A Comparison of EVT and Standard VaR Estimations

Jaroslav Baran
Czech Export Bank, Vodičkova 34, 111 21 Praha and Faculty of Finance and Accounting, University of Economics, nám. W. Churchilla 4, 130 67 Praha.
e-mail: jaroslav.baran@ceb.cz

Jiří Witzany
Faculty of Finance and Accounting, University of Economics, nám. W. Churchilla 4, 130 67 Praha
e-mail: witzanyj@vse.cz

Abstract. In this paper, Extreme value theory (EVT) is applied in estimating low quantiles of P/L distribution and the results are compared to common VaR methodologies. The fundamental theory behind EVT is built, and peaks-over-threshold method is used for modeling the tail of the distribution of losses with Generalized Pareto Distribution (GPD). The different VaR methods are then compared using backtesting procedures. Practical issues such as time varying volatility of returns, and multivariate time series (portfolio of financial instruments) are covered.

Keywords: Risk measures, Value-at-Risk, Extreme Value Theory, GARCH estimations, Expected Shortfall, backtesting.

JEL classification: C10, C22, G10

Acknowledgement: The research has been supported by the Czech Science Foundation grant no. 402/09/0732 "Market Risk and Financial Derivatives"

1. Introduction

Unanticipated moves in the financial markets require risk management tools that can model properly the instability in market prices and the implied potential losses. The rare events cause large swings in prices and thus bring significant consequences. Standard, and at the present time, benchmark method for measuring the unexpected losses is the concept of Value-at-risk (*VaR*). This popular risk management tool aggregates market risk into a single number; it measures the worst loss over a target horizon at a specified probability level. It also serves for determining of market risk capital requirements. *VaR* is often accompanied by Expected Shortfall (*ES*) which measures the average of the worst losses. There has been a lot of criticism of *VaR* and its inability to properly capture these rare losses. Empirical findings suggest modifications of the standard *VaR* and *ES* estimation methods that would capture the occurrences of unexpected rare events and measure their consequences more plausibly.

Extreme Value Theory (EVT) offers proper theoretical background for measuring the rare events. It is well-established in many fields (insurance, engineering, geology, sociology, etc.). The book Modeling Extremal Events for Insurance and Finance by Embrechts et al. is an embracing source of *EVT* for financial and insurance applications. Considering market risk, *EVT VaR* method provides better fit of rare extreme data than standard methods.

Section II presents a theoretical background on standard *VaR* and *ES* methodologies. Section III reviews *EVT* for measuring market risk. Section IV provides description of data and a sample calculation of presented methods on a given portfolio. Section V performs backtesting procedures for the given portfolio and provides the estimation results. The last section summarizes our findings.

We complete this introduction by recalling some of the basic definitions and theory about quantile based market risk measures: Value-at-risk and expected shortfall. Hereinafter, we think of market risk as a possible fluctuation of the value of an asset or a portfolio (profit or loss). A random variable *X* represents this risk and risk measure ρ quantifies it. It maps risk on real line: $\rho : X \to \mathbb{R}$. 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 measures that properly capture risks.

Value-at-risk (*VaR*) is the amount of loss that is not exceeded with a probability $1-\alpha$. Formally, it is the $(1-\alpha)$-quantile $q_{1-\alpha}(-X)$ of the continuous loss distribution

$$VaR_{\alpha}(X) = -\inf\{q \mid P(X \le q) > \alpha\} = -\sup\{q \mid P(X \le q) \le \alpha\}.$$

Equivalently, we can write $VaR_{\alpha} = -F_X^{-1}(\alpha) = -q_{\alpha}(X) = q_{1-\alpha}(-X)$, where F^{-1} is the inverse of the cumulative distribution function (cdf) $F_X(q)$, and $F_X(q) = P(X \le q)$. Parameter α often equals 0.01 or 0.05.

Expected Shortfall (*ES*) closely accompanies *VaR* as it measures the expected loss in the $100\alpha\%$ worst cases. Acerbi & Tasche [1] define *ES* for continuously distributed r.v. *X* as

$$ES_{\alpha}(X) = -\frac{1}{\alpha}\left(\mathbf{E}\left[\mathbb{I}_{X \le q_{\alpha}(X)}\right]\right) = -\mathbf{E}\left[X \mid X \le q_{\alpha}(X)\right] = \mathbf{E}[X \mid X > VaR_{\alpha}].$$

VaR deals with frequency of very large losses and *ES* complements it by taking into account the average size of these losses.

2. Estimating risk measures

In this section, we review the main assumptions, properties and estimation techniques of *VaR* and *ES*. Quantile based risk measures' accuracy depends on the assumption of portfolio return distribution. Empirically, this distribution 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). Moreover, it has been observed that down moves in the markets are more severe than the up moves, and the volatilities are clustered (Cont [6]).

Throughout the text, we work with continuously compounded returns (logarithmic price changes) of X_t,

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

where P_t is a price of a security at time *t* (business day). We simplify a portfolio return as a weighted sum of individual returns, and write $X_t^P = \sum_{i=1}^{n} w_i X_{i,t}$, where $w = (w_1, w_2, \ldots, w_n)'$ is the vector of portfolio weights and $X_{i,t}$ is the return on i-th risk factor (interest rates, foreign exchange rates, prices of underlying instruments, etc.). The common assumption when modelling future returns is that returns X_t are conditionally normally distributed, conditional on the information available at time *t* (past prices and volatilities), $X_t = \sigma_t \varepsilon_t \sim N(0, \sigma_t^2)$, 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 $Var(\varepsilon_t) = 1$. Any linear combination of the returns is also conditionally normally distributed, $X_t^P \sim N(0, \sigma_{p,t}^2)$, where $\sigma_{p,t}^2 = w'\Sigma_t w$, is the variance of the portfolio return and $\Sigma_t = \left(\sigma_{ij,t}^2\right)$ is the covariance matrix (see RiskMetrics document [13]).

Estimating volatility

To account for volatility clustering, variance of an individual asset return, and its corresponding covariances is often forecasted from historical data using autoregressive models. Commonly used is Exponentially Weighted Moving Average model (EWMA, used in RiskMetrics [13]), where more weight is put on more recent observations. The EWMA[1] variance and covariance for the next period *t+1* can be written in a recursive way

$$\sigma_{j,t+1}^2 = \mathbf{E}_t(X_{j,t+1}^2) = \lambda\sigma_{j,t}^2 + (1-\lambda)X_{j,t}^2,$$
$$\sigma_{ij,t+1}^2 = \mathbf{E}_t(X_{i,t+1}X_{j,t+1}) = \lambda\sigma_{ij,t}^2 + (1-\lambda)X_{i,t}X_{j,t},$$
$$i, j = 1, \ldots, n, \tag{1.1}$$

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*. Similarly, the correlation forecast between *i-th* and *j-th* asset return is defined as $\rho_{ij,t+1} = \frac{\sigma_{ij,t+1}^2}{\sigma_{i,t+1}\sigma_{j,t+1}}$, where $\sigma_{j,t+1} = \sqrt{\sigma_{j,t+1}^2}$ is the volatility (standard deviation) of $X_{j,t+1}$. Practically, it is convenient to choose one optimal smoothing factor λ for the whole variance covariance matrix. We determine λ for each risk factor from past return series of this factor by taking the minimum from root average squared variance forecast deviations (errors) for different

[1] by direct substitution of the equation (1.1) 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 of observations.

lambdas. Then we find optimal λ as the weighted average of individual $\lambda_i - s$ (see Baran [6]).

Calculating parametric Value-at-risk

We need to calculate the change in the value of portfolio according to the change in the value of its risk factors. Following Pichler & Selitsch [17], 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. A change in the value of portfolio $V\left(S_1,\ldots,S_n\right)$ composed of linear instruments that depend on *n* risk factors S_i over one period, ΔV, can be written as Taylor series to the first order

$$\Delta V = \sum_{i=1}^{n} \frac{\partial V}{\partial S_i} \Delta S_i \approx \sum_{i=1}^{n} \delta_i X_i,$$
$$\delta_i = \frac{\partial V}{\partial S_i} S_i, \quad X_i = \log\left(\frac{S_{i,t}}{S_{i,t-\Delta t}}\right) \approx \frac{\Delta S_i}{S_i}, \tag{1.2}$$

where δ_i is the sensitivity of the portfolio value with respect to *i*-th risk factor, or so-called *return adjusted delta*. To calculate these partial derivatives numerically, we increase relevant risk factor by small value such as 1 bp or 1 %.

When calculating *linear VaR* (also called *delta* approach) we assume linearity in the risk factors' returns, and we assume that these returns follow a multivariate normal distribution (for simplicity we can assume 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. Equation (1.2) can be written in a vector notation $\Delta V = \delta^{\mathbf{T}} \mathbf{X}$, where δ is a vector of sensitivities δ_i. Therefore, one day *VaR* of portfolio *V* is given by

$$VaR_{\alpha,t+1} = -z_\alpha \sqrt{\delta^{\mathbf{T}} \Sigma \delta}, \tag{1.3}$$

where z_α is the α-quantile of normal distribution, and the expression $\delta^{\mathbf{T}} \Sigma \delta$ 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.*

We say that a financial instrument is non-linear when the change in the value of the instrument is nonlinear in the returns of its risk factors. For non-linear instruments, we capture part of the non-linearity through second order derivatives (Taylor series to the second order)

$$\Delta V = \sum_{i=1}^{n} \delta_i X_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Gamma_{i,j} X_i X_j,$$
$$\Gamma_{i,j} = \frac{\partial^2 V}{\partial S_i \partial S_j} S_i S_j, \tag{1.4}$$

where $\Gamma_{i,j}$ is *return adjusted gamma* (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).

When calculating *non-linear VaR* (*delta-gamma*), we allow for non-linear relationship between ΔV and risk factor returns, that is, we assume that portfolio contains non-linear instruments, such as options. Equation (1.4) can be written in a matrix form

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

where Γ 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 asymmetry of the distribution of ΔV and changes its moments, thus quantile of a normal distribution is no longer appropriate. One of the methods to directly approximate the quantiles of ΔV from its moments is the **Cornish-Fisher expansion**. The first moment (expectation) and the second central moment (variance) of the distribution of ΔV are

$$\mathbf{E}(\Delta V) = \mu_{\Delta V} = \frac{1}{2} tr\left[\Gamma \Sigma\right]$$

$$Var(\Delta V) = \sigma^2_{\Delta V} = \delta^T \Sigma \delta + \frac{1}{2} tr\left[\Gamma \Sigma\right]^2,$$

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

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

where Y is the standardized value of ΔV, $Y = \frac{\Delta V - \mathbf{E}(\Delta V)}{\sqrt{Var(\Delta V)}}$. For *k=3* we get *skewness* (the third standardized moment that measures the asymmetry 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.

The desired quantile $z_{\Delta V,\alpha}$ of ΔV's distribution $F_{\Delta V}$ using its first four moments is approximately

$$z_{\Delta V,\alpha} \approx z_\alpha + \frac{1}{6}\left(z_\alpha^2 - 1\right)\mathbf{E}(Y^3) + \frac{1}{24}\left(z_\alpha^3 - 3z_\alpha\right)\mathbf{E}(Y^4) - \frac{1}{36}\left(2z_\alpha^3 - 5z_\alpha\right)\mathbf{E}(Y^3)^2.$$

The non-linear *VaR* is then given by

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

Historical Simulation

Historical simulation *HS* uses empirical distribution of portfolio returns (losses), therefore, does not depend on any distributional assumption, but we do assume that sample historical returns reasonably describe the distribution of future returns. This method is widely used in practice.

To construct the distribution of future returns, we apply last *N* days (often one or two years of history) returns on a current risk factors' values impacting the value of portfolio, and we get *N* hypothetical portfolio values. We sort these values into ascending order

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

$$VaR_\alpha^{HS} = \Delta V^{([\alpha N])}, \tag{1.5}$$

where $[\alpha N] = \max\{m \mid m \le \alpha N, m \in \mathbb{N}\}$ is the integer part of αN. If this quantile lies between two values, we can interpolate it.

Calculating Expected Shortfall

Expected shortfall[2] (*ES*) adjusts for some *VaR*'s drawbacks, namely *subadditivity* (diversifying the portfolio does not necessarily lead to diversifying (lessening) its risk when measured by *VaR*). This can induce risk manager to assume too much risk when imposing limits on traders. Artzner et al. [4] introduced four axioms for risk measures that, they argue, should hold for every effective risk measure, which is then called *coherent.* Moreover, Artztner et al. [4] propose the 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*"

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

Recalling that risk measure ρ maps the riskiness of the portfolio to required reserves to cover losses from unfavorable movements that regularly occur, this steers towards finding some kind of a weighted average of the scenarios of the worst cases of loss. We are interested both in the beginning of the tail of the underlying distribution of risk factor returns and in its shape, therefore, we are calculating both the minimum loss and the expected loss from the set of the worst losses. *VaR* ignores the tail while *ES* measures it.

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

$$ES_\alpha = \frac{\phi(z_\alpha)}{\alpha}\sqrt{\delta^{\mathbf{T}}\Sigma\delta}, \tag{1.6}$$

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, $P[Z > z_\alpha] = \alpha$. To derive the formula, we set $\sigma^2 = \delta^{\mathbf{T}}\Sigma\delta$ and we have

$$ES_\alpha = -\mathbf{E}[\Delta V \mid \Delta V \le q(\alpha)] = -\frac{1}{\alpha\sigma\sqrt{2\pi}}\int_{-\infty}^{q(\alpha)} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx$$

$$= -\frac{\sigma}{\alpha\sqrt{2\pi}}\int_{-\infty}^{z_\alpha} y \exp\left(-\frac{y^2}{2}\right)dy = \frac{\sigma}{\alpha\sqrt{2\pi}}\exp\left(-\frac{z_\alpha^2}{2}\right) = \frac{\phi(z_\alpha)}{\alpha}\sigma.$$

However, we do not have parametric expression for non-linear expected shortfall. To estimate *ES* empirically we average the $100\alpha\%$ of the largest losses obtained from *VaR*'s historical simulation

$$ES_\alpha^n(\Delta V) = -\frac{\sum_{i=1}^{[n\alpha]} \Delta V^i}{[n\alpha]}. \tag{1.7}$$

[2] In literature, Expected Shortfall is often called Conditional Value-at-risk (*CVaR*).

In both cases, calculating *ES* does not add any significant computational burden once we calculated *VaR.*

3. Extreme Value Theory

In this section, we take a closer look at the tail of the distribution (of security or portfolio returns), because the tail is where extreme losses occur. Extreme Value Theory (*EVT*) 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. The 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, already mentioned book *Modeling extremal events for insurance and finance* by Embrechts, Klüppelberg, and Mikosch, or various papers from authors such as McNeil [14], [15], [16], Danielsson & de Vries [9], 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.

Henceforth, we follow the papers of Gilli & Kellezi [12] and McNeil & Frey [16]. We focus on loss tail distribution, that is, observations that exceed some high threshold (e.g. 95-% loss quantile) and we model it with *Generalized Pareto Distribution* (*GPD*). Then we derive parametric *GPD* based formulae for *VaR* and *ES.*

Generalized Pareto Distribution

The *Generalized Pareto Distribution* describes the limit distribution of scaled excesses over high thresholds. If *X* is a random variable (daily loss) with two-parameter Generalized Pareto Distribution, then the distribution function of *X* has the form

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

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 $\lim_{\xi \to 0}\left(1-\left(1+\xi x/\beta\right)^{-1/\xi}\right) = 1-\exp\left(-x/\beta\right)$. 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*, $\mathbf{E}[X^k]$, 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 cdf of *GPD* yields the density

$$
g_{\xi,\beta}(x)=\begin{cases}\dfrac{1}{\beta}\left(1+\dfrac{\xi}{\beta}x\right)^{-1-1/\xi}, & \xi\neq 0,\\[2ex] \dfrac{1}{\beta}\exp\left(-\dfrac{x}{\beta}\right), & \xi=0.\end{cases}
$$

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.

The Distribution of Excess Losses

Let *X* be a random variable. The conditional distribution function F_u of excess losses over a threshold *u* is defined as

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

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

$$
F_u(y)=\frac{P[X-u\leq y, X>u]}{P[X>u]}=\frac{P[u<X\leq u+y]}{1-P[X\leq u]}=\frac{F(u+y)-F(u)}{1-F(u)}=\frac{F(x)-F(u)}{1-F(u)}.
$$

We are interested in estimating the extremes, that is, F_u. The following result of Balkema & de Haan [5], and Pickands [18] states that for large *u* approaching x_F, excess function F_u converges to GPD $G_{\xi,\beta}$,

$$
\lim_{u\to x_F}\sup_{0\leq y<x_F-u}\left|F_u(y)-G_{\xi,\beta(u)}(y)\right|=0, \tag{2.1}
$$

for some positive real function $\beta(u)$.

This is the key outcome in *EVT* and it 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 appropriate *u* should be high enough to approximate the convergence and low enough to leave enough extreme data. This method of modelling extreme events with *GPD* is called peaks-over-thresholds method.

Estimating *VaR* and *ES*

According to (2.1), $F_u(y)=G_{\xi,\beta(u)}(y)$ for large *u*. The expression for underlying distribution function $F(x)$ thus becomes

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

for $x>u$. We need to estimate the value $F(u)$ to find the corresponding quantile to *u*. This can be done from the empirical distribution function $\hat{F}(u)=(n-N_u)/n$, where *n* denotes the number of observations and N_u is the number of 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

$$\hat{F}(x) = \frac{N_u}{n}\left(1-\left(1+\hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}\right)+\left(1-\frac{N_u}{n}\right) = 1-\frac{N_u}{n}\left(1+\hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \tag{2.2}$$

for $x > u$.

The quantile function of the *GPD* is given by

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

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 (2.2),

$$VaR_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}\alpha\right)^{-\hat{\xi}} - 1\right). \tag{2.3}$$

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

$$ES_\alpha = \mathbf{E}[X \mid X > VaR_\alpha] = VaR_\alpha + \mathbf{E}[X - VaR_\alpha \mid X > VaR_\alpha],$$

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,

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

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 [3]

$$\mathbf{E}(X-u \mid X > u) = \int_0^\infty y\, g_{\xi,\beta}(y)\, dy = \frac{\beta}{1-\xi}, \tag{2.4}$$

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

$$e(VaR_\alpha) = \mathbf{E}(X - VaR_\alpha \mid X > VaR_\alpha) = \frac{\beta + \xi(VaR_\alpha - u)}{1-\xi},$$

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

$$e(u+z) = \mathbf{E}(X-(u+z) \mid X > u+z) = \frac{\beta+\xi z}{1-\xi}.$$

The expression for Expected Shortfall then becomes

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

$$ES_\alpha = VaR_\alpha + \frac{\beta + \xi(VaR_\alpha - u)}{1-\xi} = \frac{VaR_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi}. \qquad (2.5)$$

To understand the average excess over *VaR* in terms of *VaR*, it is sometimes convenient to work with the ratio ES_α / VaR_α,

$$\frac{ES_\alpha}{VaR_\alpha} = \frac{1}{1-\xi} + \frac{\beta - \xi u}{(1-\xi)VaR_\alpha}.$$

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

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.

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

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

where $g_{\xi,\beta}(y_i)$ is the pdf of *GPD* from (2.4) and $y = \{y_1, \ldots, y_n\}$ is the sample of observations. Equivalently, we maximize the log-likelihood function

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

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

$$l(\xi, \beta \mid y) = \begin{cases} -n\log\beta - (\frac{1}{\xi} + 1)\sum_{i=1}^{n} \log(1 + \frac{\xi}{\beta} y_i), & \xi \neq 0, \\ -n\log\beta - \frac{1}{\beta}\sum_{i=1}^{n} y_i, & \xi = 0. \end{cases} \qquad (2.6)$$

Conditional Extreme Value Theory

So far, we have considered the daily returns to have unconditional (constant) variance. Empirically, this is not the case, as returns often exhibit heteroskedasticidy and autocorrelation (of their absolute or squared values, see Engle [11]). The previous theory fails to give proper results during days of high volatility. In this section, we introduce dynamic (time-varying) volatility into *VaR* and *ES* calculations. We adopt a popular approach and work with stochastic volatility (incorporates volatility clustering). We follow the approach introduced in McNeil & Frey [16].

We work with losses as negative returns $X_t = -(\log P_t - \log P_{t-1}) = \log(P_{t-1} / P_t)$, 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

$$X_t = \mu_t + \sigma_t Z_t,$$

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 the 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 [16]). We have

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

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.

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

GARCH(1,1)[4] 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

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

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*

[4] 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 $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)\sigma_{t-1}^2 Z_{t-1}^2$.

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) 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.

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

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 $\hat{\theta} = (\hat{c}, \widehat{a_0}, \hat{a}, \hat{b})'$ 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

$$L(\theta) = \log \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left\{-\frac{\epsilon_t^2}{2\sigma_t^2(\theta)}\right\} = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n}\left(\log \sigma_t^2(\theta) + \frac{\epsilon_t^2}{\sigma_t^2(\theta)}\right). \quad (2.7)$$

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 (2.7) 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

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

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.

Estimating Conditional *VaR* and *ES*

After estimating parameters of AR(1)-GARCH(1,1) process, we calculate vector estimates of the conditional 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

$$\hat{\mu}_{t+1} = \hat{c}x_t,$$

$$\hat{\sigma}_{t+1}^2 = \widehat{a_0} + \hat{a}(x_t - \hat{\mu}_t)^2 + \hat{b}\hat{\sigma}_t^2. \quad (2.8)$$

Applying *EVT*, we fit the tail of the distribution of residuals z_t with *GPD* and calculate *VaR* and *ES* estimates as

$$\widehat{VaR^t_\alpha}(\Delta X) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} VaR_\alpha(Z)$$

$$\widehat{ES^t_\alpha}(\Delta X) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} ES_\alpha(Z), \tag{2.9}$$

where $VaR_\alpha(Z)$ denotes $(1-\alpha)$-quantile of the distribution of residuals Z_t and $ES_\alpha(Z)$ is the related expected shortfall. After ordering 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, where $k = N_u$. Again, we fit the generalized Pareto distribution to excesses above *u*, $(z_{(1)} - z_{(k+1)}, \ldots, z_{(k)} - z_{(k+1)})$ using MLE from. We use (2.2) to estimate the tail of $F_Z(z)$

$$\widehat{F_Z}(z) = 1 - \frac{k}{n}\left(1 + \hat{\xi}\frac{z - z_{(k+1)}}{\hat{\beta}}\right)^{-1/\hat{\xi}}. \tag{2.10}$$

Inverting this formula we get *VaR* estimate as in (2.3),

$$\widehat{VaR_\alpha}(Z) = z_{(k+1)} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{k}\alpha\right)^{-\hat{\xi}} - 1\right). \tag{2.11}$$

Similarly, we use $\widehat{VaR_\alpha}(Z)$, rewrite the formula (2.5) for expected shortfall, and from equation (2.9) we get the estimate of conditional expected shortfall as

$$\widehat{ES_\alpha}(Z) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\left(\frac{\widehat{VaR_\alpha}(Z)}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} z_{k+1}}{1-\hat{\xi}}\right). \tag{2.12}$$

4. Application

We construct a theoretical portfolio and calculate *VaR* and *ES* using delta, delta-gamma, historical simulation, and extreme value methods.

At a portfolio level, it is possible to use multivariate extreme value theory (modeling the tails with multivariate *GPD* and copulas), however, properly matching extreme values and accounting for their correlations still remains a challenge in real portfolios with many risk factors. Although a simplification, we believe, that it is reasonable to apply univariate *EVT* to a single risk factor represented by the returns on the portfolio. 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 [9]. We also apply conditional extreme value method: we standardize the hypothetical returns by AR(1)-GARCH(1,1) volatility estimates and apply conditional *EVT* to the residuals. We then compare *VaR* and *ES* results with parametric linear and non-linear approaches from Section 2.

Consider two equal investments, say CZK 1 million each, into Dow Jones Euro STOXX 50 Index, and PX Index, and a purchase of EUR/CZK currency put option expiring in one year, such that a domestic (Czech based) investor is protected from depreciation of Euro against Czech koruna. In the portfolio, there are following risk factors that affect its value: EUR/CZK exchange rate, , DJ Euro STOXX 50 value, PX 50 value, 1 year PRIBOR, and 1 year EURIBOR[5]. We slightly refine the data so that prices remain constant (zero returns) over holidays. Today it is January 17, 2003. The exchange rate is EUR 1 = CZK 31.458 and we are long EUR put CZK call, with expiration in one year, (artificial) contract size EUR 31,788 (CZK 1 million), and the strike price is set ATM at EUR/CZK 31.458. The 1-day exchange rate volatility is modelled by GARCH(1,1) process, and is extended to 1-year volatility by Drost-Nijman formula (see Drost & Nijman [10]). Although the EUR/CZK volatility is also a risk factor, we neglect it since it has a small impact on the computation (1-year volatility calculated with Drost-Nijman formula fluctuates insignificantly). We use closing prices of last 1000 days (from 3/22/1999 to 1/17/2003, sample size $n=1000$). We are interested in next day's, say, one chance in a hundred, and five chances in a thousand largest losses, so we set α equal 0.01 and 0.005.

Table (2.13): Portfolio specification

Instrument	**Value**
PX Index	CZK 1 million
Euro STOXX 50	CZK 1 million (EUR 31 788)
Put option (notional)	EUR 31 788
Current rate	31.458
Strike price	31.458
Option premium	0.97 % (EUR 306.87)

We use Garman-Kohlhagen formula to price FX option and arrive at a premium of 0.97 cents per 1 EUR. Next, we calculate historical log-returns for each risk factor and use the series of returns to simulate possible paths of tomorrow's returns, see (1.5), thus, we constructed the empirical distribution of portfolio returns, see Figure (2.14). 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 use the formula (1.7) and average $100\alpha\%$ largest losses.

[5] The exchange and interest rates data were downloaded from Bloomberg, STOXX 50 index is available at http://www.stoxx.com/download/chart/none/365_SX5E.csv and PX 50 at http://ftp.pse.cz/Info.bas/Cz/px.csv.

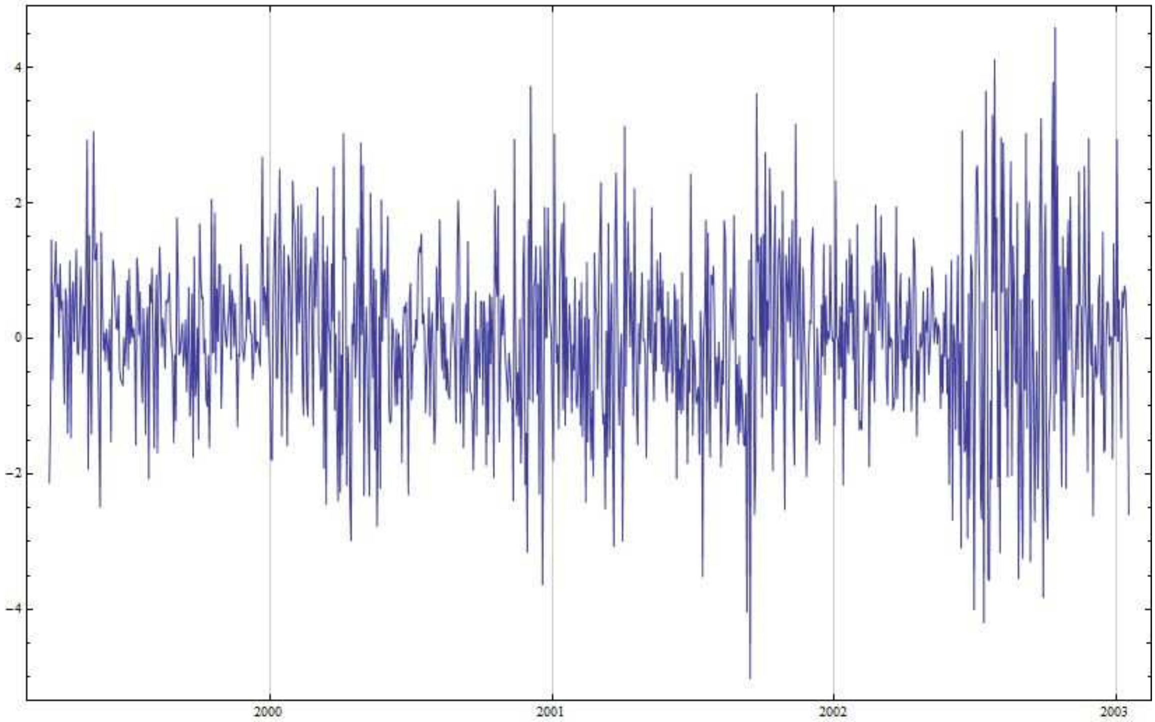

Figure (2.14): Portfolio log-returns.

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 and proceed as in section 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=1.57* (we might get a better fit if we visually chose the threshold, however, if using automated *EVT* as a risk management tool, visually choosing the threshold is impractical). We are left with satisfactory 100 *extremes*. After maximizing *GPD* log-likelihood function (2.6), we obtain the estimates $\hat{\xi} = 0.01$ and $\hat{\beta} = 0.77$ and we use (2.2) to fit the empirical tail with *GPD*. Finally, we obtain *GPD VaR* and *ES* estimates by plugging the estimated parameters into (2.3) and (2.5). In Figure (2.18) we plot different quantiles obtained from historical simulation and extreme value theory.

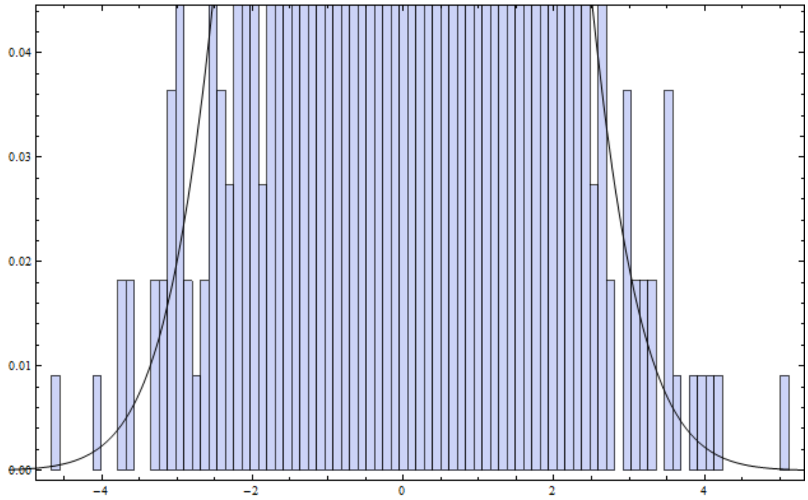

Figure (2.15): Zoom on the tails of the returns (left tail) and losses (right tail) compared to normal pdf.

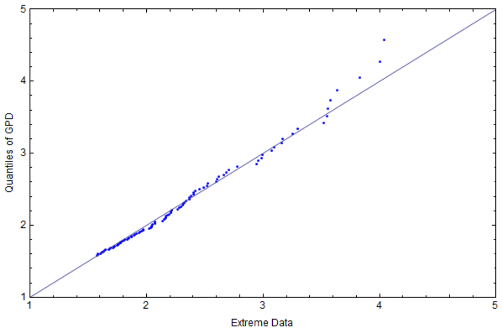

Figure (2.16): QQ-plot of sample quantiles against *GPD* quantiles.

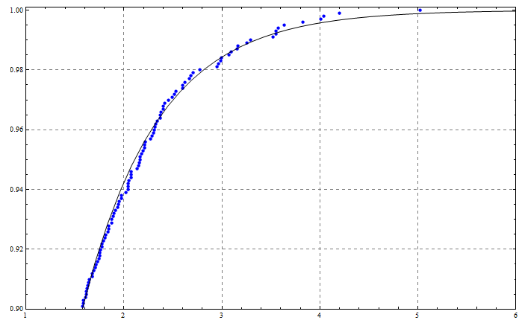
Figure (2.17): ML *GPD* fit to $N_u = 100$ tail losses for the estimates *u=1.57*, $\xi = 0.01$, $\beta = 0.77$.

Now we consider conditional *EVT*, and primarily, that the volatility of the returns is stochastic. From subsection Conditional EVT we assume that $X_t = \mu_t + \sigma_t Z_t$, and use (2.8) to estimate the next day conditional volatility σ_{t+1} and mean μ_{t+1}. We calculate the residuals (*iid* noise) Z_t and subsequently apply formulas (2.9), (2.10), (2.11), (2.12) to calculate conditional *EVT VaR* and *ES*. The results are presented in Table (2.20) and subsequently, the corresponding Figures are displayed.

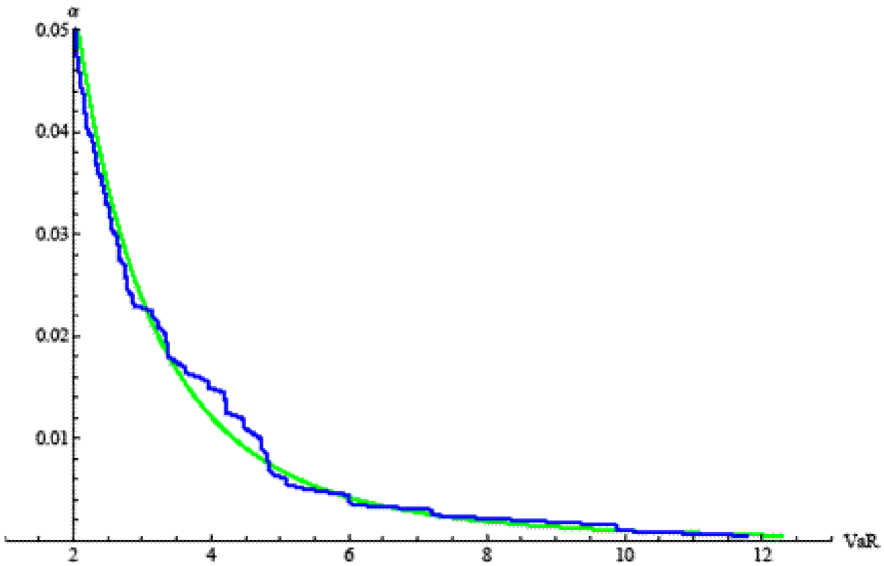

Figure (2.18): VaR estimates for different levels of α using historical simulation and HS-EVT method.

Next, we apply parametric method, we forecast the variance using EWMA (1.1) and we use prevalent RiskMetrics [13] $\lambda = 0.94$ (we 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).

Next, we calculate log-returns for indices and exchange rate and estimate variances and covariances with the help of formulas (1.1) to obtain following variance-covariance matrix

$$\begin{pmatrix} 2.02\times 10^{-5} & -2.15\times 10^{-5} & -1.18\times 10^{-8} & 7.17\times 10^{-8} & -7.98\times 10^{-8} \\ -2.15\times 10^{-5} & 3.62\times 10^{-4} & 7.51\times 10^{-5} & 2.98\times 10^{-7} & -6.77\times 10^{-9} \\ -1.18\times 10^{-8} & 7.51\times 10^{-5} & 9.43\times 10^{-5} & 1.14\times 10^{-7} & 2.40\times 10^{-7} \\ 7.17\times 10^{-8} & 2.98\times 10^{-7} & 1.14\times 10^{-7} & 4.65\times 10^{-8} & -2.64\times 10^{-9} \\ -7.98\times 10^{-8} & -6.77\times 10^{-9} & 2.40\times 10^{-7} & -2.64\times 10^{-9} & 2.13\times 10^{-8} \end{pmatrix}.$$

We calculate the vector of first derivatives numerically from (1.2)

$$\delta^{\mathrm{T}} = (556.03, 1000, 1000, -13.51, 12.39).$$

Using (1.3) we arrive at linear (delta) Value-at-Risk estimate $VaR_{0.01}^{\delta} = 2.77\%$, and exercising parametric formula for expected shortfall (1.6) we get $ES_{0.01} = 3.18\%$. We build a matrix of second derivatives using formula (1.4). Again, we numerically measure how the first derivative of each factor changes when we move each risk factor by 1 bp and we get

$$\begin{pmatrix} 4833.1 & 1000 & 0 & 131.37 & -115.67 \\ 1000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 131.37 & 0 & 0 & 3.94 & -3.48 \\ -115.67 & 0 & 0 & -3.48 & 3.07 \end{pmatrix}.$$

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 use Cornish-Fisher approximation to estimate portfolio's mean and variance; the moments of the portfolio's distribution are presented in Table (2.19)

Table (2.19): Moments of portfolio distribution.

Mean $\mu_{\Delta V}$	0.027
Variance $\sigma^2_{\Delta V}$	588.57
Skewness $\frac{\mathbf{E}\left(\Delta V-\mu_{\Delta V}\right)^3}{\sigma^3_{\Delta V}}$	-0.00172
Kurtosis $\frac{\mathbf{E}\left(\Delta V-\mu_{\Delta V}\right)^4}{\sigma^4_{\Delta V}}$	0.000016

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 $VaR^{\Gamma}_{\alpha} = 2.78$. Regarding Expected Shortfall, we do not have parametric expression for non-linear ES. Complete results are summarized in Table (2.20).

Table (2.20): 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 =1000.

	$VaR_{0.01}$	$ES_{0.01}$	$ES_{0.01}/VaR_{0.01}$	$VaR_{0.005}$	$ES_{0.005}$	$ES_{0.005}/VaR_{0.005}$
HS	3.52	3.93	1.12	3.83	4.32	1.13
EVT	3.34	4.11	1.23	3.88	4.65	1.20
EVT-GARCH	3.08	3.70	1.17	3.51	4.13	1.19
δ	2.77	3.18	1.15	3.69	4.02	1.09
$\delta-\Gamma$	2.78			3.70		

Parametric methods (based on normality of returns) give significantly lower risk estimates than historical simulation or methods based on *EVT*, especially for high 99.9% quantile.

5. Backtesting

To evaluate presented methods, we compare daily VaR and ES estimates of the given portfolio with daily actual (realized) portfolio return over a long time period and examine the number and the size of exceptions (values on days when actual loss is higher that the estimate).

We extend the application section and backtest the portfolio on historical series of risk factors' log returns from 3/22/1999 to 12/13/2010, a total of 3061 observations. We use a window of 1000 daily observations and generate 2061 99% VaR and ES forecasts. We check if the number of exceptions does not differ from theoretical number, if exceptions are not clustered in time, and finally, if the size of exceptions does not differ from estimated ES.

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

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

and we obtain the sequence of violations $\{I_{t+1}\}_{t+1}^{T}$, where T is the number of days of a backtest. We assume that $\Pr[I_{t+1} = 1] = A$, or $I_{t+1} \sim Bernoulli(A)$ *iid* and we expect that $A = \alpha$.

We test if the theoretical population fraction of *VaR* violations (*A*) from each model is significantly distinct from the expected fraction (*α=1%*), that is, we can perform one-sided or two-sided binomial test to check the null hypothesis that $A = \alpha$ (see also McNeil & Frey [16]. For 5% significance level to accept (not to reject) the hypothesis $A \le \alpha$ we need to check that $B(\ge T_1; T, \alpha) \ge 0.05$, where T_0 and T_1 are the numbers of 0s and 1s in the sequence $\{I_{t+1}\}_{t+1}^{T}$ and A is estimated from $\hat{A} = T_1 / T_0$. To test $A \le \alpha$ we look at $B(\le T_1; T, \alpha)$ and for the two-sided binomial test of $A = \alpha$ we need to have $B(\le T_1; T, \alpha) \ge 0.025$ and at the same time $B(\ge T_1; T, \alpha) \ge 0.025$. The results of the test are displayed in Table (2.23).

Independence testing

We also test if violations in different *VaR* methods are clustered in time. We prefer the VaR estimation methods where the test is rejected since a VaR estimate at time t should incorporate all the available information at that time and the probability of violation should be equal to the given constant parameter.

We assume that sequence of violations is time dependent with transition probability matrix

$$\mathrm{P} = \begin{bmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{bmatrix},$$

where A_{01} is the probability of *VaR* being violated tomorrow conditional on not being violated today and $A_{00} = 1 - A_{01}$, see Christofferson [8]. The likelihood function of P is then

$$L(\mathrm{P}) = (1 - A_{01})^{T_{00}} A_{01}^{T_{01}} (1 - A_{11})^{T_{10}} A_{11}^{T_{11}},$$

where T_{01} is the number of observations of *VaR* violations conditioned on non-violation on previous day. Maximum likelihood estimates of A_{01} and A_{11} are

$$\hat{A}_{01} = \frac{T_{01}}{T_{00} + T_{01}},\ \hat{A}_{11} = \frac{T_{11}}{T_{10} + T_{11}},$$

and the matrix of estimated transition probabilities is

$$\hat{\mathrm{P}} = \begin{bmatrix} 1 - \hat{A}_{01} & \hat{A}_{01} \\ 1 - \hat{A}_{11} & \hat{A}_{11} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{T_{00} + T_{01}} & \frac{T_{01}}{T_{00} + T_{01}} \\ \frac{T_{10}}{T_{10} + T_{11}} & \frac{T_{11}}{T_{10} + T_{11}} \end{bmatrix}.$$

If the violations are independent over time, then A_{01} should equal A_{11} (similarly $A_{10} = A_{00}$). We can test the independence hypothesis that $A_{01} = A_{11}$ by likelihood ratio test

$$LR = -2\ln\left[L(\hat{A}) / L(\hat{\mathrm{P}}) \right] \sim \chi_1^2,$$

where $L(\hat{A})$ is likelihood $L(\hat{A}) = (1 - T_1 / T)^{T_0} (T_1 / T)^{T_1}$. Figures (2.21) and (2.22) show backtests of the portfolio for different methods and the results of the backtests are displayed in Table (2.23).

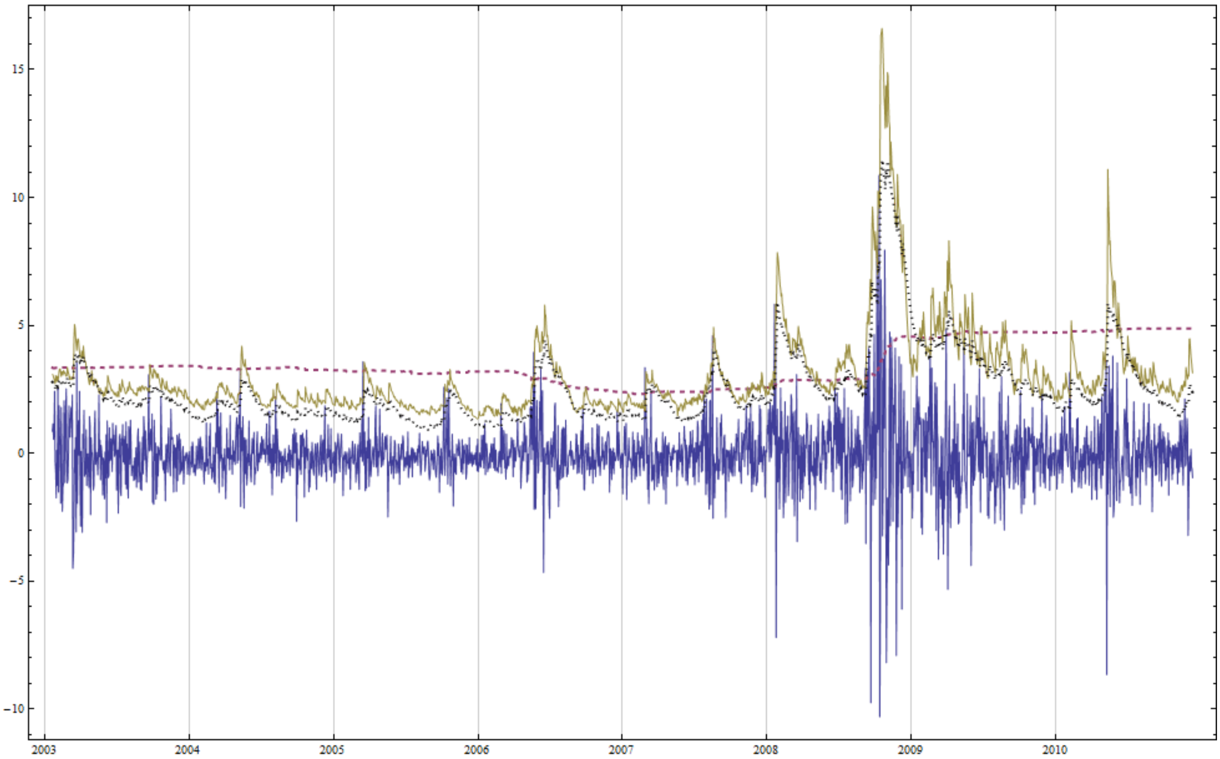

Figure (2.21): Backtesting *VaR* graph for EVT-GARCH (full line), EVT (dashed line) and delta-gamma (dotted line).

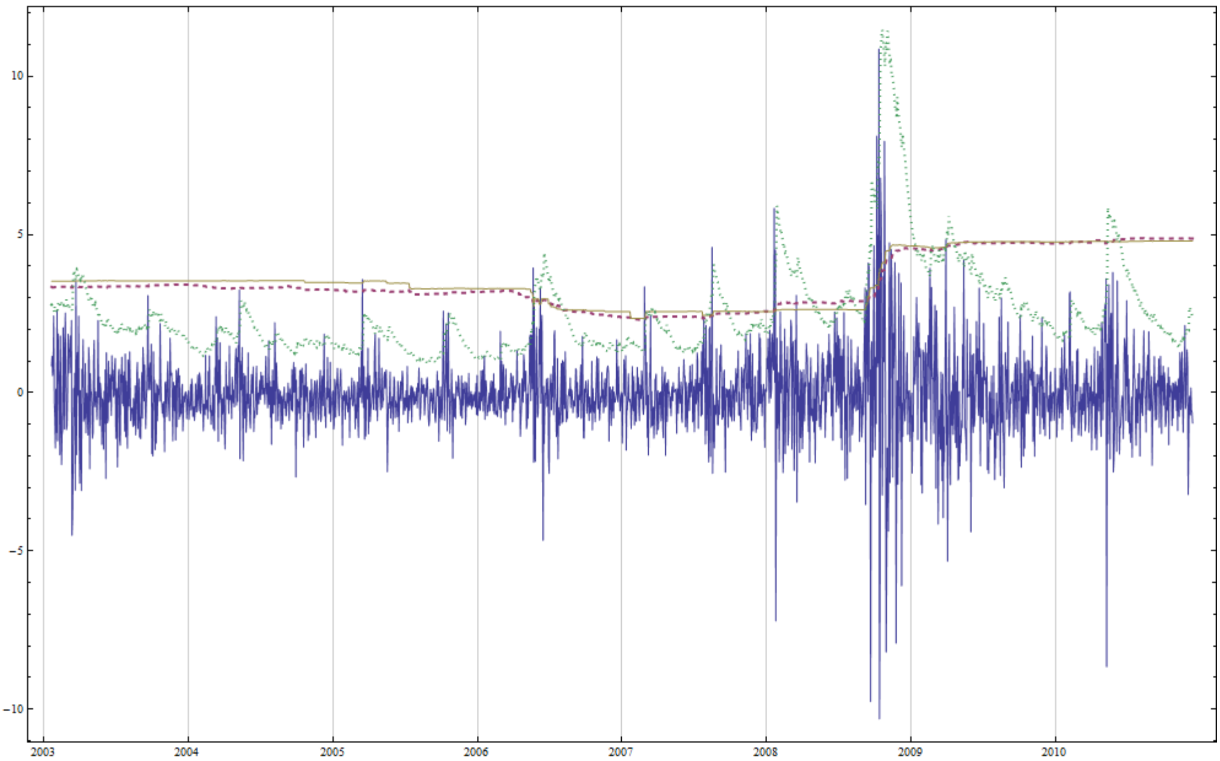

Figure (2.22): Backtesting *VaR* graph for historical simulation (full line), EVT (dashed line) and delta-gamma (dotted line).

Table (2.23): Backtesting results.

Length T=2061 α = 0.01 (99% quantile)	*Expected*	*HS*	*EVT*	*EVT-GARCH*	*δ (δ-Γ)*
# of violations	21	23	26	19	44
Exception rate	1%	1.116%	1.262%	0.922%	2.135%
$B(\geq T_1; T, \alpha)$	0.59	0.75	0.90	0.42	1.00
$B(\leq T_1; T, \alpha)$	0.50	0.33	0.14	0.67	0.00
H_0: $A \leq \alpha$ (5%)		ACCEPT	ACCEPT	ACCEPT	ACCEPT
H_0: $A \geq \alpha$ (5%)		ACCEPT	ACCEPT	ACCEPT	REJECT
H_0: $A = \alpha$ (5%)		ACCEPT	ACCEPT	ACCEPT	REJECT
LR conditional	3.84	9.98	8.54	1.93	5.96
H_0 : A_{01}=A_{11} (5%)		REJECT	REJECT	ACCEPT	REJECT

Expected Shortfall

To backtest *ES* we use approach presented in Angelidis & Degiannakis [3] and we compare actual returns with *ES* estimates for a given day by the two following loss functions

$$\Psi^1_{t+1} = \begin{cases} \left| X_{t+1} - ES_{\alpha,t+1} \right| & if \quad X_{t+1} > VaR_{\alpha,t+1}, \\ 0 & else, \end{cases}$$

$$\Psi^2_{t+1} = \begin{cases} \left(X_{t+1} - ES_{\alpha,t+1} \right)^2 & if \quad X_{t+1} > VaR_{\alpha,t+1}, \\ 0 & else. \end{cases}$$

We prefer the method which gives the lowest mean absolute error (MAE), $T^{-1}\sum_{t=1}^{T}\Psi^1_t$, and mean squared error (MSE), $T^{-1}\sum_{t=1}^{T}\Psi^2_t$. A disadvantage of the approach is that the loss function is influenced by the number of violations, i.e. it is positively biased to methods which may overestimate the ES. The results are displayed in Table (2.24) and the backtesting *ES* graph is outlined in Figure (2.25).

Table (2.24): Backtesting *ES* results (MAE and MSE).

Length T=2061 α = 0.01	*HS*	*EVT*	*EVT-GARCH*	*δ (δ-Γ)*
MAE	0.0135	0.0161	0.0060	0.0123
MSE	0.0368	0.0431	0.0074	0.0168

Total loss MAE $\sum_{t=1}^{T}\Psi_t^1$	27.81	33.10	12.38	25.31
Total loss MSE $\sum_{t=1}^{T}\Psi_t^2$	75.93	88.84	15.18	34.68
Rank	3	4	1	2

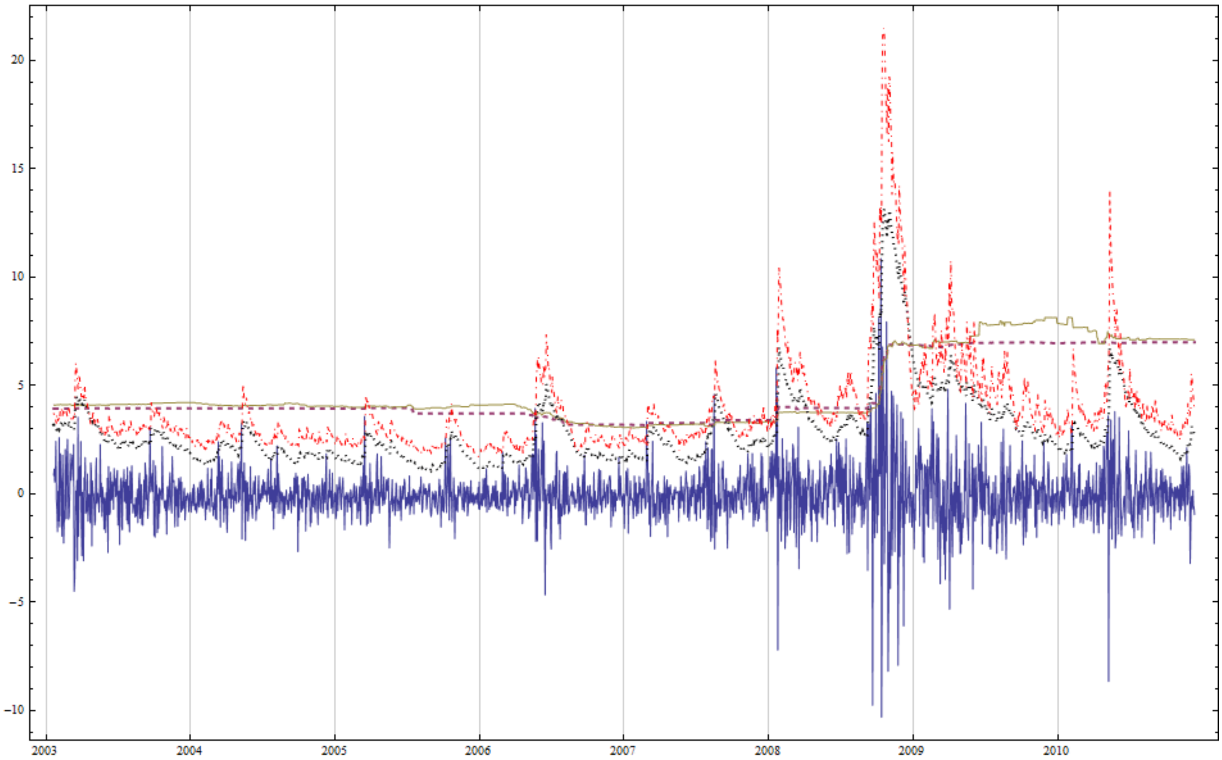

Figure (2.25): Backtesting *ES* for historical simulation (dashed), EVT (full), delta (dotted lower) and EVT-GARCH (dotted upper).

Since we test the hypothesis $E[X - ES \mid X > VaR] = 0$ we may also apply an ordinary test on $X_{t+1} - ES_{\alpha,t+1}$ provided $X_{t+1} > VaR_{\alpha,t+1}$. I.e. we test the t_{n-1} distribution statistic $t = \sqrt{n}\bar{Y} / s$ where n is the number of violations, and $\bar{Y}$ is the mean of the differences $X_{t+1} - ES_{\alpha,t+1}$. It is reasonable to assume that $X - ES$ is drawn from a distribution (determined by the dynamics of the process, the portfolio and the ES estimation method) and so by the CLT the statistic is asymptotically normal. However for smaller values of n the test may be considered doubtful since $X - ES$ is not clearly normally distributed.

Alternatively we have also applied an ordinary bootstrapping method independent on the underlying distribution as proposed in McNeil & Frey [16] in order to conduct a one-sided test against the alternative hypothesis that the residuals have mean greater than zero or, equivalently, that conditional expected shortfall is systematically underestimated, since this is the likely direction of failure.

The results are shown in Table (2.26) and confirm that EVT-GARCH, EVT, and HS methods can be accepted with best p-values shown for EVT-GARCH while the *δ (δ-Γ)* method is rejected by both tests.

Table (2.26): Backtesting *ES* results (t-test and bootstrapping).

Length T=2061 α = 0.01	*Expected*	*HS*	*EVT*	*EVT-GARCH*	*δ (δ-Γ)*
N	21	23	26	19.00	44
$\bar{Y}$	0	0.35	0.25	-0.05	0.42
S		1.86	1.89	0.92	0.90
T	0	0.90	0.67	-0.21	3.14
p-value (t-distribution, 2 sided)		0.38	0.51	0.83	0.00
H_0: $E[X - ES \mid X > VaR] = 0$		ACCEPT	ACCEPT	ACCEPT	REJECT
p-value (bootstrapping, 1-sided)		0.17	0.25	0.58	0.00
H_0: $E[X - ES \mid X > VaR] = 0$		ACCEPT	ACCEPT	ACCEPT	REJECT

In both VaR and ES backtests we show both superior performance of EVT-GARCH method and unfitness of parametric methods (based on normality of the returns). We first checked the hypothesis that the number of exceptions does not differ from theoretical number. EVT-GARCH was the only method that did not underestimate the risk while parametric methods were the only ones where we rejected the hypothesis as they clearly underestimated the probability of violation. According to independence test, EVT-GARCH was the only method that eradicates the threat of violation clustering.

We conclude that 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 we are able to estimate *VaR* and *ES* for any confidence level, in particular, the one that is out of the historical sample, see Figure (2.18)). 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. 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.

6. Conclusion

The paper is extending the work of Baran [6] and compares *EVT* to standard methods (variance-covariance, historical simulation) for calculating *VaR* and *ES*. We show how to forecast (EWMA) variance of the returns, discuss the linearity of the positions 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). Then we describe *HS* approach, which is a very popular approach for its simplicity, transparency, being free of distributional assumption and able to capture fat tails, but might not produce accurate forecasts. Next, 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 *VaR* for measuring market risk. Remarkably, Riskmetrics document [13] 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 formula (1.6).

Section III 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. 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. In section IV we apply *EVT* 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. In the last section we perform backtesting procedures for a given portfolio and confirm that EVT-GARCH is superior to other methods.

In the onward research we will further examine the performance of *EVT* on different time series representing different asset classes and investigate the correlations between the extremes through copulas.

References

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

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

[3] Angelidis T., Degiannakis S.: Backtesting VaR Models: An Expected Shortfall *Approach*, Working Paper no. 701, University of Crete, Department of Economics, 2007.

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

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

[6] Baran J.: *Calculation and Comparison of Several Value at Risk Models for Nonlinear Portfolio*, Diploma Thesis, Faculty of Mathematics and Physics, Charles University in Prague, 2009.

[7] Cont R.: Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance*, Volume 1, Issue 2, (2001), 223 – 236.

[8] Christoffersen, P. F.: *Backtesting and Stress Testing*, Chapter 8 in *Elements of Financial Risk Management*, Academic Press, 2003.

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

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

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

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

[13] Longerstaey, J.: *RiskMetrics technical document, Technical Report fourth edition*, J.P.Morgan, New York, 1996.

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

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

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

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

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