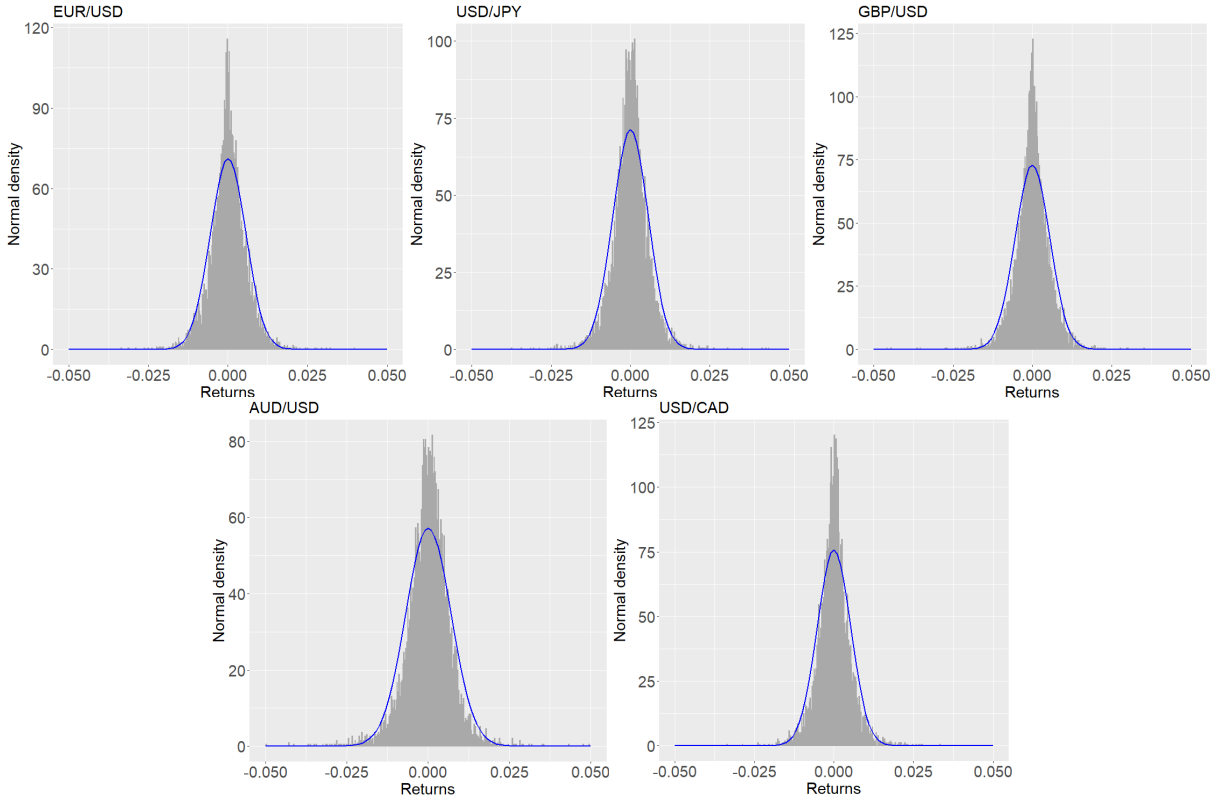

Figure 6: Histogram of daily returns with a superimposed normal density curve(blue).

Clearly the normal distribution does not characterize the empirical distributions well, neither in the peaks of the distribution or in the tails, which is the part that we are interested in. This finding supports fitting a Generalized Pareto distribution (GPD) to the tails or utilizing a different distribution that features fat tails, such as the student's t-distribution. In figure 7, the empirical distributions are compared to the best fitted student's t-distribution, represented by the black line. To examine the density, we present histograms of the daily return series with a superimposed theoretical student's t-distribution in figure 8. The degrees of freedom of these distributions were estimated using maximum likelihood. See section 6.1 in Appendix for formal definition of this probability distribution.

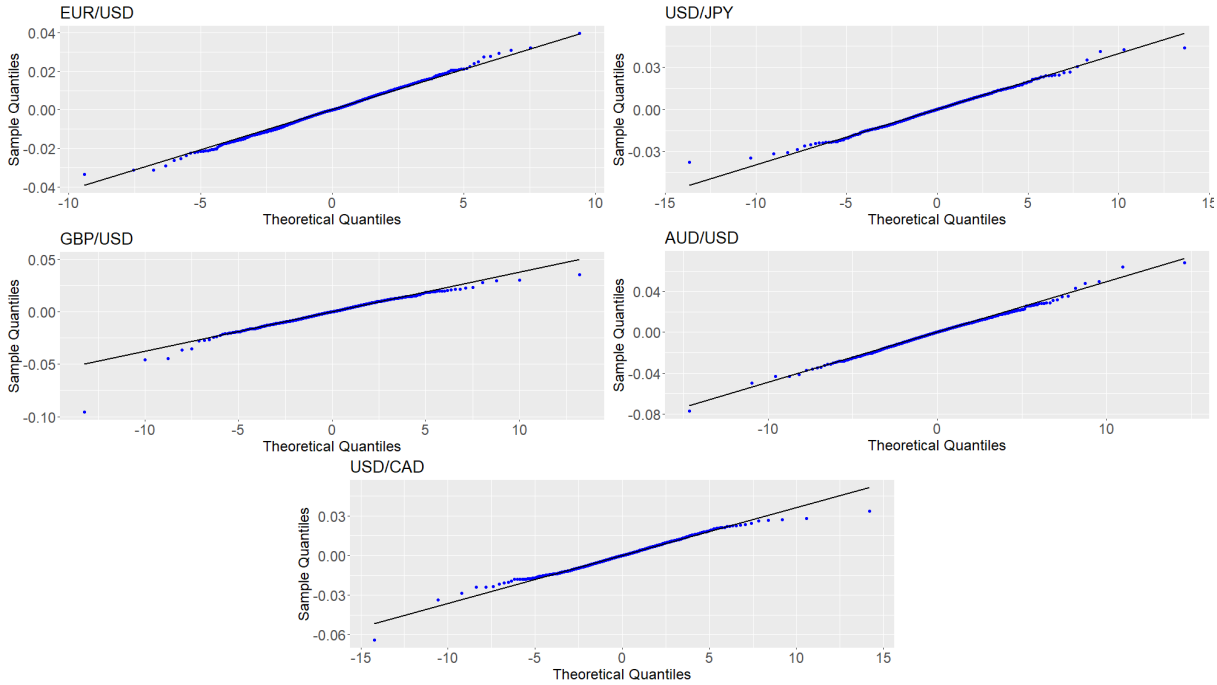

Figure 7: Q-Q plots of the empirical distribution of the returns. The black line represents the best fitted t-distribution.

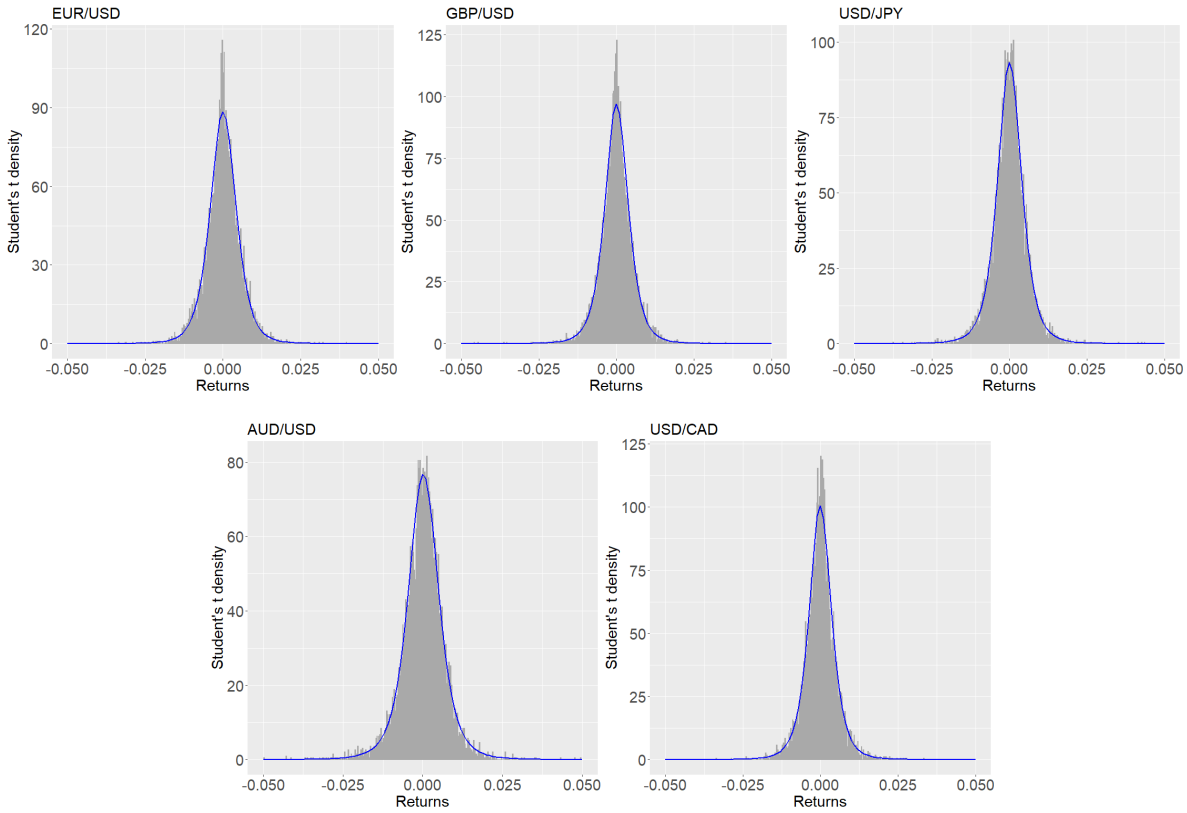

Figure 8: Histogram of daily returns with a superimposed student's t density curve(blue).

The linearity of the Q-Q plots depicted in figure 7 indicates that the student's t-distribution provides a significantly better fit for the tails of the empirical distributions than the normal distribution. By inspecting the superimposed student's t density curves of figure 8, we see that it also does a better job at capturing the "peakedness" of the empirical distribution. However, there appears to be quite a few observations in the tails that still surpass the superimposed density curve which, again, supports fitting a GPD to the tails.

To examine whether incorporating skewness improves the distributional fit, we present Q-Q plots and a superimposed density curve of the best fitted skewed student's t-distribution in figure 9 and figure 10, respectively. Again, the degrees of freedom and skewness parameter are estimated using the method of maximum likelihood. Based on an ocular inspection of these figures, it is not completely clear whether the skewed student's t-distribution provides a better fit to the series than the symmetrical student's t-distribution. They appear to be very similar for most return series, which is not too surprising given that the skewness coefficient was close to one for all currency pairs. Again, we refer the reader to section 6.1 in Appendix for more information on this distribution and the skewness coefficient.

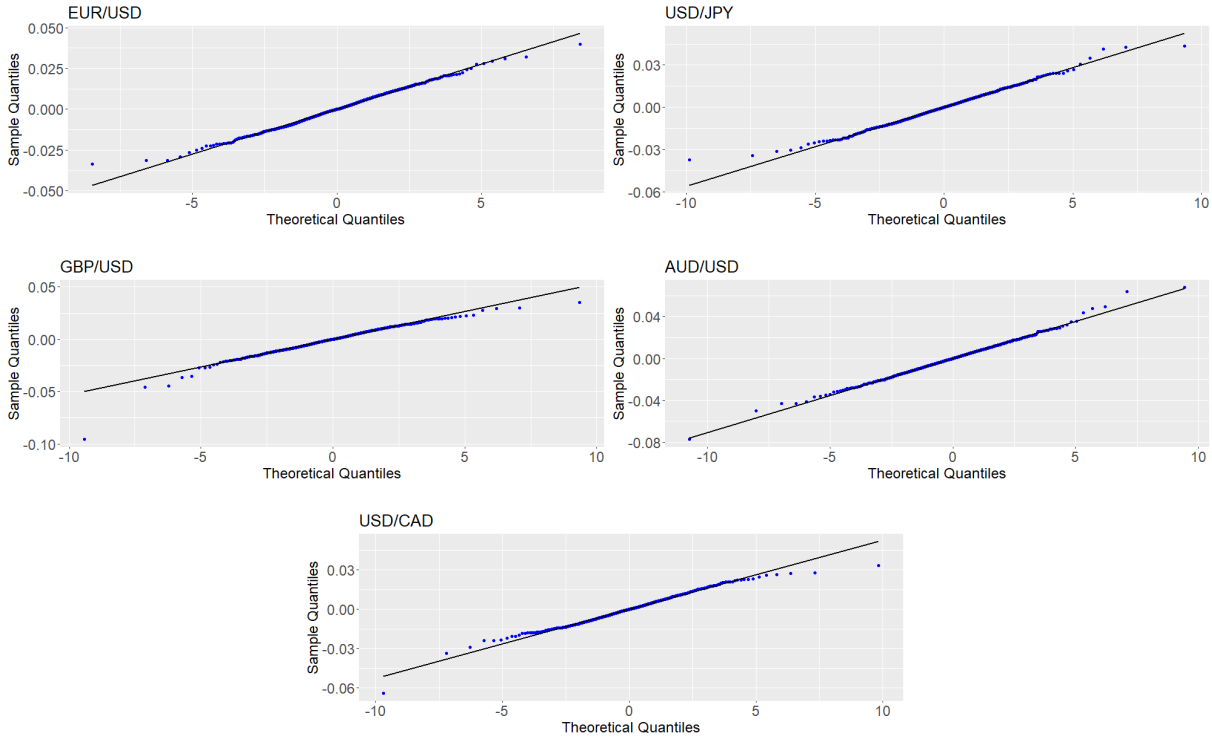

Figure 9: Q-Q plots of the empirical distribution of the returns. The black line represents the best fitted skewed t-distribution.

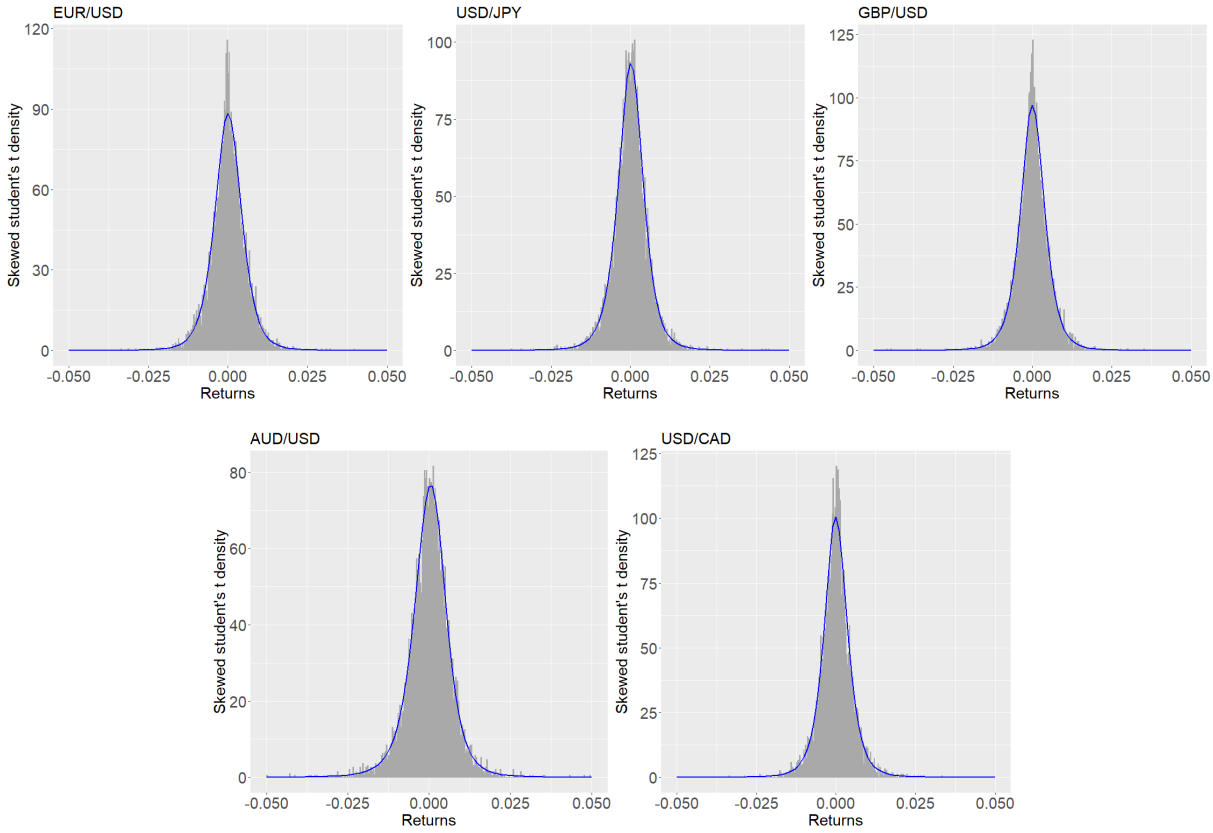

Figure 10: Histogram of daily returns with a superimposed skewed student's t density curve(blue).

3. Methodology and theoretical background

In this chapter we present the methodology that will be used in this paper as well as the underlying theory behind our models and estimates. The first section presents the reasoning behind employing a rolling window approach for out-of-sample forecasting. Section 3.2 outlines the properties of the assumed loss process, and the subsequent section presents the criteria for which the specification of the conditional mean is determined. Section 3.4 examines the various GARCH-models of interest and their different properties, including the distributional assumptions. The realized kernel estimator and its implementation will also be discussed in this section. The subsequent section provides a formal definition of the risk measures and explains how they will be forecasted. Section 3.6 outlines the Extreme Value Theory and how it will be combined with the GARCH framework. Section 3.7 examines the backtesting procedures that are employed to determine the accuracy of our forecasts. Finally, the last section shortly mentions the software that is used to carry out the calculations.

3.1 Out-of-sample forecasting

As mentioned in the introduction of this paper, this study aims to produce and evaluate 1-day ahead forecasts of VaR and ES. This method refers to so-called out-of-sample forecasting, meaning that we are using a different set of data for the fitting of the models than the data that is used for assessing the performance of the forecasts. In producing these forecasts, we will employ a rolling window approach. The basic structure is the following: let *n* denote the full sample size. The rolling window size, i.e., the number of observations that are used to fit the model, is fixed and denoted by *w*, which also is the initial forecast origin. The forecast horizon, *h*, represents the number of days to be forecasted into the future. It is fixed to one in this paper as we only wish to forecast volatility for 1-day ahead. The initial window, consisting of the first observation to observation *w*, will be used to calibrate the model and produce a one-step ahead forecast for day *w+1*. For the next forecast, the forecast origin is advanced by 1 and we now use the second observation to observation *w+1* to fit the model and produce the 1-day ahead forecast, thus keeping the window size constant. This process is repeated until the forecast origin is equal to *n*. We will then have a series of *n* - *w* forecasted values of VaR and ES, corresponding to the full out-of-sample period, which can be evaluated against the actual return series through backtesting. The rolling window approach of this paper is illustrated in figure 11 below.

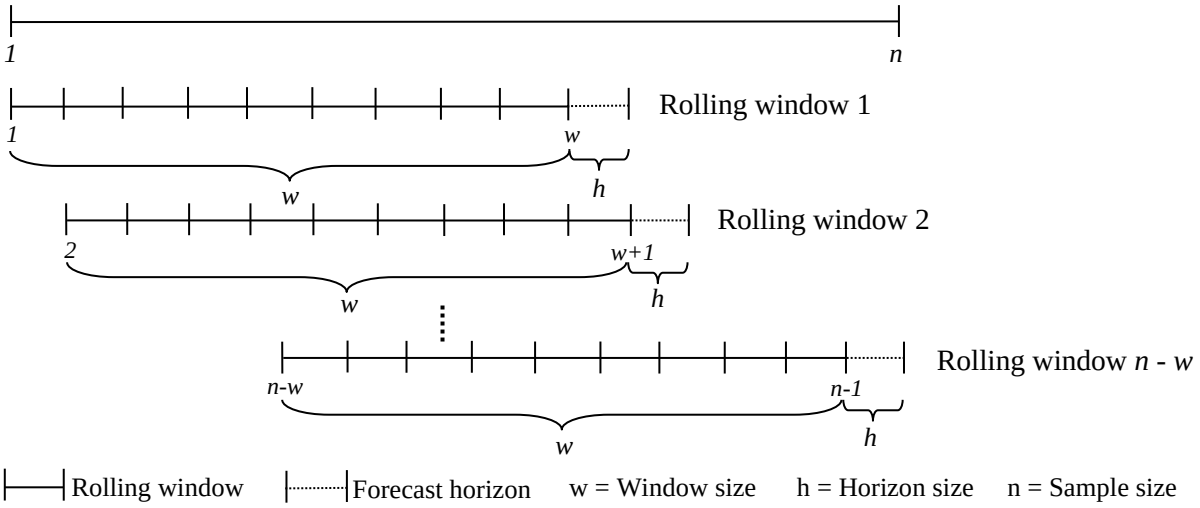

Figure 11: The rolling window approach. Note that *w* does not include the horizon size *h* using this definition.

Employing a rolling window approach is useful if the statistical properties of the data change over time as the oldest observation is dropped in each new iteration. There is, however, a trade-off when choosing the size of the rolling window. On the one hand, longer window sizes utilize more information and yield smoother estimates than windows of shorter sizes do. On the other hand, if we include too many observations in our window there is a risk that the statistical properties of the data may have changed and that the initial observations of our window adversely affect the accuracy of the forecasts. This also implies that our models will be less responsive to new changes. The optimal window length is likely dependent on the specific dynamics of the data set being used and thus not easily generalized. Therefore, we will simply use the same length that was used by McNeil & Frey (2000), i.e., a window size of $w = 1000$, for all of our data sets.

3.2 Basic structure

In equation (1) we defined the return series r_t as the log return based on the observed mid-quote prices. In the remainder of this paper, we will work with negated loss series, i.e., $X_t = -r_t$. This is done for convenience as it is the usual practice in the literature on EVT to work with the upper tails of the distributions. This transformation has no impact on the results.

We assume that the dynamics describing X_t can be characterized by the stochastic process

$$X_t = \mu_t + \varepsilon_t \tag{2}$$

$$\varepsilon_t = \sqrt{\sigma_t^2} z_t \quad z_t \sim F(0,1)\ i.i.d. \tag{3}$$

where z_t, also called the *innovations*, are random variables generated from a strict white noise process[6] with a zero mean and a unit variance stemming from a marginal distribution F. We assume that both the conditional mean, μ_t, and the conditional variance, σ_t^2, are measurable with respect to the information about the loss process up to time $t-1$, denoted by G, such that

$$\mu_t = E(X_t|G_{t-1}) \tag{4}$$

$$\sigma_t^2 = Var(X_t|G_{t-1}) = Var(\varepsilon_t|G_{t-1}) \tag{5}$$

The general idea is to model this loss process as accurately as possible. The GARCH models presented in section 3.4 are concerned with σ_t^2 and the fashion under which it evolves. However, the equation for the conditional mean should also be specified in order to accurately capture the loss process. In section 2.3 we examined the dependence in the return series and the findings suggested that the return series of some FX pairs might exhibit some minor lower order serial correlation. In fact, McNeil et al. (2015, p. 79) note that asset returns typically exhibit lower order serial correlation. Thus, to allow for potential serial correlation in the loss series we should specify a model for the conditional mean in which this is accounted for. Consequently, the GARCH models could then be estimated on the mean adjusted process ε_t.

3.3 Conditional mean and model selection

If we assume that the conditional mean follows a stationary autoregressive-moving-average (ARMA), it is described by

$$\mu_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{6}$$

[6] A white noise process is covariance stationary and serially uncorrelated with a mean equal to zero and a finite and constant variance. Moreover, a strict white noise process requires that the process is independent and identically distributed (i.i.d.).

where p is the lag order of an AR process and q is the lag order of a MA process. This representation allows for forecasting of the conditional mean. The one step ahead forecast is obtained by

$$\mu_{t+1} = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i X_{t-i+1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i+1} \tag{7}$$

Whether a process follows an AR and/or a MA process can be identified using correlograms. An AR process of order p displays a geometrically decaying ACF and p number of spikes in the PACF (Tsay, 2010, p. 46). On the contrary, a MA process of order q is characterized by a geometrically decaying PACF with q number of spikes in the ACF (Tsay, 2010, p. 60).

A common approach in the literature is to use the in-sample period, or the initial window to determine the appropriate model for the conditional mean. However, as the dynamics of the return series might change as we roll the window forward this approach might not be entirely satisfactory. As it is not feasible to make inference from ocular inspections of correlograms for every rolling window, we will instead employ an algorithm that chooses the appropriate ARMA model for the conditional mean for every window to capture any potential structural changes in the return series. More specifically, this is achieved by minimizing the Akaike information criterion. The Akaike information criterion for ARMA models is defined by Hyndman et al. (2008) as:

$$AIC = -2\ln(L) + 2(p + q + 1) \tag{8}$$

where L denotes the maximized likelihood value of the fitted model. We observe in equation (8) that the last term discourages overfitting of the model, which is a valuable property of this criterion. For more details on the Akaike criterion, see Akaike (1974).

In order to achieve stability between the windows, the model will only be updated if there is sufficient evidence of a structural change. The rule of thumb outlined by Burnham & Anderson (2004) states that if the difference between one model's AIC and the model with minimum AIC is less than 2, then there is still substantial support for the former model. If the difference is between 4-7, the support for the former is considerably reduced. If the difference is more than 10, then there is essentially no support for the model. Therefore, the model will only be updated if the minimized AIC of the identified model is at least 4 units lower than the AIC of the previously identified model (based on the data in same window). If the difference is less than 4, we will assume that both models are approximately equally good approximations of the mean process, thereby sticking with the former to achieve stability.

To examine whether model selection by the minimization of the AIC generate adequate results for our GARCH modelling purposes, the Ljung-Box test[7] was performed on the standardized residuals and the squared standardized residuals from all models based on the identified ARMA process of the initial window. These residuals should feature the properties of z_t in equation (3), i.e., they should be independent of one another. The results are presented in table 10 in Appendix. The results suggest that both the standardized residuals and the squared standardized residuals of all models are free from any autocorrelation up to lag 10, implying that the conditional mean is correctly specified and that our models are suitable for the implementation of EVT.

3.4 Conditional variance models

In this section we present all GARCH-type models that will be employed in this thesis. The parameters will be estimated using the Maximum Likelihood method. This method aims to find the most probable parameter values given the data that is observed. See section 6.2 for further details.

[7] The Ljung-box test assesses whether there is an absence of serial correlation in the data up to lag k. See section 6.4 in Appendix for more information on this test.

Standard GARCH(1,1)

The most popularized model to capture the higher order dependence of returns is the generalized autoregressive conditional heteroskedasticity (GARCH) model, introduced by Bollerslev (1986). The conditional variance of the GARCH (*p, q)* process is defined by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{9}$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, for $i = 1,\ldots,q$ and $j = 1,\ldots,p$, denoting the lagged values of the residuals and conditional variances, respectively. The first conditions are required to ensure a non-negative conditional variance, and the process is covariance-stationary since it is required that $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$. We can observe that the variance is defined as a weighted function of an intercept, the shocks from the previous periods and the conditional variances from the previous periods. The GARCH model thereby accounts for the phenomenon of volatility clustering by making the current period's volatility dependent on the last period's volatility. α measures the extent to which a shock today feeds through into next period's volatility, and β measures the degree of persistence of past observations. Given a high value of β relative to the value of α, large past conditional variances will in turn result in large values for σ_t^2, and vice versa, thus creating a clustering effect. If the opposite is true, i.e., α is large relative to β, then the conditional variance reacts more quickly to shocks, resulting in spikier volatility processes.

For all GARCH models applied in this thesis, we will set the parameters for the lagged values of the residuals and conditional variances to be equal to 1, which is the most common modelling choice in the literature. The standard GARCH(1,1) model can be used to produce a forecast of the conditional variance one period ahead by utilizing the values of the residual and conditional variance at time *t* by

$$\sigma_{t+1}^2 = \omega + \alpha \varepsilon_t^2 + \beta \sigma_t^2 \tag{10}$$

IGARCH(1,1)

Similar to the original GARCH model, the IGARCH model of Engle & Bollerslev (1986) is symmetrical in the sense that both positive and negative shocks are assumed to have the same effect on the volatility process. However, contrary to the GARCH model, IGARCH is not defined to be a covariance-stationary process. It has the same representation as the GARCH model presented in equation (9), but instead satisfies the condition $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1$. Shocks to the volatility process therefore persist, effectively giving the model infinite memory. It follows that the forecasting approach of the IGARCH model is the same as that of the standard GARCH model, given by equation (10).

GJR-GARCH(1,1)

As mentioned in the introduction of this paper, a common empirical observation among asset returns is that they exhibit the so-called leverage effect, referring to the fact that past negative shocks tend to affect current volatility to a greater extent than equally large positive shocks do. To incorporate this stylized fact into the GARCH modelling framework, Glosten et al. (1993) introduced the GJR-GARCH (*p, q)* model. The model has the following representation for the conditional variance

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i + \gamma_i I_{t-i}) \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{11}$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\gamma_i \geq 0$ for $i = 1,\ldots,q$ and $j = 1,\ldots,p$, denoting the lagged values of the residuals and conditional variances, respectively. The indicator I_{t-i} is a binary variable satisfying the condition

$$I_{t-i} = \begin{cases} 1 \text{ if } \varepsilon_{t-i} > 0 \\ 0 \text{ if } \varepsilon_{t-i} \leq 0 \end{cases} \tag{12}$$

The indicator variable thus takes the value one if there is a positive (negative return) shock, enabling the model to distinguish between positive and negative shocks. γ_i measures the magnitude to which the leverage effect impacts the volatility process. Note that the conditions in (12) would be reversed if we were to be working with a series where losses are defined as negative numbers.

For the GJR-GARCH(1,1) model, the one day ahead forecast of the conditional variance is given by

$$\sigma_{t+1}^2 = \omega + (\alpha + \gamma I_t)\varepsilon_t^2 + \beta\sigma_t^2 \tag{13}$$

EGARCH(1,1)

Another model that incorporates the leverage effect is the exponential GARCH, EGARCH (*p, q)*, of Nelson (1991). It has a somewhat different representation than GJR-GARCH, given by

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} \frac{\alpha_i \varepsilon_{t-i} + \gamma_i |\varepsilon_{t-i}|}{\sigma_{t-i}} + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2) \tag{14}$$

Using the definition of the shock given in equation (3), the process can be rewritten as

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} (\alpha_i z_{t-i} + \gamma_i |z_{t-i}|) + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2) \tag{15}$$

where γ_i captures the leverage effect. The impact of a positive shock to the logarithm of the conditional variance is (α_i + γ_i) while the impact of a negative shock is (α_i - γ_i). As we are dealing with negated return series, we expect the term γ_i to be positive, i.e., we expect there to be a leverage effect in our series. Note that we do not need to impose any restrictions on ω, α_i and β_j as we are modelling the logarithm of the conditional variance.

The one day ahead forecast of the conditional variance of the EGARCH(1,1) model is given by

$$\log(\sigma_{t+1}^2) = \omega + (\alpha z_t + \gamma |z_t|) + \beta \log(\sigma_t^2) \tag{16}$$

Realized GARCH(1,1)

The Realized GARCH model, introduced by Hansen et al. (2011), provides a framework for which the returns and the realized measure of volatility could be jointly modelled. The realized measure is estimated using high frequency intraday return data. The authors argue that realized measures of volatility provide more information about the current level of volatility than squared returns do, which in turn can be useful for modelling and forecasting purposes. The structure of the Realized GARCH (*p, q*) is as follows

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i \log(\zeta_{t-i}) + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2) \tag{17}$$

$$\log(\zeta_t) = \xi + \varphi \log(\sigma_t^2) + \tau(z_t) + u_t \tag{18}$$

$$\tau(z_t) = \eta_1 z_t + \eta_2 (z_t^2 - 1) \tag{19}$$

where ζ_t is the realized measure of volatility and $u_t \sim N(0, \sigma_u^2)$. Equation (18) provides a link between the observed realized measure to the latent volatility, and is called the measurement equation. The measurement equation can adjust the for bias caused by e.g., non-trading hours, as it is not required that ζ_t is an unbiased measure of σ_t^2. Furthermore, equation (19) is the leverage function of

the model that enables an asymmetric response in volatility to shocks. As was the case for the EGARCH model, we do not need to impose any restrictions on the model as we are modelling the logarithm of the conditional variance.

The one day ahead forecast of the conditional variance of the Realized GARCH(1,1) model is given by

$$\log(\sigma_{t+1}^2) = \omega + \alpha \log(\zeta_t) + \beta \log(\sigma_t^2) \tag{20}$$

Realized measures of volatility

To employ the Realized GARCH model we must specify the realized measure of volatility defined in equation (18). The most common measure of realized volatility is the realized variance, defined by

$$RV_t = \sum_{j=1}^{n} r_{j,t}^2 \tag{21}$$

where $r_{j,t}$ is an intraday return vector with $j = 1,\ldots,n$ on the t-th day. However, as we wish to utilize all intraday information available in the form of tick data, this measure might not be suitable. Zhou (1996) was among the first to show that the realized variance tends to be a biased and inconsistent estimator of the quadratic variation at this frequency as it is susceptible to microstructure noise. To combat this issue, we will instead employ the realized kernel estimator of Barndorff-Nielsen et al. (2009). This measure combines the intraday volatility estimation with a kernel weighting function, making it robust to microstructure noise. The realized kernel estimator is defined as:

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h \tag{22}$$

$$\gamma_h = \sum_{j=|h|+1}^{n} r_{j,t}\, r_{j-|h|,t} \tag{23}$$

where $K(X)$ is a kernel weighting function and the intraday vector $r_{j,t}$ consists of logarithmic returns calculated from mid-quote prices. We will employ the Parzen kernel function, given by

$$k(x)\begin{cases} 1 - 6x^2 + 6x^3 & \text{if } 0 \le x \le 1/2 \\ 2(1-x)^3 & \text{if } 1/2 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases} \tag{24}$$

A desired property of this kernel is that it satisfies the smoothness condition $k'(0) = k'(1) = 0$ and is guaranteed to produce non-negative values. The authors note that it is necessary to increase the bandwidth H with the sample size in order to consistently estimate the quadratic variation. In this thesis, rather arbitrarily, we choose the bandwidth $H = 100$ for all t. Barndorff-Nielsen et al. (2009) provide a method for which one could estimate the optimal bandwidth, however, implementing this approach for all t is beyond the scope of this thesis.

Prior to the estimation of the realized kernel, a cleaning algorithm is implemented to clear the data from spurious entries. Barndorff-Nielsen et al. (2009) argue that it is paramount to employ a cleaning approach when estimating volatility from tick data as a few spurious outliers can severely influence the realized kernel estimator. Specifically, the cleaning approach of this paper consists of removing every mid-quote price that deviates more than 10 mean absolute deviations from a rolling centered median of 50 observations. Furthermore, all entries for which the bid or ask quote is equal to zero are deleted.

Conditional distribution

The last moment required to fully specify the GARCH models is to determine the distributional assumption of the standardized residual - the approximation of z_t of equation (3). As mentioned previously in this section, the parameters of the GARCH models are estimated using maximum likelihood functions, however, the exact form of which depends on the parametric structure of the distribution of the innovations. We refer the reader to section 6.2 in Appendix for the structure of the different maximum likelihood functions. Thus, to accurately model the volatility process the distributional assumption is of importance. To identify the correct process for z_t is rather difficult as it is an unobservable process of the return series, but the depictions of the unconditional series presented in section 2.3 should give a general idea about the distributional suitability.

As mentioned in the introduction of this paper, we will consider three different distributions for the standardized residuals: the normal distribution, the student's t-distribution and the skewed student's t-distribution. These will be applied to all models. The density functions of the assumed distributions are presented in section 6.1 in Appendix. Note that the distributions of the standardized residuals are scaled to have a mean equal to zero and unit variance to replicate the behavior of z_t. This implies, for example, that the standardized student's t-distribution is scaled with $\sqrt{(v-2)/v}$, where *v* denotes the number of degrees of freedom.

3.5 Risk measures

For the explanations of the risk measures we follow the reasoning of McNeil et al. (2015, p. 64-72). The first risk measure that we will consider is Value-at-risk (VaR). VaR_q refers to the *q*-quintile of the loss distribution. It is defined as the smallest loss x_q such that the probability of observing a future loss $X_{t+1} > x_q$ is 1 – *q*:

$$VaR_{q,t} = \inf\{x_q \in \mathbb{R} : P(X_{t+1} > x_q) \leq 1 - q\} \tag{24}$$

Following the definition of equation (24), if a random variable *X*, with location μ and scale σ, follows some continuous location-scale distribution *F*, then VaR of *X* is defined as:

$$VaR_q(X) = \mu + \sigma F^{-1}(q) \tag{25}$$

where F refers to the standardized cumulative distribution that is scaled to have zero mean and unit variance. Assuming that the loss process X_t is described by equation (2) and (3), i.e.,

$$X_t = \mu_t + \sigma_t z_t$$

where the innovations are i.i.d. with zero mean and unit variance, the VaR of *X* at time *t* can be defined as:

$$VaR_{q,t}(X_t) = \mu_t + \sigma_t VaR_q(z) \tag{26}$$

Note that the quintile $VaR_q(z)$ is independent of *t* as we assume that the innovations are i.i.d., i.e., the probability distribution is the same for all z_t.

The second risk measure of interest is Expected shortfall (ES). ES_q refers to the expected value of the loss *X* conditional on the loss surpassing VaR_q:

$$ES_q(X) = E(X | X > VaR_q) \tag{27}$$

Again, if a random variable *X*, with location μ and scale σ, follows some continuous location-scale distribution *F*, then ES of *X* is defined as:

$$ES_q(X) = \mu + \sigma \frac{f(F^{-1}(q))}{1-q} \tag{28}$$

where f refers to the density and F is the standardized cumulative distribution scaled to have zero mean and unit variance. This density can be estimated using integrals. Analogous to the VaR definition, for the loss process X_t described by equation (2) and (3), the expected shortfall of X at time t can be defined as:

$$ES_{q,t}(X_t) = \mu_t + \sigma_t ES_q(z) \tag{29}$$

as the innovations are i.i.d. with zero mean and unit variance.

Forecasting

We try to capture the process described in equation (3) using variations of the ARMA-GARCH model. Applying the definitions described above into our modelling framework, we can predict VaR and ES by:

$$VaR_{q,t}(X_{t+1}) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} VaR_q(\hat{z}) \tag{30}$$

$$ES_{q,t}(X_{t+1}) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} ES_q(\hat{z}) \tag{31}$$

where $\hat{z}$ represents the standardized residuals - the sample counterparts of the innovations – following the standardized version of either the normal distribution, student's t-distribution or skewed student's t-distribution. $\hat{\mu}_{t+1}$ is the prediction of the conditional mean using the ARMA structure described in equation (7), and $\hat{\sigma}_{t+1}$ is the predicted conditional volatility of the different GARCH models, given by equation (10), (13), (16) and (20). Note that refitting the model every window implies that the parametric structure of the assumed distribution will be re-approximated as well. As the normal distribution is only characterized by the mean and standard deviation, the quintiles and tail densities of the standard normal distribution will be the same for all t. For the two other distributions, however, the exact form is determined by the parametric structure of the standardized residuals of each window, dictated by the degrees of freedom and skewness parameter.

We will consider five different quintiles for our forecasts of VaR and ES in this thesis: $q \in \{0.95, 0.975, 0.99, 0.995, 0.999\}$

3.6 Extreme Value Theory

The primary concern of this thesis is the events that are observed very rarely, i.e., extreme losses far out in the tail of the loss distribution. Whereas traditional parametric methods are often inadequate in capturing events of such nature, one method that has been developed specifically to model these extreme events is Extreme Value Theory (EVT). EVT only focuses on the tail of the distribution by relying on a subsample of large losses for its modelling purposes, which stands in contrast to traditional modelling approaches that focuses on the conditional moments of the entire distribution. As the tail of the empirical distribution generally differs from the tail imposed by the parametric distribution, modelling the tail separately may accommodate us in capturing the tail behavior more accurately. In order to identify the large past losses, two methods are usually applied - the block maxima approach and peak-over-threshold (POT) approach. We will focus on the latter. Following the reasoning McNeil et al. (2015, p. 146-154), large losses are defined as all observations that exceed a certain threshold, u. If we let $X = \{x_1, x_2, \ldots, x_T\}$ denote a series of i.i.d. losses that follows the distribution F and define y as the magnitude of the losses that exceed the chosen threshold u, the conditional cumulative probability function F_u is defined as:

$$F_u(y) = \Pr(X \leq u + y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)} \tag{32}$$

where $x = y + u$ with $y > 0$. The theorem by Balkema & De Haan (1974) state that, if u is sufficiently high, $F_u(y)$ will converge to the generalized Pareto distribution (GPD), i.e., $F_u(y) \approx G_{\xi,\beta}(y)$. The GPD is defined as:

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\beta}}, & \text{if } \xi = 0 \end{cases} \tag{33}$$

where ξ and β denote the shape parameter and scale parameter, respectively. A positive value of the shape parameter indicates heavy tails while a negative value indicates a short-tailed distribution. Furthermore, combining (32) and (33) yields:

$$\begin{aligned} F(x) &= \big(1 - F(u)\big) G_{\xi,\beta}(y) + F(u) \\ &\approx 1 - \frac{k}{T}\left(1 + \frac{\xi(x-u)}{\beta}\right)^{-\frac{1}{\xi}} \end{aligned} \tag{34}$$

where k is the number of exceedances over the threshold and T is the sample size. The q-quantile of $F(x)$, or VaR_q^{EVT}, can then be estimated by:

$$VaR_q^{EVT} = x_q = u + \frac{\beta}{\xi}\left[\left(\frac{T(1-q)}{k}\right)^{-\xi} - 1\right] \tag{35}$$

Following the definition of ES given in equation (27), the ES_q of X is estimated as:

$$ES_q^{EVT} = E\big(X \big| X > x_q\big) = \frac{x_q}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \tag{36}$$

There are many suggestions in the literature on how to find the optimal threshold choice. A common approach is to use the mean excess plot. For our purposes, however, this method is not feasible as it would require ocular inspections of our estimates of every rolling window. McNeil & Frey (2000) keep the number of exceedances over the threshold fixed, which implies that a threshold at the $(k + 1)$th order statistic is used for all windows. They suggest that $k = 100$, corresponding to the 90th percentile of the distribution, provides a reasonable choice after assessing the bias and MSE for different values of k through simulation. Likewise, DuMouchel (1983) suggests that the 90th percentile provides a balanced trade-off between having a sufficient number of observations to reliably estimate ξ and the theoretical need to describe the behavior of $F(x)$. We will follow these recommendations and use $k = 100$.

The two-stage approach of McNeil & Frey (2000) that is applied to combine EVT with the GARCH framework is as follows: For each window, we fit the GARCH models of section 3.4 along with an ARMA structure for the conditional mean described in sections 3.3. We then extract the standardized residuals of each model and, using a threshold corresponding to the $(k + 1)$th order statistic, fit a GPD to the k exceedances using the method of maximum likelihood. The parameters of the GPD will then be utilized to estimate VaR and ES according to equation (35) and (36), respectively. The one step ahead forecasts is finally obtained by utilizing the forecast of the conditional mean from the ARMA structure described in equation (7), and the predicted volatility of the different GARCH models, given by equation (10), (13), (16) and (20), such that:

$$VaR_{q,t}(X_{t+1}) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} VaR_q^{EVT} \tag{37}$$

$$ES_{q,t}(X_{t+1}) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} ES_q^{EVT} \tag{38}$$

3.7 Backtesting

The objective of backtesting is to evaluate the forecasting performance of our models. To backtest VaR we will perform the three coverage tests of Christoffersen (1998). These tests allows us to assess both the frequency and independence of VaR violations. To backtest ES, we will perform two tests proposed by Acerbi & Szekely (2014) which aims to evaluate whether the right tail of the loss distribution is accurately estimated.

Backtesting Value-at-Risk

In this section we will follow the reasoning and notation of Christoffersen (2011, p. 301-306). If our model is accurately capturing the loss process, the probability of seeing an exceedance of VaR would be (1 - q)·100% at each point in time. The exceedances would be unpredictable and occur independently over time. We can define an indicator variable that takes the value 1 if there is an exceedance and 0 otherwise as:

$$I_{t+1} = \begin{cases} 1 \text{ if } X_{t+1} > VaR_q^t\ (X_{t+1}) \\ 0 \text{ if } X_{t+1} \leq VaR_q^t\ (X_{t+1}) \end{cases} \tag{39}$$

The general null hypothesis of the tests is that I_{t+1} are i.i.d. Bernoulli variables. A Bernoulli variable that takes the value 1 with probability p and the value 0 with probability (1 – p) could be written as:

$$f(I_{t+1}; p) = (1-p)^{1-I_{t+1}} p^{I_{t+1}} \tag{40}$$

The first test of Christoffersen (2011, p. 302-304), called the *unconditional* coverage test, assesses whether the observed number of exceedances of VaR is in accordance with the expected number of exceedances given the choice of quintile q. If we let p denote the fraction (1 - q), and π denote the fraction of exceedances of our risk models, the null hypothesis of this test is $H_0: p = \pi$. If we further let T denote the full out-of-sample size, T_1 the number of exceedances and T_0 the number of observations below VaR, the likelihood function under the null could be written as:

$$L(p) = \prod_{t=1}^{T} (1-p)^{1-I_{t+1}} p^{I_{t+1}} = (1-p)^{T_0} p^{T_1} \tag{41}$$

The maximum likelihood estimate of the sample counterpart of π, estimated as $\hat{\pi} = \frac{T_1}{T}$, is given by:

$$L(\hat{\pi}) = \left(1 - \frac{T_1}{T}\right)^{T_0} \left(\frac{T_1}{T}\right)^{T_1} \tag{42}$$

The null hypothesis can be assessed using a likelihood ratio test:

$$LR_{uc} = -2\ln\left[L(p)/L(\hat{\pi})\right] \tag{43}$$

which is asymptotically chi-squared distributed with one degree of freedom.

The second test of Christoffersen (2011, p. 304-306) is called the *independence* test. The idea is to assess whether the exceedances are independent of each other, which is crucial for risk management purposes as multiple violations in a short period of time could imply an increased risk of insolvency. We assume that there is a dependency and that the conditional probabilities of transitions from one state to another can be described in the following Markov sequence:

$$\Pi_1 = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} = \begin{bmatrix} \Pr(I_{t+1} = 0 | I_t = 0) & \Pr(I_{t+1} = 1 | I_t = 0) \\ \Pr(I_{t+1} = 0 | I_t = 1) & \Pr(I_{t+1} = 1 | I_t = 1) \end{bmatrix} \tag{44}$$

where the first term in the subscript denotes the current state and the latter denotes the upcoming state, e.g., π_{01} refers to the probability of an exceedance tomorrow conditional on no exceedance today. The likelihood function of the Markov process can be defined as:

$$L(\Pi_1) = (1-\pi_{01})^{T_{00}}{\pi_{01}}^{T_{01}}(1-\pi_{11})^{T_{10}}{\pi_{11}}^{T_{11}} \tag{45}$$

T_{ij}, i,j = 0,1 denotes the number of observations in the out-of-sample where a j is preceded by an i. The maximum likelihood estimates are defined as:

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}} \tag{46}$$

$$\hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}} \tag{47}$$

where $\hat{\pi}_{00} = 1-\hat{\pi}_{01}$ and $\hat{\pi}_{10} = 1-\hat{\pi}_{11}$ as the probabilities have to sum to 1. The matrix of conditional probabilities of transitions can thus be described as:

$$\hat{\Pi}_1 \equiv \begin{bmatrix} \hat{\pi}_{00} & \hat{\pi}_{01} \\ \hat{\pi}_{10} & \hat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} 1-\hat{\pi}_{01} & \hat{\pi}_{01} \\ 1-\hat{\pi}_{11} & \hat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{bmatrix} \tag{48}$$

If there is a dependence we would expect a difference between the conditional state probabilities π_{01} and π_{11}, i.e., the probability of observing a violation tomorrow differs depending on whether we observe an exceedance today or not. In contrast, if the violations are independent, these conditional probabilities are expected to be the same. Therefore, the transition matrix under independence is:

$$\hat{\Pi} = \begin{bmatrix} 1-\hat{\pi} & \hat{\pi} \\ 1-\hat{\pi} & \hat{\pi} \end{bmatrix} \tag{49}$$

The null hypothesis H_0: $\pi_{01} = \pi_{11}$ can be tested using the likelihood ratio test:

$$LR_{ind} = -2\ln\left[L(\hat{\pi})/L(\hat{\Pi}_1)\right] \tag{50}$$

where $L(\hat{\pi})$ is the likelihood function from the unconditional coverage test. Again, this test statistic is asymptotically chi-squared distributed with one degree of freedom.

The third and last test that will be performed to assess the performance of our VaR forecasts is the *conditional* coverage test. It is a joint test of the unconditional coverage test and the independence test. It is defined as:

$$LR_{cc} = -2\ln\left[L(p)/L(\hat{\Pi}_1)\right] \tag{51}$$

which is asymptotically chi-squared distributed with two degrees of freedom.

Following the recommendations of Christoffersen (2011, p. 303), we will perform Monte-Carlo simulations to obtain the p-values for all of the backtests of VaR. The rationale behind this recommendation is that testing under the chi-squared distribution may give unreliable results if the number of exceedances is not sufficiently large. The simulation is performed by generating 999 samples of i.i.d. Bernoulli(p) variables, where each sample size corresponds to the out-of-sample size of our currency pairs. We then estimate the simulated test statistic of the samples, denoted by $\{\widetilde{LR}(i)\}_{i=1}^{999}$. Lastly, the p-values are obtained by:

$$\mathrm{p} = \frac{1}{1000}\left[1 + \sum_{i=1}^{999} \mathbb{1}_{\{\widetilde{LR}(i) > LR\}}\right] \tag{52}$$

where $\mathbb{1}$ denotes an indicator variable that takes the value 1 if the simulated test statistic exceeds the test statistic obtained from the actual data, and 0 otherwise.

Backtesting Expected Shortfall

In contrast to VaR, backtesting ES requires an evaluation of the whole right tail of the loss distribution. The general idea when backtesting ES is to examine the difference between the realized losses that exceeded the quintile of interest, or VaR_q^t, with our forecasts of ES. As ES is defined as the expected loss conditional on the loss surpassing VaR, the difference between these losses and our estimates of ES should preferably be small in the aggregate.

We will consider the first two of the three tests proposed by Acerbi & Szekely (2014) for the backtesting of ES. Some of the advantages of these tests are that they do not impose any distributional assumptions for the returns and can be directly evaluated through simulation. The structure of the tests is as follows: We assume that the losses X_t follow an unknown distribution F_t, which is forecasted using the predictive conditional distribution P_t. VaR_t^F and ES_t^F denotes the true measures, whereas VaR_t^P and ES_t^P denotes the estimated risk measures. The null hypothesis of the two tests is the same, defined as:

$$H_0: P_t = F_t \text{ for all } t$$

The test statistic, Z_1, of the first test is defined as:

$$Z_1 = -\frac{1}{N_T}\sum_{t=1}^{T}\frac{I_t X_t}{ES_{q,t}^P} + 1 \tag{53}$$

where $N_T = \sum_{t=1}^{T} I_t > 1$ with $I_t = \mathbb{1}_{\{X_t > VaR_q\}}$, i.e., an indicator of VaR exceedances, and T corresponds to the length of the out-of-sample period. The alternative hypothesis of the test is:

$$H_1: ES_t^P \leq ES_t^F \text{ for all t and } < \text{for some } t$$

$$VaR_t^P = VaR_t^F \text{ for all } t$$

We can observe that the alternative hypothesis is a one-sided test for underestimation of ES. Acerbi & Szekely (2014) note that this is in line with the Basel framework for VaR as it is only excesses of VaR exceptions that signal a problem. This further implies that models that overestimate ES are favored using this backtesting methodology. Furthermore, it can be seen that the test statistic is the average of the VaR exceedances over the exceedances themselves, making the test insensitive to excessive numbers of VaR violations. Furthermore, the expected value of Z_1 under H_0 is zero and negative under the H_1, implying that negative values indicate an underestimation of ES.

The second test of Acerbi & Szekely (2014) is, in contrast to the former, also sensitive to the expected number of VaR violations. This means that it also requires the quintile, VaR_q, to be correctly estimated. The test jointly evaluates both the frequency and size of the VaR exceedances, and is defined as:

$$Z_2 = -\frac{1}{T(1-q)}\sum_{t=1}^{T}\frac{I_t X_t}{ES_{q,t}^P} + 1 \tag{54}$$

The alternative hypothesis of the test is:

$$H_1: ES_t^P \leq ES_t^F \text{ for all t and } < \text{for some } t$$

$$VaR_t^P \leq VaR_t^F \text{ for all } t$$

Again, the expected value of Z_1 under H_0 is zero and negative under the H_1. The difference between Z_1 and Z_2 can be found in the denominator: N_T of the former test is replaced with $T(1-q)$ in the latter. $T(1-q)$ corresponds to the expected number of VaR violations given the quintile q, thus making the test sensitive to both magnitude and frequency of VaR violations.

As the distributions of the test statistics are unknown, we will perform Monte-Carlo simulations to obtain the p-values for our models. Using a similar simulation methodology to that of Acerbi & Szekely (2014), which also bears some resemblance to the simulation approach of Christoffersen (2011, p. 303), the p-values of the tests will be obtained as follows:

1. Simulate M number of losses $\{\tilde{X}_t^i\}_{i=1}^M$ for each t = 1,…,T, using the same predictive conditional distribution P_t that was used to forecast VaR and ES.

2. Estimate Z_1^i and Z_2^i using $\{\tilde{X}_t^i\}_{i=1}^T$ for each i = 1,…,M.

3. Estimate the p-values: $\mathrm{p}_{Z_1} = \frac{\sum_{i=1}^M \mathbb{1}_{\{Z_1^i < Z_1\}}}{M}$ and $\mathrm{p}_{Z_2} = \frac{\sum_{i=1}^M \mathbb{1}_{\{Z_2^i < Z_2\}}}{M}$

where $\mathbb{1}$ denotes an indicator variable that takes the value 1 if the test statistic obtained from the actual data exceeds the simulated test statistic, and 0 otherwise. We will use M = 20,000 when we simulate the losses for each t in the out-of-sample.

3.8 Implementation

All estimations were performed using the language `R`. The `rugarch` package of Ghalanos (2022) was used to fit the GARCH models and produce one day ahead forecasts. The package `forecast` of Hyndman et al. (2022) was used for the ARMA selection of the conditional mean. The package `tea` of Ossberger (2020) was used to fit a GPD to the standardized residuals and obtain the relevant parameters for the estimation of VaR_q^{EVT} and ES_q^{EVT}. The realized kernel estimator was computed using the `highfrequency` package of Boudt et al. (2022). Lastly, all plots presented throughout this thesis were produced using `ggplot2` of Wickham et al. (2022).

4. Results

As each of the five GARCH models presented in section 3.4 will be employed under three different distributional assumptions, where we fit a GPD to the standardized residuals of each variation, we will have a total of 30 models under evaluation. For dispositional purposes, we will use acronyms to describe these models. The upper-case letter denotes the GARCH model, and the subscript denotes the assumed distribution of the innovations. The superscript "EVT" indicates that the Extreme Value Theory of section 3.6 have been applied to the model. The acronyms can be found in table 3 below.

Acronym	Model description
S_n	Standard GARCH(1,1) with normally distributed standardized residuals
S_n^{EVT}	Standard GARCH(1,1) where EVT is applied to normally distributed standardized residuals
S_t	Standard GARCH(1,1) with t-distributed standardized residuals
S_t^{EVT}	Standard GARCH(1,1) where EVT is applied to t-distributed standardized residuals
S_{st}	Standard GARCH(1,1) with skewed t-distributed standardized residuals
S_{st}^{EVT}	Standard GARCH(1,1) where EVT is applied to skewed t-distributed standardized
I_n	IGARCH(1,1) with normally distributed standardized residuals
I_n^{EVT}	IGARCH(1,1) where EVT is applied to normally distributed standardized residuals
I_t	IGARCH(1,1) with t-distributed standardized residuals
I_t^{EVT}	IGARCH(1,1) where EVT is applied to t-distributed standardized residuals
I_{st}	IGARCH(1,1) with skewed t-distributed standardized residuals
I_{st}^{EVT}	IGARCH(1,1) where EVT is applied to skewed t-distributed standardized residuals
G_n	GJR-GARCH(1,1) with normally distributed standardized residuals
G_n^{EVT}	GJR-GARCH(1,1) where EVT is applied to normally distributed standardized residuals
G_t	GJR-GARCH (1,1) with t-distributed standardized residuals
G_t^{EVT}	GJR-GARCH(1,1) where EVT is applied to t-distributed standardized residuals
G_{st}	GJR-GARCH (1,1) with skewed t-distributed standardized residuals
G_{st}^{EVT}	GJR-GARCH(1,1) where EVT is applied to skewed t-distributed standardized residuals
E_n	EGARCH(1,1) with normally distributed standardized residuals
E_n^{EVT}	EGARCH(1,1) where EVT is applied to normally distributed standardized residuals
E_t	EGARCH (1,1) with t-distributed standardized residuals
E_t^{EVT}	EGARCH(1,1) where EVT is applied to t-distributed standardized residuals
E_{st}	EGARCH (1,1) with skewed t-distributed standardized residuals
E_{st}^{EVT}	EGARCH(1,1) where EVT is applied to skewed t-distributed standardized residuals
R_n	Realized GARCH(1,1) with normally distributed standardized residuals
R_n^{EVT}	Realized GARCH(1,1) where EVT is applied to normally distributed standardized residuals
R_t	Realized GARCH (1,1) with t-distributed standardized residuals
R_t^{EVT}	Realized GARCH(1,1) where EVT is applied to t-distributed standardized residuals
R_{st}	Realized GARCH (1,1) with skewed t-distributed standardized residuals
R_{st}^{EVT}	Realized GARCH(1,1) where EVT is applied to skewed t-distributed standardized residuals

Table 3: The acronyms of each model.

Recall that each of these models are fitted with an ARMA structure for the conditional mean following the algorithm described in section 3.3. This algorithm detected 24 structural changes in the mean process for EUR/USD, 63 for USD/JPY, 52 for GBP/USD, 23 for AUD/USD and 47 for USD/CAD.

To start off our analysis, we will look at the results of the VaR backtesting procedures. Table 4 reveals the number of Value-At-Risk exceedances of each model for each quintile and currency pair, which can be compared to the expected number of exceedances. In table 4 we observe that it is mainly the distributional assumption of the innovation process that differentiates the models from one another. For $q = 0.95$, we observe that the number of VaR violations of models assuming a normal

FX Pair	Exp	S_n	S_n^{EVT}	S_t	S_t^{EVT}	S_{st}	S_{st}^{EVT}	I_n	I_n^{EVT}	I_t	I_t^{EVT}	I_{st}	I_{st}^{EVT}	G_n	G_n^{EVT}	G_t	G_t^{EVT}	G_{st}	G_{st}^{EVT}	E_n	E_n^{EVT}	E_t	E_t^{EVT}	E_{st}	E_{st}^{EVT}	R_n	R_n^{EVT}	R_t	R_t^{EVT}	R_{st}	R_{st}^{EVT}
q = 0.95																															
EUR/USD	281	291	284	316	286	310	285	280	293	306	296	306	291	284	283	307	287	301	287	264	270	291	280	289	277	277	277	291	278	290	278
USD/JPY	282	300	297	319	295	315	292	286	301	314	294	302	294	316	309	335	307	328	306	319	314	327	296	315	295	286	283	319	286	302	286
GPB/USD	281	281	283	292	271	295	269	267	280	284	273	293	276	279	285	295	278	298	276	284	292	302	284	301	290	325	340	306	295	310	301
AUD/USD	269	302	278	322	285	296	284	295	280	315	285	286	286	303	278	320	280	293	279	296	273	313	282	287	280	291	275	307	261	269	258
USD/CAD	273	268	269	294	272	299	266	260	272	289	272	291	271	261	268	294	267	303	270	263	266	288	271	292	274	278	278	293	275	298	277
q = 0.975																															
EUR/USD	141	160	133	151	139	149	140	150	142	144	146	142	148	156	137	156	138	151	140	154	134	154	143	147	140	161	132	148	129	153	132
USD/JPY	141	177	142	164	145	154	145	173	149	156	149	153	150	198	154	181	155	164	149	190	149	172	142	157	139	175	148	165	151	159	154
GPB/USD	140	173	153	156	139	154	138	167	149	150	144	149	143	171	146	158	135	156	137	179	157	158	146	158	153	206	187	147	145	149	145
AUD/USD	134	169	129	154	129	138	131	162	127	149	132	131	131	169	139	165	136	147	138	178	139	162	136	146	138	185	141	173	135	141	136
USD/CAD	136	155	128	143	135	150	130	151	137	140	140	149	141	161	133	152	135	151	131	161	133	146	137	151	136	164	143	153	141	157	145
q = 0.99																															
EUR/USD	56	72	53	55	53	55	52	70	58	52	59	53	59	76	52	57	53	54	52	76	49	51	48	51	48	67	46	49	46	47	46
USD/JPY	56	98	53	54	50	50	51	100	57	52	49	44	51	113	66	62	57	52	57	107	61	66	60	54	57	112	68	77	62	64	61
GPB/USD	56	98	62	63	63	67	62	92	63	57	60	60	61	94	66	68	62	66	62	98	62	66	67	68	68	111	78	64	66	63	63
AUD/USD	54	91	54	73	53	55	54	88	57	64	55	58	56	95	57	70	56	59	56	102	60	73	55	60	55	98	59	69	53	60	55
USD/CAD	55	87	57	54	55	55	55	79	60	52	57	53	58	85	59	57	58	58	58	91	62	55	55	57	59	92	58	49	53	56	54
q = 0.995																															
EUR/USD	28	51	28	27	29	29	30	48	29	26	30	24	31	49	31	29	32	29	31	45	31	29	31	28	32	41	26	28	26	29	27
USD/JPY	28	63	29	28	28	26	27	59	31	24	29	23	30	68	34	34	31	31	31	70	36	30	31	29	30	83	35	33	37	30	37
GPB/USD	28	62	36	33	32	31	33	62	37	31	32	34	33	65	35	32	33	29	32	61	32	31	29	32	30	76	41	36	34	33	34
AUD/USD	27	59	25	39	26	30	28	62	30	32	30	24	30	63	28	36	26	27	26	64	28	37	27	25	26	63	27	33	24	26	25
USD/CAD	27	56	28	20	26	27	26	54	30	19	29	20	28	55	33	25	34	27	32	59	36	30	37	29	38	53	33	23	28	25	29
q = 0.999																															
EUR/USD	6	20	11	9	10	8	10	21	12	8	11	8	11	20	10	7	11	7	11	19	10	5	10	5	10	15	10	6	11	6	11
USD/JPY	6	29	14	11	13	6	12	27	13	9	12	5	12	32	13	13	14	8	14	34	14	11	11	6	11	34	10	9	10	6	10
GPB/USD	6	35	15	9	16	9	15	36	14	8	14	9	13	33	14	8	14	9	13	31	16	9	16	10	16	41	8	5	8	5	8
AUD/USD	5	25	7	8	8	6	8	21	8	6	7	6	7	22	9	7	9	6	9	21	9	7	9	8	9	26	6	6	6	6	6
USD/CAD	5	20	8	3	8	2	8	20	8	1	8	1	8	21	8	2	7	2	7	24	9	3	9	5	9	21	12	4	13	6	14

Table 4: The number of Value-At-Risk exceedances of each model. The column "Exp" denotes the expected number of exceedances.

distribution tend to be somewhat closer to the expected number of exceedances than models assuming a t-distribution or a skewed t-distribution. Models assuming the latter distributions tend to underestimate VaR at this quintile to a higher extent, as they consistently produce an excessive number of VaR violations. On the other hand, when EVT is applied to these models, the numbers appear to be much closer to those that are expected.

Moving further out to higher quintiles of the loss distribution, there appears to be a growing divergence between models assuming a normal distribution and models assuming heavy tailed distributions. Models assuming a normal distribution appear to vastly underestimate VaR for higher quintiles, more so the further out in the loss distribution we get. Applying EVT to these models seem to appropriately adjust VaR upwards, resulting in numbers closer to those that are expected. Models assuming a t-distribution or a skewed t-distribution seem to fare better further out in the loss distribution as the violations are more in line with the expected numbers. For $q \leq 0.99$, the EVT augmentation tends to slightly modify VaR upwards for these models, often resulting in numbers closer to those that are expected. This is most apparent for the lower quintiles. For the highest quintiles, however, the opposite appears to be true as the EVT augmentations produces more violations than the parent distributions. We observe that the violations of models assuming either a t-distribution or a skewed t-distribution tend to be quite similar for most data sets, although the latter produces slightly fewer in general. It is, however, not clear whether any GARCH-type model alone is yielding more accurate numbers than the others. To quantify whether the VaR violations of the models are in accordance with the expected number of exceedances, we will examine the results of the unconditional coverage test in table 5 below.

FX Pair	S_n	S_n^{EVT}	S_t	S_t^{EVT}	S_{st}	S_{st}^{EVT}	I_n	I_n^{EVT}	I_t	I_t^{EVT}	I_{st}	I_{st}^{EVT}	G_n	G_n^{EVT}	G_t	G_t^{EVT}	G_{st}	G_{st}^{EVT}	E_n	E_n^{EVT}	E_t	E_t^{EVT}	E_{st}	E_{st}^{EVT}	R_n	R_n^{EVT}	R_t	R_t^{EVT}	R_{st}	R_{st}^{EVT}
$q = 0.95$																														
EUR/USD	.553	.855	.037	.737	.086	.795	.931	.504	.131	.344	.152	.531	.864	.890	.129	.711	.227	.705	.286	.510	.555	.913	.615	.754	.783	.787	.533	.825	.595	.825
USD/JPY	.249	.328	.021	.415	.038	.509	.774	.240	.035	.452	.199	.471	.033	.087	.002	.122	.006	.119	.021	.041	.009	.377	.035	.411	.767	.909	.029	.792	.181	.793
GPB/USD	.950	.866	.480	.554	.360	.465	.371	.959	.825	.610	.416	.761	.897	.799	.369	.859	.295	.744	.829	.482	.202	.837	.216	.570	.010	.003	.111	.403	.076	.222
AUD/USD	.042	.545	.001	.311	.115	.343	.100	.476	.005	.318	.274	.244	.039	.554	.003	.455	.124	.548	.096	.782	.008	.375	.273	.475	.171	.669	.017	.632	.978	.504
USD/CAD	.769	.831	.166	.949	.110	.673	.424	.952	.309	.950	.233	.917	.445	.754	.171	.707	.064	.864	.531	.669	.334	.896	.222	.930	.731	.739	.198	.874	.112	.762
Rejections	1	0	3	0	1	0	0	0	2	0	0	0	2	0	2	0	1	0	1	1	2	0	1	0	1	1	2	0	0	0
$q = 0.975$																														
EUR/USD	.114	.490	.358	.871	.444	.934	.433	.883	.758	.637	.903	.539	.204	.725	.188	.808	.375	.931	.228	.532	.256	.850	.589	.939	.088	.464	.522	.284	.287	.479
USD/JPY	.004	.911	.060	.714	.283	.705	.006	.481	.214	.474	.299	.422	.001	.224	.004	.222	.046	.472	.001	.448	.005	.922	.169	.872	.003	.532	.047	.375	.123	.239
GPB/USD	.012	.284	.189	.893	.240	.838	.030	.442	.415	.714	.474	.801	.012	.605	.121	.644	.190	.769	.004	.151	.131	.615	.127	.273	.001	.001	.536	.645	.479	.652
AUD/USD	.005	.625	.095	.613	.732	.760	.018	.513	.227	.822	.736	.766	.006	.687	.007	.863	.281	.717	.001	.670	.020	.862	.339	.732	.001	.533	.003	.932	.540	.849
USD/CAD	.105	.487	.556	.897	.241	.582	.209	.933	.727	.739	.254	.677	.036	.757	.181	.901	.229	.626	.037	.771	.376	.936	.221	.972	.013	.530	.144	.638	.093	.441
Rejections	3	0	0	0	0	0	3	0	0	0	0	0	4	0	2	0	1	0	4	0	2	0	0	0	4	1	2	0	0	0
$q = 0.99$																														
EUR/USD	.039	.639	.839	.638	.830	.549	.061	.758	.533	.671	.631	.681	.012	.552	.893	.630	.747	.544	.011	.335	.453	.265	.457	.256	.159	.135	.321	.125	.193	.149
USD/JPY	.001	.640	.758	.391	.378	.470	.001	.893	.554	.327	.076	.469	.001	.184	.405	.891	.533	.901	.001	.505	.217	.621	.754	.891	.001	.129	.004	.423	.274	.503
GPB/USD	.001	.425	.353	.358	.123	.424	.001	.360	.900	.595	.582	.511	.001	.173	.117	.443	.180	.426	.001	.423	.201	.138	.114	.135	.001	.009	.299	.180	.353	.354
AUD/USD	.001	.947	.013	.892	.850	.930	.001	.639	.174	.839	.528	.718	.001	.651	.032	.722	.482	.725	.001	.375	.014	.847	.371	.845	.001	.463	.040	.884	.367	.851
USD/CAD	.001	.698	.894	.940	.894	.902	.002	.457	.677	.743	.785	.622	.001	.533	.737	.661	.614	.654	.001	.288	.938	.945	.739	.537	.001	.631	.409	.796	.830	.890
Rejections	5	0	1	0	0	0	4	0	0	0	0	0	5	0	0	0	0	0	5	0	1	0	0	0	4	1	2	0	0	0
$q = 0.995$																														
EUR/USD	.001	.921	.783	.867	.829	.722	.001	.862	.633	.717	.407	.564	.001	.566	.845	.473	.846	.563	.003	.521	.865	.596	.922	.500	.020	.591	.936	.645	.830	.772
USD/JPY	.001	.844	.918	.929	.633	.802	.001	.561	.386	.843	.305	.698	.001	.254	.258	.597	.572	.598	.001	.117	.713	.586	.844	.697	.001	.184	.347	.115	.715	.135
GPB/USD	.001	.141	.362	.443	.581	.341	.001	.099	.545	.465	.255	.342	.001	.193	.476	.356	.880	.429	.001	.465	.585	.856	.456	.695	.001	.023	.124	.245	.366	.268
AUD/USD	.001	.702	.032	.843	.483	.764	.001	.494	.343	.488	.554	.507	.001	.762	.079	.855	.928	.863	.001	.782	.054	.925	.695	.861	.001	.928	.264	.566	.853	.677
USD/CAD	.001	.862	.120	.780	.925	.787	.001	.554	.083	.691	.131	.834	.001	.260	.626	.246	.926	.341	.001	.107	.550	.077	.710	.058	.001	.239	.382	.832	.658	.691
Rejections	5	0	1	0	0	0	5	0	0	0	0	0	5	0	0	0	0	0	5	0	0	0	0	0	5	1	0	0	0	0
$q = 0.999$																														
EUR/USD	.001	.028	.126	.115	.286	.110	.001	.038	.319	.036	.298	.034	.001	.097	.552	.035	.571	.043	.001	.109	.669	.105	.681	.112	.003	.113	.859	.043	.837	.035
USD/JPY	.001	.008	.035	.006	.845	.028	.001	.008	.125	.033	.662	.033	.001	.006	.008	.006	.334	.004	.001	.008	.033	.052	.819	.041	.001	.109	.162	.102	.833	.102
GPB/USD	.001	.002	.137	.001	.148	.002	.001	.007	.306	.007	.134	.005	.001	.007	.300	.002	.156	.005	.001	.001	.145	.001	.106	.001	.001	.304	.710	.310	.689	.332
AUD/USD	.001	.405	.322	.269	.662	.291	.001	.308	.669	.372	.678	.391	.001	.146	.400	.157	.688	.132	.001	.153	.410	.164	.316	.160	.001	.699	.694	.670	.667	.673
USD/CAD	.001	.310	.214	.317	.082	.298	.001	.287	.015	.326	.013	.314	.001	.322	.073	.560	.079	.548	.001	.147	.204	.131	.826	.138	.001	.011	.393	.006	.654	.004
Rejections	5	3	1	2	0	2	5	3	1	3	1	3	5	2	1	3	0	3	5	2	1	1	0	2	5	1	0	2	0	2
Total Rejections	**19**	**3**	**6**	**2**	**1**	**2**	**17**	**3**	**3**	**3**	**1**	**3**	**21**	**2**	**5**	**3**	**2**	**3**	**20**	**3**	**6**	**1**	**1**	**2**	**20**	**5**	**6**	**2**	**0**	**2**

Table 5: The p-values of each model from the unconditional coverage test. The model is rejected if the p-value is less than 0.05.

In table 5 we can confirm many of the tendencies observed in table 4. Based on the p-values at $q = 0.95$, we see that models assuming a normal distribution tend to fare slightly better than those assuming a t-distribution or a skewed t-distribution. The difference is however quite marginal as seen by the numbers of rejections. At higher quintiles, we observe that practically all models assuming a normal distribution were rejected for most data sets. Conversely, models assuming either the t-distribution or the skewed counterpart performed well for all higher quintiles. Overall, the EVT extension seems to be beneficial for all quintiles for models assuming a normal distribution. For models already assuming a heavy tailed distribution for the innovation, the EVT extension seems to be helpful in adjusting VaR upwards for $q \leq 0.99$, oftentimes generating higher p-values and fewer numbers of rejections. This is particularly evident for $q = 0.95$. For the highest quintile, however, we observe that the EVT extension adversely affected the models assuming heavy tailed distributions. Overall, models with t-skewed residuals were rejected on fewer occasions than those assuming a t-distribution.

In general, all models assuming either a t-distribution or a skewed t-distribution and/or were augmented with EVT produced accurate results for the unconditional coverage test. It is not clear whether any GARCH-type model is to be preferred over the other, or if there is a general trend amongst these models. Nonetheless, it should be noted that the Realized GARCH(1,1) model with t-skewed innovations is the only model that was not rejected for any of the data sets for any quintile.

Next, we examine the models' capacities in producing independent VaR violations. The independence test of Christofferson (1998) is presented in table 6 below.

FX Pair	S_n	S_n^{EVT}	S_t	S_t^{EVT}	S_{st}	S_{st}^{EVT}	I_n	I_n^{EVT}	I_t	I_t^{EVT}	I_{st}	I_{st}^{EVT}	G_n	G_n^{EVT}	G_t	G_t^{EVT}	G_{st}	G_{st}^{EVT}	E_n	E_n^{EVT}	E_t	E_t^{EVT}	E_{st}	E_{st}^{EVT}	R_n	R_n^{EVT}	R_t	R_t^{EVT}	R_{st}	R_{st}^{EVT}
q = 0.95																														
EUR/USD	.105	.111	.123	.152	.135	.133	.234	.377	.158	.408	.194	.263	.187	.138	.312	.150	.245	.290	.164	.134	.216	.213	.309	.274	.065	.095	.148	.147	.076	.216
USD/JPY	.009	.070	.040	.017	.074	.015	.011	.074	.043	.072	.068	.072	.183	.263	.075	.175	.106	.088	.096	.168	.020	.019	.071	.027	.001	.008	.003	.053	.011	.055
GPB/USD	.303	.247	.090	.159	.160	.202	.143	.311	.090	.131	.076	.128	.217	.145	.157	.122	.234	.268	.353	.348	.499	.357	.497	.455	.078	.008	.076	.113	.064	.096
AUD/USD	.026	.008	.006	.003	.002	.002	.014	.008	.021	.001	.001	.004	.065	.077	.008	.059	.054	.051	.014	.004	.009	.004	.014	.020	.001	.001	.003	.006	.001	.001
USD/CAD	.973	.832	.843	.909	.409	.769	.711	.893	.750	.512	.533	.693	.663	.808	.817	.982	.831	.876	.939	.777	.790	.687	.909	.556	.580	.957	.458	.766	.670	.789
Rejections	2	1	2	2	1	2	2	1	2	1	1	1	0	0	1	0	0	0	1	1	2	2	1	2	2	3	2	1	2	1
q = 0.975																														
EUR/USD	.304	.331	.358	.435	.072	.239	.368	.282	.531	.170	.143	.095	.130	.425	.130	.239	.090	.254	.119	.072	.139	.078	.062	.073	.560	.631	.164	.298	.377	.329
USD/JPY	.222	.067	.188	.139	.067	.048	.106	.094	.144	.089	.230	.036	.060	.107	.070	.136	.207	.095	.075	.071	.045	.072	.071	.067	.074	.008	.043	.012	.036	.016
GPB/USD	.165	.217	.127	.126	.129	.111	.121	.086	.112	.085	.184	.080	.270	.324	.154	.097	.057	.067	.369	.474	.081	.184	.156	.211	.244	.100	.342	.303	.204	.309
AUD/USD	.002	.025	.005	.022	.003	.005	.001	.012	.001	.025	.003	.024	.035	.053	.068	.044	.068	.055	.039	.053	.016	.004	.006	.007	.001	.002	.001	.001	.002	.002
USD/CAD	.285	.608	.535	.722	.391	.632	.683	.741	.735	.828	.378	.523	.893	.717	.691	.863	.922	.957	.564	.358	.348	.741	.401	.731	.667	.703	.545	.757	.813	.658
Rejections	1	1	1	1	1	2	1	1	1	1	1	2	1	0	0	1	0	0	1	0	2	1	1	1	1	2	2	2	2	2
q = 0.99																														
EUR/USD	.476	.047	.810	.036	.807	.053	.377	.065	.753	.078	.785	.071	.043	.012	.084	.012	.053	.019	.014	.023	.040	.013	.041	.014	.229	.643	.708	.654	.683	.642
USD/JPY	.971	.795	.060	.037	.034	.040	.975	.055	.051	.043	.026	.048	.049	.014	.076	.007	.047	.008	.144	.011	.204	.006	.009	.006	.001	.009	.014	.002	.027	.018
GPB/USD	.030	.923	.927	.115	.974	.925	.045	.927	.843	.889	.878	.905	.392	.960	.263	.108	.204	.090	.587	.915	.026	.026	.029	.025	.033	.015	.010	.034	.018	.036
AUD/USD	.008	.052	.010	.051	.013	.020	.035	.076	.173	.072	.018	.016	.015	.062	.018	.020	.025	.020	.017	.109	.016	.018	.011	.022	.001	.002	.001	.001	.001	.001
USD/CAD	.861	.273	.386	.334	.322	.319	.070	.199	.430	.278	.373	.243	.844	.204	.265	.251	.260	.260	.893	.175	.843	.316	.858	.233	.860	.255	.534	.408	.324	.366
Rejections	2	1	1	2	2	2	2	0	0	1	2	2	3	2	1	3	2	3	2	2	3	4	4	4	3	3	3	3	3	3
q = 0.995																														
EUR/USD	.034	.518	.570	.449	.438	.398	.134	.428	.645	.409	.748	.344	.021	.355	.446	.098	.443	.081	.021	.107	.068	.081	.050	.103	.126	.028	.058	.030	.060	.052
USD/JPY	.977	.082	.504	.540	.645	.588	.087	.096	.785	.441	.829	.061	.138	.108	.228	.348	.346	.361	.129	.116	.415	.072	.057	.079	.018	.219	.257	.161	.422	.163
GPB/USD	.930	.116	.101	.096	.095	.099	.929	.129	.073	.090	.106	.117	.994	.112	.079	.101	.059	.109	.891	.083	.076	.077	.009	.011	.014	.111	.115	.133	.098	.119
AUD/USD	.101	.044	.012	.046	.086	.074	.112	.009	.009	.013	.012	.008	.126	.047	.015	.045	.012	.052	.124	.009	.116	.064	.007	.011	.001	.010	.004	.008	.007	.004
USD/CAD	.127	.447	.906	.588	.515	.587	.147	.371	.950	.392	.915	.460	.136	.199	.678	.217	.532	.261	.086	.160	.326	.159	.391	.146	.125	.210	.756	.482	.640	.405
Rejections	1	1	1	1	0	0	0	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	2	2	3	2	1	2	1	1
q = 0.999																														
EUR/USD	.007	.015	.062	.049	.114	.039	.003	.018	.123	.016	.117	.015	.009	.030	.209	.019	.229	.017	.008	.027	.476	.036	.489	.039	.009	.035	.325	.031	.321	.013
USD/JPY	.005	.010	.019	.007	.350	.013	.009	.009	.066	.014	.496	.012	.008	.006	.008	.013	.140	.004	.010	.012	.017	.025	.324	.025	.006	.026	.081	.031	.352	.039
GPB/USD	.012	.004	.050	.005	.060	.011	.007	.007	.136	.010	.056	.010	.013	.008	.102	.008	.008	.004	.007	.007	.005	.005	.004	.010	.007	.114	.477	.120	.494	.140
AUD/USD	.010	.178	.003	.004	.288	.006	.006	.094	.295	.183	.301	.189	.007	.062	.177	.049	.300	.052	.002	.052	.197	.068	.120	.058	.002	.322	.307	.306	.299	.304
USD/CAD	.010	.096	.766	.099	.919	.087	.006	.091	.964	.109	.981	.119	.007	.118	.904	.196	.912	.192	.012	.055	.794	.055	.486	.056	.006	.009	.637	.009	.308	.007
Rejections	5	3	2	4	0	4	5	3	0	3	0	3	5	3	1	4	1	3	5	3	2	3	1	3	5	3	0	3	0	3
Total Rejections	**11**	**7**	**7**	**10**	**4**	**10**	**10**	**6**	**4**	**6**	**5**	**9**	**10**	**6**	**4**	**9**	**4**	**6**	**9**	**7**	**9**	**10**	**9**	**12**	**14**	**13**	**8**	**11**	**8**	**10**

Table 6: The p-values of each model from the independence test. The model is rejected if the p-value is less than 0.05.

Based on the results presented in table 6, it not clear whether any of the distributional assumptions or GARCH-type models are to be preferred over the other. In general, most models obtained quite few rejections up until q = 0.999. For this quintile, the null hypothesis that the violations are independent was rejected for the majority of the models for most data sets. Furthermore, it should be noted that there appears to be a disproportionate number of rejections for the AUD/USD currency pair due to a large number of consecutive violations during the financial crisis of 08.

Next we examine the results of the joint test of the two previous tests, namely the conditional coverage test. The results are presented in table 7 below.

FX Pair	S_n	S_n^{EVT}	S_t	S_t^{EVT}	S_{st}	S_{st}^{EVT}	I_n	I_n^{EVT}	I_t	I_t^{EVT}	I_{st}	I_{st}^{EVT}	G_n	G_n^{EVT}	G_t	G_t^{EVT}	G_{st}	G_{st}^{EVT}	E_n	E_n^{EVT}	E_t	E_t^{EVT}	E_{st}	E_{st}^{EVT}	R_n	R_n^{EVT}	R_t	R_t^{EVT}	R_{st}	R_{st}^{EVT}
q = 0.95																														
EUR/USD	.162	.242	.049	.278	.102	.257	.449	.544	.120	.449	.166	.424	.348	.261	.206	.289	.262	.508	.203	.248	.404	.418	.524	.466	.111	.181	.230	.289	.116	.391
USD/JPY	.010	.085	.007	.032	.020	.030	.026	.080	.010	.106	.062	.109	.062	.155	.002	.146	.003	.067	.018	.088	.002	.033	.015	.040	.004	.024	.002	.094	.016	.099
GPB/USD	.569	.477	.157	.264	.224	.311	.199	.597	.174	.224	.108	.253	.422	.270	.214	.259	.285	.493	.613	.528	.374	.648	.387	.653	.005	.002	.064	.170	.026	.110
AUD/USD	.008	.017	.001	.006	.003	.004	.009	.014	.002	.006	.001	.006	.013	.115	.002	.082	.024	.088	.012	.008	.004	.009	.020	.042	.001	.001	.001	.010	.004	.002
USD/CAD	.971	.956	.424	.994	.240	.864	.706	.988	.597	.763	.450	.918	.708	.922	.440	.949	.218	.972	.839	.897	.627	.909	.509	.830	.838	.958	.365	.947	.294	.948
Rejections	2	1	2	2	2	2	2	1	2	1	1	1	1	0	2	0	2	0	2	1	2	2	2	2	3	3	2	1	3	2
q = 0.975																														
EUR/USD	.181	.510	.453	.746	.144	.461	.479	.490	.767	.343	.296	.184	.142	.666	.138	.441	.157	.503	.143	.109	.180	.161	.094	.130	.211	.692	.309	.373	.408	.478
USD/JPY	.015	.149	.095	.323	.093	.080	.013	.165	.153	.166	.283	.047	.001	.140	.002	.146	.096	.166	.001	.103	.003	.132	.084	.127	.001	.018	.010	.022	.014	.018
GPB/USD	.019	.256	.134	.261	.163	.240	.052	.163	.189	.160	.320	.165	.053	.518	.118	.184	.042	.109	.010	.332	.080	.328	.147	.231	.001	.002	.511	.524	.318	.518
AUD/USD	.001	.047	.006	.038	.007	.019	.001	.036	.001	.044	.008	.045	.002	.109	.006	.086	.094	.101	.001	.092	.009	.013	.011	.041	.001	.002	.001	.001	.006	.007
USD/CAD	.168	.662	.683	.917	.384	.799	.430	.933	.884	.895	.381	.765	.147	.873	.396	1.00	.507	.871	.123	.603	.449	.939	.351	.926	.084	.812	.342	.864	.251	.658
Rejections	3	1	1	1	1	1	2	1	1	1	1	2	2	0	2	0	1	0	3	0	2	1	1	1	3	3	2	2	2	2
q = 0.99																														
EUR/USD	.063	.212	.915	.219	.922	.180	.094	.324	.831	.355	.882	.348	.016	.034	.358	.037	.233	.037	.005	.034	.137	.021	.146	.025	.169	.224	.461	.223	.306	.229
USD/JPY	.001	.862	.240	.130	.108	.150	.001	.300	.202	.126	.029	.148	.001	.021	.300	.017	.177	.016	.001	.017	.199	.014	.013	.013	.001	.015	.002	.003	.079	.023
GPB/USD	.001	.833	.819	.266	.356	.838	.001	.804	.986	.899	.905	.875	.001	.431	.119	.299	.195	.275	.001	.855	.064	.045	.037	.040	.001	.004	.014	.056	.030	.085
AUD/USD	.001	.246	.001	.260	.065	.049	.001	.319	.164	.306	.056	.068	.001	.326	.007	.058	.074	.061	.001	.263	.005	.060	.018	.058	.001	.002	.001	.001	.001	.001
USD/CAD	.001	.576	.761	.685	.679	.652	.003	.397	.696	.563	.795	.470	.003	.445	.545	.489	.479	.510	.001	.260	.986	.651	.939	.447	.001	.480	.529	.803	.602	.768
Rejections	4	0	1	0	0	1	4	0	0	0	1	0	5	2	1	2	0	2	5	2	1	3	3	3	4	3	3	2	2	2
q = 0.995																														
EUR/USD	.002	.931	.875	.823	.791	.704	.002	.813	.748	.694	.518	.572	.001	.567	.794	.227	.793	.245	.005	.216	.266	.250	.233	.244	.034	.143	.213	.174	.250	.215
USD/JPY	.001	.255	.926	.942	.733	.887	.001	.239	.511	.792	.404	.283	.001	.160	.372	.588	.559	.586	.001	.120	.694	.229	.263	.291	.001	.302	.444	.142	.708	.145
GPB/USD	.001	.099	.187	.209	.262	.180	.001	.077	.220	.213	.147	.195	.001	.130	.222	.186	.265	.217	.001	.222	.210	.267	.030	.021	.001	.033	.091	.142	.192	.157
AUD/USD	.001	.161	.009	.225	.221	.266	.001	.025	.025	.034	.016	.026	.001	.220	.021	.223	.023	.207	.001	.023	.054	.228	.014	.020	.001	.024	.005	.018	.014	.008
USD/CAD	.001	.810	.222	.871	.934	.868	.001	.615	.140	.727	.250	.779	.001	.348	.679	.308	.936	.435	.001	.132	.610	.098	.753	.080	.001	.341	.467	.797	.697	.744
Rejections	5	0	1	0	0	0	5	1	1	1	1	1	5	0	1	0	1	0	5	1	0	0	2	2	5	2	1	1	1	1
q = 0.999																														
EUR/USD	.001	.033	.136	.123	.291	.115	.001	.049	.322	.040	.306	.041	.001	.101	.554	.042	.573	.046	.001	.113	.672	.107	.683	.120	.006	.120	.861	.051	.837	.040
USD/JPY	.001	.012	.043	.010	.845	.037	.001	.013	.130	.040	.663	.037	.001	.009	.013	.006	.337	.005	.001	.013	.037	.058	.819	.048	.001	.110	.167	.110	.834	.108
GPB/USD	.001	.001	.141	.001	.153	.001	.001	.003	.311	.001	.140	.002	.001	.004	.305	.002	.012	.001	.001	.001	.007	.001	.008	.001	.001	.309	.712	.312	.690	.335
AUD/USD	.001	.410	.010	.012	.663	.010	.001	.311	.671	.374	.684	.394	.001	.151	.401	.160	.689	.139	.001	.157	.411	.167	.317	.168	.001	.701	.697	.670	.667	.676
USD/CAD	.001	.312	.216	.319	.089	.305	.001	.288	.024	.333	.020	.318	.001	.324	.077	.563	.083	.548	.001	.152	.208	.141	.828	.140	.001	.015	.394	.014	.656	.005
Rejections	5	3	2	3	0	3	5	3	1	3	1	3	5	2	1	3	1	3	5	2	2	1	1	2	5	1	0	1	0	2
Total Rejections	**19**	**5**	**7**	**6**	**3**	**7**	**18**	**6**	**5**	**6**	**5**	**7**	**18**	**4**	**7**	**5**	**5**	**5**	**20**	**6**	**7**	**7**	**9**	**10**	**20**	**12**	**8**	**7**	**8**	**9**

Table 7: The p-values of each model from the conditional coverage test. The model is rejected if the p-value is less than 0.05.

In general, the results of the conditional coverage test display similar tendencies to those of the unconditional coverage test. The models with normally distributed standardized residuals are rejected for most currency pairs at quintiles higher than $q = 0.95$. It is clear that applying EVT to these models improves the accuracy of the VaR forecasts. It is not as clear whether applying EVT to models already assuming heavy tailed distributions yield more accurate results, particularly for the higher quintiles. For $q \leq 0.99$, however, several models seem to benefit from the augmentation with EVT. Nonetheless, most models assuming either a t-distribution or a skewed t-distribution and/or were augmented with EVT produced accurate results for the conditional coverage test. In this test, the standard GARCH(1,1) with skewed t-distributed innovations produced the least number of rejections.

The next part of our analysis will examine the results from the backtesting of Expected shortfall. The result from the first test of Acerbi & Szekely (2014) is presented in table 8 below.

FX Pair	S_n	S_n^{EVT}	S_t	S_t^{EVT}	S_{st}	S_{st}^{EVT}	I_n	I_n^{EVT}	I_t	I_t^{EVT}	I_{st}	I_{st}^{EVT}	G_n	G_n^{EVT}	G_t	G_t^{EVT}	G_{st}	G_{st}^{EVT}	E_n	E_n^{EVT}	E_t	E_t^{EVT}	E_{st}	E_{st}^{EVT}	R_n	R_n^{EVT}	R_t	R_t^{EVT}	R_{st}	R_{st}^{EVT}
q = 0.95																														
EUR/USD	.000	.427	.753	.431	.729	.407	.000	.333	.820	.403	.845	.276	.000	.237	.412	.300	.396	.288	.000	.193	.346	.179	.347	.117	.000	.460	.553	.478	.554	.429
USD/JPY	.000	.460	.609	.695	.903	.652	.000	.386	.926	.522	.972	.510	.000	.294	.330	.288	.709	.385	.000	.259	.401	.436	.709	.430	.000	.014	.226	.061	.467	.059
GPB/USD	.000	.012	.056	.036	.067	.026	.000	.003	.161	.028	.227	.034	.000	.011	.053	.051	.065	.048	.000	.014	.075	.022	.036	.030	.000	.022	.396	.188	.396	.259
AUD/USD	.000	.390	.098	.564	.381	.522	.000	.304	.309	.394	.605	.395	.000	.217	.081	.379	.370	.370	.000	.048	.038	.381	.239	.296	.000	.185	.054	.182	.074	.104
USD/CAD	.000	.395	.947	.391	.924	.266	.000	.316	.987	.291	.972	.306	.000	.230	.877	.172	.901	.275	.000	.045	.807	.082	.736	.111	.000	.181	.933	.280	.811	.236
Rejections	5	1	0	1	0	1	5	1	0	1	0	1	5	1	0	0	0	1	5	3	1	1	1	1	5	2	0	0	0	0
q = 0.975																														
EUR/USD	.000	.074	.434	.190	.427	.205	.000	.136	.552	.203	.509	.242	.000	.039	.356	.059	.290	.071	.000	.034	.539	.090	.291	.046	.000	.162	.625	.085	.744	.132
USD/JPY	.000	.110	.727	.397	.894	.417	.000	.239	.934	.480	.992	.497	.000	.104	.621	.114	.778	.121	.000	.017	.560	.115	.773	.100	.000	.003	.269	.031	.666	.058
GPB/USD	.000	.024	.082	.024	.052	.016	.000	.006	.180	.040	.116	.034	.000	.005	.083	.011	.051	.014	.000	.015	.077	.009	.036	.019	.000	.125	.141	.044	.128	.048
AUD/USD	.000	.043	.015	.032	.063	.046	.000	.014	.075	.031	.141	.023	.000	.113	.086	.166	.260	.223	.000	.032	.045	.143	.187	.162	.000	.167	.236	.276	.147	.256
USD/CAD	.000	.122	.955	.252	.972	.135	.000	.256	.992	.385	.995	.413	.000	.061	.964	.098	.941	.060	.000	.006	.907	.030	.915	.024	.000	.136	.986	.244	.953	.277
Rejections	5	2	1	2	0	2	5	2	0	2	0	2	5	2	0	1	0	1	5	5	1	2	1	3	5	1	0	2	0	1
q = 0.99																														
EUR/USD	.000	.017	.124	.025	.165	.020	.000	.040	.244	.074	.304	.075	.000	.004	.082	.010	.048	.006	.000	.001	.066	.002	.042	.001	.000	.003	.065	.004	.043	.004
USD/JPY	.000	.002	.137	.003	.331	.005	.000	.014	.489	.004	.549	.006	.000	.041	.143	.001	.245	.006	.000	.001	.302	.081	.375	.039	.000	.012	.642	.009	.709	.006
GPB/USD	.000	.002	.049	.021	.066	.014	.000	.001	.058	.014	.060	.017	.000	.002	.094	.011	.037	.011	.000	.000	.053	.006	.038	.008	.000	.042	.272	.065	.196	.030
AUD/USD	.000	.012	.105	.006	.030	.009	.000	.016	.100	.007	.233	.008	.000	.026	.097	.060	.152	.058	.000	.008	.110	.016	.120	.014	.000	.084	.116	.093	.205	.080
USD/CAD	.001	.271	.951	.116	.920	.117	.000	.247	.989	.219	.973	.250	.000	.092	.912	.064	.888	.077	.000	.054	.824	.005	.782	.021	.000	.057	.810	.032	.775	.027
Rejections	5	4	1	4	1	4	5	4	0	3	0	3	5	4	0	3	2	3	5	4	0	4	2	5	5	3	0	3	1	4
q = 0.995																														
EUR/USD	.000	.013	.110	.024	.196	.034	.000	.015	.229	.030	.136	.036	.000	.007	.053	.018	.061	.013	.000	.005	.167	.007	.085	.008	.000	.004	.199	.005	.237	.006
USD/JPY	.000	.001	.145	.002	.359	.001	.000	.005	.237	.004	.538	.007	.000	.009	.197	.001	.492	.007	.000	.001	.145	.025	.492	.017	.000	.005	.256	.032	.551	.036
GPB/USD	.000	.002	.042	.008	.018	.007	.000	.003	.063	.009	.089	.011	.000	.001	.021	.006	.007	.005	.000	.000	.014	.001	.011	.001	.000	.021	.417	.027	.227	.024
AUD/USD	.000	.001	.093	.001	.045	.002	.000	.004	.059	.004	.020	.004	.000	.006	.078	.003	.032	.003	.000	.000	.064	.002	.008	.001	.000	.005	.051	.005	.022	.003
USD/CAD	.006	.179	.900	.032	.966	.040	.007	.150	.955	.164	.921	.110	.001	.126	.954	.221	.930	.144	.000	.115	.951	.174	.897	.214	.000	.076	.671	.015	.517	.016
Rejections	5	4	1	5	2	5	5	4	0	4	1	4	5	4	1	4	2	4	5	4	1	4	2	4	5	4	0	5	1	5
q = 0.999																														
EUR/USD	.000	.064	.519	.052	.483	.050	.001	.085	.699	.086	.694	.081	.000	.020	.106	.041	.121	.039	.000	.010	.109	.015	.066	.011	.000	.032	.271	.040	.329	.042
USD/JPY	.001	.161	.567	.169	.719	.134	.001	.140	.740	.127	.811	.122	.003	.089	.553	.047	.763	.126	.001	.024	.527	.127	.632	.123	.001	.012	.262	.030	.384	.031
GPB/USD	.000	.013	.099	.035	.065	.026	.000	.011	.088	.024	.091	.019	.000	.005	.034	.017	.036	.014	.000	.005	.058	.014	.033	.014	.000	.005	.039	.009	.026	.008
AUD/USD	.003	.001	.014	.002	.006	.002	.001	.002	.011	.001	.017	.001	.001	.003	.017	.005	.013	.005	.000	.002	.011	.003	.034	.003	.004	.001	.014	.003	.011	.001
USD/CAD	.094	.125	.894	.082	.801	.091	.116	.141	.702	.147	.678	.131	.038	.054	.871	.046	.692	.037	.038	.037	.911	.064	.945	.065	.033	.160	.684	.201	.668	.242
Rejections	4	2	1	2	1	2	4	2	1	2	1	2	5	3	2	5	2	4	5	5	1	3	2	3	5	4	2	4	2	4
Total Rejections	**24**	**13**	**4**	**14**	**4**	**14**	**24**	**13**	**1**	**12**	**2**	**12**	**25**	**14**	**3**	**13**	**6**	**13**	**25**	**21**	**4**	**14**	**8**	**16**	**25**	**14**	**2**	**14**	**4**	**14**

Table 8: The p-values of each model from the Z1 test. The model is rejected if the p-value is less than 0.05.

By inspecting the results of table 8, it is clear that models with normally distributed standardized residuals underestimate ES. The null hypothesis is rejected for virtually all quintiles and currency pairs. The augmentation with EVT provides somewhat of a remedy to these models, although it is to a comparatively small extent. In contrast, models in which the innovations are assumed to follow either a t-distribution or skewed t-distribution appear to generate fewer rejections in general. It is not obvious whether any of these distributions is superior to the other. It appears that these models do not gain from the EVT approach as this procedure results in more rejections overall.

Similar to what was observed for the backtesting results of VaR, it is not clear whether any particular GARCH-type model alone stands out when forecasting ES. In this test, the symmetrical IGARCH model with t-distributed innovations produced the least number of rejections.

As was noted earlier, this test is not sensitive to excessive numbers of VaR violations. Therefore, it has to be looked at in conjunction with the tests for VaR presented above. Alternatively, one can turn to the second test of Acerbi & Szekely (2014), in which this is accounted for. The result of this test is depicted in table 9 below.

FX Pair	S_n	S_n^{EVT}	S_t	S_t^{EVT}	S_{st}	S_{st}^{EVT}	I_n	I_n^{EVT}	I_t	I_t^{EVT}	I_{st}	I_{st}^{EVT}	G_n	G_n^{EVT}	G_t	G_t^{EVT}	G_{st}	G_{st}^{EVT}	E_n	E_n^{EVT}	E_t	E_t^{EVT}	E_{st}	E_{st}^{EVT}	R_n	R_n^{EVT}	R_t	R_t^{EVT}	R_{st}	R_{st}^{EVT}
q = 0.95																														
EUR/USD	.048	.416	.037	.371	.067	.387	.125	.196	.130	.176	.137	.231	.053	.338	.063	.315	.105	.308	.374	.670	.250	.430	.281	.474	.235	.589	.295	.575	.313	.559
USD/JPY	.001	.182	.025	.260	.081	.306	.018	.111	.111	.238	.336	.238	.000	.037	.001	.051	.006	.043	.000	.020	.004	.188	.049	.202	.003	.247	.009	.246	.120	.239
GPB/USD	.009	.192	.114	.464	.093	.501	.061	.202	.302	.385	.171	.329	.009	.164	.080	.337	.063	.383	.005	.089	.045	.198	.032	.119	.000	.000	.069	.135	.044	.088
AUD/USD	.000	.263	.000	.176	.046	.183	.001	.209	.002	.148	.170	.137	.000	.223	.001	.226	.062	.240	.000	.242	.001	.192	.102	.210	.000	.253	.002	.549	.336	.634
USD/CAD	.178	.558	.231	.489	.142	.595	.356	.465	.401	.456	.329	.488	.242	.542	.195	.541	.087	.506	.170	.500	.270	.406	.173	.351	.050	.288	.239	.383	.124	.330
Rejections	4	0	3	0	1	0	2	0	1	0	0	0	3	1	2	0	1	1	3	1	3	0	2	0	3	1	2	0	1	0
q = 0.975																														
EUR/USD	.004	.616	.188	.468	.235	.444	.022	.316	.404	.259	.452	.216	.003	.383	.095	.431	.159	.377	.011	.555	.144	.300	.254	.359	.008	.698	.294	.750	.203	.685
USD/JPY	.000	.349	.054	.339	.263	.346	.000	.199	.244	.244	.425	.223	.000	.073	.001	.074	.037	.117	.000	.118	.008	.347	.157	.442	.000	.104	.018	.096	.093	.070
GPB/USD	.000	.057	.044	.274	.049	.288	.000	.074	.137	.177	.130	.184	.000	.112	.033	.352	.035	.312	.000	.023	.031	.110	.021	.045	.000	.000	.189	.189	.150	.196
AUD/USD	.000	.530	.013	.516	.250	.462	.000	.560	.050	.415	.504	.438	.000	.257	.002	.359	.104	.311	.000	.211	.003	.352	.122	.297	.000	.219	.000	.389	.205	.391
USD/CAD	.004	.692	.460	.491	.278	.630	.014	.429	.634	.348	.380	.321	.001	.501	.219	.446	.211	.568	.001	.419	.327	.328	.207	.349	.001	.214	.228	.294	.116	.197
Rejections	5	0	2	0	1	0	5	0	0	0	0	0	5	0	3	0	2	0	5	1	3	0	1	1	5	1	2	0	0	0
q = 0.99																														
EUR/USD	.000	.476	.454	.467	.469	.513	.001	.259	.649	.242	.615	.239	.000	.402	.328	.417	.452	.463	.000	.598	.623	.622	.605	.601	.008	.779	.737	.771	.795	.768
USD/JPY	.000	.445	.499	.616	.753	.572	.000	.304	.706	.683	.954	.591	.000	.042	.148	.221	.551	.189	.000	.094	.082	.220	.580	.322	.000	.022	.009	.101	.210	.125
GPB/USD	.000	.046	.085	.054	.031	.060	.000	.035	.254	.084	.148	.067	.000	.016	.033	.057	.032	.053	.000	.026	.038	.013	.020	.010	.000	.001	.124	.046	.131	.079
AUD/USD	.000	.318	.003	.335	.274	.293	.000	.191	.046	.241	.225	.207	.000	.200	.009	.263	.171	.259	.000	.086	.003	.266	.137	.262	.000	.167	.007	.392	.156	.334
USD/CAD	.000	.332	.664	.397	.598	.399	.000	.198	.797	.312	.733	.272	.000	.207	.489	.235	.429	.239	.000	.102	.557	.310	.445	.171	.000	.225	.820	.458	.504	.398
Rejections	5	1	1	0	1	0	5	1	1	0	0	0	5	2	2	0	1	0	5	1	2	1	1	1	5	2	2	1	0	0
q = 0.995																														
EUR/USD	.000	.294	.473	.246	.353	.204	.000	.240	.585	.202	.702	.160	.000	.119	.290	.098	.299	.127	.000	.111	.350	.112	.380	.078	.001	.395	.440	.387	.379	.319
USD/JPY	.000	.207	.388	.286	.618	.349	.000	.142	.723	.253	.840	.204	.000	.041	.099	.112	.246	.111	.000	.016	.260	.179	.428	.217	.000	.037	.145	.022	.379	.024
GPB/USD	.000	.008	.073	.042	.101	.028	.000	.006	.141	.034	.065	.025	.000	.008	.081	.022	.134	.032	.000	.014	.094	.054	.058	.038	.000	.002	.080	.052	.135	.048
AUD/USD	.000	.367	.007	.287	.162	.187	.000	.116	.091	.110	.510	.112	.000	.219	.022	.324	.323	.324	.000	.161	.013	.251	.411	.298	.000	.306	.044	.467	.390	.444
USD/CAD	.000	.387	.946	.487	.649	.498	.000	.250	.972	.314	.951	.369	.000	.106	.770	.086	.627	.144	.000	.036	.434	.027	.477	.019	.000	.096	.815	.305	.659	.242
Rejections	5	1	1	1	0	1	5	1	0	1	0	1	5	2	1	1	0	1	5	3	1	1	0	2	5	2	1	1	0	1
q = 0.999																														
EUR/USD	.000	.008	.099	.017	.169	.017	.000	.004	.209	.009	.204	.009	.000	.010	.189	.006	.194	.006	.000	.007	.494	.006	.461	.006	.000	.015	.370	.006	.391	.006
USD/JPY	.000	.000	.035	.002	.481	.005	.000	.002	.153	.005	.654	.005	.000	.002	.009	.000	.232	.001	.000	.000	.033	.012	.461	.012	.000	.013	.077	.016	.401	.017
GPB/USD	.000	.000	.040	.000	.032	.000	.000	.000	.077	.000	.036	.001	.000	.000	.050	.000	.023	.000	.000	.000	.026	.000	.008	.000	.000	.022	.369	.028	.326	.026
AUD/USD	.000	.040	.043	.017	.142	.018	.000	.020	.153	.031	.180	.033	.000	.009	.097	.009	.186	.009	.000	.005	.073	.008	.053	.008	.000	.110	.090	.091	.173	.117
USD/CAD	.000	.108	.895	.104	.948	.104	.000	.110	.985	.109	.984	.107	.000	.107	.957	.190	.947	.185	.000	.045	.889	.049	.653	.050	.000	.005	.743	.002	.428	.001
Rejections	5	4	3	4	1	4	5	4	0	4	1	4	5	4	1	4	1	4	5	5	2	5	1	4	5	4	0	4	0	4
Total Rejections	**24**	**6**	**10**	**5**	**4**	**5**	**22**	**6**	**2**	**5**	**1**	**5**	**23**	**9**	**9**	**5**	**5**	**6**	**23**	**11**	**11**	**7**	**5**	**8**	**23**	**10**	**7**	**6**	**1**	**5**

Table 9: The p-values of each model from the Z2 test. The model is rejected if the p-value is less than 0.05.

When also considering the expected number of violations for each quintile, we see that the models that were augmented with EVT do comparatively better. In the first test of Acerbi & Szekely (2014), where excessive numbers of violations were not considered, the EVT-models tended to underestimate the density of the loss distribution to a comparatively high extent in comparison to models assuming a t-distribution or skewed t-distribution. However, as was evident in table 4 and table 5, the EVT-models often generated higher quintile forecasts for $q \leq 0.99$ than those only assuming heavy tailed distributions for the innovation, resulting in more accurate estimates of VaR. The combination of these factors reflect the results of the second test of Acerbi & Szekely (2014).

Again, we observe that models in which the innovations are assumed to follow a skewed t-distribution were rejected on fewer occasions than those assuming a t-distribution. These models were generally improved by the EVT approach for $q \leq 0.99$, whereas it had an opposite effect for higher quintiles. For models assuming a normal distribution for the innovation, the EVT approach seems to be beneficial for all quintiles.

Similar to what has been observed for all backtests, it is not obvious whether any GARCH-type model is outperforming the others. In this test, the IGARCH(1,1) and Realized GARCH(1,1) with t-skewed innovations produced the fewest number of rejections.

5. Conclusion

The primary conclusion of this thesis is that the distributional assumption of the innovations is the most important determinant in producing accurate one day ahead forecasts of VaR and ES. Models in which the innovations were assumed to follow a normal distribution consistently underestimated both VaR and ES. It is clear that the two-stage EVT approach of McNeil & Frey (2000) improved the accuracy of the forecasts of these models, regardless of quintile. This approach does not, however, appear to be as effective if the innovations are assumed to follow a heavy tailed distribution, such as the t-distribution or the skewed t-distribution. Models in which the innovations were assumed to follow any of these distributions generally produced accurate forecasts of VaR and ES, particularly for higher quintiles. In applying the two-stage EVT approach to these models, the quintile forecasts for $q \leq 0.99$ were in many instances improved. For higher quintiles, however, we found that this approach tended to impair the forecasting accuracy of these models. Hence, the usefulness of the EVT approach appears to be dependent on the distributional assumption as well as the choice of quintile. Overall, models assuming a skewed t-distribution for the innovation process were found to produce the least number of rejections.

We cannot conclude that more complex extensions of the standard GARCH(1,1) model yield more accurate forecasts of VaR and ES, as no discernible trend amongst the conditional volatility models was observed. As noted earlier, however, the bandwidth parameter of the realized kernel was not chosen according to the recommendation of Barndorff-Nielsen et al. (2009). For further research it may therefore be interesting to examine whether using the optimal bandwidth enhances the performance of the Realized GARCH model of Hansen et al. (2011). Another area that may be of interest is to examine how different choices of thresholds affect the forecasting performance of models combined with EVT.

References

Acerbi, C. and B. Szekely (2014). Back-testing expected shortfall. Risk, 76-81.

Akaike, H. (1974). A new look at the statistical model identification. *Springer Series in Statistics*, 215–222.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Balkema, A. A., & de Haan, L. (1974). Residual life time at Great Age. *The Annals of Probability*, *2*(5).

Bank for International Settlement (BIS). (2020, March 27). *Governors and Heads of Supervision announce deferral of Basel III implementation to increase operational capacity of banks and supervisors to respond to Covid-19* [Press release]. Retrieved from https://www.bis.org/press/p200327.htm

Bank for International Settlements (BIS). (2013). Fundamental Review of the Trading Book: A Revised Market Risk Framework, Basel Committee on Banking Supervision. Retrieved from https://www.bis.org/publ/bcbs265.pdf

Bank for International Settlements (BIS). (2019). Foreign exchange turnover in April 2019, Monetary and Economic Department. Retrieved from https://www.bis.org/statistics/rpfx19_fx.pdf

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, *12*(3).

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

Bollerslev, T. (2008). Glossary to arch (GARCH). *SSRN Electronic Journal*.

Boudt, K., Cornelissen, J., Payseur, S., Nguyen, G., Kleen, O., Sjoerup, E. (2022). highfrequency: Tools for Highfrequency Data Analysis (version 0.9.4). Available at http://cran.r-project.org/web/packages/highfrequency/

Black, F. (1976) Studies of Stock Price Volatility Changes. *Journal of the American Statistical Association*, 177-181.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference. *Sociological Methods & Research*, *33*(2), 261–304.

Byström, H. N. E. (2004). Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, *13*(2), 133–152.

Christoffersen, P. F. (2011). *Elements of financial risk management*. Elsevier Science.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223–236.

Demsetz, H. (1968). The cost of transacting. *The Quarterly Journal of Economics*, *82*(1), 33.

DuMouchel, W. H. (1983). Estimating the Stable Index in Order to Measure Tail Thickness: A Critique. *The Annals of Statistics*, *11*(4).

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, *5*(1), 1–50.

Fernandez, V. (2005). Risk management under extreme events. *International Review of Financial Analysis*, *14*(2), 113–148.

Fernandez, C., & Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441), 359.

Gbatu, A. P., Wang, Z., Wesseh Jr., P. K., & Tutdel, I. Y. R. (2017). Causal Effects and Dynamic Relationship between Exchange Rate Volatility and Economic Development in Liberia. *International Journal of Economics and Financial Issues*, 7(4), 119–131.

Gençay, R., Selçuk, F., & Ulugülyağci, A. (2003). High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics*, *33*(2), 337–356.

Ghalanos, A. (2022). rugarch: Univariate GARCH Models (version 1.4-3). Available at http://cran.r-project.org/web/packages/rugarch/

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

Hansen, P. R., Huang, Z., & Shek, H. H. (2011). REALIZED GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.

Ho, L.-C., Burridge, P., Cadle, J., & Theobald, M. (2000). Value-at-risk: Applying the extreme value approach to Asian markets in the recent financial turmoil. *Pacific-Basin Finance Journal*, *8*(2), 249–275.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for *r*. *Journal of Statistical Software*, *27*(3).

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., Ihaka, R., Reid, D., Shaub, D., Tang, Y., Zhou, Z., R Core Team (2022). forecast: Forecasting Functions for Time Series and Linear Models (version 8.16). Available at http://cran.r-project.org/web/packages/forecast/

Islam, M. S., & Hossain, E. (2021). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters*, *3*, 100009.

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*(3-4), 271–300.

McNeil, A. J., Frey Rüdiger, & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press.

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

Ossberger, H. (2020). tea: Threshold Estimation Approaches (version 1.1). Available at http://cran.r-project.org/web/packages/tea/

Poon, S.-H., & Granger, C. W. (2003). Forecasting volatility in Financial Markets: A Review. *Journal of Economic Literature*, *41*(2), 478–539.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. J. Wiley.

Wagner, N., & Marsh, T. A. (2005). Measuring tail thickness under GARCH and an application to extreme exchange rate changes. *Journal of Empirical Finance*, *12*(1), 165–185.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics (version 3.3.6). Available at https://cran.r-project.org/web/packages/ggplot2/

Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, *29*(4), 997–1015.

Yong, Y. L., Lee, Y., Gu, X., Angelov, P. P., Ngo, D. C., & Shafipour, E. (2018). Foreign currency exchange rate prediction using neuro-fuzzy systems. *Procedia Computer Science*, *144*, 232–238.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, *14*(1), 45.

6. Appendix

6.1 Density functions

Normal distribution

The density function of the normal distribution is:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

Student's t-distribution

The density function of the student's t-distribution is:

$$f(z) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{z^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$$

where v denotes the degrees of freedom and $\Gamma(\cdot)$ is the Gamma function.

Skewed student's t-distribution

According to Fernandez & Steel (1998), the density function of the skewed student's t-distribution can be defined as:

$$f(z) = \frac{2}{\xi + \frac{1}{\xi}}\left\{f\left(\frac{z}{\xi}\right) I_{\{z\geq 0\}} + f(\xi z) I_{\{z<0\}}\right\}$$

where $f(\cdot)$ is the student's t-distribution and ξ is the skewness parameter. $I_{\{z\}}$ denotes an indicator variable.

6.2 Maximum Likelihood estimation

The Maximum Likelihood approach estimates the parameters so that they maximize the likelihood that the assumed model produced the observed data. This is done by maximizing the likelihood function with respect to the unknown parameters Θ. This function can formally be defined as:

$$L(\Theta|G_{n-1}) = \prod_{t=1}^{n} \varphi(\varepsilon_t|G_{t-1})$$

where G denotes the information set and φ is the density function of the innovation process. The form of the likelihood function depends on the assumed distribution of the innovations.

Normal distribution

If the innovations are assumed to follow a normal distribution, the log-likelihood function is:

$$\log[L(\Theta|G_{n-1})] = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right]$$

Student's t-distribution

If the innovations are assumed to follow a t-distribution, the log-likelihood function is:

$$\log[L(\Theta|G_{n-1})] = \log\left[\Gamma\left(\frac{v+1}{2}\right)\right] - \log\left[\Gamma\left(\frac{v}{2}\right)\right] - \frac{1}{2}\log\big(\pi(v-2)\big) - \frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + (1+v)\log\left(1+\frac{\varepsilon_t^2}{\sigma_t^2(v-2)}\right)\right]$$

Skewed student's t-distribution

If the innovations are assumed to follow a skewed t-distribution, the log-likelihood function is:

$$\log[L(\Theta|G_{n-1})] = \log\left[\Gamma\left(\frac{v+1}{2}\right)\right] - \log\left[\Gamma\left(\frac{v}{2}\right)\right] - \frac{1}{2}\log\big(\pi(v-2)\big) + \log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + \log(s) - \frac{1}{2}\sum_{t=1}^{n}\left[\log(\sigma_t^2) + (1+v)\log\left(1+\left(\frac{s\varepsilon_t}{\sigma_t^2(v-2)}+\frac{m}{v-2}\right)\xi^{-I_t}\right)\right]$$

where

$$m = \frac{\Gamma\left(\frac{v+1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi-\frac{1}{\xi}\right)$$

$$s = \sqrt{\left(\xi^2+\frac{1}{\xi^2}-1\right)-m^2}$$

6.3 Functions

Autocorrelation function

The autocorrelation function of a covariance-stationary process X_t is defined as:

$$\rho(h) = p(X_h, X_0) = \gamma(h)/\gamma(0)$$

where $\rho(h)$ denotes the autocorrelation of lag h.

6.4 Tests

Jarque-Bera test

The Jarque-Bera test examines if the data have the kurtosis and skewness of a normal distribution. The null hypothesis is that the data is generated from a normal distribution, while the alternative hypothesis is that it is not. The test statistic is defined as:

$$JB = \frac{n}{6}\left(S^2+\frac{1}{4}(K-3)^2\right)$$

where n is the number of observations, S refers to the skewness, and K refers to the kurtosis.

Ljung-box test

The Ljung-box test assesses whether there is absence of serial correlation in the data up to lag *k*. The null hypothesis is that there is no serial correlation, while the alternative hypothesis states that the data is dependent. The test statistic is defined as:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

where n corresponds to the sample size, $\hat{\rho}_k$ refers to the sample autocorrelation at lag k, h refers to the number of lags to be tested. The test statistic is asymptotically chi-squared distributed with h degrees of freedom. The Ljung-box test of standardized residuals and the squared standardized residuals of the initial window is presented in table 10 below.

Model	Ljung–Box	EUR/USD	JPY/USD	GBP/USD	AUD/USD	CAD/USD
S_n	Q(10)	4.64	9.97	7.53	7.76	6.86
	Q2(10)	9.75	13.06	5.33	11.91	12.81
S_t	Q(10)	4.57	10.13	7.55	8.23	6.08
	Q2(10)	9.69	13.51	5.6	12.38	13.37
S_{st}	Q(10)	4.58	10.04	7.55	9.55	6.15
	Q2(10)	9.62	13.51	5.55	13.46	13.33
I_n	Q(10)	4.84	9.8	7.57	7.29	6.61
	Q2(10)	8.63	13.69	5.61	11.03	13.7
I_t	Q(10)	4.82	9.98	7.6	7.7	6.38
	Q2(10)	8.63	13.84	6.16	11.44	13.73
I_{st}	Q(10)	4.76	9.96	7.58	8.59	6.42
	Q2(10)	8.48	13.81	6.20	12.15	13.63
G_n	Q(10)	4.66	10	7.93	7.80	6.52
	Q2(10)	10.19	12.55	5.97	12.79	12.33
G_t	Q(10)	4.68	10.08	7.91	7.43	6.17
	Q2(10)	10.24	13.28	6.77	13.32	12.52
G_{st}	Q(10)	4.62	10.09	8.03	8.57	6.23
	Q2(10)	10.11	13.12	7.38	18.07	12.5
E_n	Q(10)	5.01	9.93	7.58	7.35	6.48
	Q2(10)	11.15	11.67	6.31	12.59	11.83
E_t	Q(10)	4.84	10.06	7.64	7.71	6.31
	Q2(10)	10.8	11.81	4.73	13.28	11.8
E_{st}	Q(10)	4.82	10.03	7.64	8.5	6.31
	Q2(10)	10.7	11.75	4.75	14.58	11.8
R_n	Q(10)	6.59	9.94	7.41	8.23	6.52
	Q2(10)	8.69	12.99	4.03	15.76	13.88
R_t	Q(10)	8.08	10.37	7.40	8.42	6.76
	Q2(10)	8.94	13.08	4.03	16.54	14.1
R_{st}	Q(10)	6.56	10.36	7.42	9.41	14.09
	Q2(10)	8.67	13.08	4.04	17.7	6.75
ARMA (*p*, *q*)		ARMA (1,1)	ARMA (1,0)	ARMA (0,0)	ARMA (1,0)	ARMA (0,1)

Table 10: Ljung-box test of the standardized residuals and squared standardized residuals of the initial window. Q(10) refers to the Ljung-box test of lag length 10 of the standardized residual while Q2(10) refers to the same for the squared standardized residuals. The bottom row refers to the identified ARMA process of the initial window using the algorithm described in section 3.3. The last 5 columns reports the test statistics of the test. * and ** denote significance at the 5% and 1% levels, respectively. See Table 3 for a description of the acronyms.