Robust Prediction with Risk Measures

by

Yerlan Duisenbay

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Master of Science in Applied Mathematics

at the

NAZARBAYEV UNIVERSITY

Apr 2020

Author . . . .
Department of Mathematics
Apr 29, 2020

Certified by . . . .
Kerem Ugurlu
Assistant Professor
Thesis Supervisor

Accepted by . . . .
Daniel Pugh
Dean, School of Sciences and Humanities

Robust Prediction with Risk Measures

by

Yerlan Duisenbay

Submitted to the Department of Mathematics
on Apr 29, 2020, in partial fulfillment of the
requirements for the degree of
Master of Science in Applied Mathematics

Abstract

This thesis deals with coherent risk measures and its simulation with respect to different probability distributions. This study gives a numerical scheme to approximate any coherent risk measure via a sum of specific quantiles. We give the theoretical background on coherent risk measures in the first part and in the second part of this thesis we illustrate our findings via several simulations.

Thesis Supervisor: Kerem Ugurlu
Title: Assistant Professor

Acknowledgments

This thesis is the last and most important work of my graduate degree in applied mathematics. I would like to thank my advisor, Kerem Ugurlu, for accompanying me from the beginning to the end of writing this thesis, giving advice and helping me understanding this area in mathematics. Also, I would like to especially thank Nazarbayev University for the opportunity to gain high knowledge by providing all the necessary conditions to obtain it.

Contents

List of Figures

Introduction

Expected performance criteria are used for solving optimization problems. Starting from Bellman, dynamic programming techniques have used risk neutral performance evaluation. Risk aversive methods have been started to used for prediction the corresponding problems by utility functions because of not usefulness of the expected value to measure the criteria of the performance. Defining preferences of the risk-aversion into an axiomatic framework, by the paper of Artzner et al., then assessment of risks has a new side for random outcomes. Thus, *coherent risk measure* has been illustrated. The derivation of the dynamic programming equations of the risk-averse operations and measurement of risks is not so complicated in terms of it is not time-consuming. The reason behind this is that the Bellmann optimality principle is invalid to use in the operations in risk-averse problems. It is already known that the problem of stochastic decision in the multistage case takes more time. If the problem can be resolved at later steps, then there exist a solution for optimality on later stages. But this problem can be tackled by applying one-time Markovian dynamic risk measures. Also, this difficulty can be solved with the application of state aggregation and AVaR decomposition theorem. In this approach, a dual representation of AVaR is used, and thus this method needs optimization over a probability space when finding a solution to the equation of Bellman. Specific sums of quantiles give AVaRs, which used to calculate any coherent risk measures. It visualised as table of numerical results of simulations with quantiles of different distributions.

Chapter 1

Statistical Background

1.1 Random Variables

Consider an experiment, where we throw a die 8 times and we count even outcome. Then, we have sample space Ω,such that $w_0 = \{E, O, E, E, O, O, O, E\} \in \Omega$. In real, we are not interested about probability of coming outcome Even or Odd. We interested on functions(real-valued) of outcomes, such as the number of Even outcomes that appear among our 8 tosses, or the length of the longest run of Odd outcomes. These functions are called random variables.

Definition 1.1. *Random variable X(w) is function X:* $\Omega \to \mathbb{R}$*.*

Suppose $X(w)$ is random variable of how many even outcome will be thrown in w trials. Given 8 experiments, so $X(w)$ is finite number of values called discrete random variables. Here, the probability of the set associated with a random variable X taking on some specific value k is

$$P(X = k) := P(\{w : X(w) = k\})$$

Now, we take random variable $X(w)$ as decay time of Uranium. In this case, $X(w)$ has infinite number of values, so it is called continuous random variables. We designate probability that X takes between two real values a and b:

$$P(a \leq X \leq b) := P(\{w : a \leq X(w) \leq b\})$$

1.1.1 Expected value(mean) and variance

Definition 1.2. *For continuous random variables X with range $[a,b]$ and probability density function $f(x)$ expected value is defined by integral:*

$$\mathbb{E}[X] = \int_a^b x f(x) dx \tag{1.1}$$

Definition 1.3. *For discrete random variables X expected value is weighted sum of values x_i, where weights are probabilities $p(x_i)$:*

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i p(x_i) \tag{1.2}$$

Also, expected value is called mean with symbol μ. It has the following properties:

- $\mathbb{E}[a] = a \ \forall a \in \mathbb{R}$.
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b \ \forall a, b \in \mathbb{R}$.
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Definition 1.4. *Variance σ^2 is measure of concentration of distribution of random variables X around its mean.*

$$\begin{aligned} Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2] = \\ = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned} \tag{1.3}$$

Properties:

- $Var[a] = 0 \ \forall a \in \mathbb{R}$.
- $Var[aX + b] = a^2 Var[X] \ \forall a, b \in \mathbb{R}$.
- $Var[X + Y] = Var[X] + Var[Y]$

Definition 1.5. *Square root of variance is called standard deviation and denoted by σ.*

Definition 1.6. *p^{th} quantile of distribution X is value such that $P(X \leq q_p) = p$*

1.2 Common random variables

1.2.1 Normal distribution

Definition 1.7. *A random variable X is called normal distribution(Gausian distribution) if and only if, for σ > 0 and -$\infty < \mu < \infty$, its density function is:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1.4}$$

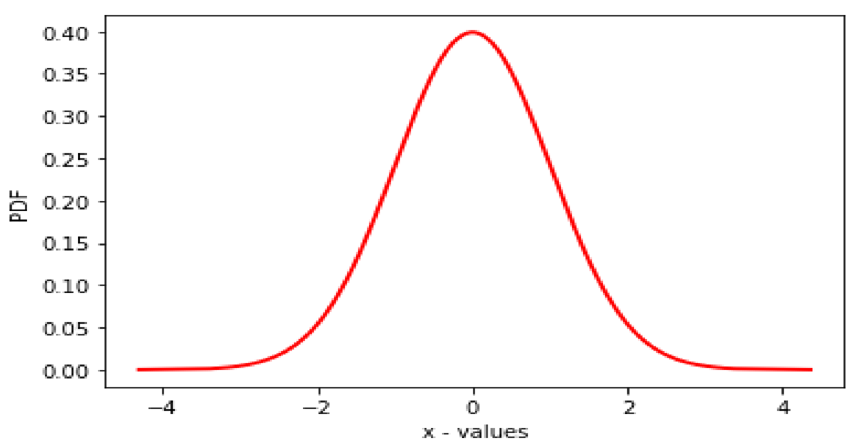


Figure 1-1: Standard Normal Distribution

1.2.2 Chi-square distribution

Definition 1.8. *The distribution which is generated by sum of squares of independent standard normal random variables $X_1, X_2 ... X_n$ is called Chi-square distribution with degree of freedom k.*

$$Q = \sum_{i=1}^{k} X_i^k$$

Probability density function of Chi-square distribution is :

$$f(x,k) = \begin{cases} \frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}, x > 0; \\ 0, otherwise. \end{cases} \tag{1.5}$$

where Γ is gamma function.

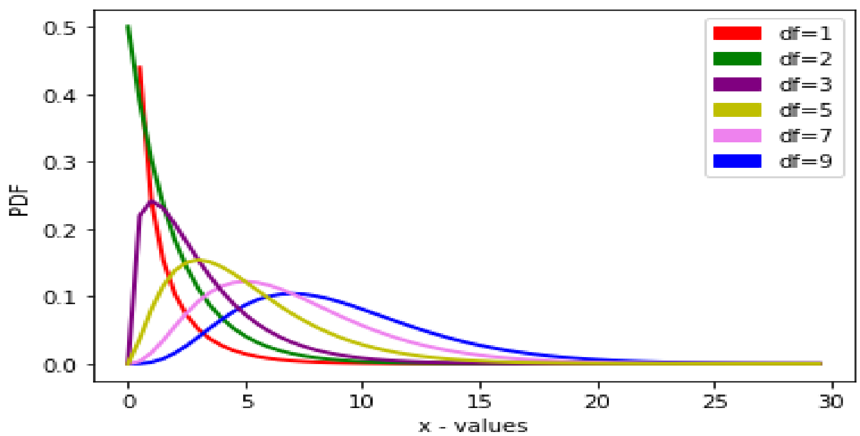


Figure 1-2: Chi square distribution with different df

1.2.3 Exponential distribution

Definition 1.9. *Exponential distribution with* λ *>0 is random variable X with density function of X:*

$$f(x,\lambda) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0; \\ 0, x < 0. \end{cases} \tag{1.6}$$

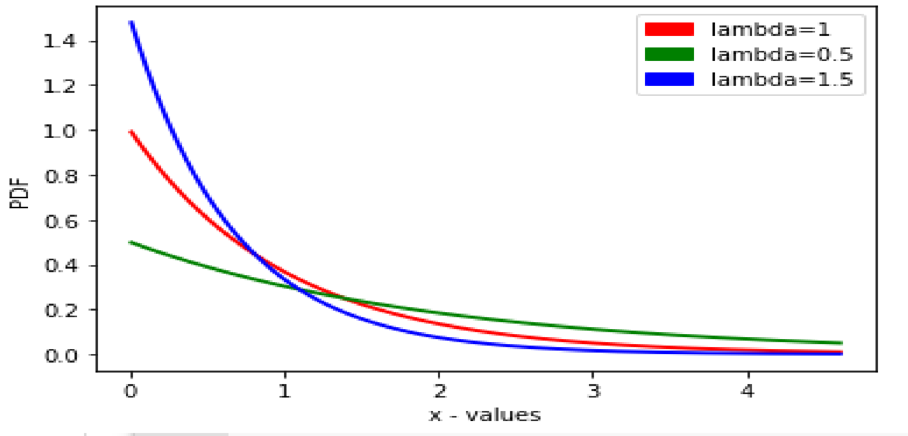


Figure 1-3: Exponential distribution with different λ

1.2.4 Student-t distribution

Definition 1.10. *Student-t distribution X is random variable with probability density function of X:*

$$f(x,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{1.7}$$

where $\nu > 0$ is degree of freedom.

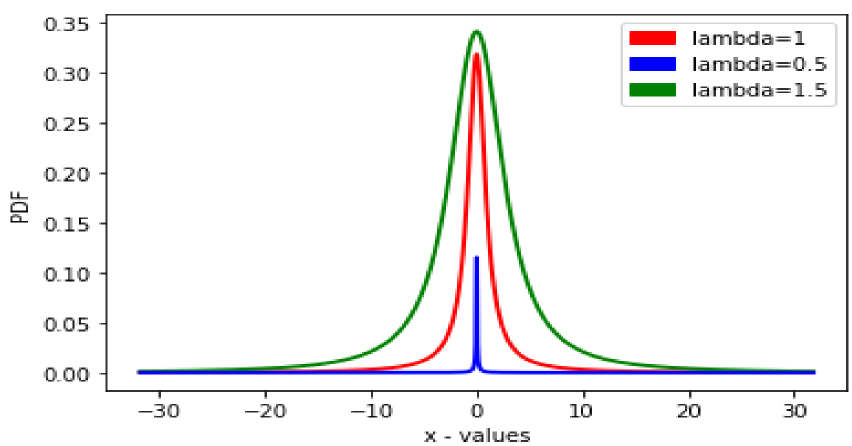


Figure 1-4: Student-t distribution with different degree of freedom

1.2.5 Weibull distribution

Definition 1.11. *Weibull distribution X is random variable with probability density function of X:*

$$f(x,\lambda,k)=\begin{cases}\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{(-\frac{x}{\lambda})^{k}}, x\geq 0;\\ 0, x<0.\end{cases} \tag{1.8}$$

where $k>0$ is shape parameter and $\lambda>0$ is scale parameter.

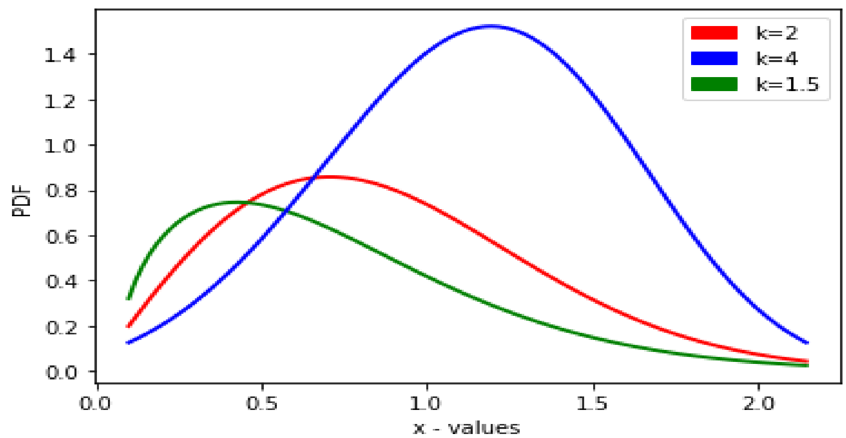


Figure 1-5: Weibull distribution with different k and λ=1

1.3 Verification of a Random Sample from a Distribution

1.3.1 Kolmogorov-Smirnov Test

Suppose that we have an i.i.d. sample $X_1, X_2, ..., X_n$ with some unknown distribution $\mathbb{P}$ and we would like to test the hypothesis that $\mathbb{P}$ is equal to a particular distribution $\mathbb{P}_0$ i.e. decide between the following hypotheses:

$$H_0:\mathbb{P}=\mathbb{P}_0,\ \ H_1:\mathbb{P}\neq\mathbb{P}_0$$

Kolmogorov-Smirnov test is a non-parametric test used for testing a hypothesis $X_1, X_2...X_n$ have a given continuous distribution function F, against the one-sided alternative $H_1^+ : \sup_{|\chi|<\infty}(EF_n(x)) - F(x)) > 0$, where EF_n is the mathematical expectation of the empirical distribution function F_n. The Kolmogorov–Smirnov test is constructed from the statistic:

$$D_n^+ = \sup_{|\chi|<\infty} (F_n(x)) - F(x)) = \max_{1\leq m\leq n} (\frac{m}{n} - F(X_{(m)}))$$

where $X_{(1)}...X_{(n)}$ is the variational series (or set of order statistics) obtained from the sample $X_1, X_2...X_n$. Thus, the Kolmogorov–Smirnov test is a variant of the Kolmogorov test for testing the hypothesis H_0 against a one-sided alternative H_1^+. By studying the distribution of the statistic D_n^+, N.V. Smirnov showed that

$$P\{D_n^+ \geq \lambda\} = \sum_{k=1}^{(1-\lambda)n} \lambda\binom{n}{k}(\lambda + \frac{k}{n})^{k-1}(1 - \lambda - \frac{k}{n})^{n-k},$$

where $0 < \lambda < 1$ and $[\alpha]$ is integer part of number α. Smirnov obtained in addition to the exact distribution of D_n its limit distribution, namely: if $n \to \infty$ and $0 < \lambda_0 < \lambda = O(n^{\frac{1}{6}})$,

$$P\{D_n^+ \geq \lambda\} = e^{-2\lambda^2}\Big[1 + O\Big(\frac{1}{\sqrt{n}}\Big)\Big],$$

where λ_0 is any positive number. By means of the technique of asymptotic Pearson transformation it has been proved that if $n \to \infty$ and $0 < \lambda_0 < \lambda = O(n^{\frac{1}{3}})$, then

$$P\{\frac{1}{18n}(6nD_n^+ + 1)^2 \geq \lambda\} = e^{-\lambda}\Big[1 + O\Big(\frac{1}{n}\Big)\Big].$$

According to the Kolmogorov–Smirnov test, the hypothesis H_0 must be rejected with significance level α whenever

$$exp\Big[\frac{1}{18n}(6nD_n^+ + 1)^2\Big] \leq \alpha,$$

where,

$$P\Big\{exp\Big[\frac{1}{18n}(6nD_n^+ + 1)^2\Big] \le \alpha\Big\} = \alpha\Big[1 + O\Big(\frac{1}{n}\Big)\Big].$$

The testing of H_0 against the alternative H_0^- : $\inf_{|\chi|<\infty}(EF_n(x)) - F(x)) < 0$ is dealt with similarly. In this case the statistic of the Kolmogorov–Smirnov test is the random variable

$$D_n^- = -\inf_{|\chi|<\infty}(F_n(x)) - F(x)) = \max_{1\le m\le n}(-\frac{m-1}{n} + F(X_{(m)}))$$

whose distribution is the same as that of the statistic D_n^+ when H_0 is true.

1.3.2 Kolmogorov Smirnov test for 2 samples

KS test for 2 sample is similar to KS test. Suppose we have first sample size of m with c.d.f $F(x)$ and second sample with size n with c.d.f $G(x)$ and we want to test:

$$H_0 : F = G \ vs \ H_1 : F \neq G$$

If $F_m(x)$ and G_n are corresponding empirical c.d.f.s then the statistic

$$D_{mn} = \big(\frac{mn}{m+n}\big)^{\frac{1}{2}} \sup_x |F_m(x) - G_n(x)|$$

and other are same.

1.4 Linear Regression

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression.

$$y \approx \hat{y} = Xw$$

As loss function we take sum of squared errors:

$$L(w) = \sum_{n=1}^{n}(x_i^T w - y_i)^2 = min_w ||Xw - \hat{y}||_2^2$$

$$w^* = argmin_w L(w) = \big(X^T X)^{-1} X^T y$$

Proof: Now, we can find the smallest w by minimizing loss function.

$$\begin{aligned} L(w) = \sum_{n=1}^{n}(x_i^T w - y_i)^2 &= min_w ||Xw - \hat{y}||_2^2 \\ &= (Xw - y)^T (Xw - y) \\ &= (Xw)^T Xw - (Xw)^T y - y^T Xw + y^T y \\ &= w^T X^T Xw - 2w^T X^T y + y^T y \end{aligned}$$

We take gradient of $L(w)$ and equate to 0:

$$\begin{aligned} \nabla L(w) &= \nabla(w^T X^T Xw - 2w^T X^T y + y^T y) \\ &= \nabla(w^T X^T Xw) - 2\nabla(w^T X^T y) + \nabla(y^T y) \\ &= 2X^T Xw - 2X^T y = 0 \\ w^* &= argmin_w L(w) = \big(X^T X)^{-1} X^T y \end{aligned}$$

Chapter 2

We refer the reader for the definitions and theorems of this chapter and for further study of risk measures to [5].

2.1 Risk measure

Definition 2.1. *A risk measure ρ is a function from random variables to real numbers: $\rho : X \to \mathbb{R}$.*
Properties:

- $\rho(0) = 0$
- $\rho(X + a) = \rho(X) + a$ *for* $\forall a \in \mathbb{R}$
- *if* $X_1 \leq X_2$ *then* $\rho(X_1) \leq \rho(X_2)$ *for* $\forall X_1, X_2 \in X$

2.2 Convex risk measures

Definition 2.2. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the space $\mathcal{Z} := L^1(\Omega, \mathcal{F}, \mathbb{P})$ of measurable functions $Z : \Omega \to \mathbb{R}$ (random variables) that have finite first order moments, i.e. $\mathbb{E}^{\mathbb{P}}[|Z|] < \infty$, where $\mathbb{E}[\cdot]$ means the expectation with respect to the probability measure $\mathbb{P}$. A mapping $\rho : \mathcal{Z} \to \mathbb{R}$ is called a* convex risk measure*, if the following axioms hold:*

- (A1)(Convexity) $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \forall \lambda \in (0,1),\ X,\ Y \in \mathcal{Z}$

- (A2)(Monotonicity) If $X \leq Y$, then $\rho(X) \leq \rho(Y)$,$\forall X, Y \in \mathcal{Z}$

- (A3)(Translation Invariance) $\rho(c + X) = c + \rho(X), \forall c \in \mathbb{R}, X \in \mathcal{Z}$

2.3 Coherent risk measures

Definition 2.3. *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and the space* $\mathcal{Z} := L^1(\Omega, \mathcal{F}, \mathbb{P})$ *of measurable functions* $Z : \Omega \to \mathbb{R}$ *(random variables) that have finite first order moments, i.e.* $\mathbb{E}^{\mathbb{P}}[|Z|] < \infty$*, where* $\mathbb{E}[\cdot]$ *means the expectation with respect to the probability measure* $\mathbb{P}$*. A mapping* $\rho : \mathcal{Z} \to \mathbb{R}$ *is said to be called a* coherent risk measure*, if the following axioms hold:*

- (A1)(Convexity) $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \forall \lambda \in (0,1),\ X,\ Y \in \mathcal{Z}$

- (A2)(Monotonicity) If $X \leq Y$, then $\rho(X) \leq \rho(Y)$,$\forall X, Y \in \mathcal{Z}$

- (A3)(Translation Invariance) $\rho(c + X) = c + \rho(X), \forall c \in \mathbb{R}, X \in \mathcal{Z}$

- (A4)(Homogeneity) $\rho(\beta X) = \beta\rho(X), X \in \mathcal{Z}, \beta \geq 0$

2.4 Value at Risk

Definition 2.4. *Let* $(\Omega, \mathcal{G}, \mathbb{P})$ *be a measurable space and let* $X \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ *be a real-valued random variable and* $\alpha \in (0,1)$*.Then VaR is:*

$$VaR_\alpha(X) = inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq \alpha\} \tag{2.1}$$

But VaR is not coherent risk measure. So we will use Average Value at Risk.

2.5 Average Value at Risk

Definition 2.5. *Let* $(\Omega, \mathcal{G}, \mathbb{P})$ *be a measurable space and let* $X \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ *be a real-valued random variable and* $\alpha \in (0,1)$*.Then AVaR is:*

$$AVaR_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_t(X)dt \tag{2.2}$$

Definition 2.6. *Suppose δ is probability measure on $\mathbb{R}$ and $\alpha \in (0,1)$.*

- *A number $q \in \mathbb{R}$ is called α-quantile of δ if:*

$$\delta((-\infty, q]) \geq \alpha \quad and \quad \delta(([q, \infty)) \geq 1 - \alpha. \tag{2.3}$$

- *A function $q_\delta : (0,1) \to \mathbb{R}$ is called a quantile function of δ if for each $\alpha \in (0,1)$, $q_\delta(\alpha)$ is an α-quantile of δ.*

Remark:

1.The set of α-quantile of δ is a non-empty bounded closed interval and end points are $q_\delta^-(\alpha)$ and $q_\delta^+(\alpha)$.
2.The set $\{\alpha \in (0,1) \mid q_\delta^- < q_\delta^+\}$ is countable. Because α-s for which $q_\delta^-(\alpha) < q_\delta^+(\alpha)$ corresponds to intervals of constancy of the cumulative distribution function of δ.

Theorem 2.5.1. *If $X \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ and $\alpha \in (0,1)$, then there exists an integral*

$$\int_0^1 VaR_\alpha d\alpha$$

and it is equal to $\mathbb{E}[X]$.

Proof:

$$\int_0^1 VaR_\alpha d\alpha = \int_0^1 q_\delta^-(\alpha) = \mathbb{E}[q_\delta^-(\alpha)] = \mathbb{E}[X]$$

Theorem 2.5.2. *Suppose q_X is quantile function for distribution X. Then*

$$AVaR_\lambda(X) = \frac{1}{\lambda}\mathbb{E}[(X - q_X(\lambda))_+] + q_X(\lambda). \tag{2.4}$$

Proof: Suppose U is standard uniform random variable. Then distributions of X

and $q_X(U)$ are same. And consequently:

$$\begin{aligned}
&\frac{1}{\lambda}\mathbb{E}[(X - q_X(\lambda))_+] + q_X(\lambda)\\
=&\frac{1}{\lambda}\mathbb{E}[(q_X(U) - q_X(\lambda))_+] + q_X(\lambda)\\
=&\frac{1}{\lambda}\int_0^1 (q_X(\alpha) - q_X(\lambda))_+ d\alpha + q_X(\lambda)\\
=&\frac{1}{\lambda}\int_{1-\lambda}^1 (q_X(\alpha) - q_X(\lambda)) d\alpha + q_X(\lambda)\\
=&\frac{1}{\lambda}\int_{1-\lambda}^1 VaR_\alpha(X) d\alpha\\
=&\mathrm{AVaR}_\lambda(X)
\end{aligned} \tag{2.5}$$

Theorem 2.5.3.

$$AVaR_\lambda(X) = \sup\left\{\mathbb{E}_{\mathbb{Q}}[X] \;\middle|\; \mathbb{Q} \ll \mathbb{P}, \frac{dQ}{d\mathbb{P}} \leq \frac{1}{1-\lambda}\right\}. \tag{2.6}$$

Proof: The supremum on the right-hand side is equal to

$$\sup\left\{\mathbb{E}[\varphi X]\middle|\varphi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}), \mathbb{E}[\varphi] = 1, 0 \leq \varphi \leq \frac{1}{1-\lambda}\right\}$$

$\mathbb{E}[\varphi X]$ is large, if φ takes large values at points, where X takes large values. Hence the supremum is attained for

$$\varphi := \begin{cases} 1/1-\lambda & \text{on } \{X > q_X(\lambda)\}, \\ 0 & \text{on } \{X < q_X(\lambda)\}, \\ k & \text{on } \{X = q_X(\lambda)\}. \end{cases}$$

where q_X is any quantile function of the distribution of X and k is such that $\mathbb{E}[\varphi] = 1$, i.e.

$$\frac{1}{1-\lambda}\mathbb{P}(X > q_X(\lambda)) + k\mathbb{P}(X = q_X(\lambda)) = 1.$$

it follows that

$$\sup\left\{\mathbb{E}[\varphi X] \;\middle|\; \varphi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}), \mathbb{E}[\varphi] = 1,\ 0 \le \varphi \le \frac{1}{1-\lambda}\right\}$$
$$= \frac{1}{1-\lambda}\mathbb{E}[X \cdot 1_{\{X > q_X(\lambda)\}}] + k\mathbb{E}[X \cdot 1_{\{X = q_+(\lambda)\}}]$$
$$= \mathrm{AVaR}_\lambda(X).$$

2.6 Tail conditional expectation

Tail conditional expectation is used to measure market and non-market risks, presumably for a portfolio of investments. It gives a measure of a right-tail risk, one with which actuaries are very familiar because insurance contracts typically possess exposures subject to "low-frequency but large-losses".

Definition 2.7. *For $\lambda \in (0,1)$ tail conditional expectation is defined by:*

$$TCE_\lambda(X) := \mathbb{E}[X | X \ge VaR_\lambda(X)] \tag{2.7}$$

Theorem 2.6.1. *For $\lambda \in (0,1)$ $TCE_\lambda(X)$ and $AVaR_\lambda(X)$ are equal if and only if $\mathbb{P}(X \ge VaR_\lambda(X)) = 1 - \lambda$. It happens if and only if X has continuous distribution.*

Proof:

Suppose $\mathbb{P}(X \ge VaR_\lambda(X)) = \lambda$. Then:

$$\begin{aligned} TCE_\lambda(X) =& \frac{\mathbb{E}[X \cdot 1_{\{X \ge VaR_\lambda(X)\}}]}{\mathbb{P}(X \ge VaR_\lambda(X))} \\ =& \frac{1}{1-\lambda}\mathbb{E}[X \cdot 1_{\{X \ge VaR_\lambda(X)\}}] \\ =& \frac{1}{1-\lambda}\mathbb{E}[(X - VaR_\lambda(X)) \cdot 1_{\{X \ge VaR_\lambda(X)\}}] + VaR_\lambda(X) \\ =& \frac{1}{1-\lambda}\mathbb{E}[(X - VaR_\lambda(X))_+] + VaR_\lambda(X) = AVaR_\lambda(X) \end{aligned} \tag{2.8}$$

2.7 Entropic risk measures

In financial mathematics, the entropic risk measure is a risk measure which depends on the risk aversion of the user through the exponential utility function.

Definition 2.8. *The entropic risk measure with the risk aversion parameter $\theta > 0$ is defined as :*

$$\rho^{ent}(X) = \frac{1}{\theta} log(\mathbb{E}[e^{\theta X}]) = \sup \left\{ \mathbb{E}^Q[X] - \frac{1}{\theta} H(Q|\mathbb{P}) \right\} \tag{2.9}$$

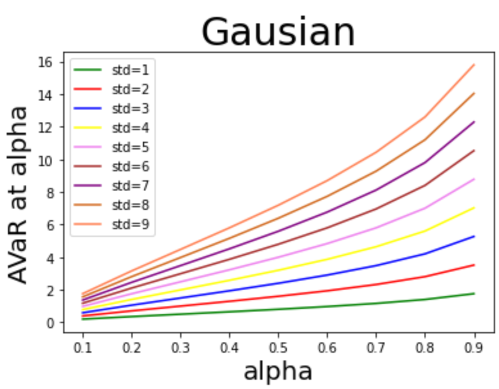
If X has standard normal distribution, then

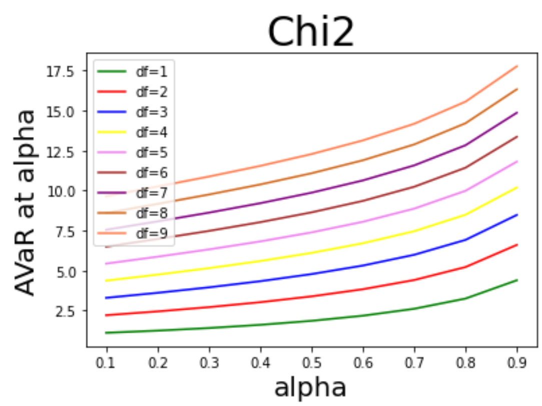
$$\begin{aligned}
\frac{1}{\theta}\log(\mathbb{E}[e^{\theta X}]) =& \frac{1}{\theta}\log(\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\theta x} e^{\frac{-x^2}{2\sigma^2}} dx) \\
=& \frac{1}{\theta}\log(\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\frac{\theta^2\sigma^2}{2} - (\frac{x}{\sqrt{2}\sigma} - \frac{\theta\sigma}{\sqrt{2}})^2} dx) \\
& \Big| u = \frac{x - \theta\sigma^2}{\sqrt{2}\sigma} \rightarrow \frac{du}{dx} = \frac{1}{\sqrt{2\pi}\sigma} \Big| \\
=& \frac{1}{\theta}\log(\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \sqrt{2}\sigma e^{\frac{\theta^2\sigma^2}{2} - u} du) \\
=& \frac{1}{\theta}\log(e^{\frac{\theta^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-u}\ du) = \frac{1}{\theta}\log(e^{\frac{\theta^2\sigma^2}{2}} erf(\infty)) = \frac{\theta\sigma^2}{2}
\end{aligned}$$

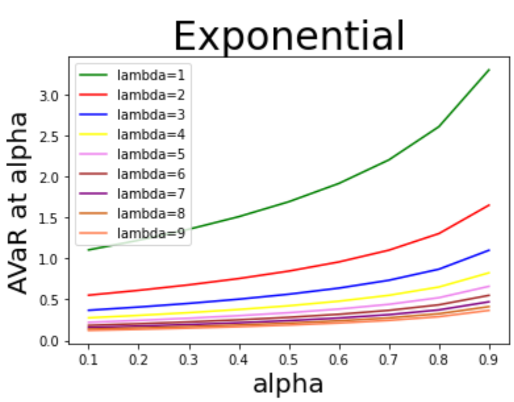
Chapter 3

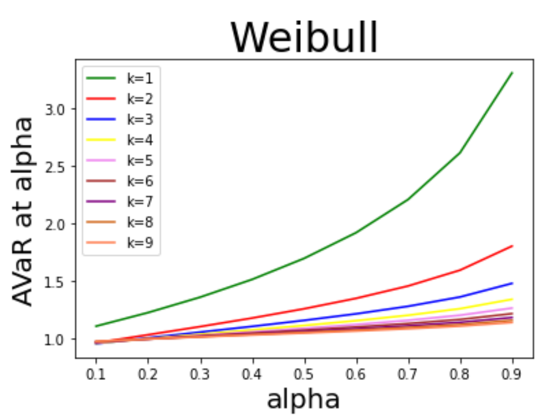
3.1 Calculating AVaR's of different distributions

We calculate AVaR's of common distributions by the formula (2.2) and present their graphs with respect to different α values.









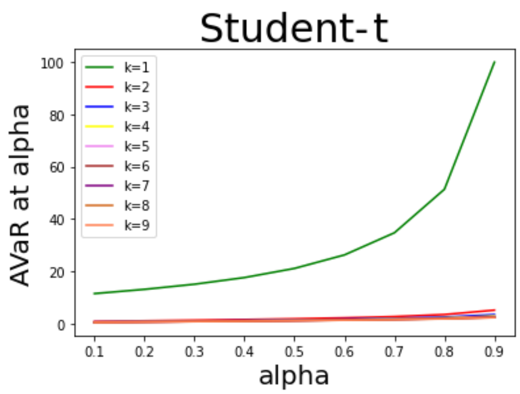


Figure 3-1: AVaRs of different distributions[Listing A.9]

3.2 Data

We use the data from 18.12.2019 to 19.12.2018 (https://http//investfunds.kz). It contains exchange rate value of USD/KZT and main factors affecting on it (BRENT price and USD/RUR exchange rate value).

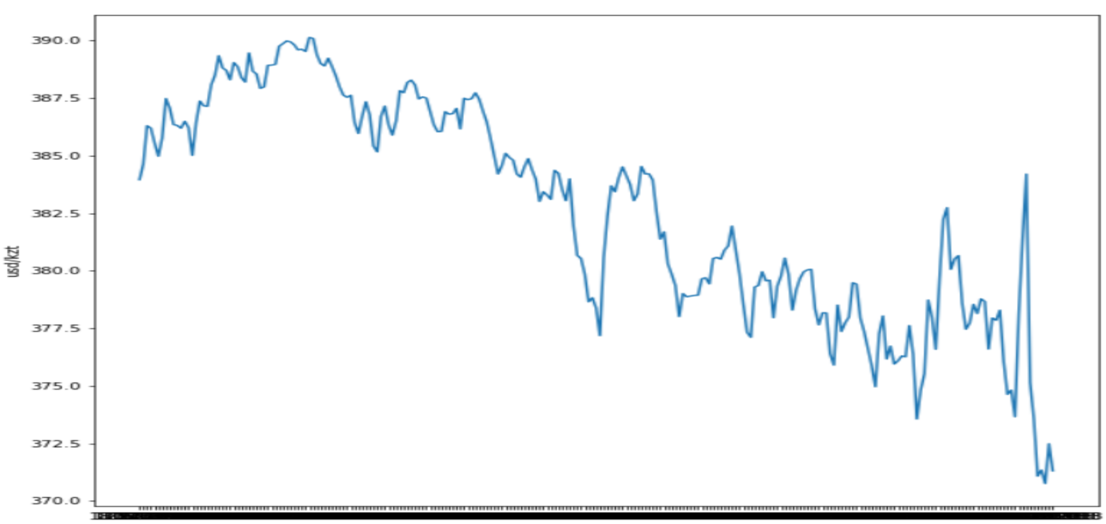


Figure 3-2: Real data

3.3 Linear regression analysis

We do linear regression to all our data. Then we predict new Y by given X values of our data. After we calculate error by subtracting exact value from predicted value.

3.4 Error analysis

From error we calculate sample mean and sample standard deviation. To define distribution of error we do Kolmogorov-Smirnov test with different distributions with same mean and standard deviation as in our error. It gives p-value equal to 0.52 with Normal distribution. Then we do histogram and qq-plot of our error :

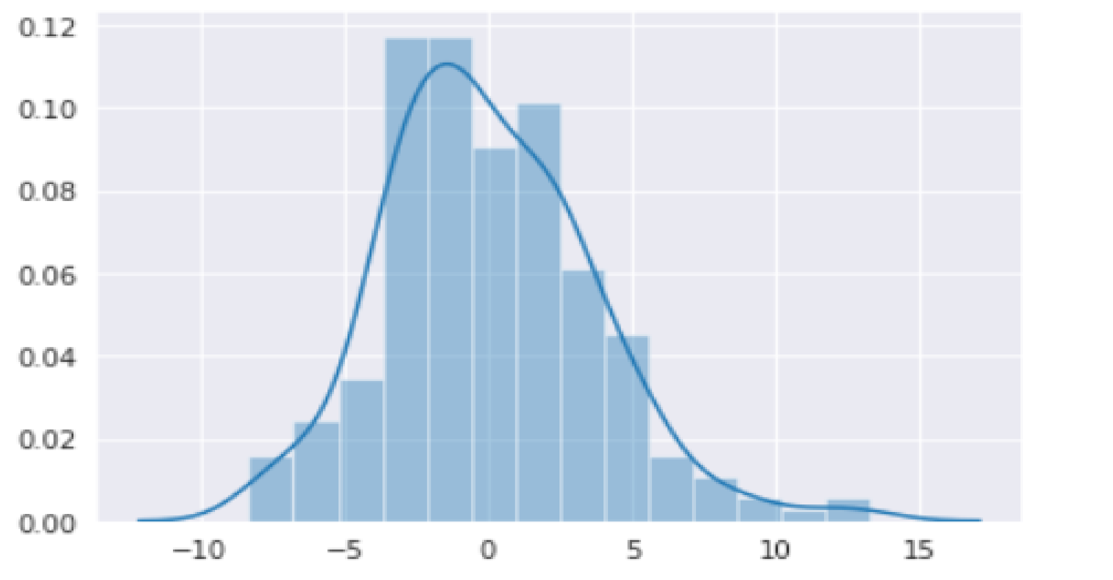


Figure 3-3: Histogram of error[Listing A.4]

It looks very similar to normal distribution.But we can see some values are not normal because plot is not exactly bell shaped.

Then, to dispel doubts, we draw qq-plot. As qq-plot is scatter plot which is plotted as one set of quantiles against second, we plot qq-plot our error against normal distribution with same mean and variance. Here we can see that most of the values are normal but in right and left sides some amount of values not normal.

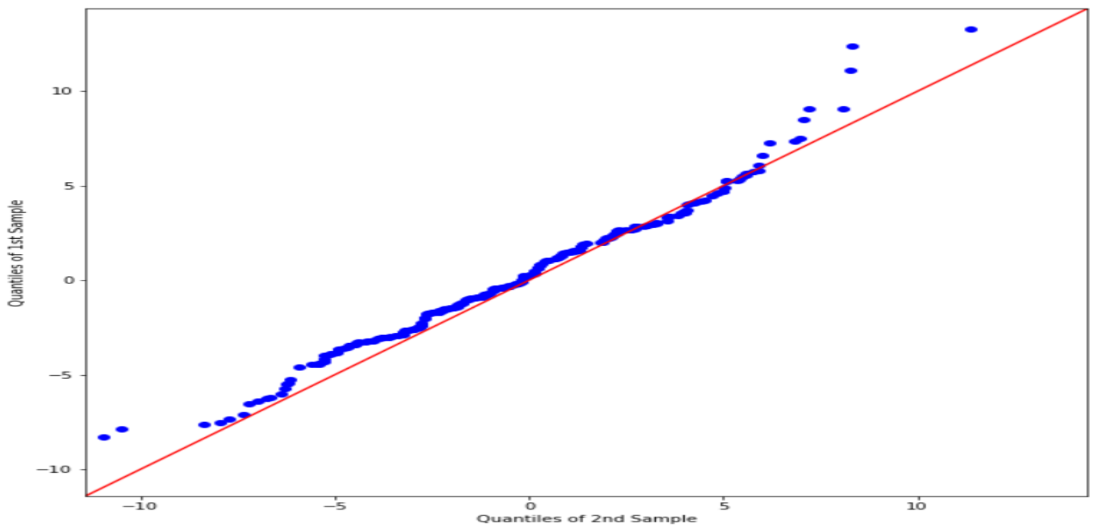


Figure 3-4: QQ-plot of error[Listing A.7]

3.5 Calculating AVaR's from error

Then we calculate AVaR's by definition, for VaR we use quantiles of normal distribution with equal mean and standard deviation with mean and standard deviation of our error. We plot bars to show amounts of AVaR at every value α.

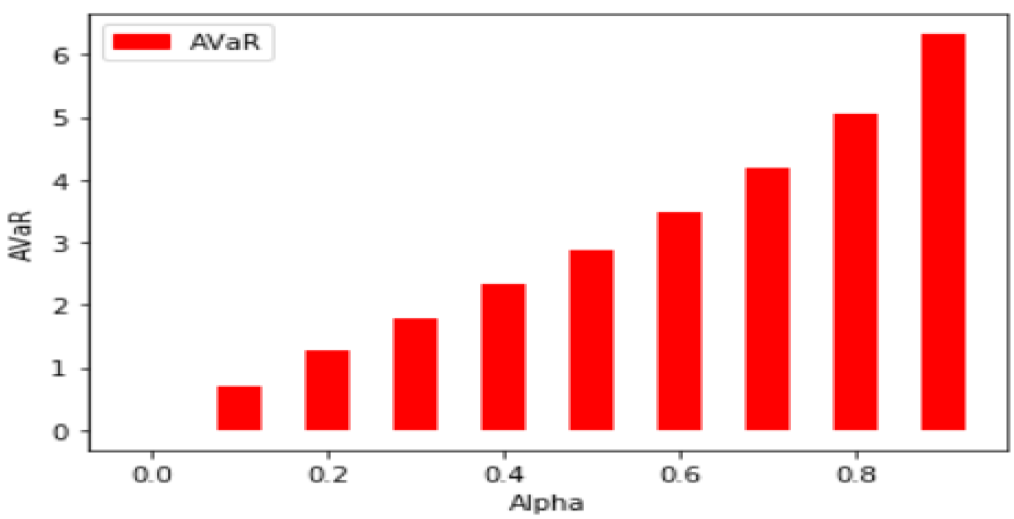


Figure 3-5: AVaRs from error[Listing A.2]

Figure shows that by rising of α values AVaR is rising too. Because by formula of AVaR, AVaR is inverse proportional to α. Then we calculate mean from last tail mean of our error such as last 10 percent then last 20 percent of our error until whole data and sort them to plot with our AVaRs. We do it to compare sorted tail means of error with AVaRs. As shown in figure it's less than AVaR but after $\alpha = 0.5$ this value become bigger than AVaR.

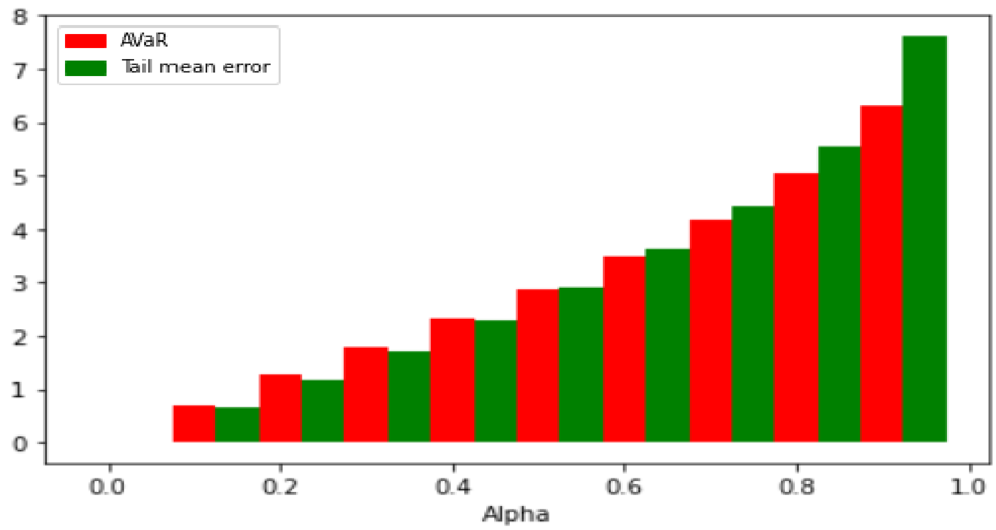


Figure 3-6: AVaRs and tail means[Listing A.8]

Then we calculate $-AVaR(-X)$ as $-X$ from error($exact - predicted$) and we see that $AVaR(X)$ and $-AVaR(-X)$ are symmetric. We do it to use AVaRs as confidence level in statistics.

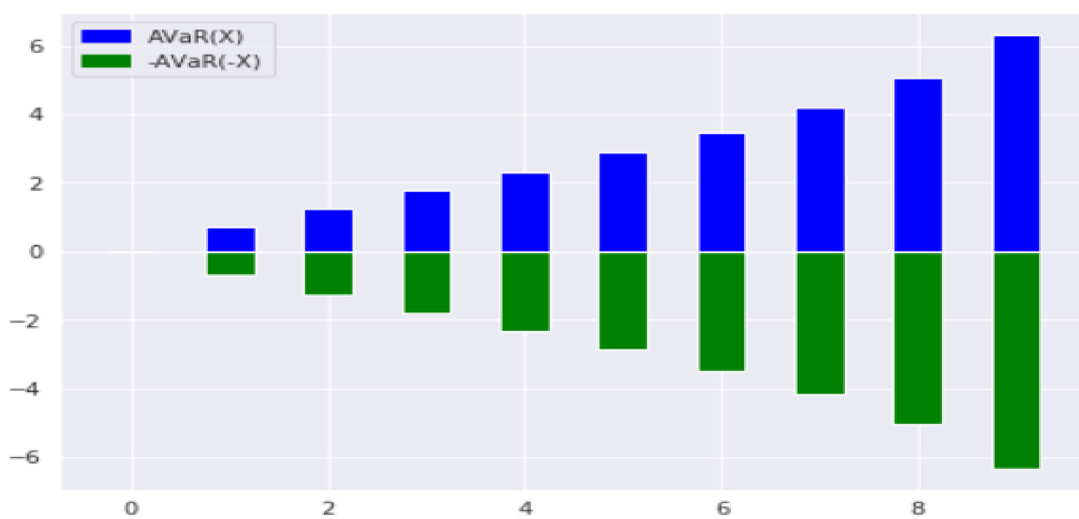


Figure 3-7: AVaR(X) and -AVaR(-X)[Listing A.3]

Then we compare our $AVaR(X)$ and $-AVaR(-X)$ values with confidence level and make plot :

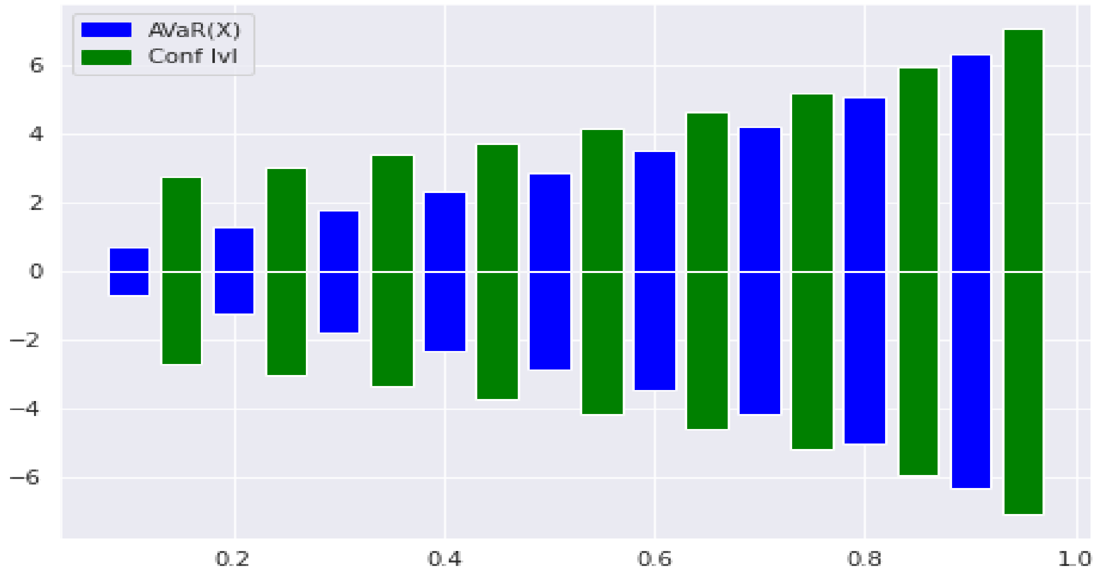


Figure 3-8: AVaR and Confidence levels[Listing A.5]

We plot $AVaR's$ by α values and confidence levels by percents. We can compare them because for positive $AVaR$, $-AVaR$ is negative values, at that time, confidence levels took negative values until 50 percent and then they become positive .

Then we compare coherent risk measure with entropic risk measure to show which will give more better. We calculate entropic risk measure with formula

$$\rho^{ent}(X) = \frac{1}{\theta} log(\mathbb{E}[e^{\theta X}]) = \sup\left\{\mathbb{E}^{Q}[X] - \frac{1}{\theta}H(Q|\mathbb{P})\right\}$$

Figure is done for equal $\theta - \alpha$ values to compare them, and we see $AVaR$ value is greater than entropic risk measure value only when $\theta - \alpha = 0.1$ and $\theta - \alpha = 0.9$. Then we compare $AVaR(X)$ and $-AVaR(-X)$ with $ERM(X)$ and $-ERM(-X)$ to show how it will be for uses as confidence level.

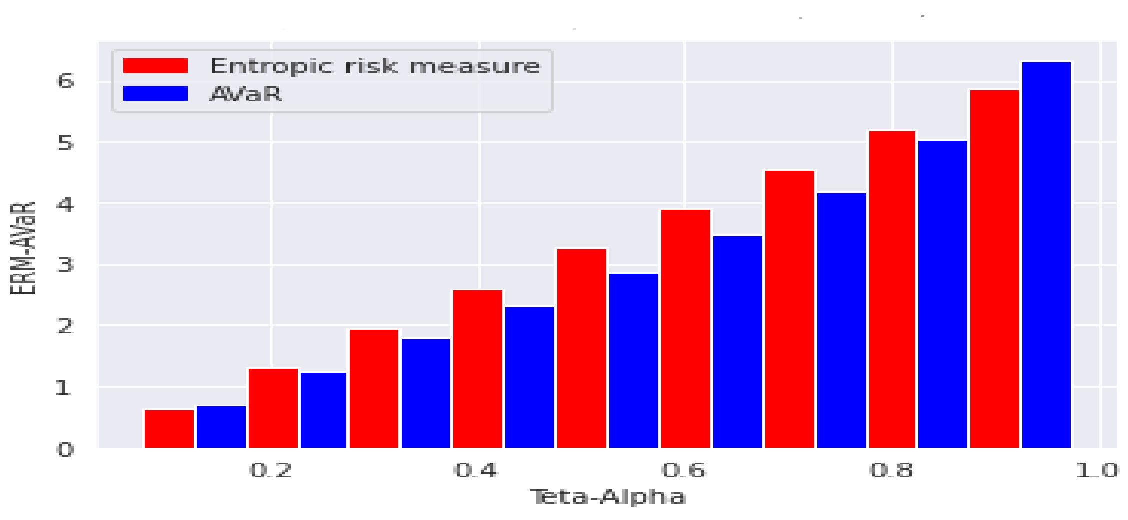


Figure 3-9: Coherent risk measure and Entropic risk measure[Listing A.6]

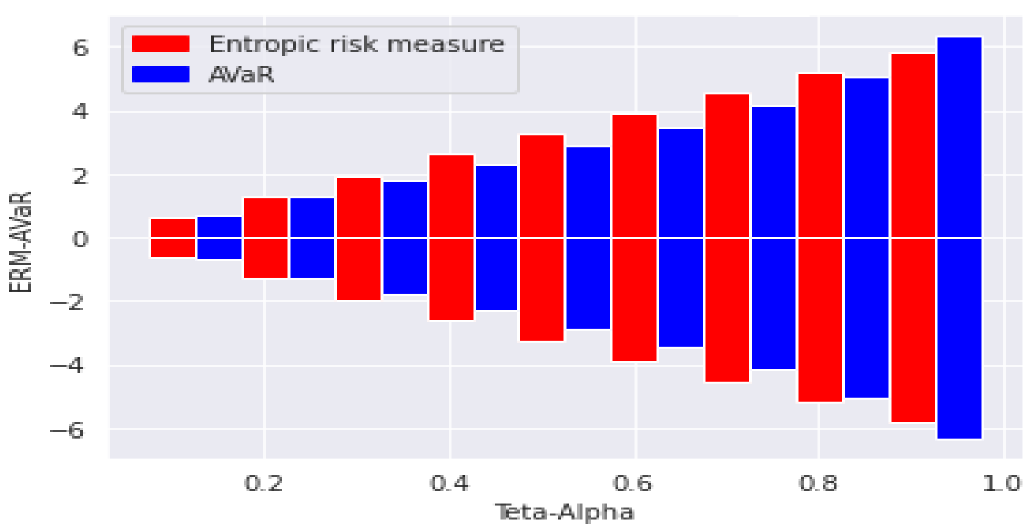


Figure 3-10: AVaR and ERM with positive/negative signs[Listing A.10]

Conclusion

The main purpose of this thesis is make prediction by risk measures. Main tool was coherent risk measure - AVaR. To achieve this,first we collected data of USD/KZT and main affecting factors from investfunds.kz and did linear regression analysis. Second, we calculated sample mean and sample standard deviation of prediction errors. Using Kolmogorov-Smirnov test, we concluded that errors are coming from normal distribuiton. Third, we calulated AVaR's and negative values of AVaR's to plot them with confidence levels. As confidence levels, AVaR's with different α values gave best intervals for errors. Every interval between AVaR and minus AVaR with different α values contains α*10 percent of errors. We also used entropic risk measure instead of AVaR and ploted AVaR confidence levels and entropic risk measure confidence levels.

Appendix A

Code

```
import numpy as np
from scipy.stats import chi2
import matplotlib.pyplot as plt
import scipy
import pandas as pd
from sklearn.linear_model import LinearRegression
import math
import random
from scipy.stats import norm
from scipy.stats import chi2
from scipy.stats import weibull_min
import seaborn as sns
import scipy.integrate as integrate
from statsmodels.graphics.gofplots import qqplot_2samples
df=pd.read_excel('data.xls')
y_1=df['kzt']
y_exact_1=np.array(y_1)
x_1=df[['brent','rur']]
reg_1 = LinearRegression().fit(x_1, y_1)
y_pred_1=reg_1.predict(x_1)
error_1=np.array (y_pred_1-y_exact_1)
std_1=np.std(error_1)
mean_1=np.mean(error_1)
def VarNorm(alpha,mean,stddev):
```

```
  s = scipy.stats.norm.ppf(alpha,mean,stddev)
  return s
def AVaR(k,mean,std):
  result = (1/ (1- k))*float(integrate.quad(lambda x: VarNorm(x,mean
    ,std),  k,1)[0])
  return result
```

Listing A.1: AVAR

```
X = np.arange(0,1,0.1)
data_avar = [[AVaR(0,mean_1,std_1),AVaR(0.1,mean_1,std_1), AVaR(0.2,
    mean_1,std_1), AVaR(0.3,mean_1,std_1), AVaR(0.4,mean_1,std_1),
    AVaR(0.5,mean_1,std_1), AVaR(0.6,mean_1,std_1), AVaR(0.7,mean_1,
    std_1), AVaR(0.8,mean_1,std_1), AVaR(0.9,mean_1,std_1)]]


plt.bar(X , data_avar[0], color = 'red', width = 0.05)

colors = {'AVaR':'red'}
labels = list(colors.keys())
handles = [plt.Rectangle((0,0),1,1, color=colors[label]) for label
    in labels]
plt.legend(handles, labels)
plt.xlabel('Alpha')
plt.ylabel('AVaR')
plt.title('AVaR whole data' )
```

Listing A.2: Calculation of AVaR

```
error_2=np.array (y_exact_1-y_pred_1)
std_2=np.std(error_2)
mean_2=np.mean(error_2)


X = np.arange(10
              )
data = -1 * np.array([[AVaR(0,mean_2,std_2),AVaR(0.1,mean_2,std_2),
    AVaR(0.2,mean_2,std_2),
```

```python
        AVaR(0.3,mean_2,std_2), AVaR(0.4,mean_2,std_2), AVaR(0.5,
    mean_2,std_2),
        AVaR(0.6,mean_2,std_2),
        AVaR(0.7,mean_2,std_2), AVaR(0.8,mean_2,std_2), AVaR(0.9,
    mean_2,std_2)]])
data1 = -1 * data
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.bar(X + 0.00, data1[0], color = 'b', width = 0.5)
ax.bar(X , data[0], color = 'g', width = 0.5)
colors = {'AVaR(X)':'blue','-AVaR(-X)':'green'}
labels = list(colors.keys())
handles = [plt.Rectangle((0,0),1,1, color=colors[label]) for label
    in labels]
ax.legend(handles, labels)
```

Listing A.3: AVaR and -AVaR

```python
sns.set_style('darkgrid')
sns.distplot(error_1)
```

Listing A.4: Histogram

```python
data = np.array([[AVaR(0.1,mean_2,std_2), AVaR(0.2,mean_2,std_2),
        AVaR(0.3,mean_2,std_2), AVaR(0.4,mean_2,std_2), AVaR(0.5,
    mean_2,std_2),
        AVaR(0.6,mean_2,std_2),
        AVaR(0.7,mean_2,std_2), AVaR(0.8,mean_2,std_2), AVaR(0.9,
    mean_2,std_2)]])
data1=np.multiply(data,-1)
data2=[[VarNorm(0.775,0,1)*std_2+mean_2,VarNorm(0.8,0,1)*std_2+
    mean_2,
        VarNorm(0.825,0,1)*std_2+mean_2,
        VarNorm(0.85,0,1)*std_2+mean_2,VarNorm(0.875,0,1)*std_2+
    mean_2,
        VarNorm(0.9,0,1)*std_2+mean_2,VarNorm(0.925,0,1)*std_2+
    mean_2,
```

```
        VarNorm(0.95,0,1)*std_2+mean_2,VarNorm(0.975,0,1)*std_2+mean_2]]
data3=np.multiply(data2,-1)
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.bar(X + 0.05, data2[0], color = 'g', width = 0.04)
ax.bar(X , data[0], color = 'b', width = 0.04)
ax.bar(X + 0.05, data3[0], color = 'g', width = 0.04)
ax.bar(X , data1[0], color = 'b', width = 0.04)
colors = {'AVaR(X)':'blue','Conf lvl':'green'}
labels = list(colors.keys())
handles = [plt.Rectangle((0,0),1,1, color=colors[label]) for label in labels]
ax.legend(handles, labels)
```

Listing A.5: AVaR and Conf lvl

```
def ERM_hand(teta,sigma):
  e=1/teta*np.log(np.exp(teta*teta*sigma*sigma/2))
  return e
 b=std_normal
X = np.arange(0.1,1,0.1)
data = [[ERM_hand(0.1,b),ERM_hand(0.2,b),ERM_hand(0.3,b),ERM_hand(0.4,b),ERM_hand(0.5,b),ERM_hand(0.6,b),ERM_hand(0.7,b),ERM_hand(0.8,b),ERM_hand(0.9,b)]]
data2 = [[AVaR(0.1,mean_normal,std_normal), AVaR(0.2,mean_normal,std_normal), AVaR(0.3,mean_normal,std_normal),
        AVaR(0.4,mean_normal,std_normal), AVaR(0.5,mean_normal,std_normal), AVaR(0.6,mean_normal,std_normal), AVaR(0.7,mean_normal,std_normal),
        AVaR(0.8,mean_normal,std_normal), AVaR(0.9,mean_normal,std_normal)]]
plt.bar(X , data[0], color = 'red', width = 0.05)
plt.bar(X+0.05 , data2[0], color = 'blue', width = 0.05)
colors = {'Entropic risk measure':'red','AVaR':'blue'}
labels = list(colors.keys())
handles = [plt.Rectangle((0,0),1,1, color=colors[label]) for label
```

```
    in labels]
plt.legend(handles, labels)
plt.xlabel('Teta-Alpha')
plt.ylabel('ERM-AVaR')
plt.title('ERM with Standard Normal dist and AVaR' )
```

Listing A.6: ERM with Standard Normal dist and AVaR

```
s=np.random.normal(mean_1,std_1,243)
fig = qqplot_2samples(error_1, s,line='45')
plt.show()
```

Listing A.7: QQ-plot

```
mean_100=np.mean(error_1)
mean_10=np.mean(error_1[225:])
mean_20=np.mean(error_1[200:])
mean_30=np.mean(error_1[175:])
mean_40=np.mean(error_1[150:])
mean_50=np.mean(error_1[125:])
mean_60=np.mean(error_1[100:])
mean_70=np.mean(error_1[75:])
mean_80=np.mean(error_1[50:])
mean_90=np.mean(error_1[25:])
data_tail=[[mean_100,mean_90,mean_80,mean_70,mean_60,mean_50,mean_40
    ,mean_30,mean_20,mean_10]]
plt.bar(X , data_avar[0], color = 'red', width = 0.05)
plt.bar(X+0.05 , data_tail[0], color = 'blue', width = 0.05)
colors = {'AVaR':'red','Tail_mean':'blue'}
labels = list(colors.keys())
handles = [plt.Rectangle((0,0),1,1, color=colors[label]) for label
    in labels]
plt.legend(handles, labels)
plt.xlabel('Tail percent-Alpha')
plt.ylabel('AVaR-Tail_mean')
plt.title('Tail mean and AVaR' )
```

Listing A.8: Tail mean and AVaR

```
def VarChi2(x,y):
  s = scipy.stats.chi2.ppf(x, df=y)
  return s
def AvarChi2(t,y):
  result = (1/ (1- t))*float(integrate.quad(lambda x: VarChi2(x,y), t, 1)[0])
  return result
k = np.arange(0.1, 1, 0.1)
y=[]
for j in range(9):
  for i in range(9):
    y.append(AvarChi2((i+1)/10,j+1))
plt.plot(k,y[:9],'g', label='df=1')
plt.plot(k,y[9:18],'r', label='df=2')
plt.plot(k,y[18:27],'blue', label='df=3')
plt.plot(k,y[27:36],'yellow', label='df=4')
plt.plot(k,y[36:45],'violet', label='df=5')
plt.plot(k,y[45:54],'brown', label='df=6')
plt.plot(k,y[54:63],'purple', label='df=7')
plt.plot(k,y[63:72],'chocolate', label='df=8')
plt.plot(k,y[72:],'coral', label='df=9')
plt.legend()
plt.xlabel('alpha',fontsize=20)
plt.ylabel('AVaR at alpha',fontsize=20)
plt.title('Chi2',fontsize=30)
plt.show()
def VarWei(x,y):
  s = scipy.stats.weibull_min.ppf(x,y)
  return s
def AvarWei(t,y):
  AVaR = (1/ (1- t))*float(integrate.quad(lambda x: VarWei(x,y), t, 1)[0])
  return AVaR
plt.figure()
k = np.arange(0.1, 1., 0.1)
y=[]
```

```
for j in range(9):
  for i in range(9):
    y.append(AvarWei((i+1)/10,j+1))
plt.plot(k,y[:9],'g', label='k=1')
plt.plot(k,y[9:18],'r', label='k=2')
plt.plot(k,y[18:27],'blue', label='k=3')
plt.plot(k,y[27:36],'yellow', label='k=4')
plt.plot(k,y[36:45],'violet', label='k=5')
plt.plot(k,y[45:54],'brown', label='k=6')
plt.plot(k,y[54:63],'purple', label='k=7')
plt.plot(k,y[63:72],'chocolate', label='k=8')
plt.plot(k,y[72:],'coral', label='k=9')
plt.legend()
plt.xlabel('alpha',fontsize=20)
plt.ylabel('AVaR at alpha',fontsize=20)
plt.title('Weibull',fontsize=30)
plt.show()
def VarStudent(x,y,z):
  s = scipy.stats.nct.ppf(x,y,z)
  return s
def AvarStudent(t,y):
  AVaR = (1/ (1- t))*float(integrate.quad(lambda x: VarStudent(x,y
   ,0.240450313312), t, 1)[0])
  return AVaR
plt.figure()
k = np.arange(0.1, 1., 0.1)
y=[]
for j in range(9):
  for i in range(9):
    y.append(AvarStudent((i+1)/10,j+1))
plt.plot(k,y[:9],'g', label='k=1')
plt.plot(k,y[9:18],'r', label='k=2')
plt.plot(k,y[18:27],'blue', label='k=3')
plt.plot(k,y[27:36],'yellow', label='k=4')
plt.plot(k,y[36:45],'violet', label='k=5')
plt.plot(k,y[45:54],'brown', label='k=6')
```

```python
plt.plot(k,y[54:63],'purple', label='k=7')
plt.plot(k,y[63:72],'chocolate', label='k=8')
plt.plot(k,y[72:],'coral', label='k=9')
plt.legend()
plt.xlabel('alpha',fontsize=20)
plt.ylabel('AVaR at alpha',fontsize=20)
plt.title('Student-T',fontsize=30)
plt.show()
def VarExp(x,y):
  s = scipy.stats.expon.ppf(x,scale=1/y)
  return s
def AvarExp(t,y):
  AVaR = (1/ (1- t))*float(integrate.quad(lambda x: VarExp(x,y), t,
   1)[0])
  return AVaR
plt.figure()
k = np.arange(0.1, 1., 0.1)
y=[]
for j in range(9):
  for i in range(9):
    y.append(AvarExp((i+1)/10,j+1))
plt.plot(k,y[:9],'g', label='lambda=1')
plt.plot(k,y[9:18],'r', label='lambda=2')
plt.plot(k,y[18:27],'blue', label='lambda=3')
plt.plot(k,y[27:36],'yellow', label='lambda=4')
plt.plot(k,y[36:45],'violet', label='lambda=5')
plt.plot(k,y[45:54],'brown', label='lambda=6')
plt.plot(k,y[54:63],'purple', label='lambda=7')
plt.plot(k,y[63:72],'chocolate', label='lambda=8')
plt.plot(k,y[72:],'coral', label='lambda=9')
plt.legend()
plt.xlabel('alpha',fontsize=20)
plt.ylabel('AVaR at alpha',fontsize=20)
plt.title('Exponential',fontsize=30)
plt.show()
def VarNorm(x,y):
```

```
  s = scipy.stats.norm.ppf(x,0,y)
  return s
def AvarNorm(t,y):
  AVaR = (1/ (1- t))*float(integrate.quad(lambda x: VarNorm(x,y), t,
    1)[0])
  return AVaR
plt.figure()
k = np.arange(0.1, 1., 0.1)
y=[]
for j in range(9):
  for i in range(9):
    y.append(AvarNorm((i+1)/10,j+1))
plt.plot(k,y[:9],'g', label='std=1')
plt.plot(k,y[9:18],'r', label='std=2')
plt.plot(k,y[18:27],'blue', label='std=3')
plt.plot(k,y[27:36],'yellow', label='std=4')
plt.plot(k,y[36:45],'violet', label='std=5')
plt.plot(k,y[45:54],'brown', label='std=6')
plt.plot(k,y[54:63],'purple', label='std=7')
plt.plot(k,y[63:72],'chocolate', label='std=8')
plt.plot(k,y[72:],'coral', label='std=9')
plt.legend()
plt.xlabel('alpha',fontsize=20)
plt.ylabel('AVaR at alpha',fontsize=20)
plt.title('Gausian',fontsize=30)
plt.show()
```

Listing A.9: AVaRs of different distributions

```
b=std_normal
X = np.arange(0.1,1,0.1)
data = [[ERM_hand(0.1,b),ERM_hand(0.2,b),ERM_hand(0.3,b),ERM_hand
    (0.4,b),ERM_hand(0.5,b),ERM_hand(0.6,b),ERM_hand(0.7,b),ERM_hand
    (0.8,b),ERM_hand(0.9,b)]]
data2 = [[AVaR(0.1,mean_normal,std_normal), AVaR(0.2,mean_normal,
    std_normal), AVaR(0.3,mean_normal,std_normal),
          AVaR(0.4,mean_normal,std_normal), AVaR(0.5,mean_normal,
```

```
std_normal), AVaR(0.6,mean_normal,std_normal), AVaR(0.7,mean_normal,std_normal),
        AVaR(0.8,mean_normal,std_normal), AVaR(0.9,mean_normal,std_normal)]]
data3= [[-ERM_hand(0.1,b),-ERM_hand(0.2,b),-ERM_hand(0.3,b),-ERM_hand(0.4,b),-ERM_hand(0.5,b),-ERM_hand(0.6,b),-ERM_hand(0.7,b),-ERM_hand(0.8,b),-ERM_hand(0.9,b)]]
data4 = [[-AVaR(0.1,mean_normal,std_normal), -AVaR(0.2,mean_normal,std_normal), -AVaR(0.3,mean_normal,std_normal),
        -AVaR(0.4,mean_normal,std_normal), -AVaR(0.5,mean_normal,std_normal), -AVaR(0.6,mean_normal,std_normal), -AVaR(0.7,mean_normal,std_normal),
        -AVaR(0.8,mean_normal,std_normal), -AVaR(0.9,mean_normal,std_normal)]]

plt.bar(X , data[0], color = 'red', width = 0.05)
plt.bar(X+0.05 , data2[0], color = 'blue', width = 0.05)
plt.bar(X , data3[0], color = 'red', width = 0.05)
plt.bar(X+0.05 , data4[0], color = 'blue', width = 0.05)
colors = {'Entropic risk measure':'red','AVaR':'blue'}
labels = list(colors.keys())
handles = [plt.Rectangle((0,0),1,1, color=colors[label]) for label in labels]
plt.legend(handles, labels)
plt.xlabel('Teta-Alpha')
plt.ylabel('ERM-AVaR')
plt.title('ERM(X) and -ERM(-X) with AVaR(X) and -AVaR(-X)' )
```

Listing A.10: AVaR and ERM with positive/negative signs

Bibliography

[1] Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. Mathematical finance, 9(3), 203-228.

[2] Uğurlu, K. (2017). Controlled Markov Decision Processes with AVaR criteria for unbounded costs. Journal of Computational and Applied Mathematics, 319, 24-37.

[3] Uğurlu, K. (2018). Robust optimal control using conditional risk mappings in infinite horizon. Journal of Computational and Applied Mathematics, 344, 275-287.

[4] Ruszczyński, A., and Shapiro, A. (2006). Optimization of convex risk functions. Mathematics of operations research, 31(3), 433-452.

[5] Föllmer, H., Schied, A. (2011). Stochastic finance: an introduction in discrete time. Walter de Gruyter.

[6] Cheridito, P., Delbaen, F., and Kupper, M. (2004). Coherent and convex monetary risk measures for bounded cadlag processes. Stochastic Processes and their Applications, 112(1), 1-22.

[7] Larsen, R. J., Marx, M. L. (2013). Introduction to Mathematical Statistics and Its Applications: Pearson New International Edition. Pearson Higher Ed.

[8] Montgomery, D. C., and Runger, G. C. (2010). Applied statistics and probability for engineers. John Wiley and Sons.

[9] Weisberg, Sanford. Applied linear regression. Vol. 528. John Wiley and Sons, 2005.

[10] Holton, G. A. (2012). Value-at-Risk: Theory and Practice, e-book 2014. URL: http://value-at-risk. net.

[11] Acerbi, C., Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. Economic notes, 31(2), 379-388.

[12] Föllmer, H., and Schied, A. (2008). Convex and coherent risk measures. Encyclopedia of Quantitative Finance, 355-363.

[13] Khokhlov, V. (2018). Conditional Value-at-Risk for Log-Distributions. Available at SSRN 3197929.

[14] Rudloff, B., Sass, J., and Wunderlich, R. (2008). Entropic risk constraints for utility maximization. Festschrift in celebration of Prof. Dr. Wilfried Grecksch's 60th Birthday, 149-180.

[15] Ahmadi-Javid, Amir. "Entropic value-at-risk: A new coherent risk measure." Journal of Optimization Theory and Applications 155.3 (2012): 1105-1123.

[16] Föllmer, H., and Knispel, T. (2011). Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. Stochastics and Dynamics, 11(02n03), 333-351.

[17] Rockafellar, R. T., and Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of risk, 2, 21-42.

[18] Rockafellar, R. T., and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. Journal of banking and finance, 26(7), 1443-1471.

[19] Andersson, F., Mausser, H., Rosen, D., and Uryasev, S. (2001). Credit risk optimization with conditional value-at-risk criterion. Mathematical Programming, 89(2), 273-291.

[20] Zou, K. H., Tuncali, K., Silverman, S. G. (2003). Correlation and simple linear regression. Radiology, 227(3), 617-628.