Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Jaroslav Baran

Analýza a porovnání různých modelů pro Value at Risk na nelineárním portfoliu

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Jiří Witzany, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční a pojistná matematika

2009

Děkuji svému vedoucímu diplomové práce RNDr. Jiřímu Witzanymu, Ph.D. za trpělivost, ochotu, konzultace a cenné připomínky.
Děkuji Katke a svým rodičům, že při mně stáli.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 30. 6. 2009 Jaroslav Baran

Contents

List of Tables

List of Figures

Název práce: Analýza a porovnání různých modelů pro Value at Risk na nelineárním portfoliu
Autor: Jaroslav Baran
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: RNDr. Jiří Witzany, Ph.D.
E-mail vedoucího: `witzanyj@vse.cz`

Abstrakt: V práci jsou popsané nástroje pro měření tržního rizika - Value-at-Risk (VaR) a Expected Shortfall (ES). Vysvětlena jsou: parametrická metoda, Monte Carlo simulace a historická simulace. V další části je podrobněji rozebrána teorie extrémních hodnot (EVT). Je vybudována základní teorie a představena metoda maxim nad prahem (peaks-over-threshold), která je následně použita pro modelování chvostu rozdělení ztrát zobecněným Paretovým rozdělením. Tato metoda je souběžně ilustrováná na výpočtu hodnoty v riziku (VaR) a očekávané podmíněné ztráty (ES) pro PX Index. Rovněž jsou rozebrány praktické otázky jako vícedenní horizont, časovo podmíněná volatilita výnosů a zpětní testování. Aplikace parametrické metody, historické simulace a teorie extrémních hodnot je následně prezentováná i na nelineárním portfoliu navrženým v programu Mathematica a výsledky jsou projednány.

Klíčová slova: hodnota v riziku (Value-at-Risk), očekávaná podmíněná ztráta (Expected Shortfall), teorie extrémních hodnot

Title: Calculation and Comparison of Several Value at Risk Models for Nonlinear Portfolio
Author: Jaroslav Baran
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Jiří Witzany, Ph.D.
Supervisor's e-mail address: `witzanyj@vse.cz`

Abstract: The thesis describes Value-at-Risk (VaR) and Expected Shortfall (ES) models for measuring market risk. Parametric method, Monte Carlo simulation, and Historical simulation (HS) are presented. The second part of the thesis analyzes Extreme Value Theory (EVT). The fundamental theory behind EVT is built, and peaks-over-threshold (POT) method is introduced. The POT method is then used for modelling the tail of the distribution of losses with Generalized Pareto Distribution (GPD), and is simultaneously illustrated on VaR and ES calculations for PX Index. Practical issues such as multiple day horizon, conditional volatility of returns, and backtesting are also discussed. Subsequently, the application of parametric method, HS and EVT is demonstrated on a sample nonlinear portfolio designed in Mathematica and the results are discussed.

Keywords: Value at Risk, Expected Shortfall, Extreme Value Theory

Introduction

Value at risk (*VaR*) has become an international standard for measuring market risk. Its position strengthened after it was adopted as a preferred measure of market risk under Basel II accord[1].

VaR measures the probable loss in a value of an investment over a specified time interval, at a given confidence level, and under normal market conditions. This risk is expressed in money units or as a percentage change in the value of a portfolio. In this work, we explain several methods for calculating *VaR* and apply them on a sample nonlinear portfolio. The work is divided into four chapters and three appendices that are placed after references.

Chapter 1 describes the theory behind parametric method, Monte Carlo, and Historical Simulation of *VaR*. Variance forecasting by exponentially weighted moving average model is explained, and portfolio non-linearity is discussed. Monte Carlo and Historical *VaR* methodologies use scenario sets, thus are non-parametric. It means that returns (losses) from each scenario are sorted and particular scenario is the estimated *VaR*.

Lately there has been *VaR* criticism about its ability to properly capture high loss quantiles. Some even say that *VaR* rather creates than reduces risk. Chapter 2 discusses the drawbacks of *VaR*, and presents and alternative quantile based measure of risk called Expected Shortfall (*ES*), which focuses on the average of the worst probable losses.

In chapter 3, a significant amount of space is devoted to *VaR* and *ES* estimates under Extreme value theory (*EVT*). In *EVT*, one does not investigate the whole distribution of returns (or losses), but only focuses on the tails of the distribution, because the tail is of primary interest. No distributional assumption for the underlying returns has to be made, and only tails are modelled. Strictly speaking, the returns (or losses) in the tails (the *extremes*) are fit with Generalized Pareto Distribution and desired quantile risk measures are then estimated. Both unconditional and conditional *EVT* methods are discussed and their use is demonstrated on Prague stock exchange PX Index.

[1]Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework: The First Pillar - Minimum Capital Requirements

Chapter 4 presents an application of parametric (delta and delta-gamma) method, historical simulation, and methods based on EVT (both conditional and unconditional) for estimating VaR and ES on a sample portfolio. Within parametric method, another section is devoted to analyze nonlinear effect of options in portfolio. The results are then discussed.

Appendix A describes Cholesky factorisation of a symmetric positive definite matrix.

Appendix B is devoted FX options. We first present Garman-Kohlhagen formula for pricing FX options and then we discuss T-day stochastic volatility estimation of option's underlying currency pair. We demonstrate the use of Drost-Nijman formula that converts one-day volatility into T-day volatility on the calculation of the premium of an FX option, which is then used in the sample portfolio in Chapter four. At the end of the appendix, we shortly mention alternative volatility estimators.

Finally, appendix C discusses the idea of mapping fixed income instruments into standardized positions, and thus reducing the number of risk factors used for calculation of risk estimates.

The enclosed compact disc contains Mathematica code, the time series, and other files that were used for related simulations and calculations, and thereby complete the thesis.

Chapter 1

Value-at-Risk

A financial risk is modelled as a random variable, which represents return on an asset or the future net worth of the asset. We view market risk as a possible fluctuation of the value of the asset or a portfolio. A risk measure quantifies this risk. It maps risk on $\mathbb{R}$. The risk measures are still being developed and risk management is an interesting and evolving field where theory meets practice as both academics and risk managers strive to construct precise risk measure. In this work, two widely used risk measures are discussed: Value-at-risk and Expected Shortfall.

Value-at-Risk is a measure of the maximum potential change in value of a portfolio of financial instruments with a given probability over a pre-set horizon, RiskMetrics - Technical Document [15].

Definition 1. *Let Δt be the time horizon, portfolio $V(t, S_1(t), \ldots, S_n(t))$ be the function of t and risk factors $S_i(t)$, and let L denote the loss in the portfolio value during Δt, that is $L = -\Delta V$ where $\Delta V = V(t + \Delta t, \boldsymbol{S}(t + \Delta t)) - V(t, \boldsymbol{S}(t))$, and $100(1-\alpha)\%$ the confidence level, $\alpha \in (0, 1)$. VaR is defined as the $(1-\alpha)$ quantile $q_L(1-\alpha)$ of the loss in portfolio value in $[t, t + \Delta t]$,*

$$VaR_{\alpha,t+\Delta t} = \inf\{q|P(L \leq q) > 1-\alpha\} = \sup\{q|P(L \leq q) < 1-\alpha\}. \tag{1.1}$$

Equivalently, we can write $VaR_{\alpha,t+\Delta t} = F_L^{-1}(1-\alpha) = q_L(1-\alpha) = -q_{\Delta V}(\alpha)$, where F^{-1} is the inverse of the cumulative distribution function (cdf) $F_L(q)$, and $F_L(q) = P(L \leq q)$. Therefore, VaR is the loss in the value of a portfolio over time Δt that is not exceeded with probability at least $1-\alpha$. Parameter α is usually equal to 0.01 or 0.05.

The Role of Distribution

Value at Risk is defined by the probability distribution of portfolio return ΔV, not by the probability distribution of the risk factors. From the definition, *VaR*'s (and other quantile based risk measures') accuracy depend on the assumption of return distribution. In this chapter, while explaining the parametric and Monte Carlo method, we assume that this distribution is normal, although empirical studies have shown that the distribution of ΔV is sometimes *skewed* (we are especially concerned with negatively skewed returns) and *leptokurtic* (with positive excess *kurtosis*), that is, empirical returns show higher probability of values around the mean than normally distributed returns (higher and sharper peaks), and higher probability of extreme values than in normal distribution (heavier tails). More, it has been observed that down moves in the markets are more severe than the up moves, volatilities are clustered, and instruments such as options include assymetry into distribution of returns. This is why we ease the assumption of normality in chapter 3, and assume Generalized Pareto distribution that fits the tail of the empirical data properly.

1.1 VaR - Parametric approach

In this section we briefly present parametric (variance-covariance) approach for calculating *VaR*.

1.1.1 Model Assumptions and Inputs

We start with a standard assumption that risk factor returns are normally distributed. We work with continuously compounded returns (logarithmic price changes) X_t,

$$X_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \tag{1.2}$$

where P_t is a price of a security at time t (business day). Similarly we write the j-day return $X_t(j)$ as

$$X_{t+j} = \ln\left(\frac{P_t}{P_{t-j}}\right), \tag{1.3}$$

which is a sum of j one day returns. For practical reasons, RiskMetrics [15] simplifies the portfolio return, and defines it as a weighted sum of individual returns

$$X_{p,t} = \sum_{i=1}^{n} w_i X_{i,t}, \tag{1.4}$$

where $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)'$ is the vector of portfolio weights and $X_{i,t}$ is the return on i-th risk factor. As mentioned above, to model future returns, we assume that returns (log prices changes) $X_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ are conditionally normally distributed, conditional on the information available at time t (past prices and volatilities),

$$X_t = \sigma_t \varepsilon_t \sim N(0, \sigma_t^2), \tag{1.5}$$

where σ_t is time dependent volatility and ε_t is independently and identically distributed (iid) random variable with $\mathbf{E}(\varepsilon_t) = 0$ and $Var(\varepsilon_t = 1)$. The expected return is assumed to be zero[1]. We use the fact that any linear combination of the returns is also conditionally normally distributed, that is

$$X_{p,t} \sim N(0, \sigma_{p,t}^2), \tag{1.6}$$

where

$$\sigma_{p,t}^2 = \boldsymbol{w}' \Sigma_t \boldsymbol{w}, \tag{1.7}$$

is the variance of the portfolio return and $\Sigma_t = \left(\sigma_{ij,t}^2\right)$ is the covariance matrix.

1.1.2 Forecasting Variance

In RiskMetrics [15], variance of an individual asset return, and its corresponding covariances is forecasted from historical data using single Exponentially Weighted Moving Average model (EWMA), where more weight is put on more recent observations. The EWMA variance and covariance forecasts[2] for the next period $t+1$ can be written in a recursive way

$$\begin{aligned} \sigma_{j,t+1}^2 &= \mathbf{E}_t(X_{j,t+1}^2) = \lambda \sigma_{j,t}^2 + (1-\lambda) X_{j,t}^2, \\ \sigma_{ij,t+1}^2 &= \mathbf{E}_t(X_{i,t+1} X_{j,t+1}) = \lambda \sigma_{ij,t}^2 + (1-\lambda) X_{i,t} X_{j,t}, \end{aligned} \tag{1.8}$$

$i, j = 1, \ldots, n$, where smoothing factor $\lambda \in (0,1)$ is optimal rate of decline over time, and the forecasts for the next period $t+1$ are conditioned on the information up to present time t. Next, the correlation forecast between $i-th$ and $j-th$ asset return is defined as

$$\rho_{ij,t+1} = \frac{\sigma_{ij,t+1}^2}{\sigma_{i,t+1} \sigma_{j,t+1}}, \tag{1.9}$$

where $\sigma_{j,t+1} = \sqrt{\sigma_{j,t+1}^2}$ is the volatility (standard deviation) of $X_{j,t+1}$.

For multiple T-day variance and covariance forecasts we can use a simple temporal rule that gives us following formulas

$$\begin{aligned} \sigma_{i,t+T}^2 = T\sigma_{i,t+1}^2 \quad \text{and} \quad \sigma_{i,t+T} &= \sqrt{T}\sigma_{i,t+1}, \\ \sigma_{ij,t+T}^2 &= T\sigma_{ij,t+1}^2. \end{aligned} \tag{1.10}$$

[1] this is to avoid inaccuracy in the estimation of the mean from past returns.

[2] by direct substitution of the equation (1.8) back into itself we get $\sigma_{ij,t+1}^2 = (1-\lambda)\sum_{n=1}^{N} \lambda^{n-1} X_{i,t+1-n} X_{j,t+1-n}$, where sum should run to ∞, but we only use finite number N of observations.

Considering correlations, Ts cancel out and correlation stays the same. The equation (1.10) can be derived with the help of basic properties of conditional expectation, concretely the "tower property" that states the following.

Theorem 1. *Let X be an integrable, real-valued random variable defined on a probability space $(\Omega, \mathcal{A}, P)$, and let $\mathcal{F}, \mathcal{G}$ be σ-algebras $\mathcal{F} \subset \mathcal{G} \subset \mathcal{A}$. Then*

$$\boldsymbol{E}(\boldsymbol{E}(X|\mathcal{G})|\mathcal{F}) = \boldsymbol{E}(X|\mathcal{F}) = \boldsymbol{E}(\boldsymbol{E}(X|\mathcal{F})|\mathcal{G}). \tag{1.11}$$

Now we can write forecasted variances over T periods conditioned to the information we have at the time t as

$$\begin{aligned}
\mathbf{E}_t\left(\sigma^2_{i,t+T}\right) &= \mathbf{E}_t\left(\lambda\sigma^2_{i,t+T-1} + (1-\lambda)X^2_{i,t+T-1}\right) \\
&= \mathbf{E}_t\left(\mathbf{E}_{t+T-2}\left(\lambda\sigma^2_{i,t+T-1}\right) + (1-\lambda)\mathbf{E}_{t+T-2}\left(X^2_{i,t+T-1}\right)\right) \\
&= \mathbf{E}_t\left(\lambda\sigma^2_{i,t+T-1} + (1-\lambda)\sigma^2_{i,t+T-1}\right) \\
&= \mathbf{E}_t\left(\sigma^2_{i,t+T-1}\right)
\end{aligned}$$

where $\mathbf{E}_t(.) = \mathbf{E}(.|\mathcal{F}_t)$ and $\mathcal{F}_t$ is σ-algebra generated by past returns $X_{i,t}$. Also, $\mathbf{E}_t(\sigma^2_{i,t+1}) = \sigma^2_{i,t+1}$. Since the T-day return is the sum of T continuously compounded daily returns, we can write

$$X_{i,t+T} = \sum_{k=1}^{T} \sigma_{i,t+k}\varepsilon_{i,t+k}$$

$$\sigma^2_{i,t+T} = \mathbf{E}_t(X^2_{i,t+T}) = \sum_{k=1}^{T} \mathbf{E}_t(\sigma^2_{i,t+k}) = T\mathbf{E}_t(\sigma^2_{i,t+1}) = T\sigma^2_{i,t+1}$$

$$\sigma_{i,t+T} = \sqrt{T}\sigma_{i,t+1},$$

thus we get a simple square root of time rule.

Finding lambda

Regarding smoothing factor λ, RiskMetrics [15] model considers following formula to determine the effective number of historical observations T

$$\alpha = (1-\lambda)\sum_{k=T}^{\infty} \lambda^k,$$

thus,

$$T = \frac{\ln\alpha}{\ln\lambda},$$

where α is the confidence level. In a portfolio with n risk factors, there are n variance and $\frac{n(n-1)}{2}$ covariance forecasts. Practically, it is convenient to choose one optimal smoothing factor λ for the whole variance covariance matrix. First, we determine λ for each risk factor from past return series of this factor. This is done by taking the minimum from root average squared variance forecast deviations (errors) for different λs. Recall that predicted variance for period $t+1$ is $\mathbf{E}_t(X_{t+1}^2) = \sigma_{t+1}^2$, therefore, our residual (estimated error) is

$$\varepsilon_{t+1} = X_{t+1}^2 - \mathbf{E}_t(X_{t+1}^2) = X_{t+1}^2 - \sigma_{t+1}^2,$$

the expectation of the error is 0 ($\mathbf{E}_t(\varepsilon_{t+1}) = \mathbf{E}_t(X_{t+1}^2) - \sigma_{t+1}^2 = 0$) and minimizing average squared errors between estimated variance and daily squared return observation gives us

$$\phi = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(X_{t+1}^2 - \hat{\sigma}_{t+1}^2(\lambda)\right)^2}.$$

This is done over different values of λ and the one with minimum ϕ is chosen. Similarly, we find ϕs for more than one day prediction.

Next, let n be the number of risk factors return series, λ_i the optimal λ for risk factor i and ϕ_i the minimum mean square error of ith risk factor, $i = 1, \ldots, n$. We can find the optimal λ as the weighted average of individual λ_is, where we put the highest weight on the lowest ϕ. Thus, the individual weight ϑ_i has the following form

$$\vartheta_i = \frac{\phi_i^{-1}}{\sum_{i=1}^{n}\phi_i^{-1}}, \tag{1.12}$$

and $\sum_{i=1}^{n}\vartheta_i = 1$. The optimal λ is then

$$\lambda = \sum_{i=1}^{n}\vartheta_i\lambda_i. \tag{1.13}$$

1.2 Calculating Value at Risk

1.2.1 Linear vs. Non-linear Positions

We dedicate this section to clarify following notions and concepts: linear, nonlinear, delta, gamma. These four words carry significant importance in market risk management.

Delta Δ - is the first derivative of the value V of an instrument or a portfolio with respect to the underlying instrument's price S,

$$\Delta = \frac{\partial V}{\partial S}.$$

Δ tells us how much the price of an instrument or a portfolio changes when the price of the underlying instrument changes by a small amount. Usually delta concept refers to derivatives but can be applied to other instruments, too. The underlying instrument can be equity, currency, fixed income instrument, commodity or a derivative. Loosely speaking, '*a delta of a derivative equal to 0.5 means that for a small change in the value of underlying instrument, the price of a derivative changes by approximately 0.5 × change of the underlying*'. With the change in the price of the underlying, Δ also changes.

Gamma Γ - is the second derivative of the value V of an instrument with respect to the underlying price S,

$$\Gamma = \frac{\partial^2 V}{\partial S^2} = \frac{\partial \Delta}{\partial S}.$$

Gamma measures the rate of change in Δ, that is, it measures how Δ changes as the price of the underlying instrument changes. If delta increases as the price of the underlying increases, then Γ is positive, more, the larger Γ, the more sensitive is Δ to the price of the underlying.

Taylor Series Expansion

The Taylor series is an infinitely differentiable function (of $k \geq 1$ variables) that can be expressed as an infinite weighted sum of its derivatives

$$f(x_1, \ldots, x_k) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} \frac{\partial^{n_1}}{\partial x_1^{n_1}} \cdots \frac{\partial^{n_k}}{\partial x_k^{n_k}} \frac{f(a_1, \ldots, a_k)}{n_1! \cdots n_k!} (x_1 - a_1)^{n_1} \cdots (x_k - a_k)^{n_k}.$$

Example. Assume that portfolio V is the function of time t and risk factor S. The Taylor series expansion of $V(t, S)$ to the second order about point (t, S) is

$$\Delta V = V(t + \Delta t, S + \Delta S) - V(t, S) \approx \frac{\partial V(t,S)}{\partial t}\Delta t + \frac{\partial V(t,S)}{\partial S}\Delta S + \frac{1}{2}\left(\frac{\partial^2 V(t,S)}{\partial t^2}\Delta t^2 + \frac{\partial^2 V(t,S)}{\partial S^2}\Delta S^2 + 2\frac{\partial^2 V(t,S)}{\partial t \partial S}\Delta t \Delta S\right).$$

Linearity

Following Pichler & Selitsch [19], a financial instrument is linear when the change in the value of the instrument (position) over time Δt is linear in the returns of its risk factors[3]. A change in the value of portfolio composed of linear instruments that depend on n risk factors S_i over one period ΔV can be written

[3] recall that market risk factors are interest rates, foreign exchange rates, prices of underlying instruments, etc.

as Taylor series to the first order

$$\Delta V = \sum_{i=1}^{n} \frac{\partial V}{\partial S_i} \Delta S_i = \sum_{i=1}^{n} \delta_i X_i,$$

$$\delta_i = \frac{\partial V}{\partial S_i} S_i, \quad X_{i,t} = \log\left(\frac{S_{i,t}}{S_{i,t-\Delta t}}\right) \approx \frac{\Delta S_i}{S_i}, \tag{1.14}$$

where δ_i is the sensitivity of the portfolio value with respect to i-th risk factor, or so-called *return adjusted delta*. This partial derivatives are calculated by increasing relevant interest rate by one basis point.

Non-Linearity

A financial instrument is non-linear when the change in the value of the instrument is nonlinear in the returns of its risk factors. To allow for this nonlinearity, we approximate the change in the portfolio value with the first two orders of Taylor series

$$\Delta V = \sum_{i=1}^{n} \delta_i X_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Gamma_{i,j} X_i X_j,$$

$$\Gamma_{i,j} = \frac{\partial^2 V}{\partial S_i \partial S_j} S_i S_j, \tag{1.15}$$

where $\Gamma_{i,j}$ is *return adjusted gamma*[4].

Example. Consider a zero-coupon bond with par value *F=100* and maturity T years. The price of the bond using continuous compounding is $P = 100e^{-r_T T}$, where r_T is the T-year spot rate (yield to maturity). The change in the value of the bond P with respect to the change in the yield r_T is approximately

$$\Delta P \approx -T 100 e^{-r_T T} \Delta r_T + \frac{1}{2} T^2 100 e^{-r_T T} \Delta r_T^2,$$

The term $-T100e^{-r_T T}$ is the bond's *Delta* and it accounts for linear change in the bond's price. The linear change in the bond's price is proportional to the *Delta* × change in the yield. The second term $T^2 100e^{-r_T T}$ is the risk exposure to the second derivative with respect to yield r_T, that is, bond's *Gamma*. Therefore, if spot rate r_T is the risk factor, then bond is a nonlinear financial instrument, and we approximated the change in the bond's price with first two derivatives, *Delta* and *Gamma*, with respect to the yield r_T. On the other hand, if we choose bond's price P as the risk factor, then the first derivative is equal to 1 and the bond is linear instrument in its price.

[4] Expressed in terms of ΔV, δ_i and $\Gamma_{i,j}$ take into account the size of the position and the change in the underlying.

Example. Consider a second simple example. An investor bought one Eurodollar futures contract (he lends one million dollars on the delivery date) at a quoted price $P = 100 - f_k$, where f_k is forward 3 months LIBOR starting in k months (futures contract expires in k months). The change in the futures price with respect to change in the forward rate is

$$\Delta P = -\Delta f_k,$$

therefore, Eurodollar futures is linear in its forward rate with *Delta* equal to -1.

After defining previous concepts, the natural idea is to model the extreme movements in the risk factors and investigate their effects on the change in the value of portfolio. For risk management, this straightforwardly leads to modelling the unfavorable changes in the value of portfolio, thus, calculating the worst expected loss.

1.2.2 Linear Value at Risk

We introduced the necessary tools to calculate *VaR*. We start with Linear *VaR* method, also called *Delta* approach. We assume linearity in the risk factors' returns, and we assume that these returns follow a multivariate normal distribution with zero mean, that is, $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$, where $\mathbf{X}$ is the vector of n risk factor returns, and Σ is $n \times n$ covariance matrix of returns. Recall that from (1.14) $\Delta V = \sum_{i=1}^{n} \delta_i X_i$. That can be written in a vector notation

$$\Delta V = \boldsymbol{\delta}^{\boldsymbol{T}} \boldsymbol{X}, \tag{1.16}$$

where $\boldsymbol{\delta}$ is a vector of sensitivities δ_i. Therefore, one day *VaR* of portfolio V is given by

$$VaR_{\alpha,t+1} = -z_\alpha \sqrt{\boldsymbol{\delta}^{\boldsymbol{T}} \Sigma \boldsymbol{\delta}}, \tag{1.17}$$

where z_α is the α quantile of normal distribution, and the expression $\boldsymbol{\delta}^{\boldsymbol{T}} \Sigma \boldsymbol{\delta}$ is the portfolio variance. Due to linearity between the change in the portfolio's value ΔV and the returns, ΔV is normally distributed, thus quantile of normal distribution can be used to calculate *VaR*.

1.2.3 Non-linear Value at Risk

This approach allows for non-linear relationship between ΔV and risk factor returns, that is, we assume that portfolio contains non-linear instruments, such as options. Equation (1.15) can be written in a matrix form

$$\Delta V = \boldsymbol{\delta}^{\boldsymbol{T}} \boldsymbol{X} + \frac{1}{2} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{\Gamma} \boldsymbol{X}, \tag{1.18}$$

where $\mathbf{\Gamma}$ is $n \times n$ matrix of *Gamma* sensitivities $\Gamma_{i,j}$. For simplicity, we neglect the terms of higher orders. Although, we still assume that individual risk factor returns are normally distributed, due to non-linear relationship, ΔV is not normally distributed. This is due to possible skewness that causes assymmetry of the distribution of ΔV and changes its moments, thus quantile of a normal distribution is no longer appropriate. We need to find the α quantile of the true distribution of ΔV. We do not know yet the distribution of ΔV, but we are able to calculate its moments from $\boldsymbol{\delta}$, $\mathbf{\Gamma}$ and Σ. One of the methods to find the quantiles of ΔV is Cornish-Fisher Expansion that directly approximates these quantiles.

Cornish-Fisher Expansion

This method approximates the desired quantile $z_{\Delta V,\alpha}$ of ΔV's distribution $F_{\Delta V}$ as a function of moments of $F_{\Delta V}$ and quantiles of the distribution of the risk factors' returns. The first moment and the second central moment of the distribution of ΔV are

$$
\begin{aligned}
Expectation: \quad \mathbf{E}(\Delta V) &= \mu_{\Delta V} = \frac{1}{2} tr\left[\mathbf{\Gamma}\mathbf{\Sigma}\right] \\
Variance: \quad Var(\Delta V) &= \sigma^2_{\Delta V} = \boldsymbol{\delta}^T \Sigma \boldsymbol{\delta} + \frac{1}{2} tr\left[\mathbf{\Gamma}\Sigma\right]^2,
\end{aligned}
\tag{1.19}
$$

where *tr(.)* is the *trace* of the $n \times n$ matrix $\mathbf{\Gamma}\Sigma$ (the sum of its eigenvalues). Higher standardized moments of ΔV are given by

$$
\mathbf{E}(X^k) = \frac{\frac{1}{2} k! \boldsymbol{\delta}^T \Sigma \left[\mathbf{\Gamma}\Sigma\right]^{k-2} \boldsymbol{\delta} + \frac{1}{2}(k-1)! tr\left[\mathbf{\Gamma}\Sigma\right]^k}{\left(\boldsymbol{\delta}^T \Sigma \boldsymbol{\delta} + \frac{1}{2} tr\left[\mathbf{\Gamma}\Sigma\right]^2\right)^{\frac{k}{2}}}, \quad k \geq 3, \tag{1.20}
$$

where X is the standardized value of ΔV, $X = \frac{\Delta V - \mathbf{E}(\Delta V)}{\sqrt{Var(\Delta V)}}$. For $k = 3$ we get *skewness* (the third standardized moment that measures the assymetry of the distribution) and for $k = 4$ we get *kurtosis* (the fourth standardized moment that measures the peak of the distribution). To a certain extent, they both describe the tails of the distributions.

In the case that the risk factors' returns are distributed normally, the expression for $z_{\Delta V,\alpha}$ using first four moments of ΔV is approximately

$$
\begin{aligned}
z_{\Delta V,\alpha} \approx{} & z_\alpha + \frac{1}{6}\left(z_\alpha^2 - 1\right)\mathbf{E}(X^3) \\
& + \frac{1}{24}\left(z_\alpha^3 - 3z_\alpha\right)\mathbf{E}(X^4) - \frac{1}{36}\left(2z_\alpha^3 - 5z_\alpha\right)\mathbf{E}(X^3)^2.
\end{aligned}
\tag{1.21}
$$

The non-linear *VaR* is then given by

$$
VaR_{\alpha,t+1} = -z_{\Delta V,\alpha}\sqrt{\sigma^2_{\Delta V}} + \mu_{\Delta V}. \tag{1.22}
$$

1.3 Monte Carlo Simulation

Monte Carlo method generates theoretical market movements (returns of the risk factors) from the statistical model of market data, in our case, from the assumption of normally distributed risk factors' returns. The objective of this approach is to repeatedly simulate risk factors' returns. After each simulation we revalue the portfolio of positions, that is, we compute the corresponding changes in portfolio value ΔV. Large sample of the simulated returns then gives a good approximation of the distribution of ΔV. We can then easily compute the empirical α-quantile of the approximated distribution of ΔV.

A standard MC approach is to use Cholesky factorisation (explained briefly in Appendix A) of the covariance matrix of returns to transform the independent random normal sequences of the returns to correlated random sample. These scenarios generated from random draws are then used to revalue the portfolio. The ordered results thus form the estimated empirical probability distribution of the changes in the value of portfolio ΔV. To calculate *VaR*, we take the desired empirical quantile from this distribution. We now describe Monte Carlo *VaR* at more length.

1.3.1 Simulating Scenarios

Simulating scenario means applying some factor to the current risk factor and obtaining a change in the risk factor value. Thus, we simulate the returns, these returns then change the value of underlying asset (portfolio) and a theoretical portfolio profit or loss is generated.

The return is modelled as in (1.2), (1.3), and (1.5), that is, at time t, the logarithmic price changes of the underlying asset (risk factor), $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$. We obtain the price of the risk factor at time T (time horizon) from the price today, P_0, and one day volatility forecast σ_1 of the return,

$$P_T = P_0 e^{\sqrt{T}\sigma_1 \varepsilon}, \tag{1.23}$$

where ε is standard normal variable. Therefore, we generate random standard normal variables ε-s to simulate the future prices P_T. These ε-s are independent but not correlated yet. To generate correlated random variables according to our covariance matrix Σ, as already mentioned, we use Cholesky factorisation.

We estimate the correlation matrix of returns Π from historical data, and then we decompose Π into Cholesky (lower triangular) matrix L and its transpose L^T, $\Pi = L\,L^T$. Next, we multiply the lower triangular matrix L with generated $n \times 1$ vector of random standard normal variables ε_i to arrive at $n \times 1$ vector $\boldsymbol{\xi}$ of standard normal random variables correlated according to Π,

$$\boldsymbol{\xi} = L\boldsymbol{\varepsilon}.$$

After this simulation we revalue the single positions and the whole portfolio, e.g., the future price of a j-th risk factor at time T is $P_{j,T} = P_{j,0}e^{\sqrt{T}\sigma_{j,1}\xi_j}$.

The disadvantage of MC is that it is computationally intensive to price each instrument each time when we revalue the whole portfolio (e.g., one has to run option pricing formula for every option in the portfolio for each simulation). It is possible to substitute full valuation method with delta-gamma approximation explained previously, however, one then loses the opportunity of the full simulation of the distribution of the change in the portfolio's value. If we use delta-gamma approximation, we revalue the portfolio to obtain the empirical distribution of ΔV by using formula $\Delta V = \boldsymbol{\delta^T r} + \frac{1}{2}\boldsymbol{r^T \Gamma r}$.

To simulate more realistic returns, it is possible to use other than normal distribution for the returns, e.g., Student t-distribution, which has heavier tails.

1.3.2 Finding Quantile

To calculate the α-quantile of the distribution of ΔV we first sort the results from the simulations of ΔV in ascending order

$$\Delta V^{(1)} \leq \Delta V^{(2)} \leq \ldots \leq \Delta V^{(N)},$$

where $\Delta V^{(i)}$ is the i-th smallest value from the total of N simulations. Value at risk is then empirical α-quantile of the distribution of ΔV, that is

$$VaR_\alpha = \Delta V^{([\alpha N])}, \tag{1.24}$$

where $[\alpha N] = \max\{m | m \leq \alpha N, m \in \mathbb{N}\}$ is the integer part of αN.

1.4 Historical Simulation

Historical simulation (*HS*) uses only empirical distribution of portfolio returns (losses), therefore, does not depend on any distributional assumption. Of course, there are other assumptions. Probably the most important one is that we assume that historical returns from our sample reasonably describe the distribution of future returns. The advantage of *HS* is that it accounts for fat tails, kurtosis, or skewness of actual distribution.

1.4.1 Simple Historical Simulation

HS uses historical returns on market variables to construct a distribution of future returns. To construct this distribution of returns, we apply last N days[5] returns on a current value of portfolio, therefore we get N hypothetical portfolio values (see e.g. Hull [13, p. 348]). We sort these values into ascending order and take empirical α-quantile z_α^t of this hypothetical distribution of changes in the portfolio value to arrive at the next day's *VaR* estimate

$$VaR_{\alpha,\,t+1}^{HS} \;=\; z_\alpha^t. \tag{1.25}$$

The frequency of large losses that occured during last N observations is thus reflected in the results. Thus, to estimate the next day's *VaR* on day t, we take sample quantile from the corresponding from last N returns). If this quantile lies between two values, we can interpolate it. To estimate extreme quantiles, obviously, we need large sample, e.g. at least $1/\alpha$ (to calculate 99.9% *VaR*, we need at least $1/0.001 = 1000$ observations).

The disadvantage of this approach arises when we estimate the extreme events. In the tails, the empirical distribution of returns is 'very' discrete. While the most returns fall within the central part of the distribution and are close to each other, there are few observations left for the tails. The intervals between nearby returns broaden as we move to the extremes, thus, estimated *VaR* for α very low might lead to either underestimation or overestimation (including or excluding few samples may lead to large swings in *VaR*, see Danielsson & De Vries [6]). More, *HS* does not take into account volatility of the returns. It assumes the distribution to be fairly constant over the sample period, and thus becomes poorer predictor of *VaR* during high volatility periods, especially when high volatilities cluster together. During these periods, such *VaR* estimate can then be exceeded several times in a row.

[5] In practice, one usually takes one-year historical period, that is, some 250 past returns.

Chapter 2

Expected Shortfall

Simultaneously as the use of *VaR* has been rapidly extending across banks, some inconsistencies and drawbacks in the model were found. This lead to modifications and extensions of the model and to the rise of alternative models that measure market risk. One of them that we present, Expected Shortfall (*ES*)[1] model, measures the expected loss of a portfolio in the α % worst cases. We turn to the work of Acerbi & Tasche [1], but first, we briefly mention the issues raised about *VaR* and the resulting need for its alternatives.

2.1 Motivation - Imperfections of Value at Risk

As a motivation, we use a simple example that we borrowed from Dowd & Blake [9].

Example[2]. Investor buys two identical bonds A, B with returns ΔA, ΔB, respectively. The probability of independent default of each bond is 4%, and there is a loss of 100 in case of default or 0 otherwise. The 95%-VaR of each bond is 0, therefore $VaR_{0.95}(\Delta A) = VaR_{0.95}(\Delta B) = VaR_{0.95}(\Delta A) + VaR_{0.95}(\Delta B) = 0$. We suffer a loss of 0 with probability $0.96^2 = 0.9216$, a loss of 200 with probability $0.04^2 = 0.0016$, and a loss of 100 with probability $1 - 0.9216 - 0.0016 = 0.0768$, therefore $VaR_{0.95}(\Delta A + \Delta B) = 100$. We see that $VaR_{0.95}(\Delta A + \Delta B) = 100 > 0 = VaR_{0.95}(\Delta A) + VaR_{0.95}(\Delta B)$. We would expect that if we diversify our portfolio by investing into two instruments instead of one, we also diversify (lessen) the risk of the portfolio. We see that if we choose *VaR* as the appropriate risk measure, this is not the case. *VaR* violates the *axiom of subadditivity* (the overall risk of two bonds is larger than the sum of risks of individual bonds, while it should be lower). In this case, risk manager may assume too much risk when imposing limits

[1]in literature, Expected Shortfall is often called Conditional Value at risk (*CVaR*).
[2]for more examples see e.g. Artzner [3].

on traders.

To manage risks efficiently, one should choose a risk measure that satisfies axioms that are essential or inevitable. Artzner et al. [3] introduced four axioms for risk measures that, they argue, should hold for every effective risk measure. These axioms are

1. *Translation (drift) Invariance:* $X \in \mathcal{G}, a \in \mathbb{R} \Rightarrow \rho(X + a) = \rho(X) - a$.
   Adding a constant return to X decreases the required reserves (risk) by that amount.

2. *Subadditivity:* $X_1, X_2 \in \mathcal{G} \Rightarrow \rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.
   The risk for the combination of two returns on instruments is less than the risk for each separate return (diversification).

3. *Positive homogeneity:* $\lambda \geq 0, X \in \mathcal{G} \Rightarrow \rho(\lambda X) = \lambda\rho(X)$.
   The risk of two returns with same relative value is linear in the scale.

4. *Monotonicity:* $X_1, X_2 \in \mathcal{G}, X_1 \leq X_2 \Rightarrow \rho(X_2) \leq \rho(X_1)$.
   If the return X_2 is always greater than X_1, then X_2 is less risky.

We think of a risk measure as a map $\rho : \mathcal{G} \to \mathbb{R}$, where $\mathcal{G}$ is the set of all risks (e.g. $\mathcal{G} = \mathbb{R}^n$). That is, ρ maps the riskiness of the portfolio to required reserves to cover losses from unfavorable movements that regularly occur. A measure that satisfies these four axioms is then called *coherent*[3]. For $\rho(X) = VaR_\alpha(X)$ we saw that *VaR* is not a coherent measure of risk as it is not subadditive. Artzner [3] proposes a general coherent risk measure as '*the supremum of the expected negative of the final net worth for some collection of generalized scenarios or probability measures $\mathcal{P}$ on states of the final net worth*',

$$\rho(X) = \sup_{P \in \mathcal{P}} \mathbf{E}_P[-X].$$

This steer towards finding some kind of a weighted average of the scenarios of the worst cases of loss. It sounds more rational to find the expected loss (*ES*) than to find the minimum loss (*VaR*) from the set of worst losses. In other words, we are interested in the shape of the tail of the underlying distribution of risk factor returns, and not only where this tail starts. *VaR* ignores the tails (large losses) while *ES* measures them.

[3] There are other risk measures that have been defined for risk measurement purposes related to coherent measures of risk or based on alternative set of axioms, e.g. convex, dynamic, distortion, spectral risk measure. For discussion, see e.g. Dowd & Blake [9].

2.2 Calculating Expected Shortfall

Recall that Monte Carlo simulation calculates VaR_α as $\Delta V^{([\alpha N])}$. But Monte Carlo method simulates the whole distribution of ΔV, thus, it allows us to find any desired quantile. When we want to estimate Expected Shortfall in the $\alpha\%$ worst cases, it naturally comes as an average of the $\alpha\%$ largest losses

$$ES_n^{(\alpha)}(\Delta V) = -\frac{\sum_{i=1}^{[n\alpha]} \Delta V^i}{[n\alpha]}. \tag{2.1}$$

Generally, ES is defined as

Definition 2 (Expected Shortfall)**.** *Let ΔV be the portfolio P/L and $\alpha \in (0,1)$ the confidence level, and $q_{\Delta V}(\alpha) = q(\alpha)$ the α-quantile of ΔV. The Expected Shortfall is defined as*

$$ES^{(\alpha)}(\Delta V) = -\frac{1}{\alpha}\Big(\boldsymbol{E}\left[\mathbb{I}_{\Delta V \le q(\alpha)}\right] + q(\alpha)\Big(\alpha - P\big[\Delta V \le q(\alpha)\big]\Big)\Big). \tag{2.2}$$

In case that ΔV is discretely distributed, in the estimate (2.1), $\Delta V^{([\alpha N])}$ can occur more than one time. We assume the underlying risk factor returns to be continuously distributed, therefore, it holds that $P\big[\Delta V \le q(\alpha)\big] = \alpha$ and the term $q(\alpha)\Big(\alpha - P\big[\Delta V \le q(\alpha)\big]\Big)$ vanishes. $ES^{(\alpha)}$ then becomes

$$ES^{(\alpha)}(\Delta V) = -\frac{1}{\alpha}\left(\boldsymbol{E}\left[\mathbb{I}_{\Delta V \le q(\alpha)}\right]\right) = -\boldsymbol{E}\left[\Delta V | \Delta V \le q(\alpha)\right]. \tag{2.3}$$

This conditional expectation of ΔV below the quantile $q(\alpha)$ is also called *tail conditional expectation*. An equivalent expression of 2.2 is given by the negative mean of $F^{-1}(u)$ on the confidence level interval $u \in (0, \alpha]$

$$ES^{(\alpha)}(\Delta V) = -\frac{1}{\alpha}\int_0^\alpha F^{-1}(u)du \tag{2.4}$$

where $F^{-1}(u)$ is the inverse function of $F(s)$, $F^{-1}(u) = \inf\{s | F(s) \ge u\}$. This is a straightforward relation to *VaR* since $VaR_u = -F^{-1}(u)$. We note that when *ES* and *VaR* are defined for all values $\alpha \in (0,1)$, they both completely determine the distribution of ΔV. *ES* is, however, much more sensitive to the model of the tail of the distribution, which is usually calibrated on historical data.

2.3 Properties of Expected Shortfall

- *ES* satisfies *subadditivity*

$$\begin{aligned} ES_n^{(\alpha)}(\Delta V_1 + \Delta V_2) &= -\frac{\sum_{i=1}^{[n\alpha]}(\Delta V_1 + \Delta V_2)^i}{[n\alpha]} \\ &\leq -\frac{\sum_{i=1}^{[n\alpha]}(\Delta V_1^i + \Delta V_2^i)}{[n\alpha]} \\ &= ES_n^{(\alpha)}(\Delta V_1) + ES_n^{(\alpha)}(\Delta V_2). \end{aligned} \tag{2.5}$$

More, *ES* satisfies all axioms of coherence, therefore, it is a coherent measure of risk[4].

- ES^α *is continuous with respect to* α. Small changes in the confidence level α may lead to large changes in *VaR* of some discontinuously distributed financial instruments (loans, derivatives), in general, *VaR* is not continuous with respect to α. ES^α is continuous and not sensitive to small changes in α.

- ES^α *is monotonous in* α. The smaller the level α, the larger the risk.

- ES^α *generalizes standard deviation as a measure of risk* in case that portfolio returns are normally distributed (*linear VaR*).

Theorem 2. *If portfolio return* $\Delta V = \boldsymbol{\delta^T r}$ *is normally distributed with zero mean and covariance matrix* $\boldsymbol{\delta^T}\Sigma\boldsymbol{\delta}$, *then*

$$ES^{(\alpha)} = \frac{\phi(z_\alpha)}{\alpha}\sqrt{\boldsymbol{\delta^T}\Sigma\boldsymbol{\delta}}, \tag{2.6}$$

where $\phi(z_\alpha)$ *is the probability density function (pdf) of standard normal distribution, and* z_α *is the* α*-quantile of the standard normal variable* Z, $P[Z > z_\alpha] = \alpha$.

Proof. Set $\sigma^2 = \boldsymbol{\delta^T}\Sigma\boldsymbol{\delta}$. We have,

$$\begin{aligned} ES^\alpha &= -\boldsymbol{E}[\Delta V | \Delta V \leq q(\alpha)] \\ &= -\frac{1}{\alpha\sigma\sqrt{2\pi}}\int_{-\infty}^{q(\alpha)} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= -\frac{\sigma}{\alpha\sqrt{2\pi}}\int_{-\infty}^{z_\alpha} y \exp\left(-\frac{y^2}{2}\right) dy \\ &= \frac{\sigma}{\alpha\sqrt{2\pi}} \exp\left(-\frac{z_\alpha^2}{2}\right) = \frac{\phi(z_\alpha)}{\alpha}\sigma. \end{aligned}$$

□

[4]For the proof of the coherence of *ES* see Acerbi [2]

Chapter 3

Extreme Value Theory

In this chapter, we take a closer look at the tail of the distribution (of individual or portfolio returns). We are especially interested in the tail of the losses because the tail is where extreme losses occur. Extreme Value Theory examines the tail area of the distribution (e.g it estimates high quantiles of a loss distribution). It studies these rare events and utilizes to the most the little information that is usually available about them. This theory has been recently widely popularized in the field of finance, although it has a vigorous history in insurance, e.g. in modelling large insurance losses. Namely, we can mention the book *Modelling extremal events for insurance and finance* by Embrechts, Klüppelberg, and Mikosch, or various papers from authors such as McNeil [16] [17] [18], Danielsson [6], de Vries, Reiss, Smith, Rootzen, Tajvidi, Longin, etc. Furthermore, we believe that many new papers on the financial applications of EVT will arise in following years due to recent extreme data available from the global financial crisis of 2008-2009.

In the following text, we follow the papers of Gilli & Kellezi [12] and McNeil & Frey [18]. We turn our focus to observations that exceed some high threshold (e.g. 95% quantile). As already mentioned, tails of the Normal distribution are often thinner than observed, therefore, we model the tails with another distribution, namely, *Generalized Pareto Distribution* (GPD), and apply it to our risk measures *VaR* and *ES*. We start with the basic theory that lies behind the study of extreme values, and we show an example how to calculate *VaR* and *ES*.

3.1 Generalized Extreme Value Distribution

Let us define maximum of sequence of *iid* random variables (observations of losses) $X_1, \ldots, X_n$ as $M_n = \max(X_1, \ldots, X_n)$. The cumulative distribution function of M_n is then

$$P[M_n < x] = P[\max(X_1, \ldots, X_n) < x] = P[X_1 < x, \ldots, X_n < x] = F^n(x),$$

where F is cdf of X_1. We notice that

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 1 & if\ F(x) = 1 \\ 0 & else, \end{cases}$$

for given x, thus the limit is degenerate. However, after normalizing this sequence, it converges to a well defined law.

Theorem 3 ((Fisher & Tippet, 1928), (Gnedenko, 1943))**.** *Let $X_1, \ldots, X_n$ be a sequence of iid random variables. If there exist real norming constant b_n, $a_n > 0$, and a non-degenerate cdf H such that*

$$\lim_{n\to\infty} P\left[\frac{M_n - b_n}{a_n} \leq x\right] = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x), \tag{3.1}$$

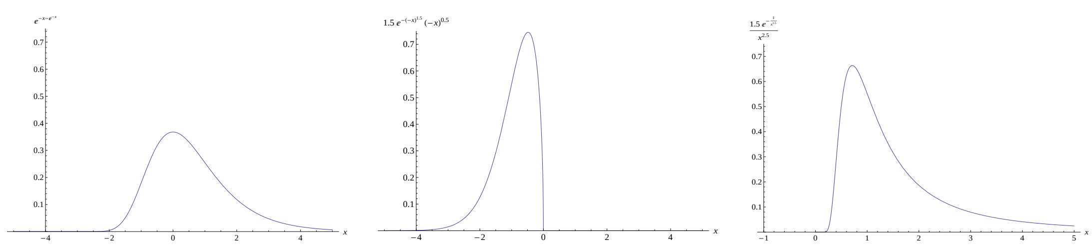
then H is one of the following cdfs:

$$Fr\acute{e}chet: \quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp\{-x^{-\alpha}\}, & x > 0, \end{cases} \qquad \alpha > 0,$$

$$Weibull: \quad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x^{\alpha})\}, & x \leq 0, \\ 1, & x > 0, \end{cases} \qquad \alpha > 0,$$

$$Gumbel: \quad \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

These three distributions are called *extreme value distributions*. We illustrate the shape of the probability density functions for *Fréchet*, *Weibull*, and *Gumbel* distributions in Figure 3.1.



(a) *Gumbel* (b) *Weibull* (c) *Fréchet*

Figure 3.1: Pdfs for (a), (b), (c), distributions, $\alpha = 1.5$.

Alternatively, these can be represented by one cdf known as *generalized extreme value distribution*

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-e^{-x}\right\}, & \xi = 0, \end{cases} \qquad x \in \mathbb{R}, \tag{3.2}$$

where $\mu \in \mathbb{R}$, and $\sigma > 0$. The parameter ξ is called the tail index. The *GEV* distribution corresponds to

$$\begin{array}{lll} Fréchet & \text{for} & \xi = \alpha^{-1} > 0 \\ Weibull & \text{for} & \xi = -\alpha^{-1} < 0 \\ Gumbel & \text{for} & \xi = 0. \end{array}$$

According to limit (3.1), the normalized maxima converge in distribution to $H(x)$ for a given F, in other words, F is in the *maximum domain of attraction* of H_ξ for some ξ. Theorem 3 (also called extreme value theorem) is a general result in extreme value theory. It is similar to central limit theorem in the way that instead of taking the average of an increasing sample, we take the sample maximum and investigate its asymptotic distribution. Since we are interested in extreme returns, the advantage of this theorem is that we know the form of the limiting distribution of extreme returns (and we can calculate extreme quantiles), although we do not need to know or assume the distribution of all returns.

3.2 Generalized Pareto Distribution

The *General Pareto Distribution* describes the limit distribution of scaled excesses over high thresholds.

Definition 3 (GPD)**.** *If X is a random variable (daily loss) with two-parameter Generalized Pareto Distribution, then the distribution function of X has the form*

$$G_{\xi,\beta} = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \xi \neq 0, \\ \\ 1 - \exp(-x/\beta), & \xi = 0, \end{cases} \tag{3.3}$$

where $\beta > 0$, and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

In case $\xi = 0$, we work with a limit $\lim_{\xi \to 0} \left(1 - (1 + \xi x/\beta)^{-1/\xi}\right) = 1 - \exp(-x/\beta)$. Parameter ξ (the tail index) accounts for the shape of the distribution and β is the parameter of the scale. The tail index ξ is the same as for generalized extreme value distribution. For $\xi \neq 0$, $G_{\xi,\beta}$ is a reparameterized Pareto distribution, for $\xi = 0$, $G_{\xi,\beta}$ is the exponential distribution. For $\xi > 0$, $G_{\xi,\beta}$ is not exponentially bounded, therefore, it is heavy-tailed. The k-th moment of *GPD*, $\boldsymbol{E}[X^k]$, is finite for $\xi < 1/k$. The *GPD* can be extended with a location parameter μ, $G_{\xi,\mu,\beta}(x) = G_{\xi,\beta}(x - \mu)$.

First derivative of (3.3) yields the density of *GPD*

$$g_{\xi,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1+\frac{\xi}{\beta}x\right)^{-1-1/\xi}, & \xi \neq 0, \\ \\ \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right), & \xi = 0. \end{cases} \tag{3.4}$$

The tail of the density fattens and the peaks are sharpening with increasing ξ while with increasing β the central part of the density gets more flat.

3.2.1 The Distribution of Excess Losses

Definition 4 (Excess Distribution)**.** *Let X be a random variable. The conditional distribution function F_u of excess losses over a threshold u is defined as*

$$F_u(y) = P[X - u \le y | X > u],$$

for $0 \le y \le x_F - u$, x_F is the right endpoint of F, that is $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} \le \infty$, and $y = x - u$ are the excesses over u.

This can be written in terms of F,

$$\begin{aligned} F_u(y) &= \frac{P[X-u \le y, X > u]}{P[X > u]} = \frac{P[u < X \le u + y]}{1 - P[X \le u]} \\ &= \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}. \end{aligned} \tag{3.5}$$

We are interested in estimating the extremes, that is, F_u. The following theorem is an important result in Extreme Value Theory.

Theorem 4 ((Balkema & de Haan, 1974), (Pickands, 1975))**.** *Let $X_1, \ldots, X_n$ be a sequence of iid random variables with distribution function F that converges to the Generalized Extreme Value distribution (GEV) H_ξ (F is in the maximum domain of attraction of H_ξ, $F \in \mathcal{D}(H_\xi)$). Then there exists positive real function $\beta(u)$, such that*

$$\lim_{u \to x_F} \sup_{0 \le y < x_F - u} \left| F_u(y) - G_{\xi,\beta(u)}(y) \right| = 0. \tag{3.6}$$

That is, for large u approaching x_F, excess function F_u converges to GPD $G_{\xi,\beta}$. All common continuous distributions satisfy the condition $F \in \mathcal{D}(H_\xi)$. This theorem allows us to model the distribution of the tails above sufficiently high thresholds. To do that, we need to choose the right u and estimate ξ and β from the extreme losses(negative returns above u) from the historical observations or simulation. The right u must be high enough to approximate the convergence and low enough to leave enough extreme data. This method of modelling extreme events under *GPD* is called *Peaks Over Thresholds* method.

3.2.2 Estimating Tails

According to (3.6), $F_u(y) = G_{\xi,\beta(u)}(y)$ for large u. The expression for underlying distribution function $F(x)$ thus becomes

$$F(x) = (1 - F(u))G_{\xi,\beta(u)}(x - u) + F(u), \tag{3.7}$$

for $x > u$. Next, we need to estimate the value $F(u)$ to find the corresponding quantile to u. This can be done from the data by empirical distribution function $\hat{F}(u) = (n - N_u)/n$, where n denotes losses and N_u are losses above threshold u. We denote the estimates of ξ and β as $\hat{\xi},\ \hat{\beta}$. The tail estimator of $F(x)$ is given by

$$\begin{aligned}\hat{F}(x) &= \frac{N_u}{n}\left(1 - \left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}\right) + \left(1 - \frac{N_u}{n}\right)\\ &= 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}},\end{aligned} \tag{3.8}$$

for $x > u$. $\hat{F}(x)$ is also *GPD* with the shape parameter ξ, scale parameter $\bar{\beta} = \beta(1 - \hat{F}(u))^{\xi}$ and location parameter $\bar{\mu} = u - \bar{\beta}((1 - \hat{F}(u))^{-\xi} - 1)/\xi$,

$$\begin{aligned}\hat{F}(x) &= 1 - \left(1 + \frac{\xi}{\beta(1-\hat{F}(u))^{\xi}}\left(x - u + \frac{\beta(1-\hat{F}(u))^{\xi}}{\xi}\left((1-\hat{F}(u))^{-\xi} - 1\right)\right)\right)^{-\frac{1}{\xi}}\\ &= 1 - \left(1 + \frac{\xi(x-u)}{\beta(1-\hat{F}(u))^{\xi}} + \left(1 - \hat{F}(u)\right)^{-\xi} - 1\right)^{-\frac{1}{\xi}}\\ &= 1 - \left(\frac{\xi(x-u)}{\beta(1-\hat{F}(u))^{\xi}} + \frac{1}{(1-\hat{F}(u))^{\xi}}\right)^{-\frac{1}{\xi}}\\ &= 1 - \left(1 - \hat{F}(u)\right)\left(1 + \frac{\xi}{\beta}(x-u)\right)^{-\frac{1}{\xi}}\\ &= 1 + \left(1 - \hat{F}(u)\right)(-1 + G_{\xi,\beta}(x-u))\\ &= (1 - \hat{F}(u))G_{\xi,\beta}(x-u) + \hat{F}(u).\end{aligned}$$

3.2.3 Estimating VaR and ES

The quantile function of the *GPD* is given by

$$G^{-1}_{\xi,\beta}(1-\alpha) = \begin{cases}\frac{\beta}{\xi}(\alpha^{-\xi} - 1), & \xi \neq 0,\\ -\beta\log(\alpha), & \xi = 0.\end{cases} \tag{3.9}$$

For probability $1-\alpha > F(u)$, we get the estimate of a quantile function (*VaR* as (1-α)-quantile of the distribution of losses) from (3.8),

$$\begin{aligned}
1-\alpha &= 1-\frac{N_u}{n}\left(1+\hat{\xi}\frac{VaR_\alpha-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} \\
\frac{n}{N_u}\alpha &= \left(1+\hat{\xi}\frac{VaR_\alpha-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} \\
\left(\frac{n}{N_u}\alpha\right)^{-\hat{\xi}}-1 &= \hat{\xi}\frac{VaR_\alpha-u}{\hat{\beta}} \\
u+\frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}\alpha\right)^{-\hat{\xi}}-1\right) &= VaR_\alpha. \qquad (3.10)
\end{aligned}$$

Expected Shortfall (expected loss if *VaR* is exceeded), can be written in terms of *VaR*,

$$ES^\alpha = \boldsymbol{E}[X|X > VaR_\alpha] = VaR_\alpha + \boldsymbol{E}[X - VaR_\alpha|X > VaR_\alpha], \qquad (3.11)$$

that is, ES^α is the sum of the threshold VaR_α and expected value of the excess distribution $F_{VaR_\alpha}(y)$ over the threshold VaR_α. This expectation is also called *mean-excess function* of VaR_α. It holds that for higher threshold than u, such as VaR_α,

$$F_{VaR_\alpha}(y) = G_{\xi,\beta+\xi(VaR_\alpha-u)}(y). \qquad (3.12)$$

Thus, the *mean-excess function* can be modelled as the expected value of a random variable following GPD.

Let the threshold excess *X*-*u* follow the GPD $G_{\xi,\beta}$. The mean excess for the GPD $G_{\xi,\beta(u)}$ (for $\xi < 1$) for the threshold u is then[1]

$$\boldsymbol{E}(X-u|X>u) = \int_0^\infty y\, g_{\xi,\beta}(y)\, dy = \frac{\beta}{1-\xi}, \qquad (3.13)$$

where $g_{\xi,\beta}(y)$ is the probability density function of $G_{\xi,\beta}(y)$, and $y = x-u$. For any higher threshold, e.g. $VaR_\alpha > u$ we define the mean-excess function $e(VaR_\alpha)$ as

$$e(VaR_\alpha) = \boldsymbol{E}(X - VaR_\alpha|X > VaR_\alpha) = \frac{\beta+\xi(VaR_\alpha-u)}{1-\xi}, \qquad (3.14)$$

or alternatively, for any $z > 0$, we have

$$e(u+z) = \boldsymbol{E}(X-(u+z)|X>u+z) = \frac{\beta+\xi z}{1-\xi}. \qquad (3.15)$$

[1]As noted earlier, k-th moment exists for $\xi < 1/k$, in this case, $\xi < 1$

Now we can write the expression for Expected Shortfall

$$ES^{\alpha} = VaR_{\alpha} + \frac{\beta + \xi(VaR_{\alpha} - u)}{1-\xi} = \frac{VaR_{\alpha}}{1-\xi} + \frac{\beta - \xi u}{1-\xi}. \tag{3.16}$$

To get the notion of average excess over VaR in terms of VaR, it is sometimes convenient to work with the ratio $\frac{ES^{\alpha}}{VaR_{\alpha}}$,

$$\frac{ES^{\alpha}}{VaR_{\alpha}} = \frac{1}{1-\xi} + \frac{\beta - \xi u}{(1-\xi)VaR_{\alpha}} \tag{3.17}$$

This ratio is largely determined by the weight of the tail, that is, by shape parameter ξ (greater $\xi > 0$, heavier tail).

3.2.4 Mean-excess function plot

We can choose the right threshold u by constructing *mean-excess plot*

$$\{(u, e_n(u)),\ X_{n:n} < u < X_{1:n}\}, \tag{3.18}$$

where $X_{i:n}$ is the i-th smallest loss from the sample and $e_n(u)$ is the sample mean excess function, an empirical estimate of the *mean-excess function*

$$e_n(u) = \frac{\sum_{i=1}^{n} (X_i - u)\, 1_{\{X_i > u\}}}{\sum_{i=1}^{n} 1_{\{X_i > u\}}}.$$

For the GDP, the mean-excess function is linear, therefore, if the plot is linear with positive slope above u, then excesses over u follow GPD with positive shape parameter. We can choose the threshold as the value on the x-axis which is located where the plot begins to be linear.

3.2.5 QQ-plot

Using quantile (QQ) plot allows us to test if the sample follows a certain distribution. To compare the sample excess distribution and e.g. a *GPD*, we plot sample quantiles exceeding u on the x-axis against quantiles (inverse of the cdf) of *GPD* on the y-axis. If the data fit to the *GPD*, then the quantiles match, and we get a roughly linear QQ-plot.

3.2.6 Maximum Likelihood Estimation

We use *MLE* to obtain the estimates of parameters ξ, β. We choose the threshold u from the *mean-excess plot*, select the observations above u, and fit the *GPD* to excess returns. Recall that *maximum likelihood estimate* selects the estimates $\hat{\xi}$ and $\hat{\beta}$ which maximize the likelihood function

$$L(\hat{\xi}, \hat{\beta}\,|y) = \max_{\xi,\beta} L(\xi, \beta\,|y) = \max_{\xi,\beta} \prod_{i=1}^{n} g_{\xi,\beta}(y_i),$$

where $g_{\xi,\beta}(y_i)$ is the pdf of *GPD* from (3.13) and $y = \{y_1, \ldots, y_n\}$ is the sample of observations. Equivalently, we maximize the log-likelihood function

$$l(\hat{\xi}, \hat{\beta}\,|y) = \max_{\xi,\beta}\; \log L(\xi, \beta\,|y) = \max_{\xi,\beta} \sum_{i=1}^{n} \log g_{\xi,\beta}(y_i).$$

The *log-likelihood* function $l(\xi, \beta\,|y)$ is the natural logarithm of the joint density $g_{\xi,\beta}(y)$ of the n observations. Using the properties of natural logarithm, $l(\xi, \beta\,|y)$ simplifies to

$$l(\xi, \beta\,|y) = \begin{cases} -n \log \beta - (\frac{1}{\xi} + 1) \sum_{i=1}^{n} \log(1 + \frac{\xi}{\beta} y_i), & \xi \neq 0, \\ -n \log \beta - \frac{1}{\beta} \sum_{i=1}^{n} y_i, & \xi = 0. \end{cases} \tag{3.19}$$

3.3 Application - PX Index

We now apply the presented theory to calculate *VaR* and *ES* from Czech equity market returns, represented by PX Index. PX Index is the official Prague Stock Exchange price index of blue chip stock issues. We analyze the daily returns from the starting day of the index (4/5/1994) to (3/20/2009)[2]. This leaves us with $n = 3685$ observations showed in Figure 3.2. The relative histogram of returns is displayed in Figures 3.3, and 3.4 (relative histogram is normalized, so that integral under the histogram is equal to 1). We use *Mathematica* program for our calculations and the code is included in the appendix.

We use former notation and work with losses as negative returns ($L = -\Delta V$). We observe deviation from normality with negative *skewness* $= -0.52$ (it is likely that extreme loss is larger than extreme return, but there are more days with positive returns than days with losses) and positive excess *kurtosis* $= 13$ (sharp peak, fat tails).

The tail of the *sample distribution function* of the losses defined for given ordered n observations $x_n^{(1)} \leq \ldots \leq x_n^{(n)}$ as

$$\hat{F}_n(x_n^{(i)}) = \frac{i}{n} \quad i = 1, \ldots, n$$

[2]The historical data can be obtained at http://ftp.pse.cz/Info.bas/Cz/px.csv.

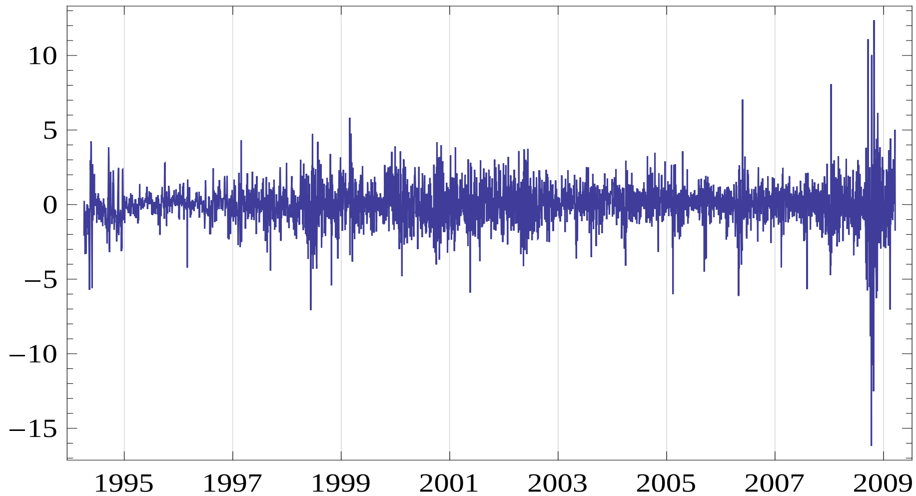


Figure 3.2: Log-returns on PX Index.

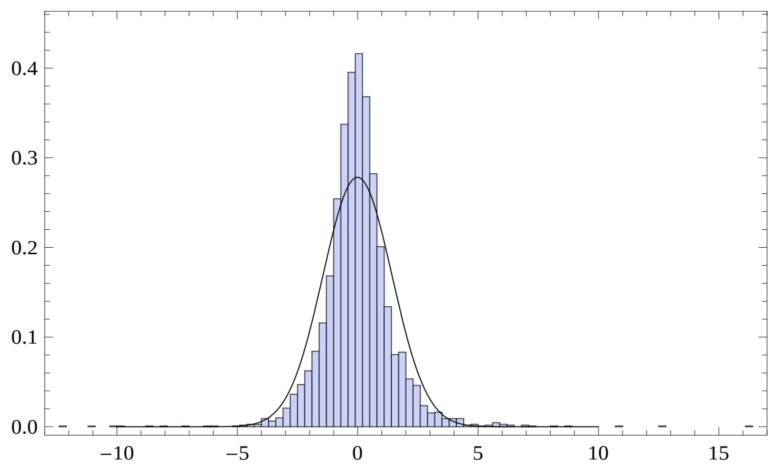


Figure 3.3: Histogram of negative returns compared to normal density.

is presented in Figure 3.5.

We fit this tail with the *GDP* with suitable parameters. First, we need to determine the appropriate threshold u. We construct the *mean-excess function* plot from (3.18) and we choose the value for u where we believe the plot starts to be linear.

We observe two values and we choose the latter, that is $u_1 = 2.57$. This leaves us with $N_{u_1} = 122$ excesses. For comparison, we also choose a different value for u, namely, a 95%-quantile of the losses, that is $u_2 = 2.2$ and corresponding $N_{u_2} = 185$.

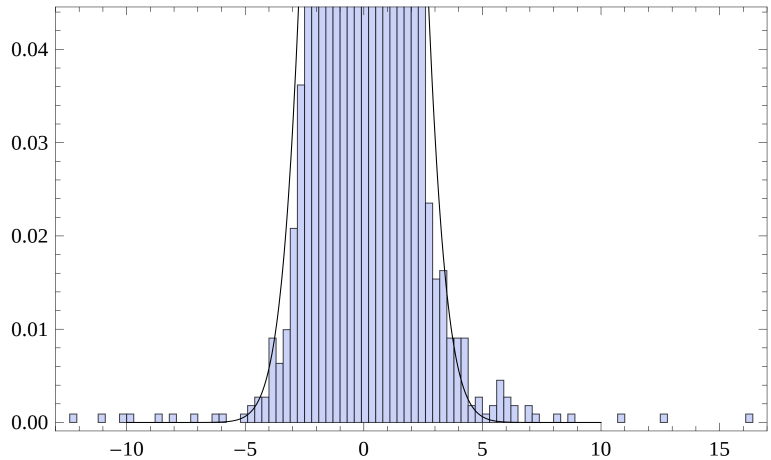


Figure 3.4: Zoom on the tails of the returns (left tail) and losses (right tail).

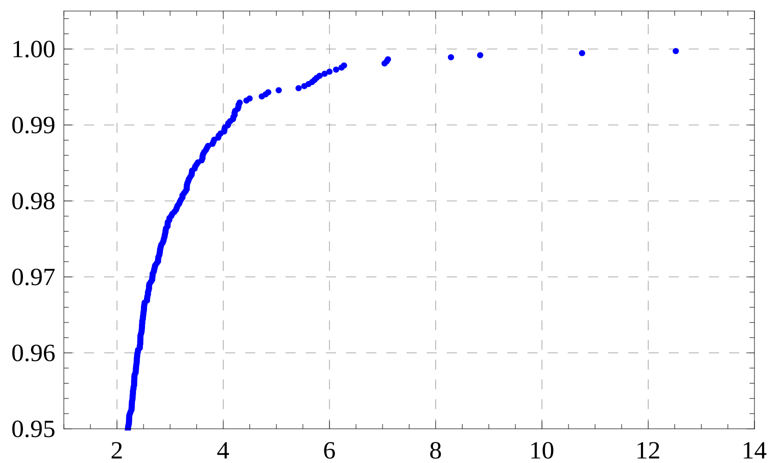


Figure 3.5: Tail of the sample distribution of losses.

Maximum Likelihood We know that losses above u follow *GPD* with parameters $\xi,\ \beta$. We estimate these parameters from (3.19). We use numerical computation of maximum (maximizing the function without using derivatives). The procedure `FindMaximum` in *Mathematica* evaluates the function at many points to find the maximum, but it returns only a local maximum, therefore, for our simulation, starting values are important. We obtain reasonable starting values from a contour plot, see Figure 3.8. For $u_1 = 2.57$ we get the estimates $\hat{\xi} = 0.25$ and $\hat{\beta} = 1.1$. For $u_2 = 2.2$ we have $\hat{\xi} = 0.31$ and $\hat{\beta} = 0.88$.

QQ-plot We check if the quantile plot is linear, see Figure (4.4). For our analysis, both figures present satisfactory fit to the *GPD*.

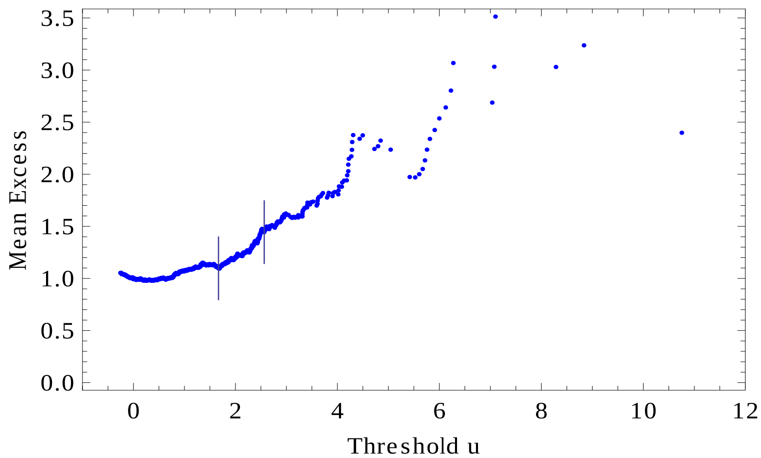


Figure 3.6: Mean Excess Function.

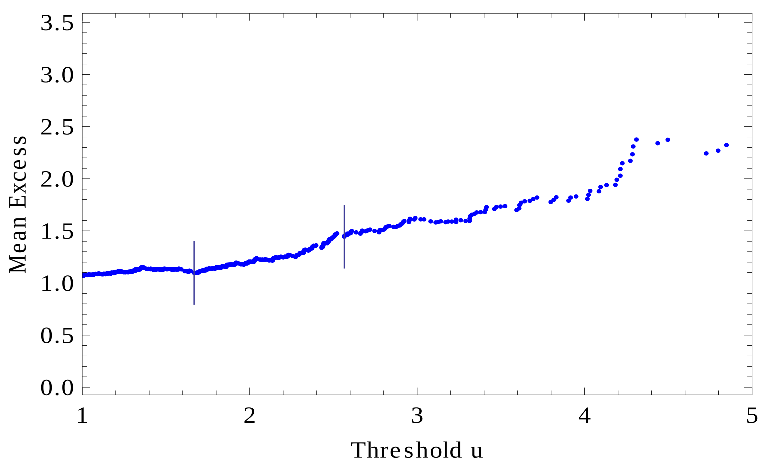


Figure 3.7: Zoom on the linear part.

We fit the empirical tail with F from (3.8), see Figure 3.10.

VaR and ES We calculate 99%-*Value-at-Risk* from the equation (3.10) and corresponding *Expected Shortfall* from the equation (3.16). The results are presented in table 3.1.

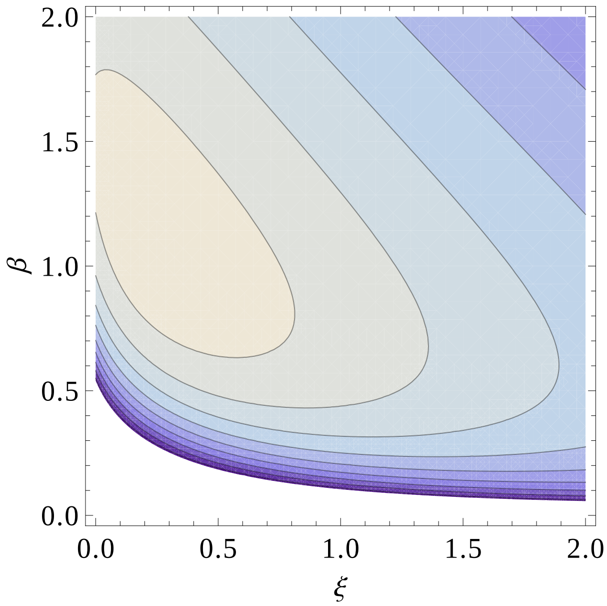
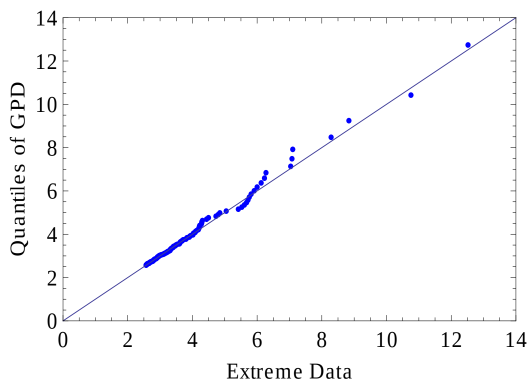
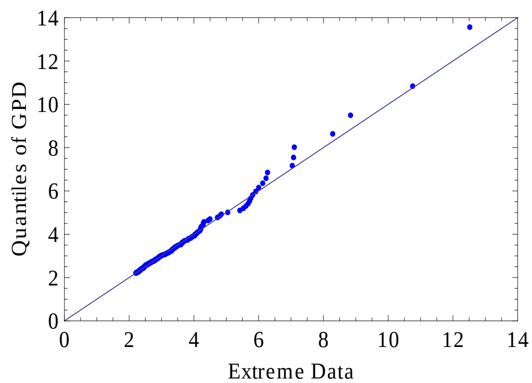


Figure 3.8: Contour plot ('topographical map') to select initial values for parameter estimates ξ and β, $u = 2.57$.



(a) $u = 2.57,\ \xi = 0.25,\ \beta = 1.1$



(b) $u = 2.2,\ \xi = 0.31,\ \beta = 0.88$

Figure 3.9: Quantile plots for estimates (a), (b).

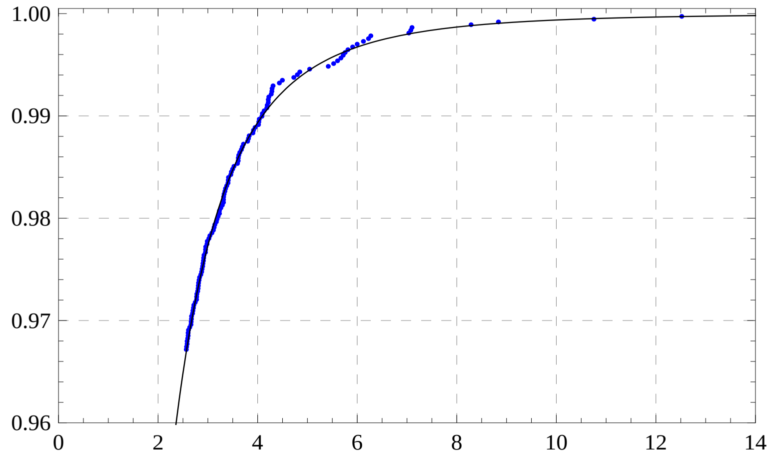


Figure 3.10: ML *GPD* fit to the empirical tail for threshold $u = 2.57$.

	$u_1 = 2.57$	$u_2 = 2.2$
$\hat{\xi}$	0.25	0.31
$\hat{\beta}$	1.1	0.88
$VaR_{0.01}$	4.09	4.04
$ES^{0.01}$	6.06	6.12
$ES^{0.01}/VaR_{0.01}$	1.48	1.52

Table 3.1: VaR and ES for $\alpha = 0.01$ (as a percentage change in the value of PX Index).

3.4 Conditional Extreme Value Theory

EVT offered us a new first insight about the risk in the tails, but we paid little attention to the volatility of the returns. In the previous example, we eased the *iid* assumption and we worked with a large sample of raw returns, which we considered some sort of residuals times unconditional (constant) variance. Empirically, however, the residuals are not *iid* (see, for example, Figure 3.11) and often exhibit heteroskedasticidy and autocorrelation (of their absolute or squared values) (Engle [11]). The previous example is static, and fails to give proper results in case of days with high volatility. There is an obvious need to capture current conditional volatility into our risk measures. This section fills in the gap by introducing dynamic (time-varying) volatility into our computations. Very popular approach is to work with stochastic volatility which takes into account volatility clustering, which means that returns cluster together (large returns are often followed by large returns or losses). While returns are uncorrelated, absolute returns (or their squares) show positive autocorrelation function. In this section, we closely follow McNeil & Frey [18].

Again, we work with losses as negative changes in the log prices

$$X_t = -(\log P_t - \log P_{t-1}) = \log \frac{P_{t-1}}{P_t},$$

where P_t is the closing value of an asset (stock index, exchange rate, etc.) or a portfolio on day t and we use last n days of data, $t = 1, \ldots, n$. A model for loss X_t that includes stochastic volatility (and eventually stochastic mean) can be written as

$$X_t = \mu_t + \sigma_t Z_t, \tag{3.20}$$

where volatility of the return σ_t and expected return μ_t are calculated from the past returns. Z_t are *iid* random variables (strict white noise) with distribution $F_Z(z)$ (with zero mean and unit variance) which bring the noise into model. This allows us to measure volatility of X_t through volatility σ_t, that is, the unit variance of Z_t ensures that σ_t^2 is the variance of X_t, conditional on past returns up to $t-1$.

We are interested in the conditional return distribution

$$F_{X_{t+1}|\mathcal{F}_t}(x),$$

with $\mathcal{F}_t$ indicating history of the process X_t up to day t (we know the past returns). This is the distribution of forecasted return over the next day and we want to come up with an estimate for the quantiles in the tails of this distribution. This is in contrast with previous section, where we worked with unconditional (time-independent) distribution $F_X(x)$. $F_X(x)$ can be seen as the marginal distribution of X_t (See McNeil & Frey pg.4) [18]. We have

$$F_{X_{t+1}|\mathcal{F}_t}(x) \;=\; P(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \le x|\mathcal{F}_t) = F_Z\left(\frac{x-\mu_{t+1}}{\sigma_{t+1}}\right).$$

Relating cdfs of a loss X_t and a noise Z_t, we can estimate quantiles of $F_{X_{t+1}|\mathcal{F}_t}(x)$ from the quantiles of the distribution of Z_t, $F_Z(z)$, which does not depend on time t. All that is left is to forecast the next day conditional volatility σ_{t+1}, mean μ_{t+1}, calculate the residuals, and apply extreme value theory to the tail of $F_Z(z)$. We work with AR(1)-GARCH(1,1) model for σ_{t+1} and μ_{t+1} predictions which is in common use in practice. We briefly introduce it.

3.4.1 AR(1)-GARCH(1,1) Process

GARCH(1,1) model is widely used stochastic model to account for volatility clustering in which the variance (expected return) depends on the variance (expected return) of the previous day

$$\begin{aligned} \mu_t &= cX_{t-1} \\ \sigma_t^2 &= a_0 + a\sigma_{t-1}^2 Z_{t-1}^2 + b\sigma_{t-1}^2, \end{aligned} \tag{3.21}$$

where $0 < a + b < 1$ is the rate of decay of the autocorrelation of σ_t (usually close to 1), $a_0 > 0$, and $|c| < 1$. Constants $a,\ b$ need to be nonnegative, and $a_0 > 0$ so that the variance is nonnegative, and $a + b < 1$ ensures the variance is finite, and after shock it eventually returns to its long-run (unconditional) average variance $a_0/(1-a-b)$ (it exhibits *mean reversion*). The notation (1,1) means that there is one autoregressive lag in the equation, and one lag in the moving average. Variance (squared volatility) of the return for this period (on day t) is forecasted as a weighted average of a constant, previous period's predicted variance, and previous period's squared error (which captures the new information). In our case, GARCH(1,1)[3] process for the conditional variance σ_t^2 of the mean-adjusted return $\epsilon_t = X_t - \mu_t = \sigma_t Z_t$ is extended with AR(1) process for the conditional mean μ_t.

3.4.2 Estimating AR(1)-GARCH(1,1) model

ARCH models in general are interesting in the way that they let the observations determine the best estimates of the parameters in the model. We use pseudo-maximum-likelihood estimation to fit the model. The parameter estimates

[3] To relate GARCH(1,1) model to EWMA model mentioned in previous chapters, we set $a_0 = 0$, $a = 1 - \lambda$, and $b = \lambda$, and we obtain $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)\sigma_{t-1}^2 Z_{t-1}^2$.

$\hat{\theta} = (\hat{c},\ \hat{a_0},\ \hat{a},\ \hat{b})'$ are obtained by maximizing normal log-likelihood function for GARCH(1,1). By normal, we mean that noise variables Z_t follow Normal distribution conditional on past history. The normal log-likelihood function of the AR(1)-GARCH(1,1) model is then given by

$$\begin{aligned} L(\theta) &= \log \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left\{-\frac{\epsilon_t^2}{2\sigma_t^2(\theta)}\right\} \\ &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n}\left(\log \sigma_t^2(\theta) + \frac{\epsilon_t^2}{\sigma_t^2(\theta)}\right). \end{aligned} \tag{3.22}$$

For computation, we can omit the first term which is a constant. Although in our case, we do not assume normality in Z_t, we can use (3.22) to obtain vector of parameter estimates $\hat{\theta}$. $L(\theta)$ is then called pseudo-log-likelihood function, since the distribution of Z_t does not need to be normal. We define pseudo-maximum-likelihood estimator (PMLE) of parameter θ as estimator $\hat{\theta}$ which maximizes the pseudo-likelihood function

$$\hat{\theta} = \arg\max_{\theta} L(\theta). \tag{3.23}$$

It can be shown that PMLE is consistent and asymptotically normally distributed. Starting values for θ need to be carefully chosen (only local maximum is calculated), for example, we can use sample mean return as a starting value for c, we can set $a_0 = 1 - a - b$, and a is usually relatively close to zero, while b is close to 1. We also set unconditional sample variance as an initial value of σ_t^2 and sample mean as initial value for μ_t.

3.4.3 Applying Conditional EVT on PX Index

We follow up with the previous example and we estimate VaR and ES of PX Index using conditional EVT. We work with a window of last $n = 1000$ negative observations, which is rouhly the most recent 4 years of negative log-returns[4]. The parameter estimates of AR(1)-GARCH(1,1) model using PMLE and the maximized value of log-likelihood function L (omitting constant) are displayed in Table 3.2.

Last 1000 daily losses and conditional volatility prediction are displayed in Figures 3.11 and 3.12.

Using the estimated parameters, we calculate vector estimates of conditional

[4] McNeil & Saladin [16] while simulating heavy-tailed data from different distributions claim that $N_u = 100$ exceedances of a threshold is a reasonable and realistic number for estimating high quantiles. In particular, one of their simulation result is that using 90%-quantile as a threshold, 100 excesses are sufficient to estimate 99%-quantile in case of Pareto distribution.

$L(a_0, a, b, c)$	a_0	a	b	c
-780	1.25×10^{-5}	0.104	0.895	0.028

Table 3.2: AR(1)-GARCH(1,1) parameter estimates for PX Index.

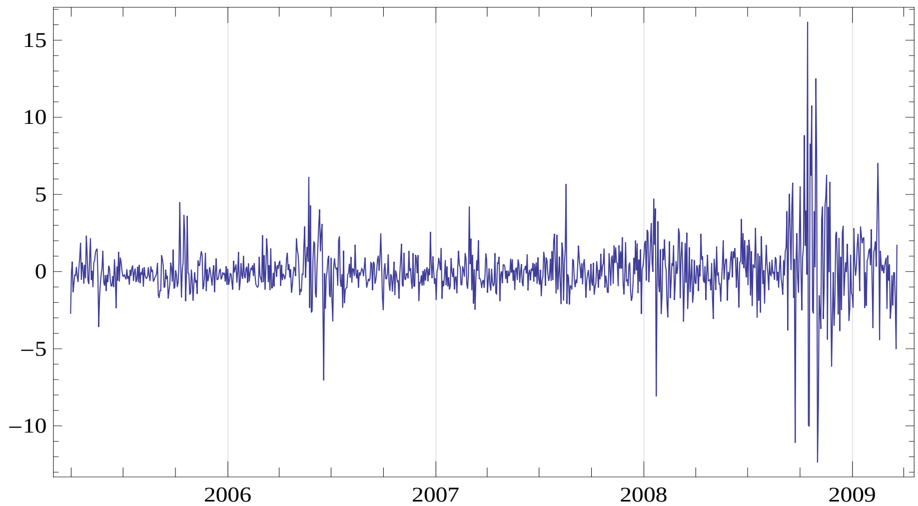


Figure 3.11: Last 1000 days of losses on PX Index from 3/31/2005 to 3/20/2009, including the stock market crash of 2008.

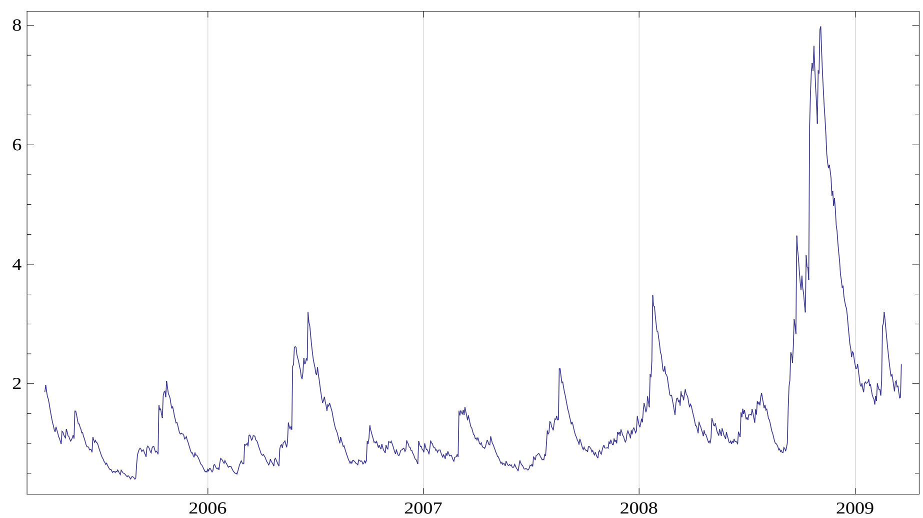


Figure 3.12: Corresponding conditional volatility prediction from AR(1)-GARCH(1,1) model.

mean $(\hat{\mu}_{t-n+1}, \ldots, \hat{\mu}_t)$, standard deviation $(\hat{\sigma}_{t-n+1}, \ldots, \hat{\sigma}_t)$, and residuals

$$(z_{t-n+1}, \ldots, z_t) = \left(\frac{x_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \ldots, \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right).$$

We consider the residuals as independent noise variables. Next, we calculate one day forecasts of the conditional mean and variance

$$\begin{aligned}\hat{\mu}_{t+1} &= \hat{c}x_t,\\ \hat{\sigma}^2_{t+1} &= \hat{a_0} + \hat{a}(x_t - \hat{\mu}_t)^2 + \hat{b}\hat{\sigma}^2_t.\end{aligned} \tag{3.24}$$

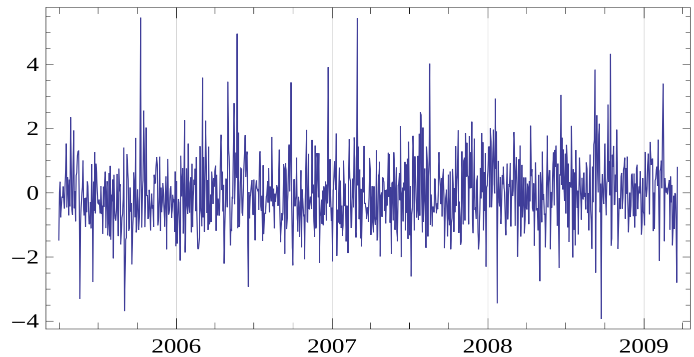


Figure 3.13: Graph of extracted standardized residuals from the sample.

We apply extreme value theory from this chapter and fit the GPD to the tails of the distribution of residuals z_t and calculate *VaR* and *ES* estimates as

$$\begin{aligned}V\hat{a}R^t_\alpha(\Delta X) &= \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} VaR_\alpha(Z)\\ E\hat{S}^\alpha_t(\Delta X) &= \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} ES^\alpha(Z),\end{aligned} \tag{3.25}$$

where $VaR_\alpha(Z)$ denotes $(1-\alpha)$-quantile of the distribution of residuals Z_t and $ES^\alpha(Z)$ is the related expected shortfall.

We set the threshold u as upper 90% quantile, which leaves us with $N_u = k = 100$ tail data. This means that when we order the residuals $z_{(1)} \geq z_{(2)} \geq \ldots \geq z_{(n)}$, the threshold $u = z_{(k+1)}$ is the $(k+1)$th order statistic. We then fit the generalized Pareto distribution to excesses above u, $(z_{(1)} - z_{(k+1)}, \ldots, z_{(k)} - z_{(k+1)})$ using MLE from (3.19).

z_{k+1}	$\hat{\xi}$	$\hat{\beta}$
1.28	0.21	0.59

Table 3.3: GPD parameter estimates for residuals.

After estimating parameters of GPD, we use (3.8) to estimate the tail of $F_Z(z)$, that is

$$\hat{F_Z}(z) = 1 - \frac{k}{n}\left(1 + \hat{\xi}\frac{z - z_{(k+1)}}{\hat{\beta}}\right)^{-1/\hat{\xi}}. \tag{3.26}$$

Inverting this formula we get VaR estimate as in (3.10),

$$V\hat{a}R_\alpha(Z) = z_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{k}\alpha\right)^{-\hat{\xi}} - 1\right). \tag{3.27}$$

Similarly, we use $V\hat{a}R_\alpha(Z)$, rewrite the formula (3.16) for expected shortfall, and from equation (3.25) we get the estimate of conditional expected shortfall as

$$E\hat{S}^\alpha(Z) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\left(\frac{V\hat{a}R_\alpha(Z)}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} z_{k+1}}{1-\hat{\xi}}\right). \tag{3.28}$$

Figure 3.14 compares GPD fit to the empirical tail of residuals with tail of the standard normal distribution. We see that the assumption of normality fails for the tails. We confirm that by constructing normal QQ-plot (Figure 3.15).

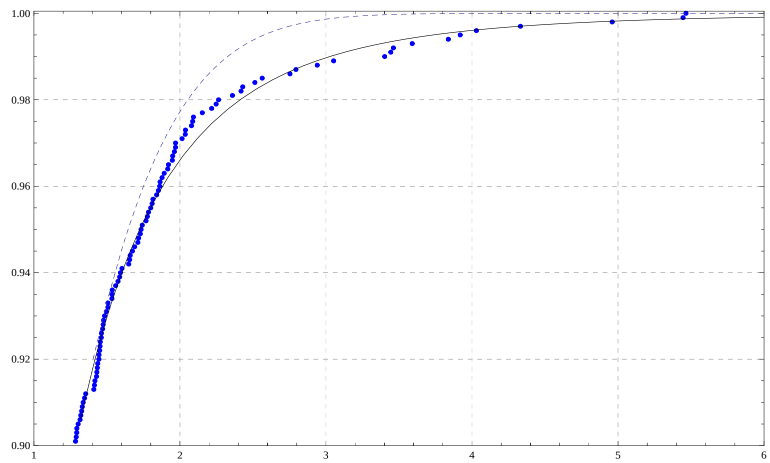


Figure 3.14: Empirical tail (dots), GPD fit to the tail (solid line), and the tail of standard normal (dashed line).

Using (3.24), (3.25), (3.27) and (3.28) we get the following results (Table 3.4 and 3.5).

$\hat{\mu}_{t+1}$	$\hat{\sigma}_{t+1}$	$VaR_\alpha(Z)$	$ES^\alpha(Z)$
0.047	2.275	3.026	4.023

Table 3.4: One-day conditional mean and volatility predictions, GPD estimate of 99%-quantile of the distribution of residuals and corresponding expected shortfall estimate.

Considering the ratio of expected shortfall to Value-at-Risk, from (3.25) for μ_{t+1} small we can write (see [18])

$$\frac{ES_t^\alpha}{VaR_\alpha^t} \approx \frac{ES_t^\alpha - \mu_{t+1}}{VaR_\alpha^t - \mu_{t+1}} = \frac{ES^\alpha(Z)}{VaR_\alpha(Z)} = 1.33. \tag{3.29}$$

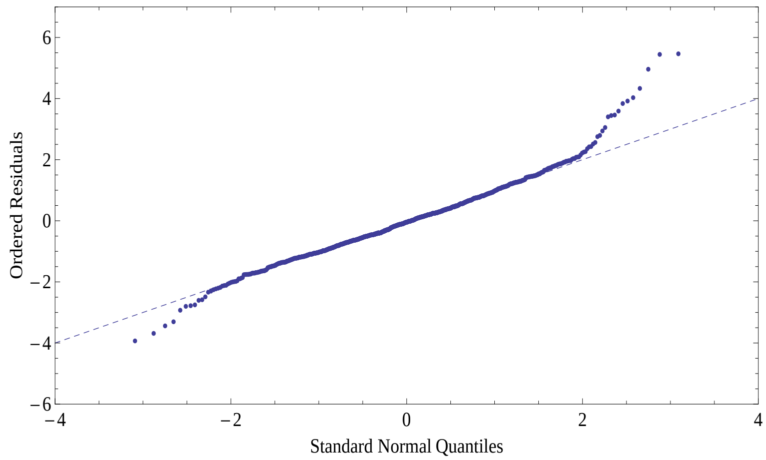


Figure 3.15: QQ-plot of ordered residuals vs. standard normal quantiles.

$VaR^t_\alpha(\Delta X)$	6.93
$ES^\alpha_t(\Delta X)$	9.65

Table 3.5: Conditional 99%-Value-at-Risk estimate under extreme value theory (as a percentage change in the value of PX Index).

3.4.4 Multi Day Prediction

It is possible to extend one day *EVT VaR* and *ES* estimates to T-day estimates as regulators usually require[5], however, we cannot use the 'square root of time' rule for non-normally distributed returns. Danielsson & de Vries [6] use theoretical results to arrive at an approximation for T-day quantile. They note that in case of *iid* r.v. X_i with a heavy-tailed distribution function F_X[6] the tail probabilities are linearly additive

$$P(X_1 + \ldots + X_T > x) \approx x^{-\lambda} T L(x) \tag{3.30}$$

for large x. They used a scaling factor $T^{1/\lambda}$ for heavy-tailed distributions for multi period quantiles. To calculate λ, they propose customized Monte Carlo simulation for future return paths. This algorithm is also used in McNeil & Frey [18], where it is applied to residuals to account for stochastic volatility, thus, obtaining different results. The algorithm takes a large sample of n residuals, randomly picks one from the sample, and if it exceeds a threshold (both tails), it samples a *GPD* distributed random variable. If it does not, the value of the residual remains unchanged. The residual is then replaced in the sample and the procedure is

[5]Basel II Framework requires 99% 1-day *VaR* scaled to 10-days (it is assumed to take 10 days to liquidate banks' portfolios)

[6]A distribution is heavy-tailed when there exists finite constant $a > 0$ such that $F(x) \approx 1 - x^{-\lambda} L(x)$, where $L(x)$ satisfies $\frac{L(tx)}{L(x)} \to 1$,for $x \to \infty$, $t > 0$.

repeated. This simulated distribution approaches the distribution of residuals for large n.

Recall that we want to estimate the next T-days return (continuous compounding) conditional distribution $F_{X_{t+1}+\ldots+X_{t+T}|\mathcal{F}_t}(x)$. The conditional quantile of this distribution is given by

$$q^t_{\alpha,T} = \inf\left\{q \in \mathbb{R} : F_{X_{t+1}+\ldots+X_{t+T}|\mathcal{F}_t}(q) \geq 1-\alpha\right\},$$

and the conditional expected shortfall by

$$ES^t_{\alpha,T} = E\left(\sum_{j=1}^{T} X_{t+j} | \sum_{j=1}^{T} X_{t+j} > q^t_{\alpha,T}, \mathcal{F}_t\right).$$

From the algorithm, high number of future return paths $(x_{t+1},\ldots,x_{t+T})$ are generated and summed to obtain realisations of $\sum_{j=1}^{T} X_{t+j}|\mathcal{F}_t$ and estimates $q^t_{\alpha,T}$ and $ES^t_{\alpha,T}$ are then calculated.
Denote q_α and $q_{\alpha,T}$ quantile of return distribution over one-day and T-days respectively. Using (3.30) for *iid* r.v. we can write

$$\alpha \approx P(X > q_{\alpha,T}) \approx (q_{\alpha,T})^{-\lambda}\, T\, L(q_{\alpha,T}),$$
$$\alpha \approx P(X > q_\alpha) \approx (q_\alpha)^{-\lambda}\, L(q_\alpha),$$

and we obtain approximate scaling law

$$q_{\alpha,T} \approx q_\alpha\, T^{1/\lambda}. \tag{3.31}$$

If we choose cdf F_X whose limiting distribution of excesses is *GPD* with shape parameter ξ as a particular heavy-tailed distribution of returns, then from (3.30) an appropriate scaling formula is

$$q_{\alpha,T} \approx q_\alpha T^\xi, \tag{3.32}$$

where $q_{\alpha,T}$ is the desired T-day quantile.

McNeil & Frey [18] adapt the scaling exponent $1/\lambda$ from (3.31) to depend on current volatility σ_t, thus obtaining

$$\frac{q^t_{\alpha,T}}{q^t_\alpha} = \frac{VaR_{\alpha,T}(X)}{VaR_\alpha(X)} \approx T^{\frac{1}{\lambda_t}}.$$

They test this empirically on S&P Index for different values of σ_t and T and find that for higher initial volatility σ_t, the scaling exponent is lower than for the average or low σ_t, that is, if the initial volatility is higher, one expects lower average volatility (a median of past volatilities) in the future, thus T-day *VaR* increases less than in case of lower initial volatility.

3.4.5 Backtesting

Backtesting procedure evaluates the risk measurement models by comparing risk estimates with realized returns using historical data. Daily risk measure estimate (*VaR* or *ES*) is tested against daily actual (realized) portfolio return (loss). We use statistical tests to verify that our model accurately captures the frequency of violations of risk estimates (we compare observed frequency of violations with expected frequency of violations according to the model).

Indicator of violations

When $VaR_{\alpha,t+1}$ estimates and actual losses X_{t+1} are compared, *VaR* violation can be defined as an indicator

$$I_{t+1} = \begin{cases} 1, & \text{when } X_{t+1} > VaR_{\alpha,t+1}, \\ 0, & \text{when } X_{t+1} < VaR_{\alpha,t+1}, \end{cases}$$

and we obtain the sequence of violations $\{I_{t+1}\}_{t=1}^{T}$, where T is the number of days of a backtest. We expect that indicator $I_{t+1} = 1$ with probability α, therefore, we are testing the null hypothesis

$$H_0 : \quad I_{t+1} \sim \text{Bernoulli}(\alpha)\ iid.$$

The iid assumption allows us to test that the expected value of indicator sequence $\frac{1}{T}\sum_{t=1}^{T} I_t = \alpha$, or that the sum of violations follows the binomial distribution with parameters T and α

$$H_0 : \quad \sum_{t=1}^{T} I_t \sim B(T, \alpha). \tag{3.33}$$

McNeil & Frey [18] carry out such backtesting of several *VaR* methods including conditional EVT on different historical return series (stock, stock index, exchange rate, gold price). They use rolling window of 1000 observations and set the threshold u as 90th percentile, $u = 100$. Each day, they compare realized loss X_{t+1} to *VaR* estimates $q_{\alpha,t+1}$ from GPD fit at different confidence levels α. They set significance level for binomial test at 5%, thus, if p-values are smaller than 0.05, the null hypothesis (3.33) is rejected.

Their results are that conditional EVT method is the best and does not lead to rejection of H_0. In the sense of binomial testing, very good results were also obtained with a GARCH model with conditionally Student-t distributed returns. They conclude that unconditional EVT estimate can be violated several times in a row during high volatility periods and the conditional normal estimate (especially at higher quantiles) is violated more often because it does not take into account leptokurtosis.

They also develop a binomial test for conditional *ES* and verify that EVT method gives better estimates. They standardize exceedance residuals that according to the model are iid with zero expectation and unit variance and test this zero mean null hypothesis. Their results show that assumption of normally distributed residuals fails and is useless for calculating *ES*. On the other hand, for standardized GPD residuals, the hypothesis is rejected only for stock index, and GPD assumption tends to underestimate the prediction for stock indices, but in overall, it gives much better estimates for *ES*.

Chapter 4

Application on a portfolio

In this chapter, we construct a theoretical portfolio and calculate *VaR* and *ES* using delta, delta-gamma, historical simulation and extreme value method. Consider two equal investments, say CZK 1 million each, into Dow Jones Euro STOXX 50 Index, and PX Index[1], and a purchase of EURCZK currency put option, such that a domestic (Czech based) investor is protected from depreciation of Euro against Czech koruna. We simplify the matter in a common way: we omit the transaction costs, dividend payments, and we work with *mid* prices observed at the close of the day. More, we assume that returns on the risk factors are *iid*.

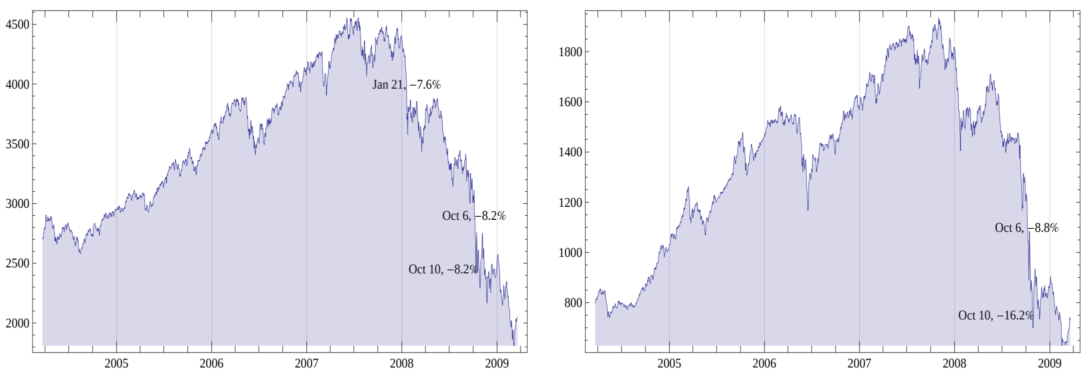
Although it is possible to use multivariate extreme value theory (modeling the tails with multivariate *GPD* and copulas) for such portfolio, in real portfolios with many risk factors, it might be difficult to properly match extreme values, and account for their correlations. Although a simplification, it is reasonable to apply univariate *EVT* to a single risk factor represented by the returns on the whole portfolio. That is, we use historical simulation to calculate hypothetical portfolio returns, and to estimate the desired quantile, we apply extreme value theory to the tail of these portfolio returns. This approach is proposed in Danielsson & De Vries [6]. We also apply demonstrate conditional *EVT* method: we standardize the hypothetical returns by AR(1)-GARCH(1,1) volatility estimates and apply conditional *EVT* to the residuals.

Next, we apply parametric linear and non-linear approach from Chapter 1. We use Cornish-Fisher expansion to arrive at correct quantile of portfolio return distribution. We then compare *VaR* and *ES* results from presented methodologies.

[1] Although indices are not directly tradable, it is convenient to work with them, because they serve as market benchmarks. Of course, there are many tradable products at different exchanges that track a certain index performance, thus, investors seeking prompt diversification usually consider exchange traded funds (e.g. FEZ etf, Lyxor etf, iShares etf, or DB x-trackers for DJ Euro STOXX 50), index certificates (e.g., PX Index Certificate or Czech Traded X-pert Index Certificate), index futures, etc.

In the subsequent section, we discuss the gamma effect of including option in the portfolio.

When using our combined *HS* and *EVT* approach, it is good to understand what caused the *extremes* and conclude how much we are concerned that these extremes will repeat. With such broader picture we get better feeling about the risk our portfolio is exposed to, than by simply looking at large changes in risk factors' returns.



(a) Dow Jones EURO STOXX 50 Index graph. (b) PX Index graph

Figure 4.1: Graphs of indices with several extreme drops highlighted. Data from 3/22/2004 to 3/20/2009.

Concretely, in Figure (4.1) we plot the graphs of the two indices and we highlight several *extreme* daily losses. On Monday January 21, 2008, DJ Euro STOXX 50 plunged 7.6% as investors feared upcoming global economic recession and the US mortgage turmoil. The sudden drop might have been also partially caused by Jérôme Kerviel incident which went public that weekend, and promptly on Monday, Société Générale started to liquidate loss-making positions in leading European indices (including Euro STOXX 50) what might have caused further sell-offs. Also on that day, German WestLB reported 1 billion euro loss for 2007. A day later, the US Federal Reserve cut rates by 75 bps, indicated possible further cuts and the markets calmed for a while.

The week October 6-10, 2008 was even more interesting: stock markets and commodities sharply fell, Iceland's banks collapse, and a number of other banks were bought, nationalized, or filed for bankruptcy, risk (or investor fear) indicators jumped at long-time highs, etc. The governments' attempts to calm the situation included simultaneous rate cuts, planned billions for bailouts, and an increase in deposit guarantees. Indeed, there are many explanations for rapid market movements, and we could continue probing into the rest of the *extremes* for a better understanding of what caused them, that is, for a better understanding of our risk. That is not our aim, and we only wanted to point out that when considering

VaR and *ES* numbers, we should take into account our concerns about the repeat of specific *extremes* in the history.

In our portfolio, there are following risk factors that affect its value: EURCZK exchange rate, 1 year PRIBOR, 1 year LIBOR, DJ Euro STOXX 50 value, PX 50 value[2]. We slightly refine the data so that prices remain constant (zero returns) over holidays. Today is March 20, 2009. The exchange rate is EUR 1 = CZK 26.628 and we are long EUR put CZK call, with expiration in one year, contract size EUR 37 555 (CZK 1 million), and a strike price set at EUR/CZK 26. The 1-day exchange rate volatility is modelled by GARCH(1,1) process, and is extended to 1-year volatility by Drost-Nijman formula. Although EURCZK volatility is also a risk factor, we neglect it since it has a tiny impact on the computation (as shown in Appendix, 1-year volatility calculated with Drost-Nijman formula fluctuates insignificantly). We use last five years of closing day prices (from 3/22/2004 to 3/20/2009) and the sample size is $n = 1287$. We are interested in next day's, say, one chance in a hundred and one chance in a thousand largest loss, so we set α equal 0.01 and 0.001.

Instrument	Value
PX Index	CZK 1 million
Euro STOXX 50	CZK 1 million (EUR 37 554)
Put option	
Notional	EUR 37 554
Current rate	26.628
Strike price	26
Option premium	1.35% (EUR 509.82)

Table 4.1: Portfolio specification.

After we set up the portfolio (see portfolio specification, Table 4.1), we price each instrument to obtain their present values. We use Garman-Kohlhagen formula to price FX option, and we get option premium today equal 0.0135 cents per 1 EUR (see Appendix B). We then calculate historical log-returns of each risk factor and use the series of returns to simulate possible paths of tomorrow's returns (see Historical Simulation section). This way, we constructed the empirical distribution of portfolio returns (see Figure 4.2). We complete the historical simulation by ordering the portfolio return sample and taking negative of α-th order statistic as a representative of historical *VaR*. To estimate historical *ES*, we

[2]The data for the exchange and interest rates were downloaded from Bloomberg, STOXX 50 index is available at http://www.stoxx.com/indices/index_information.html?symbol=SX5E, and PX 50 at http://ftp.pse.cz/Info.bas/Cz/px.csv.

use the formula (2.1) and average $\alpha\%$ largest losses.

We now use the tail of the empirical distribution of returns and apply EVT to estimate *GPD VaR* and *ES*. We treat the simulated portfolio returns as historical returns (Figure 4.2) and proceed as in chapter Extreme Value Theory. We invert the returns (loss=positive number) and set the threshold u at 90% loss quantile and obtain the value for $u = 1.35$ (we might get a better fit if we visually chose the threshold, however, if using automatized EVT as a risk management tool, visually choosing the threshold is impractical). We are left with satisfactory 129 *extremes*.

After maximizing *GPD* log-likelihood function (3.19), we obtain the estimates $\hat{\xi} = 0.27$ and $\hat{\beta} = 0.93$ and we use (3.8) to fit the empirical tail with Generalized Pareto distribution. Finally, we obtain *GPD Value-at-risk* and *Expected Shortfall* estimates by plugging the estimated parameters into (3.10) and (3.16). In Figure 4.5 we plot different quantiles obtained from historical simulation and extreme value theory.

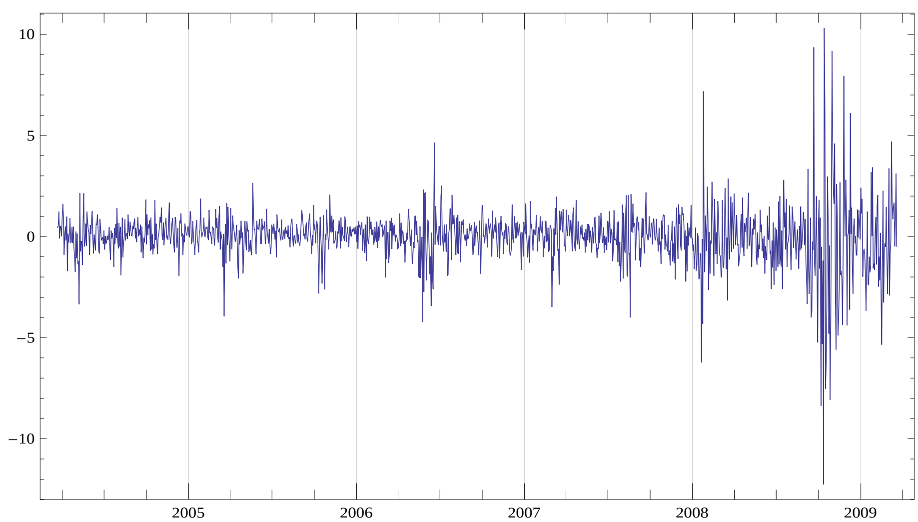


Figure 4.2: Portfolio log-returns.

We can now consider conditional EVT, and primarily, that the volatility of the returns is stochastic. As in example from previous chapter, we assume that $X_t = \mu_t + \sigma_t Z_t$, and we use AR(1)-GARCH(1,1) model to estimate the next day conditional volatility σ_{t+1} and mean μ_{t+1} using (3.24). Then we can calculate residuals (iid noise) Z_t. To calculate conditional EVT VaR and ES, we subsequently apply formulas (3.25), (3.26), (3.27), and (3.28). The results are presented in Table 4.3 and subsequently, the corresponding Figures are displayed.

Next, we apply parametric method explained in Chapter 1. We forecast the variance using EWMA (1.8) and we use prevalent RiskMetrics [15] $\lambda = 0.94$ (we

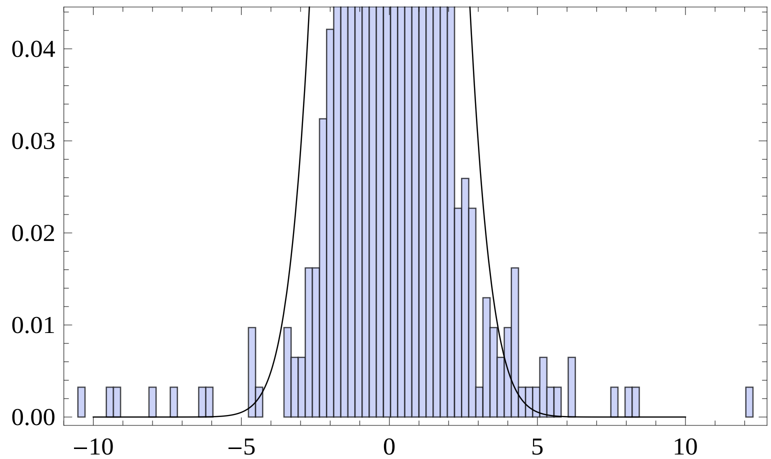
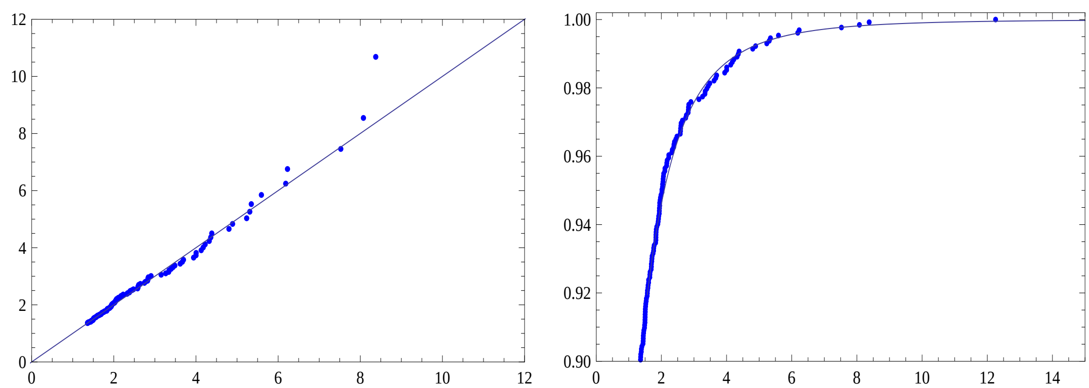


Figure 4.3: Zoom on the tails of the returns (left tail) and losses (right tail) compared to normal pdf.



(a) QQ-plot of sample quantiles against *GPD* quantiles.

(b) ML *GPD* fit to $N_u = 129$ tail losses

Figure 4.4: Quantile plot (a) and *GPD* fit to the tail (b) for the estimates $u = 1.35,\ \xi = 0.27,\ \beta = 0.93$.

also used RMSE criterion to arrive at optimal lambda for our portfolio and we obtained $\lambda = 0.91$ which in our case produces even lower *VaR* estimates, see paragraph in section 1.1.2).

Next, we calculate log-return for indices and exchange rate using formula (1.2). We estimate variance and covariance using formulas (1.8) and we obtain following

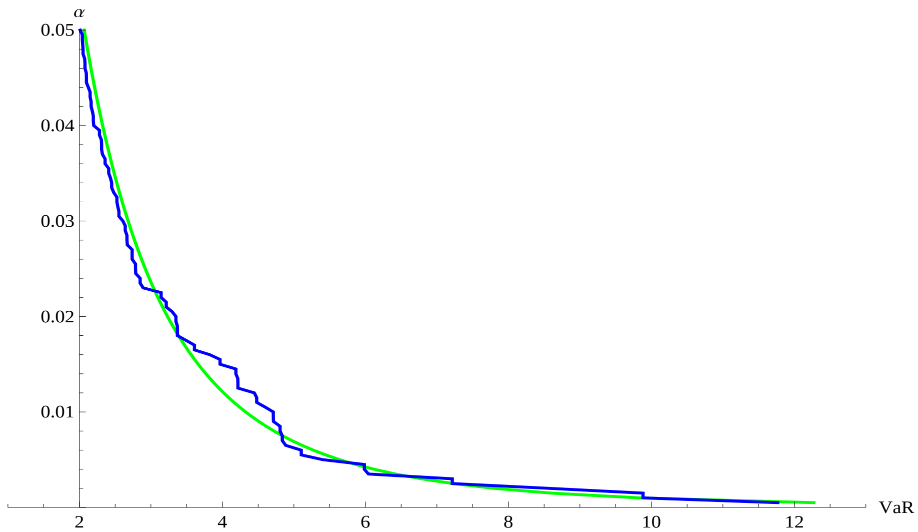


Figure 4.5: VaR estimates for different levels of α using historical simulation and generalized pareto distribution.

variance-covariance matrix

$$\begin{pmatrix} 0.0000967 & -0.0000873 & -0.0000689 & -2.56\times10^{-7} & 4.63\times10^{-8} \\ -0.0000873 & 0.000556 & 0.000247 & -4.50\times10^{-7} & 3.65\times10^{-7} \\ -0.0000689 & 0.000247 & 0.000479 & 8.74\times10^{-8} & -1.50\times10^{-7} \\ -2.56\times10^{-7} & -4.50\times10^{-7} & 8.74\times10^{-8} & 5.12\times10^{-8} & 1.63\times10^{-8} \\ 4.63\times10^{-8} & 3.65\times10^{-7} & -1.50\times10^{-7} & 1.63\times10^{-8} & 2.81\times10^{-8} \end{pmatrix}.$$

We calculate the vector of first derivatives from (1.14) numerically by increasing each risk factor by one bp

$$\boldsymbol{\delta}^{\boldsymbol{T}} = (727.52, 1000, 1000, -8.75, 5.52),$$

Using (1.17) we arrive at linear (delta) Value-at-Risk estimate $VaR^{\delta}_{0.01} = 4.25\%$, and exercising parametric formula for expected shortfall (2.6) we get $ES^{0.01} = 4.87\%$.

Next, we build a matrix of second derivatives using formula (1.15). Again, we numerically measure how the first derivative of each factor changes when we move each risk factor by 1 bp and we get

$$\begin{pmatrix} 4912.76 & 1000 & 0 & 139.39 & -86.06 \\ 1000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 139.39 & 0 & 0 & 4.21 & -2.6 \\ -86.06 & 0 & 0 & -2.6 & 1.61 \end{pmatrix}.$$

The sensitivities δ and Γ take into account the size of the position (in thousands CZK) and current levels of risk factors $S_i(t)$. Now we turn to Cornish-Fisher approximation to estimate portfolio's mean and variance. We can straightforwardly use formulas from section 1.2.3 to calculate higher moments of the distribution of portfolio's returns. The moments of the portfolio's distribution are presented in Table 4.2.

Mean $\mu_{\Delta V}$	0.091
Variance $\sigma_{\Delta V}$	1354.84
Skewness $\frac{\boldsymbol{E}(\Delta V - \mu_{\Delta V})^3}{\sigma^3_{\Delta V}}$	-0.00547
Kurtosis $\frac{\boldsymbol{E}(\Delta V - \mu_{\Delta V})^4}{\sigma^4_{\Delta V}}$	0.00012

Table 4.2: Portfolio's distribution moments.

Based on this moments, we can approximate the desired quantile, and estimate non-linear (delta-gamma) *VaR*. The impact of option on our estimates on portfolio's return is low, and our estimates change insignificantly. We obtain quantile $z_{\Delta V,\alpha} = -2.33$ and $VaR^{\Gamma}_{\alpha} = 4.27$. Regarding Expected Shortfall, we do not have parametric expression for non-linear ES.

Complete results are summarized in table 4.3.

	$\boldsymbol{VaR_{0.01}}$	$\boldsymbol{ES^{0.01}}$	$\boldsymbol{\frac{ES^{0.01}}{VaR_{0.01}}}$	$\boldsymbol{VaR_{0.001}}$	$\boldsymbol{ES^{0.001}}$	$\boldsymbol{\frac{ES^{0.001}}{VaR_{0.001}}}$
HS	4.71	6.60	1.40	9.88	11.77	1.19
EVT	4.33	6.69	1.55	9.83	14.21	1.45
EVT-GARCH	4.70	6.22	1.32	8.25	10.32	1.25
$\boldsymbol{\delta}$	4.25	4.87	1.15	5.65	6.16	1.09
$\boldsymbol{\delta}$**-**$\boldsymbol{\Gamma}$	4.27			5.67		

Table 4.3: VaR and ES estimates (as a percentage change in the value of portfolio) using Historical Simulation, Extreme Value Theory, Conditional EVT, Delta, and Delta-Gamma approaches ($\lambda = 0.94$), sample size=1287.

From the results of this simple hypothetical portfolio, we conclude that parametric methods (based on normality of returns) give lower risk estimates than historical simulation or methods based on *EVT*. Both Delta and Delta-Gamma clearly underestimate the risk for very high quantiles, namely 99.9%. Historical simulation, while capturing fat tails, is restricted to the range of the sample. This can lead to imprecise results as the high quantile estimates can be volatile (adding or dropping large observation may cause swings in the *VaR* number).

Assuming that extremes follow Generalized Pareto distribution, one can estimate any quantile measure without extra computational intensity (using EVT, we smooth the tails obtained from *HS*, and thus are able to estimate *VaR* and *ES* for any confidence level, in particular, the one that is out of the historical sample, see Figure 4.5). High quantile estimates using EVT can also be imprecise especially when using very small set of data, however, it is very useful to have an idea of how the tails behave. The proposed EVT method based on historical simulation can be seen as a suitable supplement to historical simulation in addition to stress testing and scenario analyses.

The demonstrated unconditional EVT *VaR* is more suitable for long run rather than daily forecasts because of the large sample size needed (adding new and removing old observation does not produce significant changes in *VaR* and *ES* estimates when using large sample size). *HS* and EVT thus provide stable estimates but do not update quickly when the market volatility changes (this is undesirable during periods of high or low volatility). This drawback is removed by Conditional EVT which reflects the current volatility. It is tempting to say that this makes the Conditional EVT the most appropriate method, however, extended backtesting procedures must be undertaken first. For now, we can only refer to McNeil & Frey [18] who backtested several (univariate) return series and showed that Conditional EVT is the best method for estimating high quantiles. Regarding number of observations, we can say, the larger the sample size, the better, but the size still remains an important practical issue.

Considering Expected Shortfall estimates, we observe that the ratio *ES*/*VaR* approaches 1 with decreasing α for historical simulation and parametric methods. This is a drawback of these methods because even if we believe that the VaR number they produce is reasonable, they underestimate Expected Shortfall estimates for very high quantiles. On the other hand, EVT methods due to their nature produce reasonable *ES*/*VaR* ratios.

4.1 Portfolio breakdown

In order to fully explore the impact of gamma risk from option's return on *VaR* numbers, let us investigate the option behaviour in the portfolio in these three very simple cases. First, we run the program without the option to verify that this specific hedging of Euro STOXX 50 Index investment with put option in our portfolio did not create large risk (Table 4.4).

In the second case, let us say that (Czech based) investor expects CZK to depreciate against EUR. He keeps half of his wealth in PX and Euro STOXX 50 Index (say one thousand CZK together), and goes long EUR call CZK put with the other half (another one thousand CZK). The option parameters stay the same.

VaR at 99%	**Excluding option**	**Including option**
δ	4.22	4.25
$\delta - \Gamma$	4.22	4.27

Table 4.4: Impact of FX hedging with put option on VaR number.

He does not hedge, but he uses his "play" money to speculate on the movement of exchange rate (he gambles that euro appreciates against CZK). The Table 4.5 captures *VaR* numbers in this case.

(a) Portfolio Statistics

Mean	5.128
Variance	22456.7
Skewness	0.202
Kurtosis	0.055

(b) Value-at-risk estimates

%	99%	99.9%
δ	17.41	23.13
$\delta - \Gamma$	16.55	21.23

Table 4.5: Impact of option's nonlinearity on VaR numbers (as % change in portfolio value).

As we could have expected, such option exposure notably increases our risk exposition and the impact of gamma is also in evident. We observe that including gamma risk reduces our risk estimates as the distribution becomes positively skewed.

In the third example, the investor speculates on volatility. He thinks that there is a high chance of unexpected news coming up within a year that would significantly move the exchange rates, although he is not sure about the direction of this change. He decides to establish a simple strategy called straddle, that is, he buys both a put and a call option on EURCZK at the same strike price, in the same amount, and with the same expiration date. Such portfolio consisting of only options describes the option's nonlinearity and the difference between delta and delta-gamma probably in the best way. The results are given in Table 4.6.

(a) Portfolio Statistics

Mean	0.475
Variance	18.891
Skewness	0.650
Kurtosis	0.566

(b) Value-at-risk estimates

%	99%	99.9%
δ	16.77	22.28
$\delta - \Gamma$	14.09	16.33

Table 4.6: VaR of a straddle (as % change in straddle value).

In the last case, we exhibit again positive skew in the distribution and accounting for gamma reduces our *VaR* estimates especially for very high quantiles. These results of course heavily depend on the option's specifications, for example, strike price.

Conclusion and Discussion

This thesis summarizes some of the methods used for calculation Value-at-risk and Expected Shortfall. Of course, there are other models and issues about *VaR* and *ES* that are not covered in the work.

Chapter 1 is dedicated to the original three approaches, namely, parametric, Monte Carlo, and Historical Simulation. Introducing parametric (variance-covariance) approach first, we show how to forecast (EWMA) variance of the returns, discuss the linearity of the position captured by *delta* and non-linearity captured by *delta* and *gamma*, and explain how to estimate portfolio linear and non-linear (using Cornish-Fisher expansion) *VaR*. The use of EWMA to model the variance is sometimes substituted with GARCH models. Every market crash, however, evidences failure of the assumption of normally distributed returns. In practice, normal distribution is often substituted with a distribution with heavier tails, most frequently with Student t-distribution with ν degrees of freedom obtained by maximum likelihood estimation (usually $\nu = 3$ or 4 but it does not have to be an integer). When we consider $\alpha = 0.05$, that is, 95% confidence level, then *VaR* estimate with normally distributed returns gives rather accurate results, it is the *extremes* (when $\alpha = 0.01$ or 0.001) where normality fails.

Next we discuss Monte Carlo approach that simulates returns and revalues portfolio after each simulation. Large sample of simulated returns then approximates the distribution of portfolio changes and it is easy to take empirical *VaR* and *ES* estimates from this distribution. Again, it is possible to simulate r.v. from other than normal distribution and thus allow for heavier tails. The last section describes Historical Simulation approach and completes chapter 1. *HS* is a very popular approach since it is simple, transparent, free of distributional assumption and captures fat tails, but might not produce accurate forecasts.

In chapter 2 we point out that *VaR* does not encourage diversification. If used as a risk management tool, this inefficiency can thus give misguided results and have severe consequences in terms of financial losses. We introduce Expected Shortfall which eliminates *VaR*'s deficiencies and satisfies widely accepted axioms of an effective risk measure. We show how *ES* can be (and should be) used as a complement to (or even replacement of) *VaR* for measuring market risk. Remarkably, Riskmetrics [15] document already mentions *ES* (Part V - Backtesting), where it is defined as an *expected value of a return given that it violates VaR*, and illustrated with the formula from Theorem 2.

Chapter 3 describes Extreme Value Theory. This can be seen as an improvement of the previous methodologies in a way that *EVT* particularly focuses on the tails of the distribution. In this chapter, we define Generalized Pareto Distribution and use it to model the tails, and consequently, to estimate *VaR* and *ES*. Next, we describe Conditional Extreme Value Theory which respects conditional volatility of the returns. Both unconditional and conditional *EVT* techniques are demonstrated on a stock market index example. The following section that discusses multi day *EVT VaR* and *ES* prediction completes chapter 3.

In chapter 4 we apply Extreme Value Theory to calculate *VaR* and *ES* for a nonlinear portfolio (a simple investment into local and foreign stock market indices and involved currency risk hedged with a put option) by mixing *HS* and *GPD*. We then compare this method to parametric (delta and delta-gamma) approach and historical simulation. We show how *EVT* supplements *HS* in capturing fat tails and even the tails that are out of the sample range.

Three appendices that describe Cholesky factorisation (appendix A), discuss pricing FX options (B), and explain cash flow map of fixed income instruments (C) finalise the thesis.

Value-at-risk does not describe the worst loss, and it is not designed to do so. What it does is that it evaluates the probability that a loss in the (left) tail occurs. Therefore, different approaches may produce similar *VaR* number, but different shapes of loss distribution (and its tails in particular). This fact can be seen in our results, when we moved confidence level from 99% to 99.9%, the "new" *VaR* number then varied significantly from one method to another. The confidence level and the question of sample period indicates that *VaR* is measured with some error, it is a subject to probability sampling variation.

There is, however, more criticism to *VaR*. For example, Nassim N. Taleb became an increasingly popular critic of current risk management models. His popularity spread after good (lucky?) timing of his book *The Black Swan: The Impact of the Highly Improbable* that was released on April 2007, just before the sub-prime crisis erupted. He points out the difficulty of properly assessing the probabilities of events that are out of our historical sample and high impact of estimation errors around small probabilities and argues that present models (in particular the ones described in this work) cannot estimate tail probabilities with assurance. To be fair, besides criticism, Taleb offers a proposal for estimating the tails. He often cites Mandelbrot and advocates the use of true fat tails (Paretian, power-law tails satisfying $P(X > x) \approx Kx^{-\alpha}$, which are scale invariant, see Taleb [21]) as risk management tools. As a stress test, he suggests to use power laws to measure sensitivity of errors in the tails by varying power-law exponent α and investigate its effect on the changes in *VaR* and *ES* estimates. This effect of the unseen can thus assist in making decisions. In this sense, similar stress tests can be analyzed in Extreme Value Theory by varying tail index ξ. This alternative approach is inspiring and deserves further investigation.

References

[1] Acerbi, C., Tasche, D., 2001. Expected Shortfall: a natural coherent alternative to Value at Risk, *Economic Notes*, Volume 31, 379-388.

[2] Acerbi, C., Tasche, D., 2002. On the coherence of Expected Shortfall, *Journal of Banking and Finance*, Volume 26, Issue 7, 1487-1503.

[3] Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk, *Mathematical Finance*, Volume 9, Issue 3, 203-228.

[4] Balkema, A., de Haan, L., 1974. Residual life time at great age, *Annals of Probability 2*, 792-804.

[5] Benninga, S., Wiener, Z., 1998. Value-at-Risk (VaR), *Mathematica in Education and Research*, Volume 7, Issue 4, 39 - 45.

[6] Danielsson, J., de Vries, C., 2000. Value-at-Risk and Extreme Returns, *Annales d'Economie et de Statistique*, Volume 60, 239-270.

[7] Deutsch, H., Value at Risk, University Lectures, Mathematical Institute, University of Oxford.

[8] Diebold, F., Hickman, A., Inoue, A., Schuermann, T., 1997. Converting 1-day volatility to h-day volatility: scaling by $\sqrt{h}$ is worse than you think, Working Paper - Wharton Financial Institutions Center, University of Pennsylvania, Paper no. 34.

[9] Dowd, K., Blake, D., 2006. After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures, *Journal of Risk and Insurance*, Volume 73, Issue 2, 193-229.

[10] Drost, F.C., Nijman, T.E., 1993. Temporal aggregation of GARCH processes, *Econometrica*, 61, 909-927.

[11] Engle, R., Focardi, S., Fabozzi, F., 2008. ARCH/GARCH Models in Applied Financial Econometrics, Chapter 60 in *Handbook of Finance*, Volume 3, Part 5, Hoboken, New Jersey, John Wiley and Sons.

[12] Gilli, M., Këllezi, E., 2006. An Application of Extreme Value Theory for Measuring Risk, *Computational Economics*, 27(2-3), 207-228.

[13] Hull, J., 2002. Options, futures, and other derivatives, 5th edition, New Jersey: Upper Saddle Drive, Prentice Hall Finance Series.

[14] Jondeau, E., Rockinger, M., 1999. The tail behaviour of stock returns: emerging versus mature markets, Working Paper Series - Banque de France, Paper no. 66.

[15] Longerstaey, J., 1996. Riskmetrics technical document, *Technical Report fourth edition*, J.P.Morgan, New York.

[16] McNeil, A., Saladin, T., 1997. The peaks over thresholds method for estimating high quantiles of loss distributions *Proceedings of XXVIIth International ASTIN Colloquium*, Caims, Australia: Peeters, 23-43.

[17] McNeil, A., 1999. Extreme Value Theory for Risk Managers *Internal Modelling and CAD II*, RISK Books, 93-113.

[18] McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach *Journal of Empirical Finance*, Volume 7, Issues 3-4, 271-300.

[19] Pichler, S., Selitsch, K., 1999. A Comparison of Analytical VaR Methodologies for Portfolios that Include Options, Working Paper, Technische Universitt Wien.

[20] Pickands, J., 1975. Statistical inference using extreme value order statistics, *Annals of Statistics 3*, 119-131.

[21] Taleb, N., 2007. Black Swans and the Domains of Statistics, *The American Statistician*, Volume 61, No. 3 198-200.

[22] Zangari, P., 1996. A VaR methodology for portfolios that include options, *RiskMetrics Monitor*, First Quarter 1996, 4-12.

Appendix A

Cholesky factorisation

Cholesky factorisation of matrix $\Sigma \in \mathbb{R}^{n\times n}$ is a generalisation of a square root. It decomposes a symmetric($\Sigma = \Sigma^T$) positive definite($\forall x \in \mathbb{R}^n \setminus \{0\} : x^T \Sigma x > 0$) matrix $\Sigma = (\sigma_{ij})$ into a lower triangular matrix $L = (l_{ij})$ with $l_{jj} > 0$ and its transpose L^T so that

$$\Sigma = L\,L^T. \tag{A.1}$$

We are solving the equation

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ 0 & l_{22} & \cdots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & l_{nn} \end{pmatrix},$$

that is

$$\sigma_{ij} = \sum_{k=1}^{\min(i,j)} l_{ik} l_{kj}, \quad 1 \le i,\, j \le n, \tag{A.2}$$

where $l_{i,k}, l_{k,j} = 0$ for $k > \min(i,j)$. We can find the $n!$ unknowns l_{ij} through matrix multiplication of each entry, starting from the top left column.
Concretely,

$$\begin{aligned} l_{11} &= \sqrt{\sigma_{11}} \\ \sigma_{i1} &= l_{i1} l_{11} \;\Rightarrow\; l_{i1} = \frac{\sigma_{i1}}{l_{11}}, \quad (1 < i \le n) \\ \sigma_{22} &= l_{21}^2 + l_{22}^2 \;\Rightarrow\; l_{22} = \sqrt{\sigma_{22} - l_{21}^2} \\ \sigma_{i2} &= l_{i1} l_{21} + l_{i2} l_{22} \;\Rightarrow\; l_{i2} = \frac{\sigma_{i2} - l_{i1} l_{21}}{l_{22}}, \quad (2 \le i \le n), \end{aligned} \tag{A.3}$$

in general, the solution is

$$
\begin{aligned}
l_{jj} &= \sqrt{\left(\sigma_{jj} - \sum_{k=1}^{j-1} l_{jk}^2\right)}, \quad j = 1, \ldots, n \\
l_{ij} &= \frac{\left(\sigma_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}\right)}{l_{jj}}, \quad i = j+1, \ldots, n. \qquad \text{(A.4)}
\end{aligned}
$$

Cholesky decomposition has an advantage over LU decomposition since only one triangular matrix needs to be calculated.

Appendix B

Pricing FX Options

A currency option gives the holder (buyer) right to buy (in the case of *call*) or sell (in the case of *put*) a set amount of one currency for another at a determined price (strike price) and time. To gain this right buyer needs to pay the price called option premium.

B.1 Garman-Kohlhagen Formula

Merton generalized Black-Scholes option pricing formula to price European stock or index options that pay a dividend yield (continuously compounded). In Garman-Kohlhagen formula this dividend yield is treated as the interest rate in foreign currency, thus the formula is used to price currency (FX) options. It applies only to European options. The values of the options are

$$\begin{aligned} call &= S\,e^{-r_f T} N(d_1) - K\,e^{-r_h T} N(d_2), \\ put &= -S\,e^{-r_f T} N(-d_1) + K\,e^{-r_h T} N(-d_2), \end{aligned} \tag{B.1}$$

where

$$\begin{aligned} d_1 &= \frac{\log\left(\frac{S}{K}\right) + \left(r_h - r_f + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}, \end{aligned}$$

and

$$
\begin{aligned}
N(.) &= \text{cdf for standard normal random variable} \\
S &= \text{spot exchange rate} \\
K &= \text{exercise (strike) price} \\
r_h &= \text{riskless interest rate for the home currency} \\
r_f &= \text{riskless interest rate for the foreign currency} \\
T &= \text{time to maturity} \\
\sigma &= \text{volatility of the spot exchange rate.}
\end{aligned}
$$

In the sample portfolio, we value the FX option assuming that volatility is stochastic, and follows simple GARCH(1,1) model.

B.2 T-day volatility estimate under GARCH(1,1)

We fit GARCH model to EURCZK currency pair, that is, we estimate the parameters of GARCH(1,1) model with MLE as in (3.22). Recall that for 1-day returns (X_t), simple GARCH(1,1) model has the form

$$
\begin{aligned}
X_t &= \sigma_t Z_t, \\
\sigma_t^2 &= a_0 + aX_{t-1}^2 + b\sigma_{t-1}^2,
\end{aligned}
$$

where independent $Z_t \sim N(0,1)$, $t = 1, \ldots, T$, and $a_0 > 0,\ a \geq 0,\ b \geq 0$, and $a+b < 1$. The daily long-run (unconditional) average variance is $\sigma^2 = a_0/(1-a-b)$. A simple square root of time rule is not desirable to obtain annual (or T-day) unconditional variance because returns are not *iid* (volatility clustering, fat tails, etc.) Instead, we turn to Drost-Nijman formula (see [10]), which may serve as a manual how to correctly transform the variance of 1-day returns into the variance of T-days returns under GARCH processes. Drost & Nijman showed that T-day returns also follow GARCH(1,1) process

$$
\sigma_{(T)t}^2 = a_{0(T)} + b_{(T)}\sigma_{(T)t-1}^2 + a_{(T)}X_{(T)t-1}^2, \tag{B.2}
$$

where

$$
\begin{aligned}
a_{0(T)} &= Ta_0 \frac{1-(a+b)^T}{1-(a+b)}, \\
a_{(T)} &= (a+b)^T - b_{(T)},
\end{aligned}
$$

and $|b_{(T)}| < 1$ is the root of the quadratic equation

$$
\frac{b_{(T)}}{1+b_{(T)}^2} = \frac{\alpha(a+b)^T - \beta}{\alpha(1+(a+b)^{2T} - 2\beta},
$$

where

$$\alpha = T(1-b)^2 + 2T(T-1)\frac{(1-a-b)^2(1-b^2-2ab)}{(\kappa-1)(1-(a+b)^2)} + 4\frac{(T-1-T(a+b)+(a+b)^T)(a-ab(a+b))}{1-(a+b)^2},$$

$$\beta = (a-ab(a+b))\frac{1-(a+b)^{2T}}{1-(a+b)^2},$$

and κ is the kurtosis of 1-day returns. As $T \to \infty$, then $a_{(T)} \to 0$, $b_{(T)} \to 0$, and volatility fluctuations disappear, while square root of time rule magnifies volatility fluctuations, see [8].

We implement this formula to calculate annual volatility of EURCZK exchange rate. Our model comprises of 5 years of data (3/22/2004-3/20/2009), a total of 1287 observations. First, we calculate daily log-returns and apply GARCH(1,1) model to estimate volatility of EURCZK. We obtain GARCH(1,1) parameter estimates by maximizing log-likelihood function as in (3.22). We get the following estimates

a_0	a	b
9.94×10^{-4}	0.075	0.921

Table B.1: GARCH(1,1) parameter estimates for calculating EURCZK volatility.

Using above formulas, we obtain

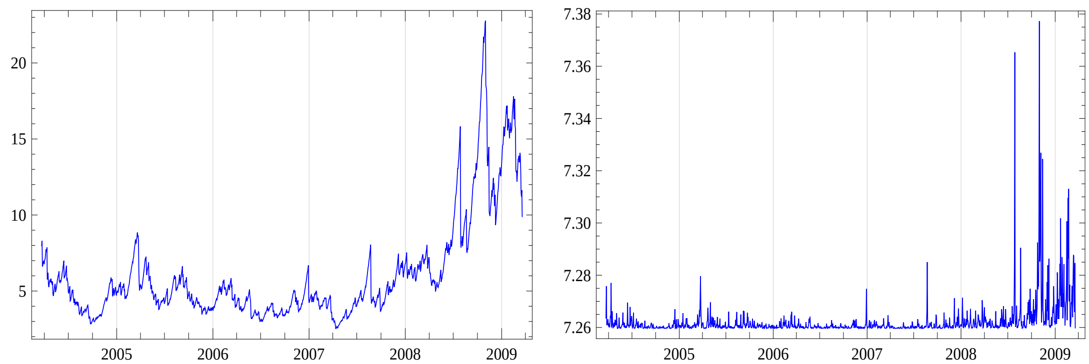
$a_{0(T)}$	$a_{(T)}$	$b_{(T)}$	α	β	κ
40.28	0.15	2.94	0.23	0.69	11.65

Table B.2: Inputs to (B.2) for calculating 1-year (T=250 days) volatility

The graphs of $\sqrt{T}$-day scaled volatility and volatility using D&N formula are displayed in Figure B.1 for $T = 250$.

B.3 Extreme-value volatility estimators

We complete this appendix by mentioning several alternatives for estimating historical volatility in addition to implied volatilities or autoregressive models. We



(a) 1-year scaled volatility (1-day volatility $\times \sqrt{250}$). (b) 1-year volatility using D&N formula (B.2).

Figure B.1: 1-year EURCZK volatility graphs (displayed in %). Scaled (a) vs. Drost & Nijman formula (b).

have not used them in our calculations and we only introduce them as a matter of interest.

A sample standard deviation as a volatility estimator is called close-to-close (CC) since it only uses the market closing prices to estimate the volatility. A more efficient approach uses daily highs (H_t) and lows (L_t), or even daily opening prices (O_t) and closing prices (C_t) on a trading day t. In this sense, daily highs and lows are seen as daily extreme values. Parkinson (1980) proposed the first extreme-value volatility estimator

$$\sigma = \sqrt{\frac{1}{4\log 2}\frac{1}{n}\sum_{t=1}^{n}\left(\log\frac{H_t}{L_t}\right)^2}. \tag{B.3}$$

Garman and Klass (1980) extended the estimator to include opening and closing prices, making the estimator even more efficient (theoretically). They assume that the process for the asset price returns P_t follows a geometric Brownian motion with zero drift and constant volatility σ (that is to be estimated), $dP_t = \sigma P_t dZ_t$, where $dZ_t = \phi\sqrt{dt}$ is an increment to a Wiener process (increments dZ_t are independent and normally distributed with zero mean and variance dt, ϕ is standard normal). This continuity in the process assumes that returns follow the process between transactions and also while the markets are closed. Garman and Klass historical volatility estimator is given by

$$\sigma = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{2}\left(\log\frac{H_t}{L_t}\right)^2 - (2\log 2 - 1)\left(\log\frac{C_t}{O_t}\right)^2\right)}. \tag{B.4}$$

Yang & Zhang derived an extension to GK-estimator that allows for opening jumps in the market

$$\sigma = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\left(\log\frac{O_t}{C_{t-1}}\right)^2 + \frac{1}{2}\left(\log\frac{H_t}{L_t}\right)^2 - (2\log 2 - 1)\left(\log\frac{C_t}{O_t}\right)^2\right)}. \quad \text{(B.5)}$$

Roger & Satchell constructed an estimator that allows for non-zero drift, but not for opening jumps

$$\sigma = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\log\frac{H_t}{C_t}\log\frac{H_t}{O_t} + \log\frac{L_t}{C_t}\log\frac{L_t}{O_t}\right)}. \quad \text{(B.6)}$$

Yang & Zhang derived historical volatility estimator that has a minimum estimation error, does not depend on the drift or opening gaps. It combines Roger & Satchell estimator, close-open volatility, and open-close volatility. The formula is

$$\sigma = \sqrt{\sigma_o^2 + k\sigma_c^2 + (1-k)\sigma_{rs}^2}, \quad \text{(B.7)}$$

where

$$\begin{aligned}
\sigma_o^2 &= \frac{1}{n-1}\sum_{t=1}^{n}\left(\log\frac{O_t}{C_{t-1}} - \mu_o\right)^2, \\
\mu_o &= \frac{1}{n}\sum_{t=1}^{n}\log\frac{O_t}{C_{t-1}}, \\
\sigma_c^2 &= \frac{1}{n-1}\sum_{t=1}^{n}\left(\log\frac{C_t}{O_t} - \mu_c\right)^2, \\
\mu_o &= \frac{1}{n}\sum_{t=1}^{n}\log\frac{C_t}{O_t}, \\
\sigma_{rs}^2 &= \frac{1}{n}\sum_{t=1}^{n}\left(\log\frac{H_t}{C_t}\log\frac{H_t}{O_t} + \log\frac{L_t}{C_t}\log\frac{L_t}{O_t}\right), \\
k &= \frac{0.34}{1+\frac{n+1}{n-1}}. \quad \text{(B.8)}
\end{aligned}$$

Appendix C

Cash Flow Mapping

We price financial instruments by discounting cash flow, in particular the fixed income instruments. The discounting is done by market interest rates and their movement creates interest rate risk. However, there is only a limited number of interest rates that are observable in the market. The idea behind cash flow mapping is to map every financial instrument's position into separate cash flows at current market rates. These positions usually generate wild combinations of cash flows at unique times. Thus observing return series and calculating variances and covariances of too many interest rates is sometimes impossible. The mapping procedure (splitting every cash flow into two closest interest rate vertices) groups all the cash flows into standardized time baskets and simplifies the *VaR* calculation. This is done in RiskMetrics [15] delta-normal method. It is possible to use only those vertices for which we have the spot rate (discount factor), variance (volatility), and correlation with all the other vertices. For example, we can restrict the actual number of interest rates into a given set of vertices

O/N	1W	1M	2M	3M	6M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	10Y	15Y

Up to 1 year, these are money market rates, and above 1 year, they are swap rates or government bond yields (treasury rates). These standard interest rates are chosen because they are liquid and available at financial data providers. The next step is to map every cash flow with maturity between two standard maturities into these standard maturities. This can be performed with different methods.

The mapping procedure

Let *CF(T)* be the expected cash flow at time T that is between two vertices T_{i-1} and T_i. We divide $CF(T)$ into two made up cash flows that mature at the previous vertex T_{i-1} and the following vertex T_i

$$CF(T) \longrightarrow \begin{cases} CF(T_{i-1}) = a\,CF(T) \\ CF(T_i) = b\,CF(T), \end{cases} \tag{C.1}$$

where *a,b* are proportions of the original $CF(T)$, and $T_1 < \ldots < T_{i-1} \leq T < T_i < \ldots < T_n, T_i \in \{$o/n, 1w, 1m, 2m, 3m, 6m, 1y, 2y, 3y, 4y, 5y, 6y, 7y, 10y, 15y$\}$. The weights a, b must satisfy two conditions:

1. The *present value* of the new cash flows is equal to the present value of original cash flow.
2. The *market risk* or the *duration* remains unchanged under the mapping.

We clear the meaning of both choices for the second condition.

Maintaining present value

This approach to determine the proportions a, b is inspired by lecture text by Deutsch [7]. We denote $IR(T_0, T) = IR(T)$ the spot interest rate at present time T_0 with maturity at time T. The condition $PV\,(CF(T)) = PV\,(CF(T_{i-1}) + CF(T_i)) = PV\,(a\,CF(T)) + PV\,(b\,CF(T))$ can be expressed by discount factors $D(T)$, $D(T_{i-1})$, $D(T_i)$,

$$\begin{aligned} PV\,(CF(T)) &= D(T)CF(T) \\ &= D(T_{i-1})CF(T_{i-1}) + D(T_i)CF(T_i) \\ &= D(T_{i-1})aCF(T) + D(T_i)bCF(T), \end{aligned}$$

thus we have

$$D(T) = aD(T_{i-1}) + bD(T_i) \tag{C.2}$$

Notice that $a + b \neq 1$. Discount factors $D(T)$, $D(T_{i-1})$, $D(T_i)$ are calculated from observed interest rates $IR(T), IR(T_{i-1}), IR(T_i)$ and interest rate *IR(T)* is interpolated from rates $IR(T_{i-1}), IR(T_i)$. It is possible to use any interpolation method and any compounding convention. For example, one can use linear interpolation and continuous compounding.

Linear interpolation of spot rate with maturity T is straightforward,

$$IR(T) = \frac{T_i - T}{T_i - T_{i-1}} IR(T_{i-1}) + \frac{T - T_{i-1}}{T_i - T_{i-1}} IR(T_i), \tag{C.3}$$

and the discount factor is calculated as $D(T) = e^{-IR(T)T}$.

Maintaining market risk

We measure market risk by variance of the risk factors. Instead of interest rates, we use discount factors directly as the risk factors. Thus the cash flow we are mapping is linear in the risk factors. The variance of the discount factor *D(T)* is therefore $D(T)^2\sigma_T^2$ and the condition for preserving market risk is

$$D(T)^2\sigma_T^2 = a^2D(T_{i-1})^2\sigma_{i-1}^2 + b^2D(T_i)^2\sigma_i^2 + 2\,a\,b\,D(T_{i-1})D(T_i)\rho_{i,i-1}\sigma_{i-1}\sigma_i \tag{C.4}$$

We already showed how to compute volatilities σ_i, σ_{i-1} and correlation $\rho_{i,i-1}$. Again, we will use linear interpolation to compute σ_T, that is

$$\sigma_T = \frac{T_i - T}{T_i - T_{i-1}} \sigma_{i-1} + \frac{T - T_{i-1}}{T_i - T_{i-1}} \sigma_i. \tag{C.5}$$

Now we can put these two conditions together. First, we substitute

$$\alpha = \frac{D(T_{i-1})}{D(T)} a, \quad \beta = \frac{D(T_i)}{D(T)} b, \tag{C.6}$$

and the two conditions become

$$\begin{aligned} 1 &= \alpha + \beta \\ \sigma_T^2 &= \alpha^2 \sigma_{i-1}^2 + \beta^2 \sigma_i^2 + 2\alpha\beta\rho_{i,i-1}\sigma_{i-1}\sigma_i. \end{aligned} \tag{C.7}$$

We substitute $\beta = 1 - \alpha$ into the second equation and solve quadratic equation with one unknown α

$$\sigma_T^2 = \alpha^2(\sigma_{i-1}^2 + \sigma_i^2 - 2\rho_{i,i-1}\sigma_{i-1}\sigma_i) + 2\alpha(\rho_{i,i-1}\sigma_{i-1}\sigma_i - \sigma_i^2) + \sigma_i^2$$

and with the solution

$$\alpha = \frac{\sigma_i^2 - \rho_{i,i-1}\sigma_{i-1}\sigma_i \pm \sqrt{\sigma_T^2(\sigma_i^2 + \sigma_{i-1}^2 - 2\rho_{i,i-1}\sigma_{i-1}\sigma_i) - \sigma_i^2\sigma_{i-1}^2(1 - \rho_{i,i-1}^2)}}{\sigma_i^2 + \sigma_{i-1}^2 - 2\rho_{i,i-1}\sigma_{i-1}\sigma_i}. \tag{C.8}$$

Finally, the cash flow *CF(T)* after mapping is

$$CF(T) \longrightarrow \begin{cases} CF(T_{i-1}) = \alpha \frac{D(T)}{D(T_{i-1})} CF(T) \\ CF(T_i) = (1-\alpha) \frac{D(T)}{D(T_i)} CF(T), \end{cases} \tag{C.9}$$

This mapping maintains present value and market risk. Alternative to market risk is to maintain duration.

Maintaining duration

This condition says that the duration must be preserved after mapping. The cash flow can be seen as a zero coupon bond and the duration of a zero coupon bond is its maturity, therefore we can write the duration condition as

$$TD(T) = T_{i-1}D(T_{i-1})a + T_iD(T_i)b \tag{C.10}$$

Thus, the two equations we need to solve are

$$\begin{aligned} 1 &= \alpha + \beta \\ T &= T_{i-1}\alpha + T_i\beta \end{aligned} \tag{C.11}$$

Similarly, we substitute $\beta = 1 - \alpha$ and we get a linear equation with the following solution for α

$$\alpha = \frac{T_i - T}{T_i - T_{i-1}}, \quad 1 - \alpha = \frac{T - T_{i-1}}{T_i - T_{i-1}}, \tag{C.12}$$

and the cash flow after duration mapping is

$$CF(T) \longrightarrow \begin{cases} CF(T_{i-1}) = \frac{T_i - T}{T_i - T_{i-1}} \frac{D(T)}{D(T_{i-1})} CF(T) \\ CF(T_i) = \frac{T - T_{i-1}}{T_i - T_{i-1}} \frac{D(T)}{D(T_i)} CF(T). \end{cases} \tag{C.13}$$