UNIVERSITY OF TURKU
Turku School of Economics

Risk Measure Estimation of European Interest Rates

Extreme Value Theory in EURIBOR Value-at-Risk

Department of Economics
Master's thesis

Author:
Peetu Paavola

Supervisor:
Ville Korpela

17.6.2026
Turku

Student’s statement regarding the use of Artificial Intelligence (AI) for preparing and/or writing this thesis:

☐ **I have not used any AI-based tools.**

☒ **I have used AI-based tools**. Their use is documented in the Appendix. The AI tools were used in a way that complies with academic integrity guidelines.

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis

Subject: Economics
Author: Peetu Paavola
Title: Risk Measure Estimation of European Interest Rates: Extreme Value Theory in Euribor Value-at-Risk
Supervisor: Professor Ville Korpela
Number of pages: 60 pages + appendices 1 page
Date: 17.6.2026

Abstract

This thesis estimates 1-day forward risk measures of Euribor tenors utilizing Extreme Value Theory and two GARCH processes (standard (1,1) and exponential (1,1) specifications). The accuracy of the approaches is measured for Value-at-Risk while the riskiness of the position is also measured with Extreme Shortfall. The examined period includes 1300 observations of each maturity exceeding 1 week (1-, 3-,6- and 12-months) ending in 26th of March 2026. This period allows for the testing of methods in multiple market regimes such as negative rates, high inflation and stabilization and times of between these regime switches. Additionally, the period includes short term market turbulences launched by the bankruptcy of Silicon Valley Bank and the 2026 Iran War.

The Thesis includes a review of relevant themes of risk measurement, interest rate dynamics and Extreme Value Theory. Based on these, the sample is discussed after which the results of the models are compared. The comparison is conducted with Kupiec's and Christoffersen's tests. In the case of Extreme Value Theory, the specific methodology chosen is that of McNeil & Frey's (2000) where the innovations of a GARCH process are used to model the tails of the underlying distribution.

Extreme Value Theory's advantage relies on modelling the tail distribution explicitly instead of the entire distribution. The two most accurate models for Value-at-Risk proved to be Extreme Value Theory and exponential GARCH with Student's t-distribution with Extreme Value Theory slightly more robust. Basic models such as exponentially weighted moving average and standard GARCH(1,1) failed the tests systematically. For Expected Shortfall, both Extreme Value Theory and exponential GARCH provided satisfactory results, with differing tendencies on over- and underestimation and exponential GARCH slightly more accurate.

Keywords: quantitative risk management, Extreme Value Theory, interest rates, financial economics, Euribor, Value-at-Rsik

Pro gradu -tutkielma

Oppiaine: Taloustiede
Tekijä: Peetu Paavola
Otsikko: Eurooppalaisten korkojen riskimittojen estimointi: ääriarvoteoria Euriborien kvantiilimitassa
Ohjaaja: Professori Ville Korpela
Sivumäärä: 60 sivua (+ liitteet 1 sivu)
Päivämäärä: 17.6.2026

Tiivistelmä

Tutkielma estimoi Euribor-korkojen riskitasoja 1 päivän eteenpäin hyödyntäen ääriarvoteoriaa sekä kahta GARCH-mallia (normaali (1,1) sekä eksponentiaalinen (1,1) malli). Lähestymistapojen tarkkuutta mitataan Value-at-Risk -mitalla (kvantiilimitta) sekä riskillisyyttä arvioidaan Expected Shortfall -mitalla (tappiomitta). Tarkasteltu aikaperiodi sisältää 1300 päivämuutosta jokaiselle maturiteetille (1-, 3-, 6- ja 12-kk). Tämä mahdollistaa menetelmien stressitestaamisen useassa eri korkoregiimissä (nollakorot, nousukausi, vakautuminen) niiden vaihdossa. Lisäksi periodi sisältää Silicon Valley Bankin konkurssin ja 2026 Iranin sodan tuomat markkinaturbulenssit.

Tutkielma sisältää taustoituksen relevantteihin teemoihin eli riskimittoihin, korkojen käyttäytymiseen ja ääriarvoteoriaan. Näiden pohjalta aineistoa kuvaillaan, jonka jälkeen metodeja vertaillaan Kupiecin sekä Christoffersenin testien kautta. Ääriarvoteorian tapauksessa lähestymistavaksi valikoitui McNeilin & Freyn (2000) kehittämä menetelmä, mikä hyödyntää GARCH-prosessin sovittamisen tuottamia residuaaleja hännän mallintamisessa.

Ääriarvoteorian etu nojaa sen kykyyn mallintaa jakauman häntäosuus erikseen koko jakauman sijasta. Kaksi tarkinta mallia kvantiilimitan arvioimiseen olivat ääriarvoteoria ja eksponentiaalinen GARCH Studentin t-jakaumalla. Ääriarvoteoria oli hieman robustimpi tuloksissaan. Perusmallit kuten eksponentiaalisesti painotettu liukuva keskiarvo ja perinteinen GARCH(1,1) epäonnistuivat systemaattisesti testeissä. Tappiomitan tapauksessa niin ääriarvoteoria kuin eksponentiaalinen GARCH tuottivat hyviä tuloksia, jotka erosivat taipumuksissaan yli- ja aliarvioida tasoa. Tappiomitan tapauksessa eksponentiaalinen GARCH oli miedosti tarkempi.

Avainsanat: kvantitatiivinen riskienhallinta, ääriarvoteoria, korot, rahoitustaloustiede, EURIBOR, kvantiilimitta

TABLE OF CONTENTS

FIGURES

TABLES

1 Introduction

Interest rates stand as the keystone of the modern financial and economic system. It comes as no surprise their dynamics have been studied for a long time. Governments, banks, insurers and the public alike all have vested interests in understanding not only their movements, but the risks associated in them. The past decade has seen significant movements both in global and European interest rates.

Low (even negative) yields of the 2010s and early 2020s rapidly rose in 2022 as inflation re-emerged into multidecade highs and central banks rushed tackle it. The past years have also seen periods of instability in yield levels such as after the bankruptcy of Silicon Valley Bank. These abnormal market conditions then gave way for a mild decline as inflation started to cool. However, this has been as of writing upended by an energy shock launched by the U.S. and Israeli attack on Iran and the subsequent blockade of the Strait of Hormuz. Comparatively large daily swings in interest rates have again been observed as the markets keep a watchful eye on the conflict and economic indicators related to inflation. It is therefore of paramount importance to study and compare methods of quantifying risk carried in the movement of interest rates.

In Europe and especially in the European Union (EU), the critical benchmark of floating interest rates are the European interbank offered rates (Euribors). With over 100 trillion euros worth of contracts and instruments using them,[1] this interest rate family constitutes the most relevant floating benchmark for the entire Finnish economy.

Dynamics of Euribors and the risk they carry is also naturally of particular interest to the Finnish policy makers. This is exacerbated by the somewhat anomalous behaviour of Finnish houseowners: they tend to prefer floating market rates in their home loans instead of fixed rates. The floating rates of choice have for years been Euribors in Finnish home loans. (Bank of Finland, 2025) This presents a somewhat unique condition for the Finnish banking sector since one of their major income streams is predominantly from variable interest rates.

Financial market and banking authorities have, for a long time, used various risk metrics to set minimum capital requirements for market operators. One of the industry's standard metrics is Value-at-Risk (VaR), which measures worst expected loss in a given quantile (confidence level). Another

[1] The European Money Market Institute (2026)

metric, which builds upon VaR, is the Expected Shortfall (ES). This gives the average loss in the event that VaR is exceeded.

These metrics have been widely implemented in the financial markets by banks, trading desks and central banks[2]. However, there is no universal methodology for calculating these metrics in use. While they may definitionally seem simple, their proper estimation often faces real-world problems and pitfalls due to dynamics observed in financial time-series. A rich econometric literature has therefore grown around their estimation, utilizing a variety of different models and methodologies.

The object of this thesis is to compare two widely used method families in calculating 1-day forward estimates VaR and ES for Euribors. The method families are the GARCH processes and Extreme Value Theory. From GARCH processes, the standard GARCH(1,1) process and the Exponential E-GARCH(1,1) process are implemented. For Extreme Value Theory, the method developed by McNeil & Frey (2000) is implemented. These accuracy of methods are reviewed with backtests developed by Kupiec (1995), Christoffersen (1998) and McNeil & Frey (2000).

The data includes 1820 observations of 3-, 6- and 12-month Euribor changes ending in 26.3.2026 (Section 3 describes the data used in more detail). Estimation is done using a rolling refit of model parameters using the previous 520 observations, leading to 1300 estimates from the models. 520 observations used to prime the models corresponds to roughly two years of market data. While a Euribor tenor of 1-week and an European overnight rate €STR (former Eonia) are also available, their movements are so closely aligned with the European Central Bank’s deposit facility rate, which is used to steer monetary policy, that a different approach would be required.

The chosen sample size was chosen not to test overall superiority of a method over decades, but to see which one provides most robust estimates during changing trends and sustained periods of instability. The markets of recent years provide exactly that. The requirement of consistency is now especially relevant. If observed exceedances of estimated the VaR levels cluster, this can point to an inability by the model to correctly and quickly adapt to new market conditions, even if the number of exceedances is within margins of error.

This thesis is also primarily concerned with quantifying general market risk created by the interest rate movements (changes, returns), not with risk to a specific financial asset tied to Euribors. Each tenor of Euribor is therefore also treated individually instead of a portfolio style approach.

---

[2] Bank of Finland

Lastly, the subject also presents a challenge in terms of theory. To properly conduct the review, understanding of risk theory, interest rate dynamics and extreme limit theory is needed. Section 2 presents theories regarding risk management, section 3 dynamics relevant when approaching interest rates (namely term structure and random walk), section 4 presents the mathematics of Exrreme Value Theory, which is foundationally unique from approaches presented in section 2. Section 5 presents the data used and a principal component analysis of the term structure, while section 6 contains the final comparison of the models. Section 7 concludes.

The result of this thesis therefore is a systematic comparison of risk management approaches in the estimation of daily European interest rate risk, utilizing theories of robust mathematical and statistical foundation. The results can be of use in financial regulation, risk management and within financial market players. The universe of Euribor linked instruments is extensive with different assets having varying sensitivities to changes in rates to one way or the other. It is therefore practical to limit the analysis in this thesis to Euribor changes as a market risk factor.

2 Measuring Risk

2.1 Risk and Randomness

There is no single, universal definition of risk. Yet inherent in it is the element of uncertainty, a chance of adverse effects or events occurring. This notion of uncertainty has naturally led the field of quantitative risk management to heavily rely on the mathematics of probability theory founded on Kolmogorov's definitions of randomness and probability (McNeil et al 2015, 4) While there are other theories which could be considered like, Minsky's financial instability hypothesis and Knightian uncertainty, modern probability theory allows for quantifying and mapping randomness in a mathematical setting.

A chance of something happening (or not happening) can also be a risk (in the negative sense of the word) for some, but a welcome development for others. This is especially true in economics and financial markets, where two parties often interact and price fluctuations can have inverse effects on these parties. A spike in Euribor yields is a negative risk for a household whose home loan is tied to one of the tenors, but the bank financing said loan will receive increased interest payments.

There are multiple categories of risk associated in financial markets such as market risk, operational risk, credit risk, liquidity risk and model risk, all of which are important to acquire a holistic risk management approach regarding financial risk. *Market risk* is the risk of changes in asset prices in the markets. (McNeil et al 2015, 5–6; Dowd 2002, 1–2) Changes in Euribor rates can be considered a market risk factor, since they affect prices of assets tied to them (be it the interbank loans themselves or instruments using them as reference rates).

Measuring risk with a mathematical toolkit requires a model of portfolio value V_t at time t, presented as a function of risk factors. Defining portfolio value as:

$$V_t = f(t, Z_t)$$

where Z_t is a vector of risk factors, and their changes are a random vector $X_{t+1} \coloneqq Z_{t+1} - Z_t$. (McNeil et al 2015, 47–48) Random variables are real-valued functions defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Ω is a nonempty set of possible events (or outcomes). $\mathcal{F}$ is a σ-algebra (sigma-algebra), which is a collection of subsets of Ω that satisfy three conditions:

i) $\emptyset \in \mathcal{F}$

ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$

iii) when $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

A probability measure for $(\Omega, \mathcal{F})$ is then defined as a function $\mathbb{P}: \mathcal{F} \to [0,1]$, which requires countable additivity (i.e. $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ for disjoint sets $A_1, A_2, \ldots$) and that $P(\Omega) = 1$. (Shreve 2004, 1–3, 7–8) A random variable essentially maps the probability space to real values $\Omega \to \mathbb{R}$ (Shreve 1996, 20-21). These conditions for a probability spaceensure that randomness and random variables will behave "nicely"; what can happen (or not) is defined, observable and measurable.

For risk measurement, the focus is often on *distributions* and *expected values* of random variables. A distribution function $F_X(x) = \mathbb{P}(X \leq x)$ assign probabilities for sets in $\mathbb{R}$, while the expected value of a random variable is defined as

$$\mathbb{E}[X] \triangleq \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

and the *conditional expectation* is marked as $\mathbb{E}[X|\mathcal{F}]$, i.e. our expectation of the value of X is based on the information available at σ-algebra $\mathcal{F}$. (Shreve 2004, 68–69; Shreve 1996, 21) For time series forward estimation (such as later in this thesis), this is convenient to combine with definitions of filtrations, which are sequences of σ-algebras, where each preceding one is contained within the next one $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$ (Shreve 1996, 58). When estimating 1-day forward VaR for day *t*, a conditional expectation $\mathbb{E}[X_t|\mathcal{F}_{t-1}]$ and an estimate of the distribution of X_t are needed.

Loss (decrease of portfolio value $\Delta V_t = V_t - V_{t-1}$), while negative, is conventionally denoted as a positive number in risk management, $L_t \coloneqq -\Delta V_t$. If the vector of risk factors z_t at time t is known, the loss function can be presented as $L_{t+1} = -\big(f(t, z_t + X_{t+1}) - f(t-1, z_t)\big)$. (McNeil et al 2015, 48–49) The value function differs between portfolios and various methods for constructing value functions exist. The concern when assessing risk carried in Euribor movements is therefore not with the value function *f*, but with the behaviour of Euribor changes as a market risk factor *X*. Since both tails are of concern in this thesis, and to avoid confusion, *loss* is not converted to a positive value for decreases in interest rates but are denoted with their original sign in the empirical section.

Some early approaches to assessing interest rate risk do not make any assumptions on the probability of a given movement and simply ask "what if interest rates rose (fell) x percent?" One such measure is (Macaulay) duration. Duration is the first order approximation of a bond's present value regarding the discounting interest rate. That is, it measures the sensitivity of a fixed income security to a parallel change (level shift across the entire term structure) in the discount interest rates. A famous approach presented by Markowitz (1952) is his portfolio theory, which treats portfolio selection as a risk-return trade-off, compared to earlier approaches focusing predominantly on just (expected) returns.

A limiting factor of portfolio theory is that it assumes a portfolio of different assets. This, while allowing for risk management through diversification, tells very little about risk dynamics of individual assets other than its variance and covariance related to other assets. Factors such as interest rate risk are not specifically considered.

These models are predominantly focused on average risk for a portfolio or on a specific event with unspecified probability. But when managing risk, one is often concerned with extreme risks. Risks with low probability of occurring, yet with potentially serious consequences. This is behaviour is not automatically considered by any of the approaches listed above. An approach considering the behaviour of interest rate changes itself – especially in the tails of the distribution– is needed.

2.2 Value-at-Risk

Bank for International Settlements defines Value-at-Risk as "*a measure of the worst expected loss on a portfolio of instruments resulting from market movements over a given time horizon and a pre-defined confidence level.*" The bank has also incorporated the measure into the Basel committee regulatory framework in areas such as minimum capital requirements of trading desks. (BIS 2019, 11) Technically, it is not practical to study maximum potential loss, since most probability distributions are infinite in the tails. Therefore, VaR presents an alternative of studying maximum loss in a given quantile. (McNeil & Frey, 2015 p. 64) The measure has gained widespread popularity, especially in financial markets, for its simplicity and various application purposes.

Unconditional Value-at-Risk, i.e. descriptive and not predictive, is presented by McNeil & Frey (2000) in the following form:

$$x_\alpha = \inf\{ x \in \mathbb{R}: F_X(x) \geq \alpha\}$$

In practice, two parameters need to be decided at discretion when utilizing VaR: time horizon and the quantile. For market risk, horizons of 1 and 10 days (2 market weeks) are often used, while for credit risk, horizons are usually 1 year. Some practical constraints and considerations when choosing the time-horizon should be I) minimum commitment period of holding the underlying asset, and II) the risk-factors changing over time, which require recalibration of the model. Some limitations may be set by government regulators or banks for their own traders.

The Basel Framework incorporates VaR measurements as part of backtesting procedures when evaluating banks' internal models for risk assessment. VaR backtesting is required for bank-wide one-day VaR at 99th confidence level ($\alpha = 0.01$), alongside other conditions for profit and loss (P&L). For trading desks specifically, backtesting must be conducted for one-day VaR both at 0.025 and 0.01 quantiles. (BIS 2023, 881–882, 885)

VaR is often used by regulators to set minimum capital requirements for financial institutions such as banks. As shown, internal models are often used provided these are sufficiently superior to regulators' version. A way to roughly calculate what a minimum capital requirement based on the VaR measure would look like is

$$RC^t = \max\{VaR_\alpha^{t,10}, \frac{k}{60}\sum_{i=1}^{60} VaR_\alpha^{t-i+1,10}\}$$

where at time *t*, the 10-day capital requirement is either the daily VaR or the average of the last 60 days of the internal model multiplied by a quality factor *k* (usually 3-4) defined by the regulator (McNeil et al 2015, 67; Dowd 2002, 26).

VaR is a natural progression in risk management from portfolio theory and offers major advantages. While portfolio theory is concerned with volatility/covariance of assets, VaR is more intuitive as a "confidence level" measure. It can also be more easily adapted to measure specific market risks with their own distribution and stochastic behaviour such as interest rate risk. VaR can also accommodate more distributions than portfolio theory, which relies on Gaussian or elliptical distributions. (Dowd 2002, 9–10)

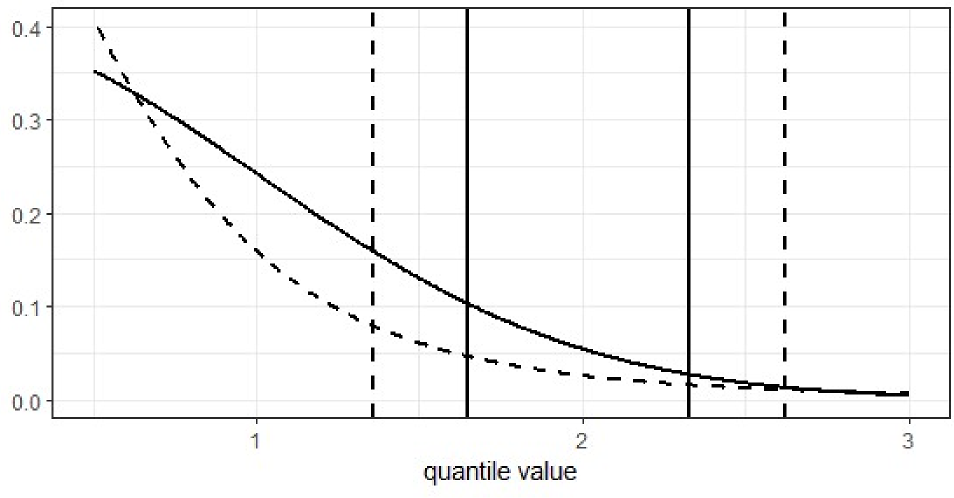

Figure 1 Value-at-Risk

Solid line is standard normal distribution and dashed is Student's t-distribution with 3 degrees of freedom. Vertical lines are 95th and 99th quantiles for each distribution.

For common distributions such as the Gaussian normal distribution and Student t distributions, VaR can be calculated easily if distribution parameters are known. For Gaussian distribution, VaR can be calculated by equation

$$VaR_{\alpha} = \mu + \sigma\Phi^{-1}(\alpha)$$

where μ is the location parameter (mean / expected value), σ is the scale parameter (standard deviation, "volatility") and $\Phi^{-1}(\alpha)$ is the quantile function for standard normal distribution. For student t the form is

$$VaR_{\alpha} = \mu + \sigma t_{v}^{-1}(\alpha)$$

where $t_{v}^{-1}(\alpha)$ is the quantile function with v degrees of freedom. Quantile functions are generalized inverses of distribution functions, where the quantile α of a random variable X with a distribution function F is given by $q_{\alpha}(X) := q_{\alpha}(F) = F^{-1}(\alpha) = \inf\{x \in \mathbb{R}: F(x) \geq \alpha\}$. (McNeil et al 2015, 65–66)

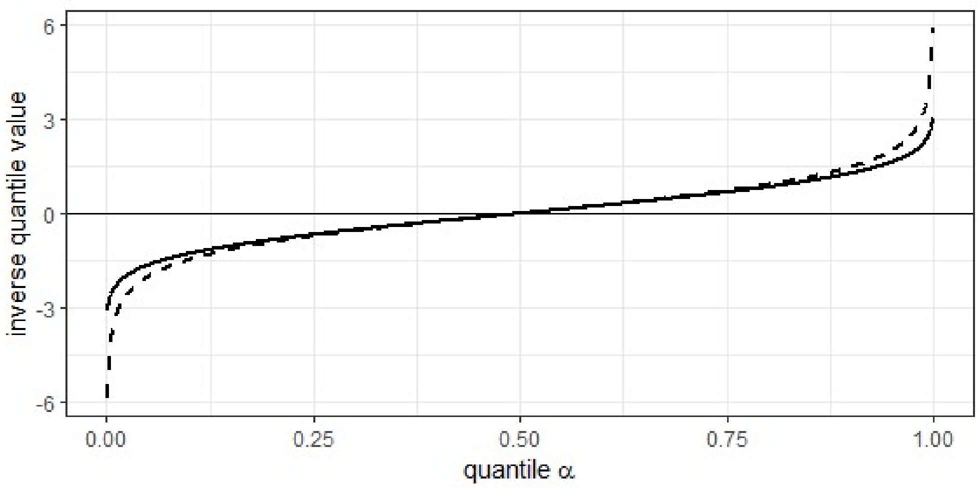

Figure 2 Quantile functions

Solid line is standard normal distribution and dashed is Student's t-distribution with 5 degrees of freedom.

There are some limitations to VaR as a risk measure. One of its deficiencies is that it's not a *coherent* risk measure. Artzner et al (1999) presented four axioms for a coherent risk measure: translation invariance, subadditivity, positive homogeneity and monotonicity. Subadditivity states that the risk of a sum is less than or equal to sum of individual risks. VaR has been shown to violate subadditivity.

However, expected shortfall, which builds upon VaR and measures the average loss if the quantile is exceeded, is a coherent risk measure satisfying subadditivity (cf. section 2.4).

A violation of the subadditivity axiom can lead to biased risk estimates or restructuring of corporations (through demergers) and accounts to reduce combined risk (Dowd 2002, 30). Non-subadditivity can be encountered in specific instances such as on low quantiles of light-tailed distributions, symmetric distributions with a specific dependence structure and very heavy-tailed distributions (McNeil et al 2015, 75–76).

VaR also tells nothing about the losses incurred if the quantile is exceeded and acts simply as a point measure. VaR might be low, while at the same time the potential losses carried in the exceeding tail considerable. This can create subpar incentive structures for traders seeking high return assets, taking on excessive tail risk, while VaRs at chosen quantile might seem relatively similar between assets. Using multiple quantile levels may help mitigate this problem as more information of quantile curvature (Figure 2). (Dowd 2002, 28–29)

2.2.1 Basic Models

The first widely known use of VaR as a risk metric in banking and finance was by RiskMetrics of JP Morgan bank in the 1990s as a measure of daily risk carried by the entire bank's trading portfolio. Most private corporations developed their risk management systems in private, but JP Morgan decided to publish their model and data, which contributed to the rapid spread of VaR within the industry. (Dowd 2002, 8)

RiskMetrics's VaR was calculated at the 95$^{\text{th}}$ percentile ($\alpha = 0.05$) and chosen model distribution was a (multivariate) normal distribution. The mean parameter was assumed to be 0 in 1-day VaR calculations. To estimate the distribution's conditional volatility, they used an exponentially weighted moving average (EWMA). The conditional volatility could be calculated from recursive variance

$$\sigma^2_{t+1|\mathcal{F}_t} = \lambda\sigma_t^2 + (1-\lambda)r_t^2$$

where $0 < \lambda < 1$ is a decay parameter and r_t are the returns/changes. For 1-day volatility, RiskMetrics used a decay parameter of $\lambda = 0.94$ while for 1-month the decay was set at 0.97. Daily observations used were 550 for both windows. However, due to the decay factor only 75 and 150 most recent observations were "effective" respectively. (Risk Metrics 1996, 6–8, 78, 81)

RiskMetrics used historical data but imposes the decay factor based on relevance in terms of time and assumes that the underlying distribution is Gaussian. Another method, which makes no distributional assumptions, is the historical method. The method takes a window of historical observations and the VaR is inferred from observations. This makes no parameter assumptions but has major deficiencies in forward estimation. It is not particularly reliable for high quantiles when available data isn't abundant. It also assumes that returns are *independent and identically distributed* (i.i.d) which contradicts observed behavior of financial time-series (cf. section 2.3.2). (Stoyanov et al 2013, 323)

One could also simulate returns / interest rate changes and infer risk from these Monte Carlo simulations. The approach requires that an (appropriate) statistical model is chosen, its parameters estimated (for example by a historical sample). A simulated sample is then created numerically and risk measures calculated from the sample.

The benefits of Monte Carlo simulation is that it does not require a closed form solution and can accommodate several statistical models (such as dynamic term structure models of section 3.2). This is especially relevant when the portfolio includes complex classes of assets such as derivatives. The deficiency in Monte Carlo simulations is the dependence on the simulated data. (Stoyanov et al 2013, 324)

As Abad et al. (2014) noted, historical simulation carries a weakness rising from dependence on the time-period. This can lead to VaR over- or undermeasuring risk if market conditions change from turbulent to calm or vice versa, since the measurement adapts to new data slowly. A more nuanced model is therefore needed to better capture non-i.i.d dynamics usually observed in financial markets time-series.

2.2.2 Conditional Heteroskedastic Models

Due to the limitations of the basic models arising from statistical behaviour observed in financial time series, models relying on ARCH and GARCH processes that capture more nuanced volatility dynamics have become essential in risk management. Abundant empirical evidence shows that returns on financial time series do not behave like an i.i.d process but show serial correlation with conditional expectations near 0. Volatility also tends to vary and cluster over time, and the return distributions exhibit heavy-tailed, leptokurtic behaviour. (McNeil et al 2015, 79–80, 97)

These facts have led to models trying to estimate or "predict" future states of the process. Serial correlation and volatility clustering suggest there is an element of predictability from recent observations and estimates can be made about the distribution of future observations. Combined with the prevalence of heavy tailed, non-Gaussian distributions has led to practical problems with analytical continuous-time pricing methods such as the Black-Scholes-Merton model, which rely both on the Gaussian distribution and an i.i.d random walk assumption. Time-series models such as GARCH processes which accommodate these dynamics have thus gained popularity. (McNeil et al 2015, 81–85, 97)

In a statistical sense, financial time series of *prices* tend not to satisfy the conditions of *stationarity*, however *returns* (changes of price) often do. (Pagan 1996, 18–21). Time series $(X_t)_{t\in\mathbb{Z}}$ is said to be strictly stationary if its joint distribution does not change over time $(X_{t_1},\dots,X_{t_n}) \stackrel{d}{=} (X_{t_{1+k}},\dots,X_{t_{n+k}})$ or weakly stationary (covariance stationary) if the first two moments exist and satisfy $\mu_t = \mu$ and $\gamma(t,s) = \gamma(t+k,s+k)$, where $\gamma(t,s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$ is the autocovariance function. (McNeil et al 2015, 98; Palma 2016, 76–77)

Heteroskedastic time series models such as ARCH and GARCH processes treat variance of the time series as conditional on past observations. An ARCH(1) process (*autoregressive conditionally heteroskedastic* process) is defined as

$$X_t = \sigma_t Z_t\,, where$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$$

with conditions that constants $\alpha_0 > 0$ and $0 \le \alpha_1 < 1$. This can be extended to include more past observations (lags) in the estimation of conditional variance as an ARCH(p) process of

$$X_t = \sigma_t Z_t\,, where$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2$$

where *p* is the number of past observations used. (Palma 2016, 243–246) Residual $(Z_t)_{t\in\mathbb{Z}}$, often called an *innovation*, is a realization of a strict white noise process, which is a series of finite-valued i.i.d. random variables. The sequence of volatilities $(\sigma_t)_{t\in\mathbb{Z}}$ needs to also be strictly positively valued to satisfy the ARCH(p) process. (McNeil et al 2015, 99, 112)

A property of the above process is that σ_t is $\mathcal{F}_{t-1}$ measurable where filtration $\mathcal{F}_{t-1} = (\{X_s : s \le t\})$ is the information set of past states of X_t. If the expectation of $|X_t|$ is also finite, the conditional expectation reduces to $\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t] = 0$ which satisfied a martingale difference property. Martingales are integrable random variable sequences adapted to some filtration $\mathcal{F}$ and whose conditional expectation of a future state of the sequence is the current state. The first order differences of martingales are thus also expected to be 0. This martingale-difference property becomes useful when using a GARCH(1,1) model in predictions as the optimal prediction for change is no change. (McNeil et al 2015, 99–100, 112, 130)

A *generalized autoregressive conditionally heteroskedastic* GARCH(p,q) process, introduced by Bollerslev (1986) builds on the ARCH process to better account for stochastic volatility in a time-series. The model can be written as (Palma 2016, 190):

$$X_t = \sigma_t Z_t, where$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-i}^2$$

where again innovations $(Z_t)_{t\in\mathbb{Z}}$ is a strict white noise process and parameters α_i and β_j are positive. To satisfy covariance stationarity, a condition of $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_i < 1$.

GARCH processes tend to have periods of persistent volatility trends since either large changes X_t and recent volatility affects the current volatility, while for ARCH processes only the first was the case. GARCH model also has many extensions and variations such as Threshold GARCH (TARCH) and Integrated GARCH (IGARCH) and have been combined with ARMA models. (McNeil et al 2015, 118–123)

Financial returns have often been noted to have some autocorrelation with strong dependence in squared returns, which often leads to the sum of squared parameters being around 1, which indicates near non-stationary behaviour. In cases where this is prevalent, approaches combining ARMA or ARFIMA with the GARCH process can be utilized to account for the autocorrelation. IGARCH processes can also be utilized to account for dependence of squared returns. (Palma 2016, 190–191)

The power of the GARCH process in financial time-series analysis is its ability to capture time-varying volatility dynamics (such as clustering, periods of instability). Volatility is treated as conditional. While the white noise process is often assumed to be normally distributed, other distributions such as Student's t can be fitted into the model to better account for economic time-series behaviour. (Marimoutou 2009, 522)

The fitting of the GARCH process to observed data is often conducted by maximum likelihood estimates (Angelidis et al 2004, 110). In practice, numerous solving techniques are available. The R package *rugarch* includes five different numerical solving techniques including the augmented Lagrangian solver (Ghalanos 2026, 25).

The basic GARCH model assumes that current volatility is symmetrically affected by both positive and negative innovations of the strict white noise process. This contradicts observed financial time series behaviour with asymmetries in conditional distributions. The observation that volatility tends to increase from negative events is called the leverage effect and models such as the *Exponential GARCH* have been constructed to better account for these asymmetries. An EGARCH(p,q) process is defined as

$$X_t = \sigma_t Z_t$$

$$\log \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, g(Z_{t-1}) + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2$$

$$g(Z_{t-1}) = \theta Z_{t-1} + \varsigma(|Z_{t-1}| - \mathbb{E}[|Z_{t-1}|])$$

where $\{\alpha_i, \beta_j, \theta, \varsigma\} \in \mathbb{R}$. This specification relaxes the positivity constraints of the basic model, accounts for asymmetricity. Conditional variance is formed with standardized past innovations instead of untreated past innovations. (Franc et al 2010, 245-247)

A GARCH process can be used to estimate the 1-day forward VaR of the distribution $F_{X_{t+1}|\mathcal{F}_t(x)}$. Once the mean and volatility parameters have been calculated, the quantile function is estimated for the strict white noise process of the model (McNeil et al 2015, 133):

$$VaR_\alpha^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} q_\alpha(Z)$$

2.2.3 Empirical evidence

In the case of diversified portfolios, Angelidis et al (2004) found that using leptokurtic distributions produce better 1-day forward estimates of portfolio VaR. Student's-t distribution produced better forecasts in GARCH and EGARCH processes compared to threshold-GARCH (TARCH) processes and improved probability values for multiple portfolios compared to normal distribution and *generalized error distribution* (GED). It was also recognized that each portfolio used had different model structures that best represented its VaR estimate. This variability in best fit distributions and model specifications also significantly relies on the chosen sample size since.

The conclusion of non-existence of an overpowering GARCH model specification in VaR estimation has also been reached in a literature review of multiple VaR methods by Abad et al (2014) and by Slim et al (2016) in a performance review of multiple GARCH family processes with different distribution specifications on global stock indexes. Slim et al found that model performances differed between developed and developing markets. Non-normal Lévy and skewed student-t distributions were, however, found to be an overall best distribution to use. Abad et al note that estimates are considerably improved when the Gaussian distribution is switched to a fat-tailed one. Additionally, they found that model specifications accounting for the leverage effect (such as EGARCH) appear to have better results.

Babis (2024) systematically compared multiple GARCH processes paired with different parametric distributions for VaR forecasting of European large stocks during the Covid-19 crisis period. The crisis allowed for testing of the methods in abnormal market conditions which differed substantially from the prior turbulent period of 2008-2009. Student-t (standard and skewed variants) distributions were again found to perform better compared to their Gaussian counterparts. Models such as IGARCH and FIGARCH were in this case found to outperform asymmetric models. However, the importance of volatility specification compared to distribution specification differed for left and right tails. The same question of volatility or distribution importance depending on the tail was also noted by Slim et al (2016).

GARCH processes have also been augmented with machine learning techniques in VaR estimation to try to improve conditional volatility estimation. Incorporating a *long short term memory* (LSTM) neural network approach into the process, Buczynski & Chlebus (2023) showed the augmented method was in some cases able to identify volatility shocks much faster and provided good VaR estimates.

Promising results have also been shown in the oil markets by Youssef et al (2015) where fractionally integrated asymmetric GARCH (FIAGARCH) was found to capture long term volatility dynamics better than FIGARCH and hyperbolic GARCH. However, this model was itself further outperformed by combining it with an EVT approach. Also studying crude oil prices, Fan et al (2008) implemented a GED-GARCH approach and found it to outperform the conventional Gaussian GARCH in both in-sample estimation and out-of-sample forecasting.

In the case of large price swings and volatilities of cryptocurrencies, standard GARCH models were found to inadequately estimate VaR levels by Fiszeder et al (2024). To account for these market dynamics, the authors combined a range-GARCH (RGARCH) with a modified bounded M-estimator to achieve better results.

The empirical literature on the use of GARCH processes to model VaR levels generally gives good results for their use but is often inconclusive in defining a single superior model from the GARCH family. Heavy-tailed distributions such as Student-t are generally found to outperform Gaussian distributions. These findings highlight the need for testing different model specifications for each Euribor tenor and expect Gaussian distributions to fail compared to heavier tails as the used quantile is tightened.

While the family GARCH processes are an improvement from traditional approaches by their ability to capture clustering and time varying volatility dynamics, the literature often finds superior performance in even more nuanced approaches. Approaches implementing Extreme Value Theory have been promoted by papers such as Abad et al (2014), Calmon et al (2021), Youssef et al (2015). One of the benefits of utilizing GARCH processes is the rich variety of specifications available. The engine of heteroskedastic volatility is shown to be easily augmented with neural networks, range-based models, skewed distributions and asymmetric volatility dynamics. The process can also be further utilized as part of an EVT approach. The question when implementing a GARCH approach is therefore which model specifications to choose.

2.3 Expected Shortfall

Expected shortfall (ES), also known as tail conditional expectation or estimated tail loss is defined by BIS as "*a measure of the average of all potential losses exceeding the VaR at a given confidence level"*. (BIS 2019, 11) The Basel Framework has started incorporating ES to risk management. In internal models, ES needs to be calculated at the 97,5th confidence level and calibrated for a period of stress alongside other specific considerations. (BIS 2023, 894–902)

ES is therefore closely related to VaR. Mathematically, it can be defined as the conditional expectation of observations exceeding a given quantile α. McNeil & Frey (2000) present it in the form:

$$ES_\alpha = \mathbb{E}[X | X > VaR_\alpha]$$

The integral version of ES can be defined through VaR and quantile functions (McNeil et al 2015, 69–70):

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^u VaR_u du = \frac{1}{1-\alpha} \int_\alpha^u q_u(F) du$$

In the case of the Gaussian (normal) distribution ES can be calculated from the equation

$$ES_{\alpha} = \mu + \sigma \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha}$$

where ϕ is the Gaussian density function and Φ^{-1} is again the Gaussian quantile function for- For Student-t distribution, the measure is given by

$$ES_{\alpha} = \frac{g_v\left(t_v^{-1}(\alpha)\right)}{1-\alpha}\left(\frac{v+\left(t_v^{-1}(\alpha)\right)^2}{v-1}\right)$$

with g_v as the density function and t_v^{-1} as the quantile function for v degrees of freedom. (McNeil et al 2015, 70–71)

There are multiple benefits to using ES instead of VaR as a risk measure. As with VaR, it also presents a consistent risk measure for different asset classes or positions. ES is also a coherent risk measure, satisfying the axioms of Artzner et al (1999). As a coherent risk measure, ES does not discourage diversification, unlike VaR in some cases (stemming from non-subadditivity) and provides well-behaved optima in risk optimization problems (coherence ensures convexity of risk surface). It is also even more intuitive than VaR, since it tells how bad the losses are expected to be in bad conditions, instead of how much it at stake at a specific level. (Dowd 2002, 33–35)

ES is, however, often quite sensitive to model specifications regarding the tail since it accounts for the entire mass exceeding α. This is why specifications relying on Gaussian distribution often fail at proper estimation. (McNeil & Frey 2000, 291–292) While VaR value might remain relatively stable over time, the mass exceeding it can vary greatly, making ES a more sensitive measure.

ES can also be extended into shortfall-to-quantile ratio $\frac{ES_{\alpha}}{VaR_{\alpha}}$, provided that μ is small. This can be used to measure how heavy the tail is past quantile α. For Gaussian distributions the ratio converges to 1 as $\alpha \to 1$. For Student-t, the ratio is $\frac{v}{v-1} > 1$ as $\alpha \to 1$. (McNeil et al 2015, 72)

3 Term Structure of Interest Rates

The stylized facts of financial time series behaviour presented in section 2 are useful for constructing models that capture econometric phenomena of single tenor interest rates. However, dynamics of interest rates go deeper than simple statistical properties of a tenor. This chapter presents some of the theories and models that aim to explain *why* interest rates behave as observed, thus allowing for a deeper discussion on observed behaviours. The models covered are i) the traditional models of term structure of interest rates and ii) affine term structure models which rely on arbitrage-free, risk neutral pricing theorems and Itô calculus. For affine models, these are the same building blocks of the famous Black & Scholes option pricing theory.

While the methods applied in section 6 are for a univariate case, that is, each tenor is considered independently from others, it is nonetheless important to consider interest rates in a wider context of the term structure. A compact review of theories of interest rate dynamics is therefore warranted. As Brigo & Mercurio (2006) note, a bad model choice for short-rate behaviour will result in a poor model for yield curve modelling.

Zero-coupon bonds (pure discount bonds) are contracts that give the holder one unit of currency at maturity T. The value of the contract at time $t < T$ is denoted as $P(t, T)$. These bonds can then be used to calculate the short rates (spot rate). These rates can in turn be used to construct the *term structure of interest rates* (yield curve), which tells how the current interest rates evolve when dependent on maturity. These short rates can be calculated in discrete or continuous time. For continuous time the rates are given by $R(t, T) = -\frac{\log P(t,T)}{T-t}$, while for discrete time the rates are calculated as $L(t, T) = \frac{1-P(t,T)}{(T-t)P(t,T)}$. The zero-coupon bonds and their corresponding term structure can be calculated from government bonds or interbank offered rates such as EURIBOR. (Brigo & Mercurio 2006, 1, 4–7, 9–10)

Rates often exhibit vector autoregression over 1-month and 12-month periods making factor models an important tool for studying the behaviour of the term structure. Around three principal components have been observed to explain almost all of the variation. The 1st factor appears to drift over time and, for U.S. Treasury bills, follows long run inflation trends and has high persistence. 2nd factor appears to move in a pattern of a business cycle (affecting the slope of the term structure), while the 3rd (affecting curvature) behaves most stochastically. These factors are also effectively linearly uncorrelated with each other. (Diebold & Rudebusch 2013, 7–13)

So, while short term changes can behave in a stochastic manner, long run dynamics of the yield curve have been observed to be closely related to macroeconomic drivers. Inflation is observed to correlate with the level of term structure, while real economic activity correlates with the slope. The third principal component does not seem to have a close relationship with any major macroeconomic phenomenon. Additionally, the macroeconomic variables have been found to have a stronger causality on the yields than vice versa. The factors have also been shown to relate to central bank policy. Level factor has been shown to have linkages to central bank inflation targets while the slope factor relates to the cyclical dynamics stemming from central bank policy mandates (U.S. Federal Reserve has a mandate of both target inflation and economic stability). (Diebold & Rudebusch 2013, 19–22)

Nelson & Siegel (1987) presented a factor model for the yield curve, which fits available rate information to achieve a smooth yield curve. The model takes four parameters and allows the yield curve to take multiple different shapes. The short rate curve is given by the model by equation

$$R(t,T) = \beta_0 + (\beta_1 + \beta_2)\frac{\tau}{T}(1 - e^{-T/\tau}) - \beta_2 e^{-T/\tau}$$

The model has been extended for even more complicated shapes of yields curves, such as Svensson's (1995) extension of a 4th term, with 2 additional parameters. The spot rates are in this case given by the equation:

$$R(0,T) = \beta_0 + (\beta_1 + \beta_2)\frac{\tau_1}{T}\left(1 - e^{-T/\tau_1}\right) - \beta_2 e^{-T/\tau_1} + \beta_3\left(\frac{\tau_2}{T}\left(1 - e^{\{-T/\tau_2]}\right) - e^{-T/\tau_2}\right)$$

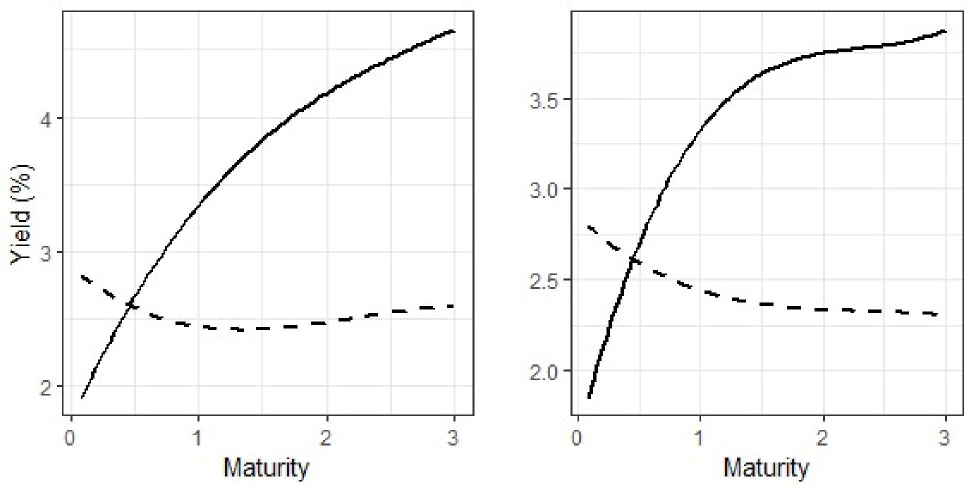

Figure 3 Yield Curve fits

Nelson-Siegel and Svensson methods implemented for Euribors. First observations of Januarys 2023 and 2024. Solid being 2023 and dashed 2024.

The Nelson-Siegel model can be extended into a dynamic form which accounts for time-varying behaviour in the factors. In such a model, the three β_i parameters again describe the level slope and curvature, but can also be described as long-, short- and term-term factors respectively due to their weight on different parts of the term structure. (Diebold & Rudebusch 2013, 26–29)

While term structure models such as Nelson & Siegel (1987) and Svensson (1995) above are well suited to explain shape and behaviour of the term structure in terms of factors driving the shape of the curve, they aren't the first nor only theory aiming to explain interest rate dynamics. Another theory is that of short rate models which rely on random walk processes and aim to explain interest rate movements in the context of arbitrage-free pricing of assets.

In these models the short rate dynamics are driven by a *stochastic differential equation* (SDE), under probability measure $\mathbb{P}$ of a form:

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dW_t,$$

where W_t is a random walk Wiener process (a Brownian motion) and μ (drift) and σ (diffusion) are well behaved functions. Thus, the rate evolves based on initial state r_0, random walk, and properties of the two functions. A technical problem with Brownian motion is that it is not Riemann-integrable, which leads to the use of Itô calculus. Mainly the Itô integral. (Björk 2019, 272, 45)

Affine term structure (ATS) models more specifically are interest rate models where the zero-coupon bond price dynamics can be represented in the form of $P(t,T) = e^{A(t,T)-B(t,T)r(t)}$ with deterministic functions *A* and *B* (Cairns 2018, 74–75; Brigo & Mercurio 2006, 68–69). Filipović (2001) characterized sufficient and necessary conditions that yield a non-negative one-factor ATS, which can in some cases be sufficed by non-continuous processes and processes which also exhibit *jumps*. Additionally, Chirstensen et al (2010) showed that the dynamic Nelson-Siegel model also satisfies affine term structure.

Some of the SDE dynamics produce explicitly non-negative rates and not all produce analytical bond prices or option prices. The distribution dynamics for many are based on Gaussian properties, but some follow a chi-squared distribution. The Dothan Model, for example is a simple lognormal dynamic of $dr_t = ar_t dt + \sigma r_t dW_t$, while some models such as CIR++ are significantly more complex. (Brigo & Mercurio 2006, 57, 102–104)

Vasiček's (1977) model $dr_t = k(\theta - r_t)dt + \sigma dW_t$ relies on an Ornstein-Uhlenbeck process, which produces a walk of interest rate with a tendency to revert to a mean rate. This process has a Gaussian distribution and allows for negative interest rates. The Cox, Ingerson and Ross (CIR) model is an extension from the work of Vasiček, which yields solely non-negative rates. The model's dynamics are driven by an SDE of the form $dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t$ with a condition of $2k\theta > \sigma^2$. (Brigo & Mercurio 2006, 58–59, 64–65)

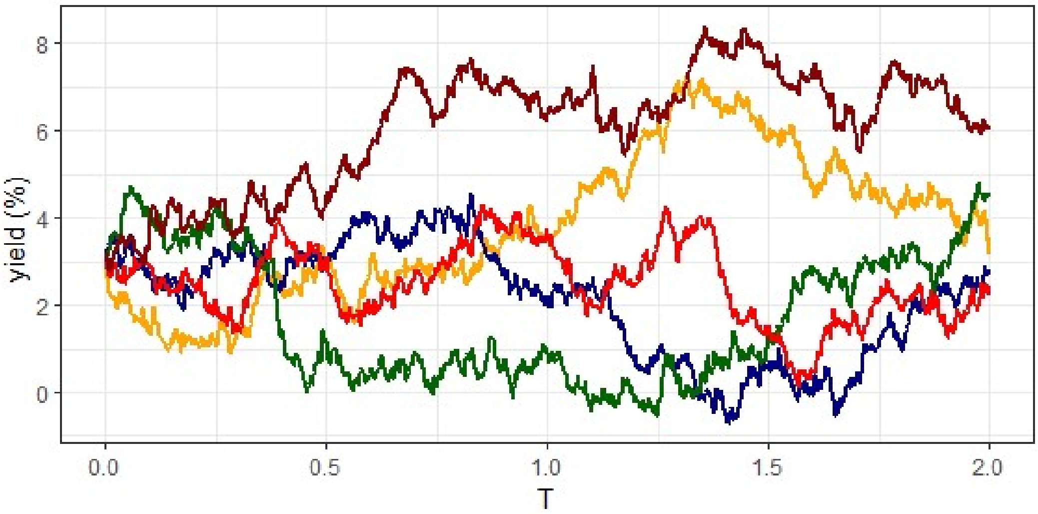

Figure 4 Vasiček model

Simulation with Euler-Maruyama method for T=2, $r_0 = 3$ and 1000 steps.

Some of the critiques of early short rate dynamics models is that they cannot be well calibrated to available market data. Furthermore, the term structures generated by these models (such as Vasiček and standard CIR) hardly match those observed. Some of the shapes such as an inverted yield curve are hard to produce with such models. Therefore, the success they've generated has largely been on the side of analytical asset pricing. (Brigo & Mercurio 2006, 57–58)

If one relaxes the requirement of continuous-time and allows for discrete time in ATS (DTATS) models, a GARCH style volatility dynamic can be fitted such as by Realdon (2025). This relies on transformations where volatility can be heteroskedastic under probability measure $\mathbb{P}$ but homoscedastic under risk neutral measure $\mathbb{Q}$ for bond pricing. It was noted that unspanned (i.e. not affecting yield or bond price calculations) GARCH DTATS models can improve VaR calculations for fixed income assets.

4 Extreme Value Theory

Extreme Value Theory (EVT) focuses on modelling tail distributions of extreme events. This has naturally led to widespread use of the theory in financial risk management. While other models usually focus on estimating the entire distribution of underlying returns, EVT focuses solely on tail events and their distribution shape. This approach is naturally attractive for risk management as it does not "waste" time modelling the center bulk of the distribution and uses relevant information (tail events) more powerfully. There is plentiful evidence in the literature for the superior performance of EVT in higher quantiles compared to GARCH models. Indeed, EVT methods can even implement these volatility models as part of their process, thus building on their estimates.

Two main branches of extreme value theory methods exist. 1st is the "block maxima" method and the 2nd is the "threshold exceedances" or "peaks over threshold" (POT) method. While the methods are in some parts very similar, there are mathematical differences that affect, for example, extreme limit converge. This paper implements the threshold exceedances method since it often better accounts for statistical properties observed in financial time series than the block maxima. It is also considered to be more efficient in its usage of extreme values than block maxima (Abad et al 2014, 24). However, the mathematical foundations of EVT will be presented within the literature review of Block Maxima. A review of the Hill estimator (a possible alternative to block maxima and threshold exceedances) is included as part of review of threshold selection, yet it only works in one class of extreme value distributions (McNeil & Frey, 2000 p. 287; Nieto & Ruiz, 2016, 482).

This chapter discusses extreme right tail events in terms of maxima, that is, threshold exceedance to above instead of below. When discussing extreme left tail events, we can utilize these theories by conversion between minima and maxima, $\min(X_1, X_2, \dots X_n) = -\max(-X_1, -X_2, \dots, -X_n)$. As a further note, while referencing multiple sources, this chapter uses the notation style of McNeil et al (2015) where applicable to maintain consistency and avoid confusion.

4.1 Block Maxima

The foundational mathematical theorems behind EVT were laid by Gnedenko (1943) and earlier (but with non-rigorous proofs) by Fisher and Tippett (1928) and von Mises (1936) in defining three types of distributions of extreme values and their conditions for domain of attraction (Smith, 1991 p.186). Central limit theorem (CLT) states that appropriately normalized sums of random variables converge in their distribution to standard normal distribution $\frac{(S_n - n\mu)}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$. In the case of limit behaviour of extreme values, we can note that nondegenerate (i.e. not concentrated on a single point) limit distributions of maxima converge to the general extreme value distribution, which consists of the three distributions (as shown by Gnedenko). The driving mechanism in block maxima is essentially to partition the underlying data into blocks, pick the maximum observation from each block and then to model the distribution of these maximum observations. (McNeil et al 2015 p. 136)

4.1.1 Method

Supposing we have a sequence of n i.i.d. random variables X with an underlying distribution F. We note that $\max(X_1, X_2, \dots, X_n) \xrightarrow{\mathbb{P}} \sup\{x: F(x) < 1\}$ and $\mathbb{P}(\max(X_1, X_2, \dots, X_n)) = F^n(x)$. If we then have two standardizing sequences of constants $a_n > 0$ and b_n with which the equation

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n}$$

has a limit distribution

$$\lim_{n\to\infty} F^n(a_n x + b_n) = H(x)$$

which is nondegenerate, then distributions F satisfying these conditions are in the *maximum domain of attraction of H*. (de Haan & Ferreira, 2006 p. 3-4) As Gnedenko showed, there are only three distribution types of H. While exact distributions may differ, they are said to be of the same type if there is a possible transformation $F(x) = G(Ax + B)$ with constants $A > 0$ and B. Additionally, if two different distributions F and G are reached with different sequences of constants, they are said to be of the same distribution type. (Smith, 1991 p.186) This is called converge to types (cf. Ebrechts et al. 1997 p. 554).

These possible types of distribution are the Fréchet, Weibull and Gumbel distribution families:

1. $H_{Gumbel}(x) = \exp(-e^{-x}), \quad -\infty < x < \infty$
2. $H_{Fréchet}(x) = \exp(-x^{-\xi}), \; x > 0, \xi > 0$
3. $H_{Weibull}(x) = \exp\left(-(-x)^{\xi}\right), \; x \le 0, \; \xi > 0$

Each of these families have their own maximum domains of attraction (MDA). Most continuous distribution functions have an MDA. Finding the MDA for a distribution function is sometimes a simple matter as in the case of a uniform distribution, yet for some distribution functions the process can be extensive due to MDA criteria. (Leadbetter, 2016 p. 12-14)

In the case of the Gumbel type, distributions such as the exponential, normal, lognormal and hyperbolic are all in its MDA, leading to great variety within the distributions, with some tails heavier than others. However, a common element in this type is exponential decay in the tail and the right endpoints satisfy $x_F \leq \infty$ which allows for a finite endpoint. The distribution functions in the MDA of the Fréchet type, however, stem from slowly or regularly varying functions. More specifically, the condition can be written as $F \in MDA(H_{Fréche}\ ) \Leftrightarrow \bar{F} = x^{-1/\xi}L(x),\ \ \xi > 0,\ L \in \mathcal{R}_0$. These distributions are heavy tailed with infinite moments and an infinite right endpoint. This family includes student-t, Pareto and inverse gamma distributions. Lastly, distributions in the MDA of Weibull type all have finite right endpoints. These properties make the Fréchet type most common in financial applications, while the Weibull type is rarer in market risk applications and more common in credit risk management. (McNeil et al 2015 p. 138-141)

These distribution types can be further combined to achieve the Generalized Extreme Value Distribution (GEV):

$$H_{\xi,\mu,\sigma}(x) = H_\xi((x-\mu)/\sigma) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right), & \xi \neq 0 \\ \exp\left(-e^{-x}\right), & \xi = 0 \end{cases}$$

The value of our shape parameter ζ defines the distribution type of our tail distribution. With ζ < 0 we get a short-tailed Weibull distribution, with ζ > 0 a Fréchet distribution and in the case of ζ = 0 a Gumbel distribution. (McNeil et al 2015 p. 136)

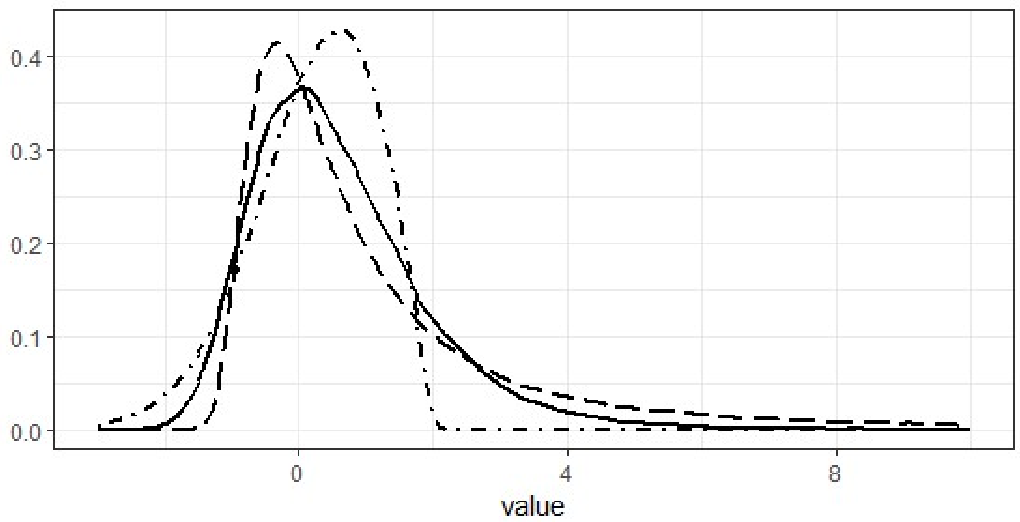

Figure 5 GEV distribution

Solid = Gumbel, dashed = Fréchet, dot-dashed = Weibull. Parameters $\mu = 0,\ \sigma = 1,\ \xi\ =\ \{-0.5,0,0.5\}$.

In the case of interest rates and financial time series more generally, this i.i.d definition needs to be expanded into stationary time series. McNeil et al (2015 p. 141-142) note that the maxima of an associated strict white noise process $(Z_i)_{i\in\mathbb{Z}}$ of a strictly stationary time series $(X_i)_{i\in\mathbb{Z}}$ converges to the same type of distribution type (and with the same ζ parameter) as the normalized maxima of the time series provided there exists an extremal index $\theta \in (0,1]$. This extremal index is provided for many processes as follows:

$$\lim_{n\to\infty} \mathbb{P}((\max(Z_1, Z_2, \dots, Z_n) - d_n)/c_n) \le x) = H(x)$$

if and only if

$$\lim_{n\to\infty} \mathbb{P}((\max(X_1, X_2, \dots, X_n) - d_n)/c_n) \le x) = H^{\theta}(x)$$

And where $H(x)$ is nondegenerate. Furthermore, in the case of ARCH and GARCH processes, the extremal index is $\theta < 1$.

In estimating the distribution, two methods are widely used with block maxima: the probability weighted moments and maximum likelihood (de Haan & Ferrera, 2015 p. 277). To fit the GEV distribution using maximum likelihood, each maximum M_{nm} of *m* blocks (sized *n*) is taken and the log-likelihood is calculated by maximizing subject to parameter constraints (Jakata & Chikobvu, 2025 p. 59):

$$l(\xi,\mu,\sigma | \textstyle\sum_{i=1}^{m} M_{ni}) = -m \ln \sigma - (1 + 1/\xi) \displaystyle\sum_{i=1}^{m} (1 + \xi(M_{ni} - \mu)/\sigma) - \sum_{i=1}^{m} (1 + \xi(M_{ni} - \mu)/\sigma)^{-1/\xi}$$

To obtain Value-at-Risk from block maxima for a stationary time series, the extremal index is used again. A property given by the limit condition for stationary series converge given above is (Bloss et al, 2025 p. 9; McNeil et al 2015 p. 141):

$$\mathbb{P}(M_n \le u) \approx \mathbb{P}^{\theta}(Z_n \le u) = F^{n\theta}(u)$$

provided the blocks are large enough. This is used to estimate the VaR for a given quantile. Denoting α-quantile of the GEV distribution as q_α we have

$$q_\alpha = H^{-1}(\alpha) = \mu - \frac{\sigma}{\xi}(1 - (1 - \log(\alpha))^{-\xi})$$

noting that as $n \to \infty$, $H(q_\alpha) = \mathbb{P}(Z_n \leq q_\alpha) = F^{n\theta}(q_\alpha)$, with a VaR of $H(VaR_\alpha) = F^{n\theta}(VaR_\alpha) = \alpha^{n\theta}$, which satisfies the VaR definition for distribution F. After this, the expected shortfall can be solved by analytical integration. (Bloss et al, 2025 p. 8-9)

An alternative approach is to calculate the return levels R_n^k from the distribution, which are the k quantiles of returns. A return level $R_{365}^{10} = 0.20$ would mean a 20% daily return occurring every 10 years. Return levels are estimated by $\widehat{R_n^k} = H_\xi^{-1}\left(1 - \frac{1}{k}\right)$. (Idirizov, 2025 p. 3-5) An explicit solution for this in the case of the GEV distribution is $\widehat{R_n^k} = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(-\ln\left(1 - \frac{1}{k}\right)\right)^{-\hat{\xi}} - 1\right)$ and corresponds to distribution H's 1-1/k quantile (McNeil et al, 2015 p . 144-145).

4.1.2 Practical use

A key weakness of the block maxima is its requirement for large amounts of data, which needs to be divided into blocks large enough to get credible observations of maxima. This also means that block maxima wastes a lot of data. Additionally, the extremal index can sometimes be hard to accurately estimate. (McNeil et al, 2015 146)

There are certain conditions under which block maxima however is preferrable to threshold exceedances as the method of choice. The data of interest might already be naturally divided into large blocks (years, months) or the data available might already be block maxima, such as yearly maximums of underlying observations. Statistically, the data might also not be i.i.d within blocks but satisfy the condition between blocks. In practice, block maxima seems to have comparable performance with threshold exceedances when the sample size is large. Threshold exceedances also often requires exceedances to be more numerous than number of blocks to beat block maxima. At the theoretical level, the block maxima level is also rather efficient when comparing with threshold exceedances. (de Haan & Ferreira, 2015 p. 276-278)

Using block maxima in economic and financial risk analysis is therefore suitable when one has a lot of structured data and clustered instability (peaks) are not central to the research. For example, Li (2024) studied risk profiles of three U.S. stock market indexes (S&P 500, Nasdaq Composite and DJIA) using monthly maxima from 360 months, showing that the Nasdaq Composite index is the riskiest of the three and upside risk of DJIA the least risky. Idirizov (2025) used block maxima in the commodity markets to assess extreme behaviour of silver prices between 1983 and 2024 using annual maxima and minima. Using return levels, it was found that daily price gains exceeded 9.11% once every ten years, while for daily price drops the level was 10.24%.

4.2 Threshold Exceedances

Due to the limitations of block maxima, threshold exceedance (also known as peaks-over-threshold) methods have gained popularity over it in quantitative risk management in finance. Threshold exceedance uses all data points which exceed a certain high threshold.

However, in threshold exceedance the extreme distributions do not converge to GEV as is the case in block maxima. Instead, these distributions converge towards a *generalized pareto distribution* (GPD). The GPD uses the same shape parameter ζ as GEV distributions.

$$G_{\xi,\beta}(x) = \begin{cases} 1-(1+\xi x/\beta)^{-1/\xi}, & \xi \neq 0 \\ 1-\ \exp(-x/\beta), & \xi = 0 \end{cases}$$

In the case of $\xi < 0$ we again get a short-tailed distribution type, which is the Pareto type II, and has a finite right endpoint. With type $\xi = 0$, there is an exponential distribution, while distributions of type $\xi > 0$ are again heavy tailed such as Fréchet, power functions and Student's t. The additional parameter β is a scale parameter. (McNeil et al, 2015 p. 147; McNeil & Frey, 2000 p. 281)

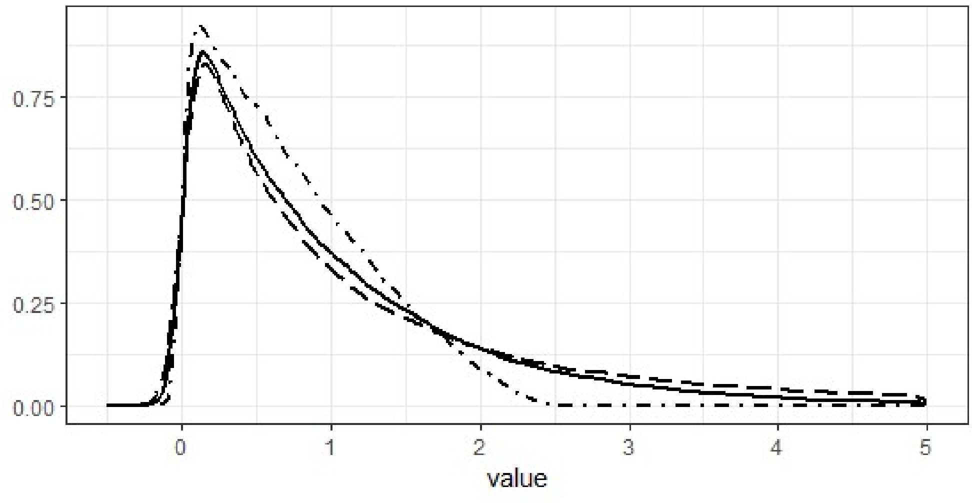

Figure 6 Generalized Pareto Distribution

Solid: $\xi = 0$, dashed: $\boldsymbol{\xi} = \mathbf{0.4}$, dot-dashed: $\xi = -\mathbf{0.4}$, with $\beta = 1$ for each.

The threshold exceedances relies on the Pickands-Balkema-de Haan theorem to show converge to GPD, which also utilizes MDA to GEV types as in block maxima. Defining the excess distribution above threshold u as:

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

where u is our chosen threshold. The theorem states that if and only if $F \in MDA(H_\xi), \xi \in \mathbb{R}$ there is a function $\beta(u)$ such that:

$$\lim_{u \to \infty} \sup_{0 \leq x < x_F - u} \left| F_u(x) - G_{\xi, \beta(u)}(x) \right| = 0$$

and the function $\beta(u)$ is positive-measurable. Noting that the theorem uses GEV distribution MDA criteria leads to a conclusion that the same set of distributions converging to GEV types also converge to GPD as the threshold is raised. The shape parameter ζ also stays the same in GPD as in GEV. (Bloss et al, 2025, 10; McNeil et al, 2015 p.147-149; McNeil & Frey, 2000 p. 280-281; Scarlott & MacDonald, 2012 p. 35)

4.2.1 Method

The threshold exceedances approach presented in this section, and later implemented in section 5, was developed by McNeil & Frey (2000), which is a variation of the peaks-over-threshold (POT) method. This is a 2-step approach. In step 1, a GARCH model is fitted to the returns to estimate mean, volatility and implied innovations (residuals) (cf. section 2.2.1). The innovations are then used in step 2 to model the tail using EVT. A GARCH(1,1) makes minimal assumptions about the underlying process of returns and thus allows for concentration on the actual EVT model. McNeil & Frey used a pseudo maximum likelihood (PML) approach with normal innovations to fit the GARCH-model into data. (McNeil & Frey, 2000 p. 278).

The method utilizes the mean and volatility given by the GARCH(1,1) model, which leaves only the innovation (residual) distribution to be calculated using EVT. That is, from $VaR_\alpha = \mu_{t+1} + \sigma_{t+1} z_\alpha$ our interest after GARCH model fitting is z_α. The underlying assumption again is that daily returns can be represented by a process $X_t = \mu_t + \sigma_t Z_t$ where Z_t is a strict white noise process. Additionally, mean and volatility are $\mathcal{F}_{t-1}$ measurable. The conditional 1-day forward VaR is then given by $VaR_\alpha = \inf\{x \in \mathbb{R}: F_{X_{t+1|\mathcal{F}_t}}(x) \geq \alpha\}$. (McNeil & Frey, 2000 p. 275-276)

Noting $N_u > 0$ as the number of observations from sample *n* that exceed threshold u and assuming they are i.i.d and taking into consideration the equality

$$1 - F(x) = \big(1 - F(u)\big)\big(1 - F_u(x - u)\big)$$

where $x > u$, a tail estimator is obtained:

$$\widehat{F(x)} = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}(x-u)/\hat{\beta}\right)^{-1/\hat{\xi}}$$

It is also noted that the estimates for GPD parameters $(\hat{\xi}, \hat{\beta})$ are asymptotically normal and consistent when $\xi > -\frac{1}{2}$ and asymptotically normal under a weaker assumption. Asymptotic unbiasedness is achieved as long as u approaches x_F fast enough, which is dependent on the converge rate of Pickands-Balkema-de Haan theorem's equation. This leads to a trade-off between bias and variance: a high u reduces bias while a large N_u keeps the variance low. (McNeil & Frey, 2000 p. 281)

While N_u is in theory random, in practice McNeil & Frey (2000 p. 282) fix this number to a constant $k \ll n$. Next, the residuals are ordered, where $z_{(1)} \geq z_{(2)} \geq \ldots \geq z_{(n)}$. Here, observation $z_{(k+1)}$ represents the threshold u, which is in this case random. The excess amounts over threshold are then used to obtain the tail estimator (McNeil & Frey, 2000 p.281-282):

$$\widehat{F_Z(Z)} = 1 - \frac{k}{n}\left(1 + \hat{\xi}\left(z - z_{(k+1)}\right)/\hat{\beta}\right)^{-1/\hat{\xi}}$$

From this, we get the quantile (as long as $\alpha > 1 - k/n$) for innovations:

$$\widehat{z_\alpha} = z_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{\hat{\xi}} - 1\right)$$

When calculating the left tail (chance of interest rate drop) we invert the innovations and calculate their right tail. This way, both α and $1 - \alpha$ quantiles are modelled with GPD extending to the right for easier calculations.

Combining this result with the conditional mean and volatility estimates from step 1, the following Value-at-Risk representation is obtained:

$$VaR_\alpha = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\widehat{z_\alpha}$$

The expected shortfall can be then solved analytically in threshold exceedances as (McNeil et al, 2015 p. 154):

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_x\,dx = \frac{VaR_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi}$$

which has, in the case of McNeil & Frey's approach, an explicit conditional expectation form of (McNeil & Frey, 2000 p. 292-293):

$$ES_t^{\alpha} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\left(\frac{\hat{z}_{\alpha}}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} z_{(k+1)}}{1-\hat{\xi}}\right)$$

This approach makes minimal assumptions on the underlying data, yet two choices are subjective to user's discretion. The prominent one is choosing a proper threshold level u (in this case the ratio of exceedances to sample k/N). This presents a dilemma since picking a low threshold u risk biasing the results (capturing non-extreme observations) while a too high threshold might lead to large variance in estimations (Nieto & Ruiz 2016, 482-483; Benito et al 2023 p. 6). McNeil & Frey (2000 p. 282-283) used a ratio of 1:10 by choosing a k of 100 for their sample size $n = 1000$.

A limitation of fixing the threshold u (k in McNeil & Frey method) is that once chosen, it will not respond to possible changes in the behaviour of the underlying process. Multiple methods have been developed to better decide a proper threshold level u. There are both graphical and numerical methods to use. Graphical approaches include methods such as mean residual life plots (also known as the mean excess plot) where mean excess over u is graphically compared over u. Once a threshold is sufficiently high, the mean excess function $\mathbb{E}[X-u|X>u] = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi}u$ should be roughly a straight line. (Benito et al, 2023 p. 6-7; Coles, 2001 p. 79; Davidson & Smith, 1990 p. 396-397;)

A convenient property of GPD is that for any threshold v higher than u, the excesses still follow a GPD of the same shape. Only the scale shifts by $\beta_v = \beta_u + \xi(v-u)$. Additionally, when using a modified scale of $\beta^* = \beta_v - \xi v$ the result is constant once the tail approximation is adequate. When trading bias for variance in threshold selection, an approach implementing this method is to fit GPD over a range of threshold candidates and assess at which level u does the scale and the modified scale parameters stabilize, choosing then the lowest point deemed appropriate (Coles 2001 p.83-85; Scarlott & MacDonald, 2012 p. 36)

The numerical methods include the Hill estimator (Hill 1975), which has been shown to be in general an efficient estimator of ζ with low mean squared errors. The Hill estimator for shape parameter is

$$\hat{\xi}_k^{(H)} = k^{-1}\sum_{j=1}^{k} \log z_{(j)} - \log z_{(k+1)}$$

and this can be used to calculate an alternative quantile estimator to the GPD method:

$$\hat{z}_{\alpha k}^{(H)} = z_{(k+1)}\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\xi}^{(H)}}$$

However, in simulations, the efficient range of k for Hill estimator was lower than GPD based methods. Additionally, the Hill estimator only works for heavy tailed distributions $\xi > 0$, requiring the data to be consistently heavy tailed and for this to be known beforehand. The GPD method also allows for better estimates of expected shortfall and is at least as effective in high quantiles as the Hill estimator. (McNeil & Frey, 2000 p. 284-287)

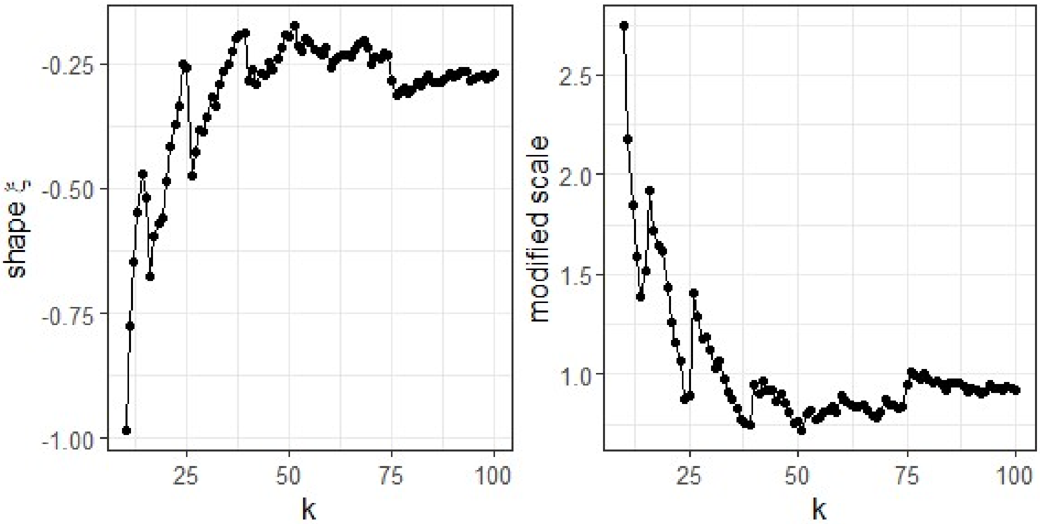

Figure 7 Threshold sensitivity

Finnish Government 10-year bond's parameter sensitivity to threshold *k* choice. Sample of 520 observations from 2019.

The second discretionary choice is regarding GARCH fitting to returns. McNeil & Frey (2000 p. 278) used a pseudo maximum likelihood method to fit GARCH(1,1) to their data using an assumption of normal distributions for innovations. The authors note that, while this makes an assumption they do not believe, the PML still delivered reasonable estimates for parameters.

McNeil & Frey's method is well suited for non-i.i.d. time series data, where the associated white noise process can be considered i.i.d. Under normal conditions, the VaR_{α} estimate would be of the form

$$VaR_{\alpha} = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right)$$

as long as $\xi < 1$. (McNeil et al, 2015 p. 154)

4.2.2 Econometric literature

McNeil & Frey (2000) concluded that their threshold exceedance method provided more accurate tail distribution estimates when compared with GARCH models using normal and student's t-distributions. Backtesting the method was conducted by fitting an AR(1)-GARCH(1,1) to US (S&P 500, 1/1960-6/1993) and German (DAX, 1/1973- 7/1996) market indexes, USD/GBP exchange rate (1/1980-5/1996), and prices of gold (1/1980-12/1997) and BMW shares (1/1973- 7/1996). After this, residuals were again calculated and used to fit the GPD. The method was then rolled over observations over each time series, providing a 1-day forward conditional VaR and ES. Robustness of the model against GARCH models was compared by a binomial test, in which 11 out of 15 times EVT proved most accurate in 99th quantiles.

Abad et al (2014) present a review of multiple papers comparing different methodologies for estimating VaR levels. Of the reviewed papers, EVT was found to be the best approach with FHS coming out 2nd and CaViaR showing promising results. Importantly for this thesis, the review included GARCH methods, historical simulation, RiskMetrics EWMA and the parametric methods, all of which the EVT outperfomed.

In a review of different VaR estimation methods, Calmon et al (2021) replicated the threshold exceedance method proposed by McNeil & Frey as one of the compared VaR approaches. Using 80 assets from a variety of sources and additional simulations based on their behaviour, threshold exceedance was compared against filtered historical simulation (FHS), 3 conditional autoregressive VaR (CAViaR) methods and time-varying higher order conditional moments (HOM). Threshold exceedances was found to be the most accurate.

Gencay & Selcuk (2006) implemented EVT to study tail behaviour of interbank offered rates in Turkey for a period covering extreme market turbulence. This data is from an economy where overnight interbank borrowing rates reached 4000% (annualized), exhibiting volatility quite uncommon for more developed financial markets such as within the European Monetary Union. They concluded that GPD fits well to the data even during a significant period of volatility. The authors acknowledge the niche nature of this dataset and conduct an additional test the method also on U.S. Federal Funds Rate (FFR) from 1954 to 2000. FFR was found to have a thin tailed ($\xi < 0$) right tail (increases in interest rates), which indicates that such moves as observed in Turkey would be (in a statistical sense) an impossibility in the U.S. This finding is somewhat significant regarding this thesis, since thin-tailed type distributions are rare in financial time series. Interbank interest rates can present a deviation from this rule of thumb.

EVT has also provided good results in the commodities sector. Studying EVT performance in oil markets can be beneficial when thinking about application in interest rates. Both share similar dynamics such as sensitivity to policy decisions (central banks and OPEC), are closely monitored as global economic indicators and exhibit periods of high and low volatilities. Marimoutou et al (2009) implemented McNeil & Frey's method in crude oil prices (Brent and West Texas Intermediate) and found that EVT provided the most accurate VaR estimates along FHS. The method clearly outperformed HS, and GARCH with normal distribution. GARCH with student's t-distribution was found to be as accurate as FHS and EVT yet tended to overestimate right tail events. This, while accurate in VaR level, can present a bias problem when calculating ES levels by providing overestimated results.

While these papers explore the fitness of EVT in financial applications, Szubzda & Chlebus (2019) focused on comparing the two main approaches to EVT implementation. Using stock indexes of emerging and developed countries in Europe from before, during, and after the great financial crisis (GFC) of late 2000s. Their results were mainly inconclusive of superiority between the two methods. They however emphasized deficiencies when using static EVT methods and noticed that block maxima tended to overemphasize VaR levels during calm periods, while threshold exceedance underestimated the levels during GFC. The authors also did not use McNeil & Frey's approach, preferring the standard approach.

Bloss et al (2025) also implemented both methods when studying gold price extremes spanning almost 50 years of data (1975-2024). Their findings underlined the superiority of EVT methods against legacy methods such as the normal distribution approach as explicit tail modelling allowed for more accurate calculations of VaR and ES levels and to discern the differences between left and right tails. However, they focused on comparing block maxima and threshold exceedance methods against realized historical periods instead of accuracy in forward estimation. In this regard, block maxima was found to be more accurate (although more volatile when comparing decades) while threshold exceedance was more stable over time periods, yet the GARCH process led to large emphasis on recent volatility environments when calculating VaR levels. This result, while unsatisfactory for historical comparison, is precisely one of the dynamics motivating the use of threshold exceedance in this thesis.

Yi et al (2014) combined McNeil & Frey's method with the quantile autoregression model developed by Xiao & Koenker (2009) to bypass the normality assumption of GARCH fitting described above to estimate quantile levels. The results gave good results for VaR levels of non-gaussian simulations.

In conclusion, both literature reviews and tests of different methods on the same time series support using extreme value theory in estimation of Value-at-Risk levels. The use of threshold exceedance method is further supported by a series of papers into interest rates, commodities and equities. The main drivers for use of EVT are grounded on robust mathematical theorems of extreme limit behaviour, explicit focus on tail events instead of overall distribution and (in the case of threshold exceedance approach) the efficient use of extreme observation data. Additionally, one or both tails can be of the thin-type GPD with a finite right endpoint.

5 Data

The European Interbank Offered Rates (EURIBORs) are defined by The European Money Market Institute (EMMI), which is responsible for calculating and publishing the rates, as *“the rate at which wholesale funds in euro could be obtained by credit institutions in current and former European Union and European Free Trade Association countries in the unsecured money market.”* The data is gathered from a panel of credit institutions (such as deposit banks) that form a representative sample of euro money market, considering geographic diversity and capability of panel members to handle significant volumes of euro denominated instruments. (EMMI 2024)

The rates are calculated for every TARGET day (Eurosystem operated financial infrastructure) for tenors of 1 week, 1-, 3-,6- and 12-months. The rates are annualised with a conversion ratio of $\frac{actual}{360}$ and represent the average of panel members’ lending rate following a set of hierarchical calculation methods. As part of the process, top and bottom 15% of panel banks’ lending rates are eliminated when calculating the final rates. The final rates are averaged arithmetically and rounded to 1 basis points (0.01 %) and published in the morning of the following TARGET day shortly after 11:00 Central European Time. (EMMI 2024)

The Euribor daily data used is obtained from Bank of Finland’s public statistics for each of the tenors with a maturity exceeding 1 week. Risk measures are calculated for the last 1300 observations, ending in 26.3.2026. However, the models the previous 520 observations on a rolling basis, meaning that, in total, 1820 observations are used in the analysis phase. In the case of the 3-month tenor, a few daily observations were missing from Bank of Finland’s statistics over the past 1300 TARGET days. These omissions were not sequential nor did the changes in any of the other tenors suggest abnormal market movements during the missing days. These missing days are treated simply as non-TARGET days in the following calculations. A few extra days are therefore included in the 3-month tenor’s sample beginning, so that each tenor has 1300 estimates. When analysing the entire term structure, the days of missing variables are omitted entirely from each tenor.

5.1 Observed behaviour

The case of 1-week tenor is somewhat special, since it closely follows European Central Bank’s deposit facility rate (DFR). The deposit facility rate is the main instrument with which the ECB steers its monetary policy stance. DFR is the rate which the ECB pays for overnight deposits in the national central banks of the system. DFR thus acts as a floor rate for interbank market rates. (ECB 2026) This can lead to large jumps sparked by changes in ECB's monetary policy implementation, while between ECB deposit rate hikes the daily changes are very modest for the tenor. This leads the 1-week Euribor to exhibit jumping behaviour in its level. This also means that proper risk measurement of extreme events that are not induced by ECB monetary policy would require significant trimming of data to include only the periods between DFR jumps and omitting periods around the jumps.

However, this also presents a problem, since DFR changes are also layered in uncertainty. Monetary policy is decided by the ECB's Governing Council, which votes on the issue. A DFR change is therefore not certain in a deterministic sense and market players have coined ways to try to measure probabilities of policy decisions. In a way therefore, monetary policy triggered events are the true short term extreme events of 1-week Euribor. This is why the 1-week tenor is dropped from the analysis in this thesis.

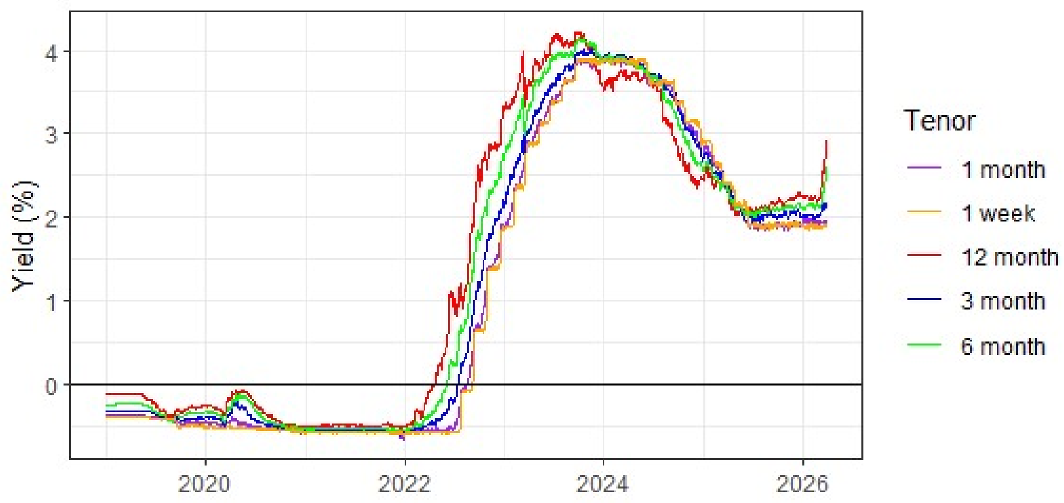

Figure 8 Euribors

Movements of all Euribor tenors between 2019 and March 2026.

Unsurprisingly, all the tenors move roughly in the same direction. However there remains some notable changes, especially related to interest rate term structure. The longest tenor, 12-month Euribor, remains elevated above others for the most part, yet in December 2023 it drops below the shorter tenors, resulting in what's known as an inverse yield curve. This type of yield curve can be formed when long term inflation expectations are lower than current inflation.

Using factor analysis, it can be observed that the three principal components driving the term structure behave roughly in line with the expectations of term structure theories. The first component appears to drive the level of the term structure, moving also roughly in tandem with inflation levels observed during the period. The second component moves slightly more stochastically than the first and seems to drive movements in the slope of the term structure. The third component is the most stochastic, with periods of large volatility seen especially during the time of rising rates and monetary policy hikes.

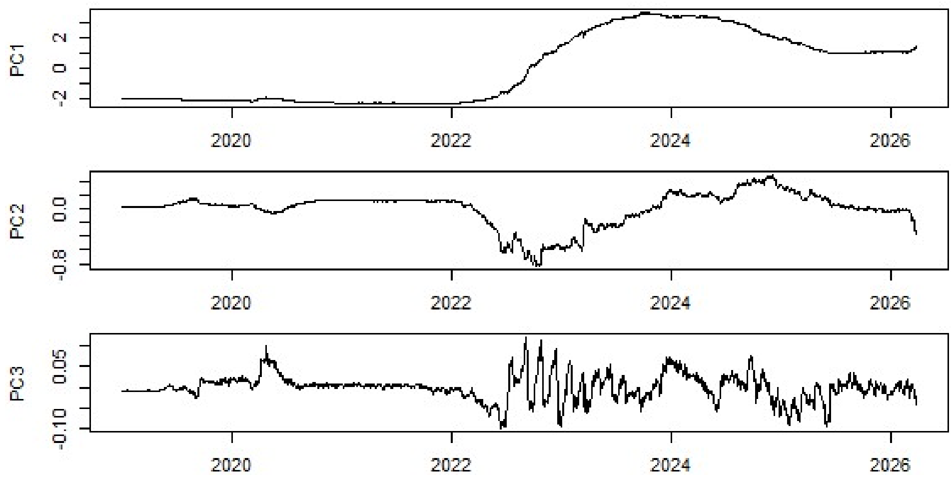

Figure 9 Principal Components

Movements of the term structure's principal components over the sample period.

5.2 Statistical tests

To assess the statistical properties of the tenors, a series of tests are conducted for the first 520 observations with which the models are primed for the out-sample rolling estimations. The tests are the Augmented Dickley-Fuller, Jarque-Bera and Ljung-Box. Each test is conducted for every tenor. These are done to assess the statistical properties of the sample and to explore whether the GARCH approach can be justified for Euribors and if the distribution satisfies Gaussian properties.

To test whether the individual tenors are stationary, an Augmented Dickley-Fuller (ADF) test is used. ADF tests for a unit root in the time series, which would indicate non-stationary behaviour and if using differences (in this case, rate changes) would alternatively yield a stationary time-series. (Said & Dickley, 1984)

For each of the tenors, the null hypothesis of non-stationarity is clearly accepted. When taking the differences (changes) of each tenor and conducting the same test again, the null hypothesis of non-stationarity is clearly rejected at $p < 0.01$. Based on these results, the use of GARCH family processes to model the changes in tenors can be proceeded with.

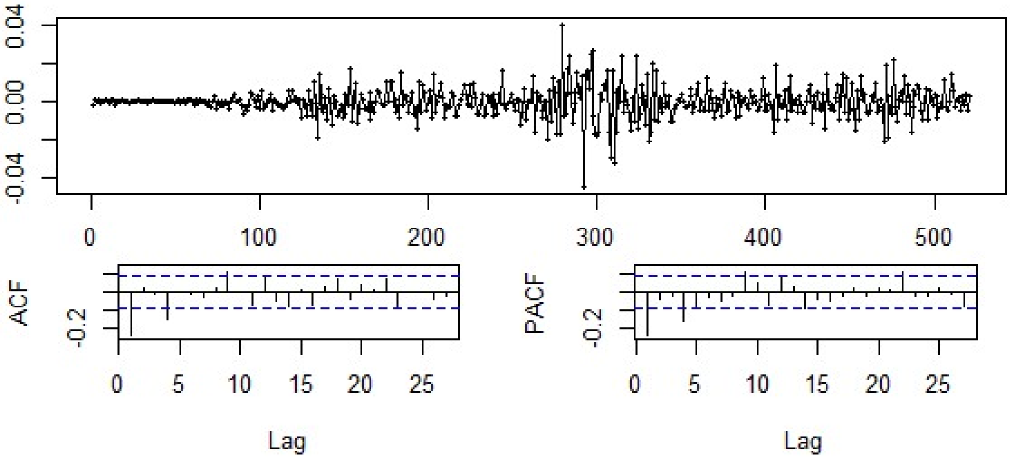

Figure 10 Summary Statistics

Summary Statistics for 1-month Euribor tenor changes for the first 520 observations.

To test whether the samples have Gaussian distributions, a Jarque-Bera test is conducted. The test uses 3rd and 4th moments of the distribution (skewness and kurtosis) to produce a test statistic of Gaussian behaviour of these higher moments. A large test statistic indicates that levels of skewness and kurtosis significantly differ from those of normal distribution (0 and 3). (Jarque & Bera 1987; Taeger & Kuhnt 2014, 150–152)

The Ljung-Box test (Ljung & Box 1978) measures possible independence of time series observations. The test is often used as a goodness-of-fit test for time-series models by applying it to model residuals. Here, however, it is implemented on the interest rate changes to assess whether the changes exhibit independence. A p-value can be obtained from the test statistic to assess the validity of the null hypothesis. For each of the tenors, independence is rejected at $p < 0.01$.

In conclusion, each of the Euribor tenors exhibit non-stationary behaviour at the yield level, but stationary behaviour of daily changes. Observations of changes are serially correlated, but the direction and magnitude of autocorrelation for a lag of 1 differs between tenors. Furthermore, the changes do not appear to follow a Gaussian distribution. The volatility of changes also appears time varying. Implementation of GARCH models is therefore proceeded with and a heavier tailed, Student’s t-distribution tested.

6 Result Comparison

The models are evaluated using two techniques, Kupiec's (1995) unconditional coverage test and Christoffersens' (1998) conditional coverage test. Both are widely used in literature when evaluating VaR method validity. Kupiec notes that, regardless of the order in which failures (exceedances) appear, the probability of observing an x amount of them from a sample size of n follows a binomial distribution. The method therefore tests *proportion of failures* (exceedances) is within defined a confidence level from the expected one. The test statistic is given by

$$LR_{uc} = 2\log\left(\left(1-(x/n)\right)^{n-x}(x/n)^{x}\right) - \log((1-p^{*})^{n-x}(p^{*})^{x}),$$

where p^{*} is the expected number of exceedances (given by quantile α) and p is the observed exceedances. The test statistic follows a chi-squared distribution, which can be used to calculate the p-value.

Christoffersen argues that merely the frequency of exceedances is not enough, but that a good model should also account for independence of exceedances (in the case of financial time series, while the exceedance frequency might be correct, if the exceedances are clustered the model perhaps fails to account for some market conditions). Christoffersen's conditional coverage test builds on Kupiec's unconditional coverage by adding in testing of independence of exceedances. The independence is tested by a null hypothesis that today's probability of an exceedance should not depend on whether an exceedance happened yesterday. In practice, this is done through a first-order Markov chain. The likelihood ratio LR_{ind} follows again a chi-squared distribution. Combining both the unconditional and independence test statistics yields a test statistic with a chi-squared distribution with 2 degreees of freedom

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^{2}(2).$$

When reporting results, * will signify a rejection of the unconditional coverage hypothesis at $p < 0.05$ level and ** when the conditional coverage hypothesis is rejected at $p < 0.05$ level.

6.1 Legacy Models

From the legacy models, the EWMA approach presented by RiskMetrics is tested. The decay parameter is set at $\lambda = 0.94$. From the results it's clear that the method systematically fails to properly estimate the VaR measures for both tails at both quantiles $\alpha = \{0.01, 0.05\}$. At 1st and 99th quantiles, the both unconditional and conditional coverage tests are failed for each of the tenors, while at 5th and 95th quantiles, one or both tests are passed depending on the tenor, although at 6- and 12-month tenors both tests still fail.

Table 1. VaR results with EWMA

Using parameters presented in the original RiskMetrics (1996) method.

RiskMetrics EWMA				
quantile α	**1 Month**	**3 Month**	**6 Month**	**12 Month**
5 %	5.20 %	5.69%^{cc}	4.23%^{cc}	5.8^{cc}
95 %	6.0%^{cc}	5.38 %	6.31%$^{uc+cc}$	5.6^{uc+cc}
1 %	1.8^{uc+cc}	2.08%$^{uc+cc}$	1.77%$^{uc+cc}$	2.2%$^{uc+cc}$
99 %	2.2^{uc+cc}	1.62%$^{uc+cc}$	2.54%$^{uc+cc}$	1.9%$^{uc+cc}$

Percentage of violations in rolling estimates for each target quantile. Superscript: uc = failure of unconditional coverage test, cc = failure of conditional coverage test, both at $p < 0.05$.

Interestingly, for left tail of the 1-month and right tail of the 3-month tenors, both tests fail and the hit rate is fairly accurate. However, given the wider context of systematic failure, this early approach to measuring VaR cannot be considered as a valid general method to capture Euribor dynamics in fluctuating yield curve conditions.

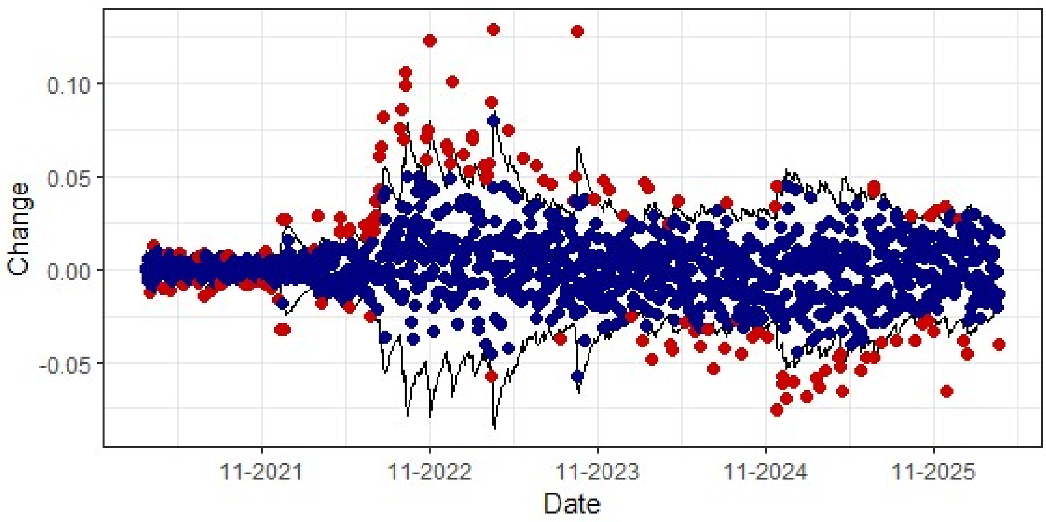

Figure 11 EWMA VaR

1-month tenor's exceedance of estimates for 5th and 95th quantiles.

6.2 GARCH models

Fitting of the GARCH model is done using the last 520 observations, giving the innovation distribution and the 1-day forward estimates of both mean and volatility parameters. The specific fitting method will be determined by the rolling estimation algorithm itself. The formal representation of the measure provided by the model is the following:

$$\hat{V}_{t+i}^{\alpha} = \mathbb{E}[V_{t+i}^{\alpha} | \mathcal{F}_t] = \hat{\mu}_t + \hat{\sigma}_t q_{\alpha}(Z)$$

Here $\mathcal{F}_t$ is the information set available at time t, $\widehat{\sigma_t}$ is the estimated volatility given by the model and $\hat{\mu}_t$ is the mean change. $q_{\alpha}(Z)$ is the quantile function of the innovation distribution and α is the chosen quantile. Later, estimated Expected Shortfall will be given by $\mathbb{E}\left[R_{t+1} \middle| R_{t+1} < \hat{V}_{t+i}^{\alpha}\right]$ for left tail and by $\mathbb{E}\left[R_{t+1} \middle| R_{t+1} > \hat{V}_{t+i}^{1-\alpha}\right]$ for the right tail. The distributions used are the student-t distribution and the Gaussian (normal) distribution. Gaussian distribution is included in the tests so that a comparison can be made whether a fatter tail distribution such as t-distribution is indeed more accurate when modelling tail behaviour of interest rate data.

The model results for GARCH(1,1) and E-GARCH(1,1) are listed in Table 2 and Table 3 respectively. In the case of Gaussian distribution, both standard and exponential versions fail unconditional and conditional coverage tests more systematically than the Student's t-distribution. This is especially prevalent in higher quantiles such as 1st and 99th. Here though, E-GARCH presents slightly more improved results from the standard GARCH. Failures are especially prevalent in the 99th quantile, right tail of the distribution, which indicates that interest rate increases are underestimated more often than decreases.

Table 2. VaR results for GARCH

Table lists realized percentage of exceedances in each target quantile.

GARCH (1,1)				
quantile α	**1 Month**	**3 Month**	**6 Month**	**12 Month**
Student-t				
5 %	5.5 %	4.8 %	4.2 %	6.1 %
95 %	6.5%^{uc}	5.6 %	5.8%^{cc}	6.8%$^{uc+cc}$
1 %	1.2 %	1.0 %	1.0 %	1.3 %
99 %	1.6%^{uc}	1.4%^{cc}	1.8%$^{uc+cc}$	1.6%^{uc}
Gaussian				
5 %	4.9 %	4.3 %	3.9 %	5.4 %
95 %	6.0 %	5.0 %	5.8%^{cc}	5.2 %
1 %	1.6%^{uc}	1.5 %	1.2%^{cc}	1.6%^{uc}
99 %	2.2%$^{uc+cc}$	1.6%$^{uc+cc}$	2.2%$^{uc+cc}$	2%$^{uc+cc}$

Percentage of violations in rolling estimates for each target quantile. Superscript: uc = failure of unconditional coverage test, cc = failure of conditional coverage test, both at p < 0.05.

E-GARCH with Student's t-distribution presents a significant improvement from GARCH with Student's t, resulting in reduced UC and CC failures, expect in the case of 1 Month tenor at 95th quantile, where the results deteriorate significantly further. Student's t overall produces markedly better results from the Gaussian counterpart. However, here too, the right tail exhibits persistent underestimation of VaR levels. Overall, Student's t standard GARCH fails the unconditional coverage test 5 times and the conditional coverage test 4 times, in two cases for the same scenario. E-GARCH reduces these numbers to 3 and 3 with 1 case of both failing simultaneously.

Table 3. VaR results for E-GARCH

Table lists realized percentage of exceedances in each target quantile.

Exponential-GARCH (1,1)				
quantile α	**1 Month**	**3 Month**	**6 Month**	**12 Month**
Student-t				
5 %	5.80 %	5.10 %	3.90 %	5.70 %
95 %	7.3%$^{uc+cc}$	5.70 %	5.8%^{cc}	5.4%^{cc}
1 %	1.20 %	0.90 %	1 %	1.10 %
99 %	1.6%^{uc}	1.20 %	1.6%^{uc}	1.30 %
Gaussian				
5 %	5.40 %	4.50 %	3.2%$^{uc+cc}$	5.20 %
95 %	6.9%$^{uc+cc}$	5 %	5%^{cc}	4.8%^{cc}
1 %	1.7%$^{uc+cc}$	1.50 %	1.20 %	1.30 %
99 %	2.3%$^{uc+cc}$	1.8%$^{uc+cc}$	2%$^{uc+cc}$	1.7%^{uc}

Percentage of violations in rolling estimates for each target quantile. Superscript: uc = failure of unconditional coverage test, cc = failure of conditional coverage test, both at $p < 0.05$.

From these results, it's clear that a method incorporating a heavy tailed distribution specification and with properties that allow to capture asymmetric dynamics in volatility is preferable to a basic model with a Gaussian distribution. Interestingly neither of the models using Student's t incurred any failures of the tests for the left tail VaR on either 5th or 1st quantiles. This could indicate that interest rate decreases follow a more predictable dynamic, while the shocks are driven by upwards momentum. Some credence for this hypothesis can be discerned from the data. As rates began to rise, left tail exceedances became rarer for a while. When the conditions changed and the term structure level started to decline, left tail exceedances started to appear again.

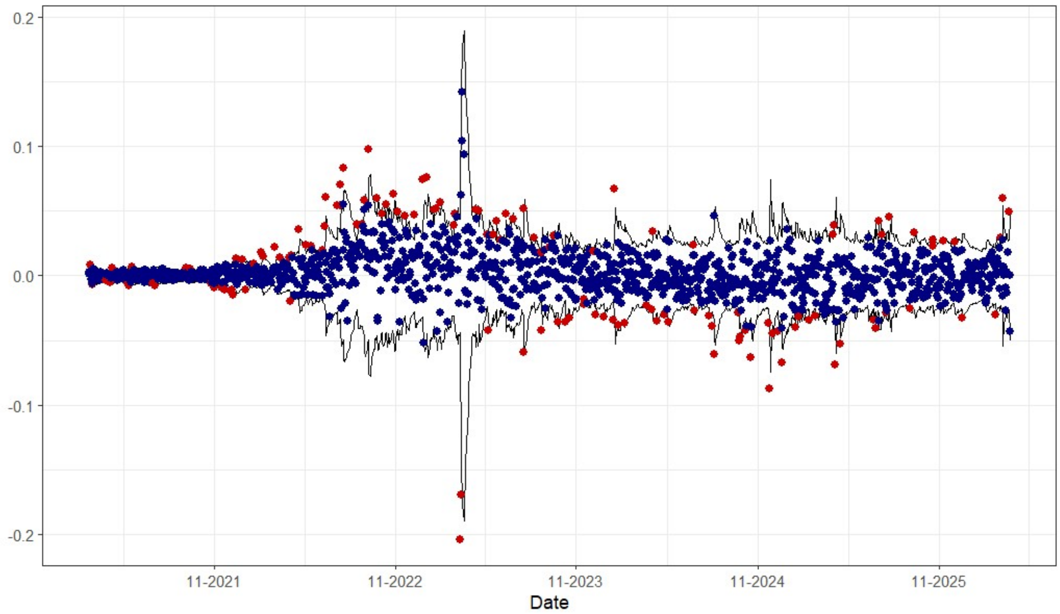

Figure 12 GARCH VaR

3-month tenor's exceedances of estimates in 5th and 9th quantiles for GARCH(1,1) utilizing a Student's t-distribution specification.

6.3 Threshold Exceedances

The approach developed by McNeil & Frey (2000), presented in section 2.2. is adopted. Again, the evaluation is done daily, with refits of model parameters every 10 and 1 days. Forward estimation for location and scale parameters (mean and volatility) is done for 1-day. Innovations (residuals) of the GARCH(1,1) fit are treated as a white noise process, which is used to model the extreme tail. One additional mechanism that McNeil & Frey (2000) used is an autoregressive process of lag one in modelling the returns, which helps account for autocorrelation in returns. This AR(1)-GARCH(1,1) fit is tested against the simple GARCH(1,1) fit to see whether allowing for autocorrelation in rate changes improves the estimates.

A part of the process of implementing the method is the choice of *k*, the number of "exceedances" used to model the tail. As mentioned in section 4.2, a variety of both analytical and graphical methods are available when choosing a suitable threshold *u* or exceedance number *k*. The method described treats the threshold level *u* as random and number of exceedances *k* as fixed. A graphical method of fitting a generalized pareto distribution over several potential *k* candidates is used in this thesis to pick a suitable threshold number *k*. A threshold is deemed suitable once the shape parameter ζ starts to stabilize (i.e. changes begin to linearize).

Each of the 1- to 12-month tenors appear to have a roughly linear k for the left tail exceedances starting around 55–70 exceedances depending on the tenor. With fewer exceedances, the shape parameter starts to destabilize. After manually conducting the graphical analysis for each of the tenors, a threshold choice of $k = 60$ was adopted for all the tenors.

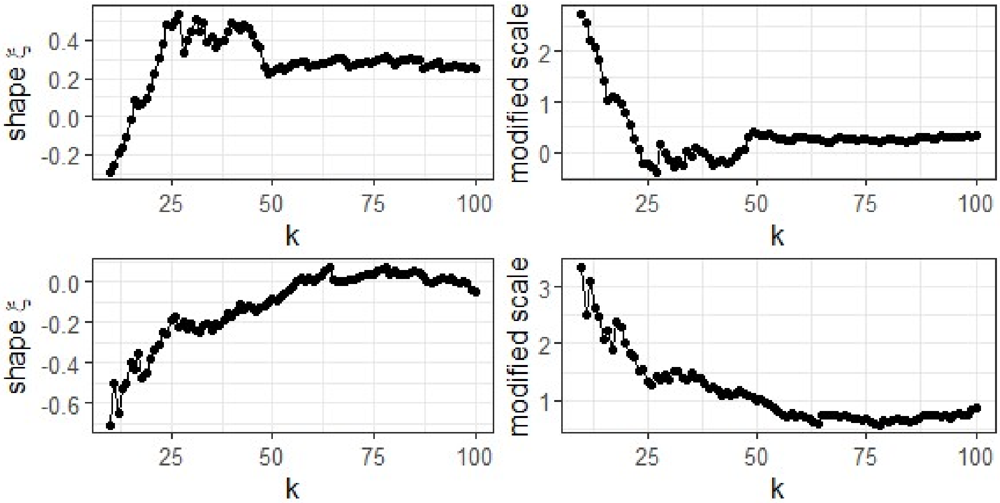

Figure 13 Threshold Sensitivities

Shape parameter's sensitivity to threshold level choice. Illustrated for the cases of 3-month (top) and 12-month (bottom) Euribors and using the first 520 observations.

Again, both unconditional and conditional coverage tests are implemented. The model results are listed on Table 4. It is discerned that the threshold choice 60 yields good results with threshold exceedance method compared to earlier GARCH models. EVT clearly beats standard GARCH with either Gaussian or Student's t-distributions and E-GARCH with Gaussian distribution. When comparing with E-GARCH with a Student's t specification, the results are very similar in terms of accuracy.

The results suggest that GARCH(1,1) fitted EVT presents a good approach to estimating VaR levels across term structure, with a modest amount of coverage test failures. Using an AR(1)-process as part of the model significantly improves results for 3-, 6-, and 12-month tenors, but deteriorates the model for 1-month tenor. Looking at exceedances and VaR levels for 1-month tenor more granularly, the phenomenon seems to be caused by central bank monetary policy. Exceedance clustering seems to happen during periods when the ECB either raises or lowers their rates.

Table 4. VaR results for EVT

Table lists realized percentage of exceedances in each target quantile. Threshold k = 60

EVT Threshold Exceedances				
quantile α	**1 Month**	**3 Month**	**6 Month**	**12 Month**
No AR(1)				
5 %	6.5%^{uc}	5.5 %	5.5 %	5.2 %
95 %	5.3 %	5.4 %	6.0%^{cc}	6.2%$^{uc+cc}$
1 %	1.6%^{uc}	1.0 %	1.0%^{cc}	1.2 %
99 %	1.2 %	1.5%^{cc}	1.5 %	1.3 %
Including AR(1)				
5 %	6.6%$^{uc+cc}$	5.6 %	5.2 %	5.0 %
95 %	5.5%^{cc}	5.3 %	5.8%^{cc}	5.9 %
1 %	1.7%^{uc}	0.9 %	0.9%^{cc}	1.2 %
99 %	1.1 %	1.4 %	1.5 %	1.3 %

Percentage of violations in rolling estimates for each target quantile. Superscript: uc = failure of unconditional coverage test, cc = failure of conditional coverage test, both at $p < 0.05$.

AR(1) approach comes on par with E-GARCH in terms of number of test failures (6), with 2 failures of UC and 4 failures of CC. AR(0) fails 7 times (3 and 4 respectively). When looking at test statistics, EVT presents a modestly more robust approach than E-GARCH when looking at conditional coverage. However, their overall VaR level accuracy based on unconditional coverage is almost the same, with a very slight edge to E-GARCH. When tallying up the differences between realized exceedances and target exceedances, EVT also presents on aggregate a slightly more accurate model.

An advantage of EVT is that it performs better on the right tail than GARCH and E-GARCH, with 3 violations of tests instead of 6 for E-GARCH. Since EVT models each tail separately, it has been able to capture the nuanced dynamics underlying the term structure change better than its counterparts. This can also be discerned by looking at the shape parameters of the tail distributions in table 5. It can be observed that the tails do not follow a uniform dynamic over the term structure. The average tail distribution types differ depending on the tenors.

For 12-month, the types are different for left and right tails, with right having a tendency for heavier distribution unlike the left tail with shorter distribution types. 1-, 3- and 6-month tenors have on average same types of distribution for both left and right tails over the period. In the case of 1-month tenor, the distributions are short-tailed, while 3- and 6-month tenors tend to have heavier type tails. A possible explanation for this is that as a rather short maturity, 1-month changes are naturally more conservative.

Table 5. Tail shapes

Average tail distribution's shape parameter given by EVT for both left and right tail.

EVT Tail Shapes				
	1 Month	**3 Month**	**6 Month**	**12 Month**
No AR(1)				
Left tail	-0.074	0.033	0.050	-0.098
Right tail	-0.123	0.078	0.008	0.191
Including AR(1)				
Left tail	-0.060	0.040	0.068	-0.113
Right tail	-0.101	0.039	0.002	0.150

Average shape parameter over sample

Overall, EVT presents the most consistent method of the approaches covered. Exponential GARCH with a heavy tail can be used to produce almost as good results, but EVT's ability to capture more nuanced tail distributions seems to give it a slight edge in the long run.

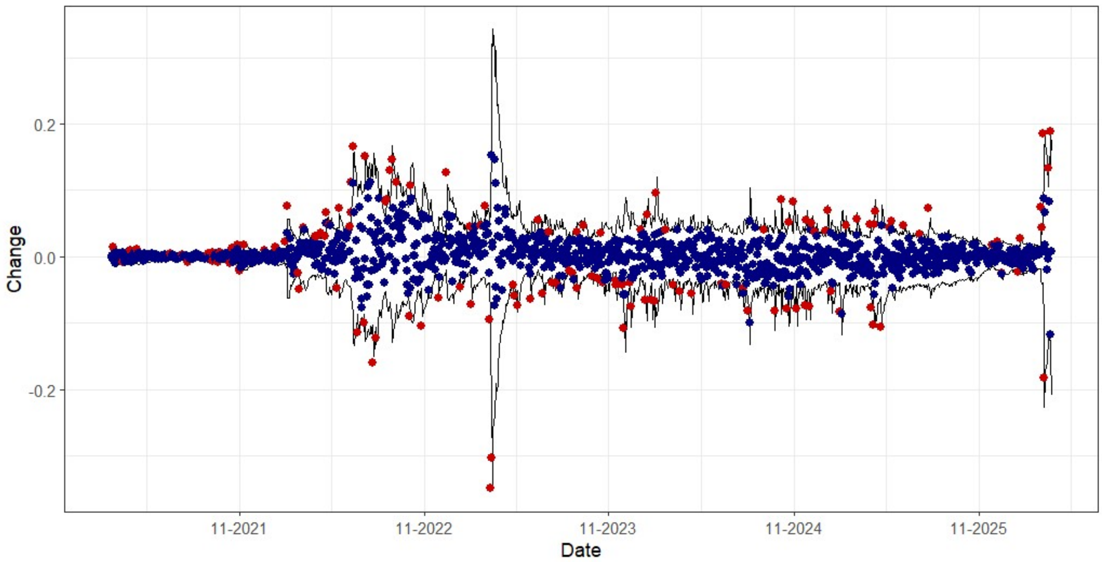

Figure 14 Threshold Exceedance VaR

12-month tenor's exceedances of estimates for 5th and 95th quantiles. Results for AR(1)-GARCH(1,1) specification.

6.4 Expected Shortfall

Expected Shortfall accuracy is tested for the two methods assessed to be the most accurate in Value-at-Risk estimation. Of the covered models and model specifications, Threshold Exceedance (with autoregression) and E-GARCH(1,1) using Student-t distribution proved to be the most accurate. The test presented by McNeil & Frey (2000, 294–296) for ES is implemented. This approach rests on measuring the difference between estimated ES and actual change when the VaR given by the model is breached. This is done with exceedance residuals

$$r_{t+1} = \frac{x_{t+1} - \widehat{ES}_{\alpha}^{t}}{\hat{\sigma}_{t+1}}$$

where r_{t+1} are standardized with estimated volatility. The null hypothesis assumes that both mean, and volatility estimates of the model correctly estimate the dynamics and that the innovations Z should behave like an i.i.d process with mean zero. This is tested by using a one-sided test of whether ES is systematically underestimated. In practice this is done with bootstrapping with no assumptions on underlying distribution of exceedance residuals.

If the mean residuals for left quantile are negative, the model underestimates ES. Conversely, for right tail, positive residuals indicate underestimation of ES. In the case of EVT the results on underestimation are mixed. For 1- and 12-month tenors, the results indicate slight underestimation in both tails. 3- and 6-month tenors appear to have no clear trend on over- or underestimation, with perhaps slight overestimation in the case of 3-month tenor. E-GARCH on the other hand underestimates on both tails at 6- and 12-month tenors but appears to have no clear trend on 1- and 3-month tenors.

Table 6. Expected Shortfall

Mean of exceedance residuals

Expected Shortfall residuals				
quantile α	**1 Month**	**3 Month**	**6 Month**	**12 Month**
Extreme Value Theory				
5 %	-0.089	0.066	0.007	-0.072
95 %	0.070	0.040	0.050	0.178
1 %	-0.208	-0.061	-0.225	-0.195
99 %	0.356	-0.055	-0.039	0.983*
Exponential-GARCH(1,1)				
5 %	0.008	0.010	-0.142	-0.255
95 %	0.043	-0.041	0.161	1.553
1 %	-0.065	-0.043	-0.123	-0.039
99 %	0.125	0.034	0.179	0.417

Mean standardized residuals of actualized change vs. expected shortfall in the cases that VaR is breached. * p<0.05 ** p<0.01

Overall, the results indicate that both methods give very reliable estimates for ES levels for both tails, in each quantile and for all the tenors. For EVT, in only 1 of the covered scenarios (12 month, 99th quantile) does it fail the test at 5% confidence level, while for E-GARCH the number of failures was 0. Giving a slight edge to E-GARCH. These results support the conclusion that conditional Expected Shortfall is a more reliable measure of risk as long as the model estimating the corresponding conditional VaR level is systematically accurate.

7 Conclusions

The goal was to compare which models for measuring Value-at-Risk and Expected Shortfall measures proved to be the most accurate for European Interbank Offered Rates (Euribors). The chosen time-period contained periods of significant market volatility, which constituted a natural stress test environment for the models. The models were compared with unconditional coverage and conditional coverage tests, which are common tests for exceedance models measuring correctness and independence of exceedances.

Models incorporating a Gaussian distribution as the assumed underlying distribution consistently failed one or both backtests used at 1st and 99th quantiles. The early approach to VaR of exponentially weighted moving average, which relies on a Gaussian distribution, failed the tests systematically even at lower quantiles. The standard GARCH with Gaussian specification yielded improved but still very unsatisfactory results. Heavier tail specification improves model accuracy considerably.

The two most accurate models proved to be Threshold Exceedance (Extreme Value Theory) and Exponential GARCH (E-GARCH) with Student's t-distribution, with E-GARCH providing almost as good results in terms of coverage tests and raw aggregate accuracy. Both models still encounter problems mainly on the 1-month tenor due to apparent influence of central bank policy changes on short maturities.

Extreme Value Theory's advantage relies on modelling the tail distribution explicitly instead of the entire distribution. This is more economical when the focus is on rare event risk management and important if the tails are asymmetrical. For 12-month tenor, the average shape parameter values differed for left and right tails. The left tail (interest rate decreases) being of the heavy-tailed type, while the right tail (increases) being short tailed. For 3- and 6-month tenors, both tails were on average heavy-tailed, while 1-month tenor had on average short tails. However, it should be noted that as the yield curve evolved over time, the tail parameters evolved alongside new conditions.

Extreme Value Theory presents a good approach to estimating 1-day forward VaR and ES measures in European Interbank money markets. To achieve the same standard of accuracy with a GARCH process, an asymmetric, exponential variation with a fat tailed distribution was needed. The use of threshold exceedance yields good results especially for Expected Shortfall, although E-GARCH provided to be at least as good at measuring Expected Shortfall. Further comparisons between the two approaches could be done by implementing more advanced methods of threshold selection for

Extreme Value Theory and by using a skewed distribution for E-GARCH. Another point of further research would be a multivariate analysis using

References

Abad, Pilar –Benito, Sonia – Lopez, Carmen (2014) A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*, Vol. 12 (1) 15–32.

Angelidis, Timotheos – Benos, Alexandros – Degiannakis, Stavros (2004) The use of GARCH models in VaR estimation. *Statistical Methodology*, Vol. 1 (1) 105–128.

Artzner, Philippe – Delbaen, Freddy – Eber, Jean-Marc – Heath, David (1999) Coherent Measures of Risk. *Mathematical Finance*, Vol. 9 (3) 203–228.

Babis, Alex (2024) Performance Evaluation of a Family of GARCH Processes Based on Value at Risk Forecasts: Data Envelopment Analysis Approach. *Computational Economics*, Vol. 66 (2) 1379–1411.

Benito, Sonia – López-Martin, Carmen – Navarro, Ángeles (2023) Assessing the importance of the choice of threshold in quantifying market risk under the POT approach (EVT). *Risk Management*, Vol. 25 (1), Article 6.

Björk, Thomas (2019) *Arbitrage Theory in Continuous Time, 4th Edition*. Oxford University Press, Oxford, United Kingdom.

Bloss, Michael – Ernst, Dietmar – Geisinger, Leander (2025) Extreme Value Theory and Gold Price Extremes, 1975-2025: Long-Term Evidence on Value-at-Risk and Expected Shortfall. *Commodities*, Vol. 4 (4), 1–35.

Brigo, Damiano – Mercurio, Fabio (2006) *Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit*. Springer Finance, Berlin, Heidelberg.

Buczynski, Mateusz – Chlebus, Marcin (2023) GARCHNet: Value-at-Risk Forecasting with GARCH Models based on Neural Networks. *Computational Economics*, Vol. 63 (5) 1949–1979.

Cairns, Andrew G. (2018) *Interest Rate Models: An Introduction*. Princeton University Press, New Jersey, United States of America.

Calmon, Wilson – Ferioli, Eduardo – Lettieri, Davi – Soares, Johann – Pizzinga, Adrian (2021) An Extensive Comparison of Some Well-Established Value at Risk Methods. *International Statistical Review*, Vol. 89 (1) 148–166.

Christensen, Jens H.E. – Diebold, Francis X.– Rudebusch, Glenn D. (2011) The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*, Vol. 164 (1) 4–20.

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

Coles, Stuart (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Springer-Verlag, London.

Davidson, A. C. – Smith, R. L. (1990) Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society*, Vol. 52 (3) 393–442.

Diebold, Francis X. – Rudebusch, Glenn D. (2013) *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press, New Jersey, United States of America.

Dowd, Kevin (2002) *Measuring Market Risk.* John Wiley & Sons, England.

Duffie, Darrell – Gârleanu, Nicolae (2001) Risk and Valuation of Collateralized Debt Obligations. *Financial Analysts Journal*, Vol. 57 (1) 41–59.

de Haan, Laurens – Ferreira, Ana (2006) Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering, Springer, New York.

Embrechts, Paul – Klüppelberg, Claudia – Mikosch, Thomas (1997) *Modelling Extremal Events: For Insurance and Finance*. Springer, Berlin.

Euribor Governance Framework (2024) The European Money Market Institute, Benchmark Determination Methodology for Euribor D0016F-2019.

European Central Bank (2026) The Eurosystem's instruments. <https://www.ecb.europa.eu/mopo/implement/html/index.en.html/>, searched 15.4.2026.

Fan, Ying – Zhang, Yue-Jun – Tsai, Hsien-Tang – Wei, Yi-Ming (2008) Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics*, Vol. 30 (6) 3156–3171.

Filipović, Damir (2001) A general characterization of one factor affine term structure models. *Finance and Stochastics*, Vol. 5 (3) 389–412.

Fisher, Ronald A. – Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical proceedings of the Cambridge Philosophical Society*, Vol. 24 (2), 180–190.

Fiszeder, Piotr – Małecka, Marta – Molnár, Peter (2024) Robust estimation of the range-based GARCH model: Forecasting volatility, value at risk and expected shortfall of cryptocurrencies. *Economic Modelling*, Vol 141 (1) Article 106887.

Francq, Christian – Zakoian Jean-Michel (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, New Jersey, United States of America.

Gençay, Ramazan – Selçuk, Faruk (2006) Overnight borrowing, interest rates and extreme value theory. *European Economic Review*, Vol. 50 (3) 547–563.

Ghalanos, Alexios (2026) Introduction to the rugarch package (Version 1.4-3). [Technical guide] <https://cran.rstudio.com/web//packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf> searched 8.4.2026.

Gnedenko, Boris (1943) Sur La Distribution Limite De Terme Maximum D'Une Série Aléatoire. *Annals of Mathematics*, Vol. 44 (3), 423–453.

Hill, Bruce M. (1975) A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, Vol. 3 (5) 1163–1174.

Idirizov, Byulent (2025) Application and development of the block maxima method in analysis of silver price. *Journal of Physics: Conference Series*, Vol. 3145 (1) 012014.

Jakata, Owen – Chikobvu, Delson (2025) Modelling Financial Risk in the South African Stock Market: An Application of the Generalised Extreme Value Distribution (GEVD). *Econometric Research in Finance*, Vol. 10 (2), 53–70.

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

Leadbetter, M. R. (2017) Extremes Under Dependence – Historical Development and Parallels with Central Limit Theory. In: *Extreme events in finance: a handbook of extreme value theory and its applications*, eds. Longin, François M., 48–69. John Wiley & Sons, New Jersey.

Li, Simon (2024) Assessing stochastic dominance of downside and upside financial risk profiles using the block maxima method in extreme value theory. *Journal of Risk Management in Financial Institutions*, Vol. 17 (4) 426–438.

Ljung, G. M – Box, G. E. P. (1978) On a Measure of Lack of Fit in Time Series Models. *Biometrika*, Vol. 65 (2), 297–303.

Marimoutou, Velayoudoum – Raggad, Bechir – Trabelsi, Abdelwahed (2009) Extreme Value Theory and Value at Risk: Application to oil market. *Energy Economics*, Vol. 31 (4), 519–530.

Markowitz, Harry (1952) Portfolio Selection. *The Journal of Finance*, Vol. 7 (1) 77–91.

McNeil, Alexander J. – Frey, Rüdiger (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance,* Vol. 7 (3), 271–300.

McNeil, Alexander J. – Frey, Rüdiger – Embrechts Paul (2015) *Quantitative Risk Management: Concepts , Techniques and Tools.* Princeton Series in Finance, Princeton University Press, Princeton and Oxford.

Nawalkha, Sanjay K. – Soto, Gloria M. – Beliaeva Natalia A. (2005) *Interest Rate Risk Modeling: The Fixed Income Valuation Course*. John Wiley & Sons, New Jersey, United States of America.

Nelson, Charles R. – Siegel, Andrew F. (1987) Parsimonious Modeling of Yield Curves. *The Journal of Business*, Vol. 60 (4) 473–489.

Nieto, Maria R. – Ruis, Esther (2016) Frontiers in VaR forecasting and backtesting. *International Journal of Forecating,* Vol. 32 (2), 471–501.

Palma, Wilfredo (2016) *Time Series Analysis*. John Wiley & Sons, New Jersey, United States of America.

Realdon, Marco (2025) Affine term structure with GARCH volatility. Risk Management, Vol. 27 (4) Article 28.

RiskMetrics (1996) RiskMetrics —Technical Document, Fourth Edition. < https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a>, searched 8.4.2026.

Said, Said E. – Dickley, David A. (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, Vol. 71 (3) 599–607.

Scarlott, Carl – MacDonald, Anna (2012) A review of extreme value threshold estimation and uncertainty quantification. *Revstat – Statistical Journal*, Vol. 10 (1), 33–60.

Shreve, Steven E. (2004) *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance, Springer, United States of America.

Shreve, Steven E. (1996) *Stochastic Calculus for Finance I: The Binomial Asset-Pricing Model*. Springer Finance, Springer, United States of America.

Slim, Skander – Koubaa, Yosra – BenSaïda, Ahmed (2016) Value-at-Risk under Lévy GARCH models: Evidence from global stock markets. *Journal of International Financial Markets, Institutions & Money*, Vol. 46 (1) 30–53.

Smith, Richard L. (1991) Introduction to Gnedenko (1943) On the Limiting Distribution of the Maximum Term in a Random Series. In: *Breakthroughs in Statistics: Foundations and Basic Theory,* eds. Kotz, Samuel – Johnson, Norman L., 185–194. Springer, New York.

Stoyanov, Stoyan V. – Rachev, Svetlozar T. – Fabozzi, Frank J. (2013) Value-at-Risk. In: *Encyclopedia of Financial Models, Volume III*, eds. Fabozzi, Frank J., John Wiley & Sons, New Jersey, United States of America.

Svensson, L. E. O. (1995) Estimating Forward Interest Rates with the Extended Nelson & Siegel Method. Quarterly Review, Central Bank of Sweden. 1995-3, 13–26.

Szubzda, Filip – Chlebus, Marcin (2019) Comparison of Block Maxima and Peaks Over Threshold Value-at-Risk models for market risk in various economic conditions. *Central European Economic Journal*, Vol. 6 (53), 70–85.

Vasiček, Oldrich (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics*, Vol. 5 (2) 177–188.

Yi, Yanping – Feng, Xingdong – Huang, Zhuo (2014) Estimation of extreme value-at-risk: An EVT approach for quantile GARCH model. *Economics Letters*, Vol. 124 (3) 378–381.

Youssef, Manel – Belkacem, Lotfi – Mokni, Khaled (2015) Value-at-Risk estimation of energy commodities: A long-memory GARCH-EVT approach. *Energy Economics*, Vol. 51 (1) 99–110.

Appendices

7.1 Explanation of the use of AI

No AI tools have been used to produce any of the text in this thesis. AI has been used as an additional check for validity of the codes used in the econometric section. These checks have been prompted by providing the LLM with the code in question and asking if it correctly implements the method of the relevant paper. The LLM chosen was OpenAI's ChatGPT v5.2.