FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Bc. Jana Junová

Quality of Stochastic Dominance Approximation Based on the Probability Distribution

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	doc. RNDr. Ing. Miloš Kopa, Ph.D.
Study programme:	Mathematics
Study branch:	Probability, Mathematical Statistics and Econometrics

Prague 2022

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I would like to thank doc. RNDr. Ing. Miloš Kopa, Ph.D. for suggesting me the interesting topic of this thesis, and for his valuable advice throughout my work. I am also grateful to my family for supporting me in various ways, which allowed me to write this work.

Title: Quality of Stochastic Dominance Approximation Based on the Probability Distribution

Author: Bc. Jana Junová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Ing. Miloš Kopa, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This work focuses on measuring the quality of stochastic dominance approximation. A measure of non-dominance is developed to quantify the error caused by assuming that a stochastic dominance relationship holds even when it does not. It is computed exactly for uniform, normal, and exponential distribution, and a numerical study is performed to estimate its values for log-normal and gamma distribution. Portfolio optimization problems involving stochastic dominance constraints are also presented. They are applied to real-life data using monthly returns of twelve assets captured by the German stock index DAX. The end of this work focuses on the computation of the measure of non-dominance for the optimal portfolio with respect to the second-order stochastic dominance.

Keywords: stochastic dominance, approximation, non-dominance, portfolio optimization

Contents

Introduction

There is a wide range of approaches to evaluation of investments under uncertainty. It is generally agreed that investors prefer investments with higher returns and lower risks. A usual measure of return is the expected value of the random variable, which represents the unknown return of an investment. It is, however, unclear what the right measure of risk is. Moreover, every investor may have a different risk attitude so it is impossible to state the maximum acceptable level of risk in general.

Stochastic dominance is a concept that allows the comparison of investment opportunities without the precise knowledge of a particular investor's preferences. It accepts the fact that each investor may have a different utility function and works with whole classes of them. It allows us to compare investments under the only assumption that the correct utility function is in a particular class of them. The widest class includes all non-decreasing and continuous utility functions.

We follow Levy [2006] in introducing stochastic dominance of the first, the second, and the third order and the related theory. Levy [2006] briefly mentions also decreasing absolute risk aversion stochastic dominance which Post et al. [2015] amongst others further explore. Whitmore [1989] is concerned with the infinite-order stochastic dominance. We use the results regarding stochastic dominance between variables with particular distributions presented in Ali [1975], Levy [2006], and Mikulka [2011].

To define a measure that assesses the quality of stochastic dominance approximation, we need to find an appropriate measure of distance between two random variables. We follow Kozmík [2019] in using the Wasserstein distance for it. Its properties presented in Pflug and Pichler [2014] are applied in further computations. We assess the quality of stochastic dominance approximation also in practical portfolio optimization problems. We use the formulations of these problems presented in Kuosmanen [2004] and Dentcheva and Ruszczynski [2006].

This work is structured in the following way. The general theory regarding stochastic dominance is explained in the first chapter. We define the first-order, the second-order, the third-order, decreasing absolute risk aversion, and the infinite-order stochastic dominance by imposing assumptions on the utility functions. When available, we provide alternative definitions of the described types of stochastic dominance using the distribution functions of the compared random variables.

We define a measure of non-dominance in the second chapter, and we explore its general properties. We use it to assess the quality of stochastic dominance approximation based on the probability distribution. We derive its values exactly for uniform, normal, and exponential distribution in the third chapter, and we present its estimations for log-normal and gamma distribution in Chapter 4.

Chapter 5 is concerned with its application in portfolio optimization problems. Portfolio optimization problems involving the first-order and the second-order stochastic dominance constraints are presented. Two approaches to finding the closest dominating portfolios are then introduced. We apply this theory to real-life data regarding the monthly returns of twelve assets from the German stock index DAX.

1. Stochastic Dominance

Stochastic dominance compares investments using utility functions. Utility function assigns utility, expressed by a number, to a given wealth. We assume that it is non-decreasing and continuous. Investments with higher expected utility are preferred over those with lower expected utility. We generally do not know the utility function of a particular investor exactly, and stochastic dominance allows us to work with whole classes of them. Each type of stochastic dominance imposes different assumptions on the utility functions. They will be described and compared in the following sections.

The return of an investment is usually unknown and can be described as a random variable. We use the following notation regarding random variables. The distribution function of a random variable X is $F_X(x) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$. The integrated distribution function of a random variable X is $F_X^{(2)}(x) = \int_{-\infty}^{x} F_X(t)dt$, $x \in \mathbb{R}$. The twice integrated distribution function of a random variable X is $F_X^{(3)}(x) = \int_{-\infty}^{x} F_X^{(2)}(t)dt$, $x \in \mathbb{R}$. Definitions and theorems in this chapter follow Levy [2006] unless stated differently, and proofs can be found there as well.

1.1 First-Order Stochastic Dominance

A very general assumption regarding utility functions u is that $u' \geq 0$, which ensures that the utility function is non-decreasing. We define a set of all utility functions that satisfy this condition: $U_1 = \{u \text{ utility function}, u' \geq 0\}$. The first-order stochastic dominance (FSD) aims to compare the expected utility of random variables assuming the utility function is in U_1.

Definition 1.1 (First-Order Stochastic Dominance). *We say that a random variable X dominates a random variable Y by the first-order stochastic dominance ($X \succeq_{(1)} Y$) if*

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \text{ for all } u \in U_1 \text{ such that these expected values exist.}$$

We say that X dominates Y by the first-order stochastic dominance strictly ($X \succ_{(1)} Y$) if

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \text{ for all } u \in U_1 \text{ such that these expected values exist}$$
$$\text{with a strict inequality holding for at least one } u_0 \in U_1.$$

FSD can be equivalently described using the distribution functions of the random variables.

Theorem 1.1 (Alternative Definition of FSD). *$X \succeq_{(1)} Y$ if and only if $F_X(x) \leq F_Y(x)$ for all $x \in \mathbb{R}$. $X \succ_{(1)} Y$ if and only if $F_X(x) \leq F_Y(x)$ for all $x \in \mathbb{R}$, and there is at least one x for which the inequality is strict.*

We are going to apply this theory to discrete random variables with equiprobable atoms in the fourth and fifth chapter. A simple equivalent rule for FSD in such distribution exists so we present it here. Suppose X and Y are discretely distributed random variables with equiprobable atoms, which we also call empirically distributed random variables. The following notation will be used for discrete random variables throughout this work unless specified differently. A random variable X attains values $x_1 \leq \cdots \leq x_T$ with probabilities $1/T$, and a random variable Y attains values $y_1 \leq \cdots \leq y_T$ with probabilities $1/T$. Then Theorem 1.2 follows from Theorem 1.1.

Theorem 1.2 (FSD for Empirical Distributions)**.**

$$X \succeq_{(1)} Y \iff x_t \geq y_t,\ t = 1, \ldots, T.$$

$$X \succ_{(1)} Y \iff x_t \geq y_t,\ t = 1, \ldots, T,$$

and there is a t for which the inequality is strict.

1.2 Second-Order Stochastic Dominance

The second-order stochastic dominance (SSD) assumes a narrower set of utility functions: $U_2 = \{u \text{ utility function}, u' \geq 0, u'' \leq 0\}$. These assumptions ensure that the considered utility functions are increasing and concave. It is a reasonable assumption because it corresponds to the assumption that an investor is risk-averse.

An investor is risk-averse when the following holds for his utility function u, his current wealth w and a random variable X, which represents an investment's return: $\mathbb{E}\ u(w + X) < u(w + \mathbb{E}\ X)$. Let $W \subseteq \mathbb{R}$. If an investor is risk-averse at every level of wealth $w \in W$, he is globally risk-averse on W and his utility function is concave on W. Second-order stochastic dominance is defined as follows.

Definition 1.2 (Second-Order Stochastic Dominance)**.** *We say that a random variable X dominates a random variable Y by the second-order stochastic dominance ($X \succeq_{(2)} Y$) if*

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \text{ for all } u \in U_2 \text{ such that these expected values exist.}$$

We say that X dominates Y by the second-order stochastic dominance strictly ($X \succ_{(2)} Y$) if

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \text{ for all } u \in U_2 \text{ such that these expected values exist}$$

$$\text{with a strict inequality holding for at least one } u_0 \in U_2.$$

SSD can be equivalently described using cumulative distribution functions of the random variables.

Theorem 1.3 (Alternative Definition of SSD)**.** *$X \succeq_{(2)} Y$ if and only if $F_X^{(2)}(x) \leq F_Y^{(2)}(x)$ for all $x \in \mathbb{R}$. $X \succ_{(2)} Y$ if and only if $F_X^{(2)}(x) \leq F_Y^{(2)}(x)$ for all $x \in \mathbb{R}$, and there is at least one x for which the inequality is strict.*

It is possible to formulate equivalent conditions for SSD for the special case of discrete random variables with equiprobable atoms.

Theorem 1.4 (SSD for Empirical Distributions)**.**

$$X \succeq_{(2)} Y \Longleftrightarrow \sum_{i=1}^{t} x_i \geq \sum_{i=1}^{t} y_i, t = 1, \ldots, T.$$

$$X \succ_{(2)} Y \Longleftrightarrow \sum_{i=1}^{t} x_i \geq \sum_{i=1}^{t} y_i, t = 1, \ldots, T,$$

and there is a t for which the inequality is strict.

The following relationship between the first-order and the second-order stochastic dominance holds.

Theorem 1.5. *Let X and Y be random variables. Then $X \succeq_{(1)} Y \Longrightarrow X \succeq_{(2)} Y$.*

It can be seen from the definition because $U_2 \subset U_1$. Therefore, if $\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y)$ for all $u \in U_1$, then the inequality holds also for all $u \in U_2$.

1.3 Third-Order Stochastic Dominance

The third-order stochastic dominance (TSD) assumes an even narrower set of utility functions $U_3 = \{u \text{ utility function}, u' \geq 0, u'' \leq 0, u''' \geq 0\}$. Justifying this restriction is slightly more complicated than the previous ones. We will show that it corresponds to the situation when investors prefer to be exposed to the possibility of unlikely but high returns over unlikely but high losses. This means that they prefer positive skewness. We now provide a definition of skewness and an explanation of the relationship between preference for positive skewness and positive third derivative of utility function similarly to Levy [2006].

Definition 1.3 (Skewness)**.** *Skewness of a distribution of a random variable X is defined as*

$$\mu_3 = \mathbb{E}\,(X - \mathbb{E}\ X)^3.$$

Suppose that an investor's wealth is w. For an investment X, and a utility function u, the function $u(w+X)$ can be approximated based on Taylor expansion as follows:

$$u(w+X) \approx u(w+\mathbb{E}\ X) + u'(w+\mathbb{E}\ X)(X-\mathbb{E}\ X) + \frac{u''(w+\mathbb{E}\ X)}{2}(X-\mathbb{E}\ X)^2 + \frac{u'''(w+\mathbb{E}\ X)}{3!}(X-\mathbb{E}\ X)^3.$$

By taking the expected value from both sides and denoting the variance of X as σ^2 we receive the following equation.

$$\mathbb{E}\ u(w+X) \approx u(w+\mathbb{E}\ X) + \frac{u''(w+\mathbb{E}\ X)}{2}\sigma^2 + \frac{u'''(w+\mathbb{E}\ x)}{3!}\mu_3.$$

Assuming a constant σ^2 and a preference for positive skewness, the overall expected utility should increase with increasing μ_3. Therefore, we need $u''' \geq 0$. This provides an explanation to the definition of U_3 and leads to the definition of the third-order stochastic dominance.

Definition 1.4 (Third-Order Stochastic Dominance)**.** *We say that a random variable X dominates a random variable Y by the third-order stochastic dominance ($X \succeq_{(3)} Y$) if*

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \textit{ for all } u \in U_3 \textit{ such that these expected values exist.}$$

We say that X dominates Y by the third-order stochastic dominance strictly ($X \succ_{(3)} Y$) if

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \textit{ for all } u \in U_3 \textit{ such that these expected values exist}$$

$$\textit{with a strict inequality for at least one } u_0 \in U_3.$$

TSD can be equivalently described also by the twice integrated cumulative distribution functions of the random variables.

Theorem 1.6 (Alternative Definition of TSD)**.** $X \succeq_{(3)} Y$ *if and only if*

$$F_X^{(3)}(x) \leq F_Y^{(3)}(x) \textit{ for all } x \in \mathbb{R},$$

and

$$\mathbb{E}\ X \geq \mathbb{E}\ Y.$$

$X \succ_{(3)} Y$ *if and only if*

$$F_X^{(3)}(x) \leq F_Y^{(3)}(x) \textit{ for all } x \in \mathbb{R},$$

and

$$\mathbb{E}\ X \geq \mathbb{E}\ Y,$$

and there is at least one strict inequality.

The following relationship between the second-order and the third-order stochastic dominance holds.

Theorem 1.7. *Let X and Y be random variables. Then* $X \succeq_{(2)} Y \implies X \succeq_{(3)} Y$.

Similarly as in Theorem 1.5, it follows from the fact that $U_3 \subset U_2$. A natural consequence of the theorem above and Theorem 1.5 is that $X \succeq_{(1)} Y \implies X \succeq_{(3)} Y$.

The rules for FSD and SSD in empirically distributed variables presented in Theorem 1.2 and Theorem 1.4 cannot be directly extended for the third-order stochastic dominance. The reasons for it are precisely explained in Levy [2006]. Briefly described, it is caused by the fact that $F_Y^{(3)}(x) - F_X^{(3)}(x)$ is not linear on each interval between neighbouring x_i and x_{i+1} or y_j. Therefore, it is not sufficient to compare the two considered distributions only at the points $x_1, \ldots, x_T, y_1, \ldots, y_T$. Levy [2006] suggests a more complicated algorithm for verifying TSD relationship between two empirically distributed random variables, which is based on checking the positiveness of $F_Y^{(3)}(x) - F_X^{(3)}(x)$ at points $x_1, \ldots, x_T, y_1, \ldots, y_T$ as well as at certain points between them.

1.4 Decreasing Absolute Risk Aversion Stochastic Dominance

Decreasing absolute risk aversion stochastic dominance (DARA SD) assumes that risk aversion of an investor decreases as his wealth increases. Following Arrow [1965] and Pratt [1964], risk aversion is measured by the Arrow-Pratt absolute risk aversion function which is defined as

$$r(w) = -\frac{u''(w)}{u'(w)},$$

where $w \in W$ is the wealth of an investor.

It is related to risk premium $\pi(X)$, which was defined by Pratt [1964]:

$$u(w + \mathbb{E}\ X - \pi(X)) = \mathbb{E}\ u(w + X).$$

It represents the amount of money that an investor is willing to pay to avoid undertaking the risk of an investment whose returns are represented by X. It was shown in Pratt [1964] that $\pi(X)$ can be approximated using the absolute risk aversion function

$$\pi(X) \approx 1/2 \cdot \sigma^2 \cdot r(w + \mathbb{E}\ X),$$

where σ^2 is the variance of X.

It is reasonable to assume that many investors are less afraid of losing some amount of money when they are richer. Hence, they are less motivated to pay risk premium to avoid undertaking the risk of an investment. So, the risk premium decreases with increasing wealth and so does the absolute risk aversion function. It is therefore reasonable to define decreasing absolute risk aversion stochastic dominance.

Following Post et al. [2015], we define the set of utility functions satisfying decreasing absolute risk aversion as $U_D = \{u \in U_3, u' > 0, r' \leq 0\}$. The condition $u' > 0$ is necessary to ensure that $r(w)$ is well defined. The conditions $r' \leq 0$ and $u' > 0$ imply that $u''' \geq 0$ because $u'''(w) \cdot u'(w)$ has to be positive to allow the condition below to be fulfilled:

$$0 \geq r'(w) = \left(-\frac{u''(w)}{u'(w)}\right)' = \frac{-u'''(w) \cdot u'(w) + (u''(w))^2}{(u'(w))^2}.$$

We should note that the condition $r' \leq 0$ in fact ensures non-increasing absolute risk aversion. Nevertheless, the term decreasing absolute risk aversion is usually used. Decreasing absolute risk aversion stochastic dominance is defined as follows.

Definition 1.5 (Decreasing Absolute Risk Aversion Stochastic Dominance)**.** *We say that a random variable X dominates a random variable Y by decreasing absolute risk aversion stochastic dominance ($X \succeq_{(D)} Y$) if*

$$\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y) \text{ for all } u \in U_D \text{ such that these expected values exist.}$$

The definition of U_D is not as straight-forward as in the previous cases. Unfortunately, describing it by imposing rules directly on the shape of utility functions results in a complicated condition which is described in Levy [2006]. For arbitrary constants c_1, c_2, it is following:

$$u \in U_D \iff u(w) = \int_{-\infty}^{w} e^{-\int_{-\infty}^{z} r(v)dv + c_1} dz + c_2.$$

We do not know of an alternative definition of DARA SD using the distribution functions of the random variables.

A sufficient condition for DARA SD is clear directly from the definition.

Theorem 1.8. *Let X and Y be random variables. Then $X \succeq_{(3)} Y \implies X \succeq_{(D)} Y$.*

The fact that TSD is a necessary condition for DARA SD gives an important application to TSD, which is easier to prove than DARA SD.

Moreover, under the assumption of equal means, TSD and DARA SD coincide.

Theorem 1.9. *Let X and Y be random variables satisfying $\mathbb{E}\, X = \mathbb{E}\, Y$. Then $X \succeq_{(3)} Y \iff X \succeq_{(D)} Y$.*

We have seen a sufficient rule for DARA SD, and under the special condition of equal means also a necessary and sufficient rule. We, however, do not know of an applicable general necessary and sufficient rule. Therefore, we introduce also a general necessary rule for DARA SD, which is infinite-order stochastic dominance presented in the following section.

1.5 Infinite-Order Stochastic Dominance

The infinite-order stochastic dominance (ISD) assumes the narrowest set of utility functions presented so far. We define $U_\infty = \{u \text{ utility function}, (-1)^{n-1} \cdot u^{(n)} \geq 0, n \in \mathbb{N}\}$ according to Whitmore [1989], who calls such utility functions completely monotonic. This class of completely monotonic functions contains important and widely used utility functions such as $\log(x)$, $\frac{x^\alpha}{\alpha}$ for $\alpha < 1$, or $1 - e^{-\alpha x}$ for $\alpha > 0$. Yet, a precise economical justification of these conditions on utility functions is not as clear as in the previous cases. Nevertheless, we define the infinite-order stochastic dominance following Whitmore [1989].

Definition 1.6 (Infinite-Order Stochastic Dominance)**.** *We say that a random variable X dominates a random variable Y by infinite-order stochastic dominance ($X \succeq_{(\infty)} Y$) if*

$$\mathbb{E}\, u(X) \geq \mathbb{E}\, u(Y) \textit{ for all } u \in U_\infty \textit{ such that these expected values exist.}$$

Unlike for DARA SD, there is a useful equivalent rule for ISD. It was proved in Whitmore [1989].

Theorem 1.10 (Alternative Definition of ISD)**.** *$X \succeq_{(\infty)} Y$ if and only if $\mathbb{E}\, e^{-aX} \geq \mathbb{E}\, e^{-aY}$ for all $a > 0$.*

Whitmore [1989] states that all $u \in U_\infty$ exhibit non-increasing risk aversion. Therefore, $U_\infty \subseteq U_D$, and the necessary condition for DARA SD follows from it.

Theorem 1.11. *Let X and Y be random variables. Then $X \succeq_{(D)} Y \implies X \succeq_{(\infty)} Y$.*

This gives an important application to ISD as it can be used to approximate DARA SD.

2. Measure of Non-Dominance

We explained in the previous chapter that $FSD \Rightarrow SSD \Rightarrow TSD \Rightarrow DARA\ SD \Rightarrow ISD$. The opposite implications generally do not hold. We are interested in quantifying how incorrect would it be to approximate the stricter types of stochastic dominance by the weaker ones.

We develop a way to assess how accurate the approximation is. Suppose $X \succeq_{(n)} Y$ does not hold[1]. Our goal is to measure how inaccurate would it be to assume that X dominates Y by nSD ($X \succeq_{(n)} Y$) even though X non-dominates Y by nSD ($X \not\succeq_{(n)} Y$). We measure how much must X change in order for it to dominate Y with respect to nSD, or, in other words, how far X is from $X \succeq_{(n)} Y$. For dist(X, Y) being a distance of the random variables X and Y, we define a measure of n$^{\text{th}}$ non-dominance (n-ND) as follows.

Definition 2.1 (Non-Dominance)**.** *Let X and Y be random variables. The measure of n^{th} non-dominance of X and Y, n-ND(X, Y), is computed by solving the following program:*

$$\begin{aligned} n\text{-}ND(X, Y) = \min_{\hat{X}} \ dist(X, \hat{X}) \\ \text{subject to } \hat{X} \succeq_{(n)} Y. \end{aligned} \tag{2.1}$$

We can see that if $X \succeq_{(n)} Y$, then $\hat{X} = X$ and n-ND$(X, Y) = 0$.

When we apply the measure of non-dominance to assess the quality of approximations by higher order stochastic dominance, it is natural to assume that $X \succeq_{(n+1)} Y$. It is, however, not necessary for the definition.

2.1 Convexity of the Feasible Set

It is advantageous in solving optimization problems to have a convex feasible set. Let as define the set

$$A_n(Y) = \{X : X \succeq_{(n)} Y\}.$$

Dentcheva and Ruszczynski [2004] show by a simple example, that $A_1(Y)$ is not convex. Suppose that X_1, X_2, Y are independent identically distributed random variables with the following distribution: $P(Y = 0) = P(Y = 1) = 1/2$. Define X as a convex combination of X_1 and X_2 as follows:

$$X = \frac{X_1 + X_2}{2}.$$

Then $P(X = 0) = 1/4$, $P(X = 1/2) = 1/2$ and $P(X = 1) = 1/4$. $X_1 \succeq_{(1)} Y$, $X_2 \succeq_{(1)} Y$, but $X \not\succeq_{(1)} Y$ because $F_X(x) > F_Y(x)$ for $x \in [1/2, 1)$. Fortunately, we will see later that for some particular distributions of X and Y, the feasible set of (2.1) is convex even for 1-ND.

We will show that $A_n(Y)$ is convex for $n \geq 2$. Suppose $X_1 \succeq_{(n)} Y$, $X_2 \succeq_{(n)} Y$, and define $X = \lambda X_1 + (1 - \lambda)X_2$, $\lambda \in (0, 1)$. We want to show that $X \succeq_{(n)}$

[1]For the sake of simpler notation we mean DARA SD by $n = 4$ in this context.

Y which means that $\mathbb{E}\ u(X) \geq \mathbb{E}\ u(Y)$ for all $u \in U_n$, $n \geq 2$. The fact that $u \in U_n$, $n \geq 2$, implies that $u'' \leq 0$, which implies that u is concave. Therefore $u(\lambda X_1 + (1-\lambda)X_2) \geq \lambda u(X_1) + (1-\lambda)u(X_2)$. It follows that

$$\begin{aligned}\mathbb{E}\ u(X) &= \mathbb{E}\ u(\lambda X_1 + (1-\lambda)X_2) \geq \mathbb{E}\left(\lambda u(X_1) + (1-\lambda)u(X_2)\right) \\ &= \lambda\,\mathbb{E}\ u(X_1) + (1-\lambda)\,\mathbb{E}\ u(X_2) \geq \lambda\,\mathbb{E}\ u(Y) + (1-\lambda)\,\mathbb{E}\ u(Y) = \mathbb{E}\ u(Y)\end{aligned}$$

for all $u \in U_n$. The inequality on the first line follows from concavity of u. The inequality on the second line follows from the fact that $X_1 \succeq_{(n)} Y$ and $X_2 \succeq_{(n)} Y$. Therefore $X \succeq_{(n)} Y$, and A_n is convex for $n \geq 2$.

2.2 Measure of Distance

Following Kozmík [2019], we use the Wasserstein distance as the measure of distance of two random variables. We provide its definition from Pflug and Pichler [2014]. To measure the distance of random variables, we measure the distance of the probability measures induced by them.

Definition 2.2 (Wasserstein Distance)**.** *Let there be two probability spaces* $(\Omega_1, \mathcal{F}_1, P_1)$ *and* $(\Omega_2, \mathcal{F}_2, P_2)$*. The Wasserstein distance of order* $r \geq 1$ *of probability measures* P_1 *and* P_2 *is defined as*

$$d_r(P_1, P_2) = \left(\inf_{\pi} \int\int_{\Omega_1 \times \Omega_2} d(\omega_1, \omega_2)^r \pi(d\omega_1, d\omega_2)\right)^{\frac{1}{r}},$$

where the infimum is among all probability measures π *on* $\Omega_1 \times \Omega_2$ *such that*

$$\pi(A \times \Omega_2) = P_1(A) \text{ and } \pi(\Omega_1 \times B) = P_2(B) \text{ for all } A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2,$$

and $d(\omega_1, \omega_2)$ *is inherited distance between elements of* Ω_1 *and* Ω_2*.*

Wasserstein distance of order $r = 1$ or $r = 2$ is used the most. We will use the Wasserstein distance of order 2 throughout this work.

Following Pflug and Pichler [2014] we present the following theorem which simplifies the computation of the Wasserstein distance of two measures on the real line.

Theorem 2.1. *The Wasserstein distance of order* $r \geq 1$ *of measures* P_1 *and* P_2 *on the real line is*

$$d_r(P_1, P_2)^r = \int_0^1 \mid F_{P_1}^{-1}(\alpha) - F_{P_2}^{-1}(\alpha) \mid^r d\alpha,$$

where $F_P^{-1}(\alpha)$ *is the inverse distribution function, which is defined as* $F_P^{-1}(\alpha) = \inf\{y : F_P(y) \geq \alpha\}$*.*

To simplify the computation of the measure of non-dominance, we use $d_r(P_1, P_2)^r$ instead of $d_r(P_1, P_2)$ when solving the program (2.1). Because raising a non-negative number to power $r \geq 1$ is a strictly increasing function, it does not change the optimal solution of the program (2.1). We can extract the r$^{\text{th}}$ root of the solution to receive the actual Wasserstein distance.

2.3 Simplification for Empirical Distributions

We will now focus on computing the measure of non-dominance in empirical distributions (discrete distributions with equiprobable atoms), which will be crucial in the fourth and fifth chapter. Pflug and Pichler [2014] describe how the definition of the Wasserstein distance can be adjusted for random variables with a discrete distribution.

Theorem 2.2 (Wasserstein Distance for Discretely Distributed Variables). *Let there be two discrete random variables. X attains values $x_1, \ldots, x_T$ with probabilities $p_1, \ldots, p_T$, and Y attains values $y_1, \ldots, y_T$ with probabilities $q_1, \ldots, q_T$. Then their Wasserstein distance of order $r \geq 1$ can be computed by solving the following program.*

$$\begin{aligned} d_r(X,Y)^r = \min_{\xi_{ts}} & \sum_{t=1}^{T} \sum_{s=1}^{T} \xi_{ts} \mid x_t - y_s \mid^r && && (2.2) \\ \textit{subject to} & \sum_{s=1}^{T} \xi_{ts} = p_t, && t = 1, \ldots, T, \\ & \sum_{t=1}^{T} \xi_{ts} = q_s, && s = 1, \ldots, T, \\ & \xi_{ts} \geq 0, && s = 1, \ldots, T, \quad t = 1, \ldots, T. \end{aligned}$$

The computation of the Wasserstein distance of two discrete random variables (2.2) can be significantly simplified for the case of two discrete random variables with equiprobable atoms. Firstly, we can substitute all p_t and q_t in the above program by $1/T$. Moreover, we will show in Theorem 2.3 that if the equiprobable atoms are ordered ($x_1 \leq \cdots \leq x_T$ and $y_1 \leq \cdots \leq y_T$), then for $r > 1$ the optimal solution of the program (2.2) is $\xi^*_{tt} = 1/T$ for all t and $\xi^*_{st} = 0$ for all $s \neq t$. There can be more optimal solutions for $r = 1$ but $\xi^*_{tt} = 1/T$ for all t and $\xi^*_{st} = 0$ for all $s \neq t$ is always among them.

We state and prove the following theorem only for the Wasserstein distance of integer order, which is sufficient for the usual use of Wasserstein distance of order 1 and 2. Considering only positive integer orders simplifies the computations in the proof of Lemma 2.4 in the third case.

Theorem 2.3 (Wasserstein Distance for Empirically Distributed Variables). *Let there be two discrete random variables. X attains values $x_1 \leq \cdots \leq x_T$ with probabilities $1/T$ and Y attains values $y_1 \leq \cdots \leq y_T$ with probabilities $1/T$. Then their Wasserstein distance of integer order $r \geq 1$ is*

$$d_r(X,Y)^r = \frac{1}{T} \sum_{t=1}^{T} \mid x_t - y_t \mid^r .$$

To prove this theorem, we will use the following lemma, which will be proved first.

Lemma 2.4. *Let $x_a \leq x_b$, $y_a \leq y_b$ be real numbers. Then for integer $r \geq 1$*

$$\mid x_a - y_a \mid^r + \mid x_b - y_b \mid^r \leq \mid x_a - y_b \mid^r + \mid x_b - y_a \mid^r . \qquad (2.3)$$

Proof. We assume without loss of generality that $x_a \leq y_a$. There are then three possible orderings of these four numbers:

1. $x_a \leq y_a \leq y_b \leq x_b$,
2. $x_a \leq y_a \leq x_b \leq y_b$,
3. $x_a \leq x_b \leq y_a \leq y_b$.

We will now work with each of these orderings separately. It enables us to eliminate the absolute values from equation (2.3) because we know when the differences are positive and when they are negative.

1. $(y_a - x_a)^r + (x_b - y_b)^r \leq (y_b - x_a)^r + (x_b - y_a)^r$ because both the first and the second summands are lower on the left-hand side.

2. $y_b - x_a \geq 0$ and $y_b - x_b + y_a - x_a \geq 0$ but $-x_b + y_a \leq 0$, so $(y_b - x_a)^r \geq (y_b - x_b + y_a - x_a)^r$. It holds for two non-negative numbers $(y_a - x_a)^r$ and $(y_b - x_b)^r$ that $(y_a - x_a + y_b - x_b)^r \geq (y_a - x_a)^r + (y_b - x_b)^r$. Putting these results together, we receive that $(y_b - x_a)^r \geq (y_b - x_b + y_a - x_a)^r \geq (y_a - x_a)^r + (y_b - x_b)^r$. Adding the non-negative number $(y_b - x_a)^r$ to the left-hand side of this inequality does not invalidate it, and we receive that $(y_b - x_a)^r + (x_b - y_a)^r \geq (y_a - x_a)^r + (x_b - y_b)^r$.

3. We will use the following notation to prove this case: $a = x_b - x_a$, $b = y_a - x_b$, $c = y_b - y_a$. Then $a, b, c \geq 0$. It follows from the binomial theorem[2] that $(y_b - x_a)^r + (y_a - x_b)^r = (a+b+c)^r + b^r = \sum_{k=0}^{r} \binom{r}{k} (a+b)^k \cdot c^{r-k} + b^r$. Similarly, $(y_a - x_a)^r + (y_b - x_b)^r = (a+b)^r + (b+c)^r = (a+b)^r + \sum_{k=0}^{r} \binom{r}{k} b^k \cdot c^{r-k}$. We compute as follows:

$$
\begin{aligned}
(a+b+c)^r + b^r &= \sum_{k=0}^{r} \binom{r}{k} (a+b)^k \cdot c^{r-k} + b^r \\
&= \sum_{k=0}^{r-1} \binom{r}{k} (a+b)^k \cdot c^{r-k} + (a+b)^r + b^r \\
&\geq \sum_{k=0}^{r-1} \binom{r}{k} b^k \cdot c^{r-k} + (a+b)^r + b^r \\
&= \sum_{k=0}^{r} \binom{r}{k} b^k \cdot c^{r-k} + (a+b)^r = (a+b)^r + (b+c)^r,
\end{aligned}
$$

where the inequality holds because a is non-negative. This proves that $(y_a - x_b)^r + (y_b - x_a)^r \geq (y_a - x_a)^r + (y_b - x_b)^r$.

□

We now return to the proof of the original theorem.

[2]The binomial theorem holds only for non-negative integer r, which is why this lemma is not stated and proved for any real $r \geq 1$. A generalization of the binomial theorem exists, but since the Wasserstein distance of integer order 1 or 2 is used the most, we do not consider it in this context.

Proof. We first add an assumption that $\xi_{ts} \in \{0, 1/T\}$ to the problem (2.2). Then there is exactly one $\xi_{ts} = 1/T > 0$ for each t and for each s. There cannot be more positive ξ_{ts} for given t and s because the equalities $\sum_{t=1}^{T} \xi_{ts} = 1/T$ and $\sum_{t=s}^{T} \xi_{ts} = 1/T$ would be violated because the sums would be higher than $1/T$. Similarly, there has to be at least one positive ξ_{ts} because the sums would be 0 otherwise. If $\xi_{ts} = 1/T$, then $\mid x_t - y_s \mid^r$ is added to the objective function $1/T$ times and it is the only time x_t and y_s are included in the objective function. The distance of x_t and y_s was assigned the weight $1/T$ while the distance of x_t and y_r, $r \neq s$, was assigned the weight 0. This allows us to formulate the problem in a different way: each y_s is assigned to one x_t, and their distance is added to the objective function.

Let k be a permutation $k : \{1, \dots, T\} \longrightarrow \{1, \dots, T\}$ which represents which y_s is assigned to which x_t. The objective function of problem (2.2) can be rewritten in the following way. Our goal is then to minimize it over all existing permutations k:

$$\frac{1}{T} \sum_{t=1}^{T} \mid x_t - y_{k(t)} \mid^r . \tag{2.4}$$

We show that k being identity, $k(t) = t$, is an optimal solution. There may be more optimal solutions but this is always one of them. We prove it by showing that by changing any k to an identity, the function (2.4) does not increase.

Suppose k_1 is not an identity. Then there are at least two numbers $a, b \in \{1, \dots, T\}$ such that $a < b$ and $k_1(a) > k_1(b)$. We call such situation *an inversion*. It follows from the lemma above that $\mid x_a - y_{k_1(a)} \mid^r + \mid x_b - y_{k_1(b)} \mid^r \geq \mid x_a - y_{k_1(b)} \mid^r + \mid x_b - y_{k_1(a)} \mid^r$. We therefore define k_2 in a following way: $k_2(a) = k_1(b)$, $k_2(b) = k_1(a)$ and $k_2(t) = k_1(t)$ for all other $t \in \{1, \dots, T\}$. The value of (2.4) using k_2 is lower than or equal to its value using k_1.

k_2 has strictly less inversions than k_1. If there are still inversions in k_2, we construct k_3 from k_2 in the same way as we constructed k_2 from k_1. We proceed in the same way until we arrive at k_n which does not have any inversions - it is an identity. Such n exists and is finite because the number of inversions in a finite sequence is finite, and we keep decreasing their number as we construct subsequent k's. The value of (2.4) keeps decreasing (or not increasing) so its value for $k_n = \textit{identity}$ is lower than or equal to its value for all k_i, $i < n$.

Any permutation k can be transformed to an identity in the described way and we have shown that the value of (2.4) does not increase throughout this process. Therefore identity leads to the lowest value of (2.4) and it is an optimal solution.

We return to the optimization problem (2.2) with the added constraint that $\xi \in \{0, 1/T\}$. The value of (2.4) is lowest for k being an identity, which corresponds to $\xi_{tt} = 1/T$ for all t and $\xi_{ts} = 0$ for all $t \neq s$.

We now loose the assumption that $\xi \in \{0, 1/T\}$. Then $\xi_{tt} = 1/T - \varepsilon$, $\varepsilon \in (0, 1/T)$ is also a feasible solution. However, it corresponds to decreasing the weight of the optimal assignment of x's to y's by ε and increasing the weight of a different assignment by ε. The assignment of x_t to y_t is an optimal one, so decreasing its weight in the objective function and increasing the weight of different assignments of x's to y's does not lead to a better solution.

The optimal solution of (2.2) for empirical distributions gives the maximum weights ξ_{ts} to the optimal assignment of x's to y's. The optimal assignment is an identity. Therefore the optimal solution of (2.2) is $\xi^*_{tt} = 1/T$ for all t and $\xi^*_{st} = 0$ for all $s \neq t$. This simplifies the objective function of (2.2) to $\frac{1}{T}\sum_{t=1}^{T} \mid x_t - y_t \mid^r$, and no more constraints are needed. So, the Wasserstein distance of positive integer order of two empirically distributed random variables is $\frac{1}{T}\sum_{t=1}^{T} \mid x_t - y_t \mid^r$. □

Using this result, we now return to the original problem of computing the measure of non-dominance.

Theorem 2.5 (Non-dominance for Empirically Distributed Variables)**.** *Let X and Y be discrete random variables with ordered equiprobable atoms $x_1 \leq \cdots \leq x_T$, $y_1 \leq \cdots \leq y_T$. Then the measure of n^{th} non-dominance of X and Y is computed, using the Wasserstein distance of integer order $r \geq 1$, as follows:*

$$n\text{-}ND(X,Y)^r = \min_{\hat{X}} \frac{1}{T}\sum_{t=1}^{T} \mid x_t - \hat{x}_t \mid^r$$
$$\text{subject to } \hat{X} \succeq_{(n)} Y.$$

Proof. It follows from Definition 2.1. We have seen in Theorem 2.3 that the Wasserstein distance of positive integer order of empirically distributed random variables with ordered atoms can be computed as $\frac{1}{T}\sum_{t=1}^{T} \mid x_t - \hat{x}_t \mid^r$. The atoms of the optimal solution $\hat{X}^*$ are ordered for the following reasons. Suppose $a < b$. Then $x_a \leq x_b$ as follows from the assumptions of this theorem. If $\hat{x}_c \leq \hat{x}_d$, then Lemma 2.4 states that $\mid x_a - \hat{x}_c \mid^r + \mid x_b - \hat{x}_d \mid^r \leq \mid x_a - \hat{x}_d \mid^r + \mid x_b - \hat{x}_c \mid^r$. Because the objective function is minimized, the optimal solution satisfies that $\hat{x}_c \leq \hat{x}_d$ whenever $c < d$. Because the condition $\hat{X} \succeq_{(n)} Y$ does not impose rules which would imply that $\hat{x}_a$ must be greater than $\hat{x}_b$ for $a < b$, the atoms of $\hat{X}$ are ordered. So, we can use the Theorem 2.3 to compute the distance of X and $\hat{X}$. □

We saw in Chapter 1 that the condition $\hat{X} \succeq_{(n)} Y$ can be rewritten for $n = 1$, or $n = 2$ as $\hat{x}_t \geq y_t$ for all t, or $\sum_{j=1}^{t} \hat{x}_j \geq \sum_{j=1}^{t} y_j$ for all t respectively.

3. Approximation of Stochastic Dominance for Particular Distributions

The measure of non-dominance can be computed exactly for some distributions. This chapter presents its values for uniform, normal and exponential distributions.

3.1 Uniform Distribution

Suppose $a < b$, $c < d$ are real numbers and X and Y are uniformly distributed random variables: $X \sim U(a, b)$, $Y \sim U(c, d)$. Using the alternative definitions of FSD, SSD and TSD, one quickly receives necessary and sufficient conditions for FSD and SSD and necessary conditions for TSD under uniform distribution.

Theorem 3.1. *Suppose $X \sim U(a, b)$ and $Y \sim U(c, d)$. Then*

$$X \succeq_{(1)} Y \iff a \geq c, b \geq d.$$
$$X \succeq_{(2)} Y \iff a \geq c, a + b \geq c + d.$$
$$X \succeq_{(3)} Y \implies a \geq c, b - a \geq d - c.$$

Proof. The Alternative Definition of FSD states that $X \succeq_{(1)} Y \iff F_X(x) \leq F_Y(x)$ for all x. $a \geq c$ is a necessary condition for $F_X(x) \leq F_Y(x)$ because $F_X(x) = 0$ for $x \leq a$ and $F_Y(x) > 0$ for $x > c$. Similarly, $b \geq d$ is a necessary conditions for $F_X(x) \leq F_Y(x)$ because $F_X(x) = 1$ for $x \geq b$ and $F_Y(x) < 1$ for $x < d$. They are also sufficient conditions because the distribution functions are straight lines on the support of the distribution and under these conditions they do not intersect on it. So, if $F_X(a) \leq F_Y(a)$ and $F_X(d) \leq F_Y(d)$, then $F_X(x) \leq F_Y(x)$ for all x.

The Alternative Definition of SSD states that $X \succeq_{(2)} Y \iff F_X^{(2)}(x) \leq F_Y^{(2)}(x)$ for all x. $F_X^{(2)}(x) = 0$ for $x \leq a$ and $F_Y^{(2)}(x) > 0$ for $x \geq c$. This leads to a necessary condition for $X \succeq_{(2)} Y$: $a \geq c$. If $b \geq d$, then $X \succeq_{(1)} Y$ and therefore also $X \succeq_{(2)} Y$.

Suppose now that $b < d$. For $x \geq d$, both $F_X(x) = 1$ and $F_Y(x) = 1$ so the integrated distribution functions increase by the same rate. So, it is necessary to check that $F_X^{(2)}(x) \leq F_Y^{(2)}(x)$ only for $x \in (a, d)$. F_X and F_Y are straight lines on the interval (a, b). Since we suppose that $c \leq a < b < d$, they intersect exactly once in the interval. For x lower than the point of the intersection $F_Y(x) \geq F_X(x)$ so $F_Y^{(2)}(x) \geq F_X^{(2)}(x)$, too. $F_Y(x) < F_X(x)$ for all x greater then point of intersection and lower than d, therefore, $F_X^{(2)}(x)$ increases faster than $F_Y^{(2)}(x)$ between the point of intersection and d. It is therefore sufficient to check that $F_Y^{(2)}(x) \geq F_X^{(2)}(x)$ at $x = d$ because if the inequality is violated for any x from the interval, it is violated at $x = d$, too. The integrated distribution functions at point d are in this case following:

$$F_X^{(2)}(d) = \frac{(b-a)^2}{2(b-a)} + (d-b) = d - \frac{b+a}{2},$$
$$F_Y^{(2)}(d) = \frac{(d-c)^2}{2(d-c)} = \frac{d-c}{2}.$$

Therefore $F_X^{(2)}(d) \leq F_Y^{(2)}(d)$ if and only if $\frac{d-c}{2} \geq d - \frac{b+a}{2} \Leftrightarrow a+b \geq c+d$. We supposed that $b < d$ but the case when $b \geq d$ also fulfills the condition $a+b \geq c+d$ so we have a general necessary rule for $X \succeq_{(2)} Y$.

It is also a sufficient rule. $a \geq c$ implies that $F_X^{(2)}(x) - F_Y^{(2)}(x) \leq 0$ for $x \leq c$, $a+b \geq c+d$ implies that $F_X^{(2)}(x) - F_Y^{(2)}(x) \leq 0$ for $x \in (a, \max(b,d))$. For $x \geq \max(b,d)$, the integrated distribution functions do not intersect so the condition cannot be violated.

The Alternative Definition of TSD states that $X \succeq_{(3)} Y \iff F_X^{(3)}(x) \leq F_Y^{(3)}(x)$ for all x. $F_X^{(3)}(x) = 0$ for $x \leq a$ and $F_Y^{(3)}(x) > 0$ for $x \geq c$. This leads to a necessary condition for $X \succeq_{(3)} Y$: $a \geq c$. The other necessary condition follows from the fact that $\mathbb{E}\, X = \frac{b-a}{2} \geq \mathbb{E}\, Y = \frac{d-c}{2}$ is a necessary condition for TSD. □

To prove that $X \succeq_{(3)} Y$ by showing that $F_Y^{(3)}(x) - F_X^{(3)}(x) \geq 0$ for all x, one must check that $a \geq c$ and that either $b \geq d$ or $F_Y^{(3)}(x) - F_X^{(3)}(x) \geq 0$ at points b, d, and that the minimum of $F_Y^{(3)}(x) - F_X^{(3)}(x)$ on the intervals (a,b) and (b,d) is greater than 0, and that $\mathbb{E}\, X \geq \mathbb{E}\, Y$. There is, however, a necessary and sufficient condition for TSD, which follows from the necessary conditions presented above.

Theorem 3.2. *Suppose $X \sim U(a,b)$ and $Y \sim U(c,d)$. Then*

$$X \succeq_{(3)} Y \iff X \succeq_{(2)} Y \iff a \geq c, a+b \geq c+d.$$

Proof. We will show that the necessary conditions for TSD presented in Theorem 3.1 imply the equivalent conditions for SSD. In both cases $a \geq c$. If also $b-a \geq d-c$ holds, we add $2a$ to the left side of this inequality and $2c$ to the right side of this inequality. The inequality is preserved because $a \geq c$. We receive $b+a \geq d+c$ which is the second condition for SSD.

So, using Theorem 3.1 and the sufficient condition for TSD from Theorem 1.7, we obtain

$$X \succeq_{(3)} Y \Rightarrow a \geq c, b-a \geq d-c \Rightarrow a \geq c, a+b \geq c+d \Leftrightarrow X \succeq_{(2)} Y \Rightarrow X \succeq_{(3)} Y,$$

which proves the equivalence of SSD and TSD for uniform distributions. □

We now proceed to compute the measure of non-dominance. The following theorem shows how the Wasserstein distance of order 2 of two uniformly distributed random variables is computed.

Theorem 3.3. *The Wasserstein distance of order 2 of two uniformly distributed random variables $X \sim U(a,b)$, $Y \sim U(c,d)$ is*

$$d_2(X,Y)^2 = (a-c)(b-d) + \frac{1}{3}(a-b-c+d)^2.$$

Proof. We use Theorem 2.1 to compute it. Note that $F_X^{-1}(\alpha) = a + \alpha(b-a)$. Therefore:

$$\begin{aligned} d_2(X,Y)^2 &= \int_0^1 (F_X^{-1}(\alpha) - F_Y^{-1}(\alpha))^2 d\alpha \\ &= \int_0^1 (a + \alpha(b-a) - c - \alpha(d-c))^2 d\alpha \\ &= \int_0^1 \alpha^2(-a+b+c-d)^2 + 2\alpha(a-c)(-a+b+c-d) + (a-c)^2 d\alpha \\ &= \frac{1}{3}(-a+b+c-d)^2 + (a-c)(-a+b+c-d) + (a-c)^2 \\ &= (a-c)(b-d) + \frac{1}{3}(a-b-c+d)^2. \end{aligned}$$

□

Using this results and supposing that $\hat{X} \sim U(\hat{a}, \hat{b})$, we can compute the 1-ND for uniform distributions by solving the following program:

$$\begin{aligned} \min_{\hat{a},\hat{b}} & (a-\hat{a})(b-\hat{b}) + \frac{1}{3}(a-b-\hat{a}+\hat{b})^2 \\ \text{subject to } & \hat{a} \geq c, \\ & \hat{b} \geq d. \end{aligned} \tag{3.1}$$

The following theorem describes the results of the program (3.1) above.

Theorem 3.4. *The measure of first non-dominance of two uniformly distributed random variables, $X \sim (a,b)$, $Y \sim (c,d)$, is computed by solving the problem (3.1). It leads to the following results.*

1. *If $a \geq c$ and $b \geq d$, then $\hat{a} = a$ and $\hat{b} = b$ and 1-ND$(X,Y) = 0$.*
2. *If $2a + b \geq 2c + d$ and $b < d$, then $\hat{a} = a - \frac{d-b}{2}$ and $\hat{b} = d$. The optimal value of the objective function is $\frac{1}{4}(d-b)^2$ and 1-ND$(X,Y) = \frac{d-b}{2}$.*
3. *If $a < c$ and $a + 2b \geq c + 2d$, then $\hat{a} = c$ and $\hat{b} = b - \frac{c-a}{2}$. The optimal value of the objective function is $\frac{1}{4}(c-a)^2$ and 1-ND$(X,Y) = \frac{c-a}{2}$.*
4. *If $2a + b < 2c + d$ and $a + 2b < c + 2d$, then $\hat{a} = c$ and $\hat{b} = d$. The optimal value of the objective function is $(a-c)(b-d) + \frac{1}{3}(a-b-c+d)^2$ and 1-ND(X,Y) is equal to its square root, which is also $d_2(X,Y)$.*

Proof. It can be easily seen that the feasibility set in this case is convex. The objective function is also convex, which can be seen from its Hessian matrix. The second derivatives of the objective function are following:

$$\frac{\partial^2 (a-\hat{a})(b-\hat{b})+\frac{1}{3}(a-b-\hat{a}+\hat{b})^2}{\partial\hat{a}\partial\hat{a}} = \frac{2}{3},$$
$$\frac{\partial^2 (a-\hat{a})(b-\hat{b})+\frac{1}{3}(a-b-\hat{a}+\hat{b})^2}{\partial\hat{a}\partial\hat{b}} =$$
$$\frac{\partial^2 (a-\hat{a})(b-\hat{b})+\frac{1}{3}(a-b-\hat{a}+\hat{b})^2}{\partial\hat{b}\partial\hat{a}} = \frac{1}{3},$$
$$\frac{\partial^2 (a-\hat{a})(b-\hat{b})+\frac{1}{3}(a-b-\hat{a}+\hat{b})^2}{\partial\hat{b}\partial\hat{b}} = \frac{2}{3}.$$

So, the Hessian matrix is

$$\frac{1}{3}\cdot\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

which is a positive semi-definite matrix. Because the objective function and the feasibility set are convex, the Karush-Kuhn-Tucker point is the optimal solution.

The Lagrange function of this problem is

$$L(\hat{a},\hat{b},u_1,u_1) = (a-\hat{a})(b-\hat{b}) + \frac{1}{3}(a-b-\hat{a}+\hat{b})^2 + u_1(c-\hat{a}) + u_2(d-\hat{b}).$$

We now derive the Karush-Kuhn-Tucker conditions for this problem. The stationarity conditions are following:

$$\frac{\partial L(\hat{a},\hat{b},u_1,u_1)}{\partial\hat{a}} = -b+\hat{b}-\frac{2}{3}(a-b-\hat{a}+\hat{b}) - u_1 = 0, \tag{3.2}$$

$$\frac{\partial L(\hat{a},\hat{b},u_1,u_1)}{\partial\hat{b}} = -a+\hat{a}+\frac{2}{3}(a-b-\hat{a}+\hat{b}) - u_2 = 0. \tag{3.3}$$

The feasibility conditions are $\hat{a} \geq c$ and $\hat{b} \geq d$.

The dual feasibility conditions are $u_1 \geq 0$ and $u_2 \geq 0$.

The complementary slackness conditions are $u_1(c-\hat{a}) = 0$ and $u_2(d-\hat{b}) = 0$.

Depending on the assumptions we make about the positiveness of u_1 and u_2, we receive the following four solutions. Each of these four assumptions implies that the dual feasibility conditions are satisfied.

Firstly, we suppose that $u_1 = 0$ and $u_2 = 0$. Then the stationarity conditions mean that $-b+\hat{b}-\frac{2}{3}(a-b-\hat{a}+\hat{b}) = 0$ and $-a+\hat{a}+\frac{2}{3}(a-b-\hat{a}+\hat{b}) = 0$ which implies that $\hat{a} = a$ and $\hat{b} = b$. The primal feasibility conditions mean that $a \geq c$ and $b \geq d$. The complementary slackness conditions are satisfied. Therefore the optimal solution is $\hat{a} = a$ and $\hat{b} = b$ if $a \geq c$ and $b \geq d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.1), we receive $(a-a)(b-b)+\frac{1}{3}(a-b-a+b)^2$, which equals 0. So the optimal value of the objective function is 0 in this case.

Secondly, we suppose that $u_1 = 0$ and $u_2 > 0$. Then the first complementary slackness condition is satisfied and the second one yields that $\hat{b} = d$. The first stationarity condition (3.2) yields for $u_1 = 0$ and $\hat{b} = d$ that $-b+d-\frac{2}{3}(a-b-\hat{a}+d) = 0$, which implies that $\hat{a} = a - \frac{d-b}{2}$. Then the second stationarity condition 3.3 yields that $-a+a-\frac{d-b}{2}+\frac{2}{3}(a-b-a-\frac{d-b}{2}+d) > 0$, which implies that $b < d$. The first feasibility condition yields that $a - \frac{d-b}{2} \geq c \Leftrightarrow 2a+b \geq 2c+d$, the second one is always satisfied. Therefore $\hat{a} = a - \frac{d-b}{2}$ and $\hat{b} = d$ if $a \geq c$ and $2a+b \geq 2c+d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.1), we receive $(a - a + \frac{d-b}{2})(b-d) + \frac{1}{3}(a - b - a + \frac{d-b}{2} + d)^2 = -\frac{(d-b)^2}{2} + \frac{1}{3}(\frac{3d-3b}{2})^2 = \frac{1}{4}(d-b)^2$.

Thirdly, we suppose that $u_1 > 0$ and $u_2 = 0$. Then the first complementary slackness condition yields that $\hat{a} = c$, the second one is satisfied. The stationarity condition (3.3) implies that $\hat{b} = b - \frac{c-a}{2}$. The second feasibility condition yields that $\hat{b} = b - \frac{c-a}{2} \geq d \Leftrightarrow a + 2b \geq c + 2d$. After substituting for $\hat{a}$ and $\hat{b}$ in (3.2), it implies that $-b + b - \frac{c-a}{2} - \frac{2}{3}(a - b - c + b - \frac{c-a}{2}) > 0$, which is equivalent to the condition $a < c$. As a result, $\hat{a} = c$ and $\hat{b} = b - \frac{c-a}{2}$ if $a < c$ and $a + 2b \geq c + 2d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.1), we receive $(a-c)(b-b+\frac{c-a}{2}) + \frac{1}{3}(a-b-c+b-\frac{c-a}{2})^2 = -\frac{c-a}{2} + \frac{1}{3}(\frac{3a-3c}{2}) = \frac{1}{4}(c-a)^2$.

Fourthly, we suppose that $u_1 > 0$ and $u_2 > 0$. Then the complementary slackness conditions yield that $\hat{a} = c$ and $\hat{b} = d$. The feasibility conditions are satisfied. After substituting for $\hat{a}$ and $\hat{b}$, the stationarity conditions imply that $-b + d - \frac{2}{3}(a - b - c + d) > 0$ and $-a + c + \frac{2}{3}(a - b - c + d) > 0$. By solving these inequalities, we receive that $2a + b < 2c + d$ and that $a + 2b < c + 2d$. Therefore $\hat{a} = c$ and $\hat{b} = d$ if $2a + b < 2c + d$ and $a + 2b < c + 2d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.1), we receive $(a-c)(b-d) + \frac{1}{3}(a - b - c + d)^2$.

□

The conditions of the first case, $a \geq c$ and $b \geq d$, imply that $X \succeq_{(1)} Y$ so it is correct that 1-ND$(X, Y) = 0$ in this case.

The conditions defining the second case imply that $a \geq c$. It can be seen by adding $-b$ to the left-hand side of $2a + b \geq 2c + d$ and $-d$ to the right-hand side of it. It can be done because $-b \geq -d$ is implied by the second condition defining this case. As a result, $X \succeq_{(2)} Y$ is implied by the conditions of the second case. The optimal solution may be surprising because $\hat{a}$ may be lower than a in order to minimize the Wasserstein distance between X and $\hat{X}$.

It can be seen similarly that the conditions defining the third case imply that $b \geq d$. The condition $a < c$ violates a necessary condition for $X \succeq_{(2)} Y$.

By adding up the conditions that define the fourth case, we receive that $a+b < c+d$. This violates a necessary condition for $X \succeq_{(2)} Y$ which is a necessary condition for $X \succeq_{(1)} Y$. As a result, if the conditions of the fourth case are satisfied, then $X \not\succeq_{(2)} Y$.

Only the first two cases correspond to computing the measure of first non-dominance when second-order stochastic dominance holds. If we approximate FSD by SSD in uniform distribution, the measure of non-dominance is going to be $\frac{d-b}{2}$, which is lower than the measure of non-dominance if the fourth case hold. But, the measure of non-dominance if the third case hold may actually be lower in some cases than when SSD holds.

All possible orderings of a, b, c, d are covered by the four described cases, and for a given ordering only the conditions of one of the four cases are satisfied. If $a \geq c$ and $b \geq d$, the conditions of the first case are satisfied. If $a < c$ and $b \geq d$, then depending on whether $a + 2b \geq c + 2d$ holds, either the conditions of the third or of the fourth case are satisfied. If $a \geq c$ and $b < d$, then depending on whether $2a+b \geq 2c+d$ holds, either the conditions of the second or of the fourth case are satisfied. If $a < c$ and $b < d$, then the conditions of the fourth case are satisfied.

We will now focus on the computation of 2-ND. We are going to be solving the following program.

$$\begin{aligned} &\min_{\hat{a},\hat{b}} (a-\hat{a})(b-\hat{b}) + \frac{1}{3}(a-b-\hat{a}+\hat{b})^2 \\ \text{subject to } &\hat{a} \geq c, \\ &\hat{a}+\hat{b} \geq c+d. \end{aligned} \tag{3.4}$$

The following theorem describes the results of the above program.

Theorem 3.5. *The measure of second non-dominance for two uniformly distributed random variables, $X \sim (a,b)$, $Y \sim (c,d)$, is computed by solving the problem (3.4). It leads to the following results.*

1. *If $a \geq c$ and $a+b \geq c+d$, then $\hat{a}=a$, $\hat{b}=b$ and 2-ND$(X,Y) = 0$.*
2. *If $a+b < c+d$ and $a-b \geq c-d$, then $\hat{a} = \frac{a-b+c+d}{2}$, $\hat{b} = \frac{-a+b+c+d}{2}$. The optimal value of the objective function is $\frac{1}{4}(a+b-c-d)^2$ and 2-ND$(X,Y) = \frac{-a-b+c+d}{2}$.*
3. *If $a < c$ and $a+2b \geq c+2d$, then $\hat{a}=c$ and $\hat{b} = b - \frac{c-a}{2}$. The optimal value of the objective function is $\frac{(c-a)^2}{4}$ and 2-ND$(X,Y) = \frac{c-a}{2}$.*
4. *If $a-b < c-d$ and $a+2b < c+2d$, then $\hat{a}=c$, $\hat{b}=d$. The optimal value of the objective function is $(a-c)(b-d)+\frac{1}{3}(a-b-c+d)^2$ and 2-ND(X,Y) is equal to its square root, which is also $d_2(X,Y)$.*

Proof. It can be easily seen that the feasibility set in this case is convex. The objective function is the same as in the problem (3.1) and therefore also convex. So, we use the Karush-Kuhn-Tucker conditions to receive the solution of this convex program. The Lagrange function of this problem is

$$L(\hat{a},\hat{b},u_1,u_1) = (a-\hat{a})(b-\hat{b}) + \frac{1}{3}(a-b-\hat{a}+\hat{b})^2 + u_1(c-\hat{a}) + u_2(c+d-\hat{a}-\hat{b}).$$

We derive the Karush-Kuhn-Tucker conditions for this problem. The stationarity conditions are following:

$$\frac{\partial L(\hat{a},\hat{b},u_1,u_1)}{\partial \hat{a}} = -b+\hat{b} - \frac{2}{3}(a-b-\hat{a}+\hat{b}) - u_1 - u_2 = 0, \tag{3.5}$$

$$\frac{\partial L(\hat{a},\hat{b},u_1,u_1)}{\partial \hat{b}} = -a+\hat{a} + \frac{2}{3}(a-b-\hat{a}+\hat{b}) - u_2 = 0. \tag{3.6}$$

The feasibility conditions are $\hat{a} \geq c$ and $\hat{a}+\hat{b} \geq c+d$.

The dual feasibility conditions are $u_1 \geq 0$ and $u_2 \geq 0$.

The complementary slackness conditions are $u_1(c-\hat{a}) = 0$ and $u_2(c+d-\hat{a}-\hat{b}) = 0$.

Depending on the assumptions we make about u_1 and u_2, we receive the following solutions. Each of these four assumptions implies that the dual feasibility conditions are satisfied.

Firstly, we suppose that $u_1 = 0$ and $u_2 = 0$. Then the stationarity conditions imply that $-b+\hat{b}-\frac{2}{3}(a-b-\hat{a}+\hat{b}) = 0$ and $a+\hat{a}+\frac{2}{3}(a-b-\hat{a}+\hat{b})-u_2 = 0$, which implies that $\hat{a} = a$ and $\hat{b} = b$. The feasibility conditions mean that $a \geq c$ and $a + b \geq c + d$. The complementary slackness conditions are satisfied. Therefore, $\hat{a} = a$ and $\hat{b} = b$ if $a \geq c$ and $a + b \geq c + dd$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.4), we receive $(a - a)(b - b) + \frac{1}{3}(a - b - a + b)^2 = 0$.

Secondly, we suppose that $u_1 = 0$ and $u_2 > 0$. Then the second complementary slackness condition implies that $-\hat{a} - \hat{b} + c + d = 0$. By subtracting the second stationarity condition (3.6) from the first one (3.5), we receive that $-a + \hat{a} + b - \hat{b} = 0$. Combining these two equalities, we receive that $\hat{a} = \frac{a-b+c+d}{2}$ and $\hat{b} = \frac{-a+b+c+d}{2}$. The first as well as the second stationarity condition imply for $u_2 > 0$ that $c + d > a + b$. The first feasibility condition implies that $a - b \geq c - d$. The second feasibility condition is satisfied. Therefore, $\hat{a} = \frac{a-b+c+d}{2}$ and $\hat{b} = \frac{-a+b+c+d}{2}$ if $a + b < c + d$ and $a - b \geq c - d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.4), we receive $(a-\frac{a-b+c+d}{2})(b-\frac{-a+b+c+d}{2})+\frac{1}{3}(a-b-\frac{a-b+c+d}{2}+\frac{-a+b+c+d}{2})^2 = \frac{1}{4}(a+b-c-d)^2+\frac{1}{3}\cdot 0 = \frac{1}{4}(a + b - c - d)^2$.

Thirdly, we suppose that $u_1 > 0$ and $u_2 = 0$. Then the first complementary slackness condition yields that $\hat{a} = c$. The second stationarity condition (3.6) implies for $u_2 = 0$ and $\hat{a} = c$ that$-a+c+\frac{2}{3}(a-b-c+\hat{b}) = 0$, which implies that $\hat{b} = b+\frac{a-c}{2}$. To satisfy the first stationarity condition for $u_1 > 0$, it must hold that $a < c$. The first feasibility condition is always satisfied in this case. To satisfy the second one, it must hold that $a+2b \geq c+2d$. Therefore, $\hat{a} = c$ and $\hat{b} = b+\frac{a-c}{2}$ if $a < c$ and $a+2b \geq c+2d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.4), we receive $(a-c)(b-b-\frac{a-c}{2})+\frac{1}{3}(a-b-c+b+\frac{a-c}{2})^2 = -\frac{(a-c)^2}{2}+\frac{1}{3}(\frac{3a-3c}{2})^2 = \frac{(a-c)^2}{4}$.

Fourthly, we suppose that $u_1 > 0$ and $u_2 > 0$. Then the complementary slackness conditions yield that $\hat{a} = c$ and $-\hat{a} - \hat{b} + c + d = 0$, which implies that $\hat{b} = d$. The feasibility conditions are therefore satisfied. The second stationarity condition (3.6) means that $-a + c + \frac{2}{3}(a - b - c + d) > 0$, which implies that $a+2b < c+2d$. Because both (3.5) and (3.6) equal 0, it holds that $-b+\hat{b}-\frac{2}{3}(a-b-\hat{a}+\hat{b})-u_1 = -a+\hat{a}+\frac{2}{3}(a-b-\hat{a}+\hat{b})$. Because u_1 is positive in this case, we receive that $-b+\hat{b}-\frac{2}{3}(a-b-\hat{a}+\hat{b}) > -a+\hat{a}+\frac{2}{3}(a-b-\hat{a}+\hat{b})$. Substituting for $\hat{a}$ and $\hat{b}$, we receive that $c-d > a-b$. Therefore, $\hat{a} = c$ and $\hat{b} = d$ if $a-b < c-d$ and $a + 2b < c + 2d$. Substituting for $\hat{a}$ and $\hat{b}$ in the objective function (3.4), we receive $(a - c)(b - d) + \frac{1}{3}(a - b - c + d)^2$. □

The conditions of the first case, $a \geq c$, $a + b \geq c + d$, imply that $X \succeq_{(2)} Y$ so it is correct that 2-ND$(X, Y) = 0$ in this case. The second and the third case violate the SSD rules apparently. The conditions of the fourth case imply that $a < c$ so the SSD rules are violated in this case as well. So, it is correct that 2-ND$(X, Y) \neq 0$ in these cases.

All possible orderings of a, b, c, d are covered by the four described cases, and for a given ordering only the conditions of one of the four cases are satisfied. If $a \geq c$ and $a + b \geq c + d$, the conditions of the first case are satisfied. If $a \geq c$

and $a+b<c+d$, then $-a-b>-c-d$ and $a-b\geq c-d$, so the conditions the second case are satisfied. If $a<c$ and $a+b\geq c+d$, then $-a>-c$ and $2a+2b\geq 2c+2d$, which implies that $a+2b\geq c+2d$, so the conditions of the third case are satisfied. If $a<c$ and $a+b<c+d$, then the conditions of the second, or the third or the fourth case hold. If $a-b\geq c-d$, then the conditions of the second case hold. These conditions imply that $b\geq d$ so the conditions of the third and the fourth case are violated. If $a-b<c-d$, then depending on whether $a+2b\geq c+2d$ holds, the conditions of the third or of the fourth case are satisfied.

Because we have shown that SSD and TSD are equivalent for uniform distribution, the 3-ND is also equivalent to 2-ND.

3.2 Normal Distribution

Suppose μ_X and μ_Y are real numbers, σ_X^2 and σ_Y^2 are real positive numbers and X and Y are normally distributed random variables: $X \sim N(\mu_X,\sigma_X^2)$ and $Y \sim N(\mu_Y,\sigma_Y^2)$.

It is shown in Levy [2006] that the following holds:

Theorem 3.6. *Let $X\sim N(\mu_X,\sigma_X^2)$ and $Y\sim N(\mu_Y,\sigma_Y^2)$. Then*

$$X \succeq_{(1)} Y \iff \mu_X \geq \mu_Y \text{ and } \sigma_X^2 = \sigma_Y^2,$$

$$X \succeq_{(2)} Y \iff \mu_X \geq \mu_Y \text{ and } \sigma_X^2 \leq \sigma_Y^2. \tag{3.7}$$

Mikulka [2011] showed that (3.7) is a necessary and sufficient condition also for $X \succeq_{(\infty)} Y$. Because ISD is a necessary condition for TSD and DARA SD, and SSD is a sufficient condition for them, (3.7) is a necessary and sufficient condition for $X\succeq_{(3)} Y$ and $X \succeq_{(D)} Y$ as well. So, under the assumption of normal distribution, SSD, TSD, DARA SD, and ISD are equivalent. We, therefore, compute only 1-ND(X,Y), and 2-ND(X,Y).

Pflug and Pichler [2014] state that the Wasserstein distance of order 2 of two normally distributed variables can be computed as follows. σ denotes the square root of σ^2.

Theorem 3.7. *The Wasserstein distance of order 2 of two normally distributed variables $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$ is*

$$d_2(X,Y)^2 = (\mu_X-\mu_Y)^2 + (\sigma_X-\sigma_Y)^2.$$

Therefore, assuming $\hat{X} \sim N(\hat{\mu},\hat{\sigma}^2)$, the measure of first non-dominance of two normally distributed random variables is computed as follows:

$$\begin{aligned} \min_{\hat{\mu},\hat{\sigma}} & (\mu_X-\hat{\mu})^2 + (\sigma_X-\hat{\sigma})^2 \\ \text{subject to } & \hat{\mu} \geq \mu_Y, \\ & \hat{\sigma} = \sigma_Y. \end{aligned} \tag{3.8}$$

The following theorem describes the results of the program (3.8) above.

Theorem 3.8. *The measure of first non-dominance for two normally distributed random variables, $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, is computed by solving the problem (3.8). It leads to the following results.*

1. *If $\mu_X \geq \mu_Y$, then $\hat{\mu} = \mu_X$ and $\hat{\sigma} = \sigma_Y$. The optimal value of the objective function is then $(\sigma_X - \sigma_Y)^2$ and 1-ND(X, Y) is its square root, which is $\mid \sigma_X - \sigma_Y \mid$.*

2. *If $\mu_X < \mu_Y$, then $\hat{\mu} = \mu_Y$ and $\hat{\sigma} = \sigma_Y$. The optimal value of the objective function is then $(\mu_X - \mu_Y)^2 + (\sigma_X - \sigma_Y)^2$ and 1-ND(X, Y) is its square root, which is $d_2(X, Y)$.*

Proof. Because $\hat{\sigma} = \sigma_Y$ due to the second condition, the problem remaining to be solved is following:

$$\min_{\hat{\mu}} (\mu_X - \hat{\mu})^2$$
$$\text{subject to } \hat{\mu} \geq \mu_Y.$$

The objective function is quadratic, hence convex, and the feasibility set is also convex. So, we will solve this problem using the Karush-Kuhn-Tucker conditions. The Lagrange function of this problem is

$$L(\hat{\mu}, u) = (\mu_X - \hat{\mu})^2 + u(\mu_Y - \hat{\mu}).$$

The stationarity condition is

$$\frac{\partial L(\hat{\mu}, u)}{\partial \hat{\mu}} = 2 \cdot (\mu_X - \hat{\mu}) - u = 0.$$

The feasibility condition is $\hat{\mu} \geq \mu_Y$, and the dual feasibility condition is $u \geq 0$. The complementary slackness condition is $u \cdot (\mu_Y - \hat{\mu}) = 0$. We immediately receive that if $u = 0$, then $\hat{\mu} = \mu_X$, and if $u > 0$, then $\hat{\mu} = \mu_Y$. The feasibility condition and the stationarity condition state when each of these solutions holds. □

The situation when $X \succeq_{(2)} Y$ falls into the first case of Theorem 3.8. The situation when $X \succeq_{(1)} Y$ is as a special case of it when $\sigma_X^2 = \sigma_Y^2$. Then 1-ND$(X, Y) = 0$ as desired.

The measure of second non-dominance of two normally distributed random variables is computed as follows:

$$\min_{\hat{\mu}, \hat{\sigma}} (\mu_X - \hat{\mu})^2 + (\sigma_X - \hat{\sigma})^2 \tag{3.9}$$
$$\text{subject to } \hat{\mu} \geq \mu_Y,$$
$$\hat{\sigma} \leq \sigma_Y.$$

We could add the assumption that $\hat{\sigma} > 0$ to the program (3.9) to ensure that $\hat{X}$ satisfies the definition of normal distribution. However, we will see in the following theorem that the optimal solutions of the program (3.9) satisfy this anyway.

The results of the program above are described in the following theorem.

Theorem 3.9. *The measure of second non-dominance for two normally distributed random variables, $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, is computed by solving the problem (3.9). It leads to the following results.*

1. *If $\mu_X \geq \mu_Y$ and $\sigma_X \leq \sigma_Y$, then $\hat{\mu} = \mu_X$ and $\hat{\sigma} = \sigma_X$. The optimal value of the objective function is then 0 so 2-ND(X, Y)=0.*

2. *If $\mu_X \geq \mu_Y$ and $\sigma_X > \sigma_Y$, then $\hat{\mu} = \mu_X$ and $\hat{\sigma} = \sigma_Y$. The optimal value of the objective function is then $(\sigma_X - \sigma_Y)^2$ and 2-ND$(X, Y) = \sigma_X - \sigma_Y$.*

3. *If $\mu_X < \mu_Y$ and $\sigma_X \leq \sigma_Y$, then $\hat{\mu} = \mu_Y$ and $\hat{\sigma} = \sigma_X$. The optimal value of the objective function is then $(\mu_X - \mu_Y)^2$ and 2-ND$(X, Y) = \mu_Y - \mu_X$.*

4. *If $\mu_X < \mu_Y$ and $\sigma_X > \sigma_Y$, then $\hat{\mu} = \mu_Y$ and $\hat{\sigma} = \sigma_Y$. The optimal value of the objective function is then $(\mu_X - \mu_Y)^2 + (\sigma_X - \sigma_Y)^2$ and 2-ND(X, Y) is its square root, which is $d_2(X, Y)$.*

Proof. Program (3.9) is convex so the Karush-Kuhn-Tucker point is its optimal solution. Its Lagrange function is

$$L(\hat{\mu}, \hat{\sigma}, u_1, u_2) = (\mu_X - \hat{\mu})^2 + (\sigma_X - \hat{\sigma})^2 + u_1(\mu_Y - \hat{\mu}) + u_2(\hat{\sigma} - \sigma_Y).$$

We derive the Karush-Kuhn-Tucker conditions. The stationarity conditions are following:

$$\frac{\partial L(\hat{\mu}, \hat{\sigma}, u_1, u_2)}{\partial \hat{\mu}} = 2(\hat{\mu} - \mu_X) - u_1 = 0,$$

$$\frac{\partial L(\hat{\mu}, \hat{\sigma}, u_1, u_2)}{\partial \hat{\sigma}} = 2(\hat{\sigma} - \sigma_X) + u_2 = 0.$$

The feasibility conditions are $\hat{\mu} \geq \mu_Y$ and $\hat{\sigma} \leq \sigma_Y$. The dual feasibility conditions are $u_1 \geq 0$ and $u_2 \geq 0$. The complementary slackness conditions are $u_1(\mu_Y - \hat{\mu}) = 0$ and $u_2(\hat{\sigma} - \sigma_Y) = 0$.

Depending on the assumptions we make about u_1 and u_2, we receive the following four solutions. The dual feasible conditions are always satisfied due to the assumptions.

Firstly, we assume that $u_1 = 0$ and $u_2 = 0$. The stationarity conditions then yield that $\hat{\mu} = \mu_X$ and $\hat{\sigma} = \sigma_X$. The feasibility conditions then imply that $\mu_X \geq \mu_Y$ and $\sigma_X \leq \sigma_Y$. The complementary slackness conditions are satisfied. The optimal value of the objective function is then 0.

Secondly, we assume that $u_1 = 0$ and $u_2 > 0$. Using the first stationarity condition and the second complementary slackness condition, we receive that $\hat{\mu} = \mu_X$ and $\hat{\sigma} = \sigma_Y$. The remaining conditions imply that this holds if $\mu_X \geq \mu_Y$ and $\sigma_X > \sigma_Y$. The optimal value of the objective function is then $(\sigma_X - \sigma_Y)^2$.

Thirdly, we assume that $u_1 > 0$ and $u_2 = 0$. Using the second stationarity condition and the first complementary slackness condition, we receive that $\hat{\mu} = \mu_Y$ and $\hat{\sigma} = \sigma_X$. The remaining conditions imply that this holds if $\mu_X < \mu_Y$ and $\sigma_X \leq \sigma_Y$. The optimal value of the objective function is then $(\mu_X - \mu_Y)^2$.

Fourthly, we assume that $u_1 > 0$ and $u_2 > 0$. We receive from the complementary slackness conditions that $\hat{\mu} = \mu_Y$ and $\hat{\sigma} = \sigma_Y$. The stationarity conditions then imply that $\mu_X < \mu_Y$ and $\sigma_X > \sigma_Y$. The feasibility conditions are satisfied. The optimal value of the objective function is then $(\mu_X - \hat{\mu})^2 + (\sigma_X - \hat{\sigma})^2$. □

It was already mentioned that SSD, TSD, and DARA SD are equivalent for normally distributed random variables. Therefore also 2-ND = 3-ND = 4-ND for normally distributed random variables.

3.3 Exponential Distribution

Suppose λ_X and λ_Y are real positive numbers, and X and Y are exponentially distributed random variables: $X \sim Exp(\lambda_X)$, $Y \sim Exp(\lambda_Y)$.

Using the infinite-order stochastic dominance, it follows from Mikulka [2011] that for exponentially distributed random variables the first-order, the second-order, the third-order, and DARA stochastic dominance are equivalent to each other.

Theorem 3.10. *Let X and Y be exponentially distributed random variables. $X \sim Exp(\lambda_X)$, $Y \sim Exp(\lambda_Y)$. Then*

$$X \succeq_{(1)} Y \Leftrightarrow X \succeq_{(2)} Y \Leftrightarrow X \succeq_{(3)} Y \Leftrightarrow X \succeq_{(D)} \Leftrightarrow X \succeq_{(\infty)} Y \Leftrightarrow \lambda_X \leq \lambda_Y.$$

In consequence, the 1-ND = 2-ND = 3-ND = 4-ND for exponentially distributed random variables. Under the assumption that $X \succeq_{(n)} Y$ for any n, 1-ND = 2-ND = 3-ND = 4-ND = 0. If $X \not\succeq_{(n)} Y$, then 1-ND = 2-ND = 3-ND = 4-ND > 0. We use again the Wasserstein distance of order 2 to compute the distance of two exponentially distributed random variables.

Theorem 3.11. *The Wasserstein distance of order 2 of two exponentially distributed random variables $X \sim Exp(\lambda_X)$, $Y \sim Exp(\lambda_Y)$ is*

$$d_2(X,Y)^2 = 2 \cdot \left(\frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}\right)^2.$$

Proof. We use Theorem 2.1 to compute it. Note that $F_X^{-1}(\alpha) = -\frac{\ln(1-\alpha)}{\lambda_X}$. Therefore:

$$\begin{aligned} d_2(X,Y)^2 &= \int_0^1 (F_X^{-1}(\alpha) - F_Y^{-1}(\alpha))^2 d\alpha \\ &= \int_0^1 \left(-\frac{\ln(1-\alpha)}{\lambda_X} + \frac{\ln(1-\alpha)}{\lambda_Y}\right)^2 d\alpha \\ &= \left(\frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}\right)^2 \cdot \int_0^1 \ln^2(1-\alpha) d\alpha \\ &= \left(\frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}\right)^2 \cdot 2. \end{aligned}$$

The last equality holds because $\int_0^1 \ln^2(1-\alpha)d\alpha = 2$, which can be seen using the substitution $t = \ln(1-\alpha)$ and integration by parts.

$$\begin{aligned} \int_0^1 \ln^2(1-\alpha) d\alpha &= \int_{-\infty}^0 t^2 \cdot e^t dt \\ &= 2 \cdot \int_{-\infty}^0 e^t dt \\ &= 2 \end{aligned}$$

□

Suppose $\hat{X} \sim Exp(\hat{\lambda})$. To satisfy the definition of exponential distribution, we must add the assumption that $\hat{\lambda} > 0$. Such assumption did not exist when we computed the measure of non-dominance for uniform distributions and it was unnecessary for the computation of the measure of non-dominance for normal distributions. To find the measure of non-dominance for exponential distributions, we solve the following program:

$$\begin{aligned} \min_{\hat{\lambda}} \; & 2 \cdot \left(\frac{\lambda_X - \hat{\lambda}}{\lambda_X \hat{\lambda}} \right)^2 \\ \text{subject to } & \hat{\lambda} \leq \lambda_Y. \\ & \hat{\lambda} > 0. \end{aligned} \tag{3.10}$$

The results of the above program are summarized in the following theorem.

Theorem 3.12. *The measure of non-dominance of two exponentially distributed random variables, $X \sim Exp(\lambda_X)$, $Y \sim Exp(\lambda_Y)$, is following.*

1. *If $\lambda_X \leq \lambda_Y$, then $\hat{\lambda} = \lambda_X$ and n-$ND(X,Y) = 0$ for any n.*
2. *If $\lambda_X > \lambda_Y$, then $\hat{\lambda} = \lambda_Y$. The optimal value of the objective function (3.10) is then $2 \cdot \left(\frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}\right)^2$, and the measure of non-dominance of any order n is its square root. So, n-$ND(X,Y) = \sqrt{2} \cdot \frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}$.*

Proof. First, we lose the assumption that $\hat{\lambda} > 0$. The objective function of this problem is not convex. But, the constraint function $\hat{\lambda} - \lambda_Y \leq 0$ is affine, so the Karush-Kuhn-Tucker conditions are necessary conditions for a locally optimal solution. We will first find the Karush-Kuhn-Tucker points, and then verify whether they are the optimal solutions. The Lagrange function is

$$L(\hat{\lambda}, u) = 2 \cdot \left(\frac{\lambda_X - \hat{\lambda}}{\lambda_X \hat{\lambda}} \right)^2 + u \cdot (\hat{\lambda} - \lambda_Y).$$

We derive the Karush-Kuhn-Tucker conditions. The stationarity condition states that

$$\frac{\partial L(\hat{\lambda}, u)}{\partial \hat{\lambda}} = 4 \cdot \frac{\hat{\lambda} - \lambda_X}{\lambda_X \cdot \hat{\lambda}^3} + u = 0.$$

The feasibility condition is $\hat{\lambda} \leq \lambda_Y$. The dual feasibility condition is $u \geq 0$. The complementary slackness condition is $u \cdot (\hat{\lambda} - \lambda_Y) = 0$.

If we set $u = 0$, we receive from the stationarity condition that $\hat{\lambda} = \lambda_X$. The feasibility condition implies that this holds if $\lambda_X \leq \lambda_Y$. The dual feasibility and the complementary slackness conditions are satisfied.

If we set $u > 0$, we receive from the complementary slackness condition that $\hat{\lambda} = \lambda_Y$. The stationarity condition implies for $u > 0$ that $4 \cdot \frac{\hat{\lambda} - \lambda_X}{\lambda_X \cdot \hat{\lambda}^3} < 0$, which means that $\lambda_Y < \lambda_X$. The feasibility and dual feasibility conditions are satisfied.

We will use the second order sufficient conditions (SOSC) to show that these points are local minimums of this problem. We are going to need the second derivative of Lagrange function, which is

$$\frac{\partial^2 L(\hat{\lambda}, u)}{\partial\hat{\lambda}\partial\hat{\lambda}} = \frac{12\lambda_X - 8\hat{\lambda}}{\lambda_X \cdot \hat{\lambda}^4}.$$

In the first case, when $u = 0$ and $\hat{\lambda} = \lambda_X$, the second derivative of the Lagrange function is equal to $4/\lambda_X^4$. So, it is always positive and $z \cdot 4/\lambda_X^4 \cdot z$ is also always positive for any $z \in \mathbb{R}$, $z \neq 0$. Therefore, the SOSC hold in this case, and $\hat{\lambda} = \lambda_X$ is a locally optimal solution of the problem if $\lambda_X \leq \lambda_Y$. The objective function is then equal to 0.

In the second case, when $u > 0$ and $\hat{\lambda} = \lambda_Y$, the constraint $\hat{\lambda} - \lambda_Y \leq 0$ is active. The set of z, for which the inequality $z \cdot \frac{\partial^2 L(\hat{\lambda},u)}{\partial\hat{\lambda}\partial\hat{\lambda}} \cdot z > 0$ must hold, is empty because $z \cdot \frac{\partial(\hat{\lambda}-\lambda_Y)}{\partial\hat{\lambda}} = z \cdot 1 = 0$ does not hold for any $z \in \mathbb{R}$, $z \neq 0$. So, the SOSC hold in this case, too. $\hat{\lambda} = \lambda_Y$ is a locally optimal solution of the problem if $\lambda_X > \lambda_Y$. The objective function is then equal to $2 \cdot \left(\frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}\right)^2$.

Both of these solutions satisfy that $\hat{\lambda} > 0$ so they are feasible also when we add the second constraint of the problem (3.10). Each of these solutions is the only locally optimal solution for a given ordering of λ_X and λ_Y, so they are also the globally optimal solutions for the given ordering if the global solution exists. To check that, we have to check that the objective function does not decrease to lower values at the left open end of the feasible set. Because $\lim_{\hat{\lambda}\to 0} 2 \cdot \left(\frac{\lambda_X - \hat{\lambda}}{\lambda_X \hat{\lambda}}\right)^2 = \infty$, we have found the globally optimal solution for each ordering of λ_X and λ_Y. □

The first case of Theorem 3.12 corresponds to the situation when $X \succeq_{(n)} Y$ and the second case corresponds to the situation when $X \not\succeq_{(n)} Y$ for any n.

4. Estimation of Non-Dominance

We have derived the measure of non-dominance for certain distributions exactly in the previous chapter. We will now focus on other widely used distributions for which we have not derived the exact values of n-ND. To gain some understanding of non-dominance in these distributions, we will perform a numerical study.

We are going to use empirical distributions to estimate the distributions of interest. We randomly generate T numbers from the distributions of X and Y and use them as atoms of two discrete distributions with equiprobable atoms. We then explore the relationship of these two empirical distributions, which we denote $\tilde{X}$ and $\tilde{Y}$, and compute their measure of non-dominance. We use the Wasserstein distance of order 2 to compute the n-ND$(\tilde{X}, \tilde{Y})$ of the empirical distributions.

Following Theorem 2.5 about non-dominance in empirical distributions, 1-ND of two empirical distributions $\tilde{X}$, $\tilde{Y}$ with ordered atoms $x_1 \leq \cdots \leq x_T$, $y_1 \leq \cdots \leq y_T$ is computed as follows:

$$\begin{aligned} \text{1-ND}(\tilde{X}, \tilde{Y})^2 = \min_{\hat{x}_1, \ldots, \hat{x}_T} & \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2 \\ \text{subject to } & \hat{x}_t \geq y_t, \qquad t = 1, \ldots, T. \end{aligned} \tag{4.1}$$

We will show that this optimization problem has a solution that simplifies its application.

Theorem 4.1. *Let $\tilde{X}$ and $\tilde{Y}$ be discrete random variables with equiprobable atoms such that $x_1 \leq \cdots \leq x_T$ and $y_1 \leq \cdots \leq y_T$. Then*

$$\text{1-ND}(\tilde{X}, \tilde{Y})^2 = \frac{1}{T} \sum_{t=1}^{T} (\min(0, x_t - y_t))^2.$$

Proof. Problem (4.1) is convex so the Karush-Kuhn-Tucker point is its optimal solution. Its Lagrange function is

$$L(\hat{\boldsymbol{x}}, \boldsymbol{u}) = \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2 + \sum_{t=1}^{T} u_t (y_t - \hat{x}_t).$$

We derive the Karush-Kuhn-Tucker conditions. The stationarity conditions state for every $t = 1, \ldots, T$ that

$$\frac{\partial L(\hat{\boldsymbol{x}}, \boldsymbol{u})}{\partial \hat{x}_t} = \frac{2}{T}(x_t - \hat{x}_t) - u_t = 0$$

The feasibility conditions are $\hat{x}_t \geq y_t$ for all t. The dual feasibility conditions are $u_t \geq 0$ for all t. The complementary slackness conditions are $u_t \cdot (y_t - \hat{x}_t) = 0$ for all t. These conditions hold for all $t = 1, \ldots, T$ but they restrict each t separately. This enables us to find the solution for each t separately, which simplifies the solution substantially.

For each t the conditions lead to the following solutions.

- If $x_t \geq y_t$, then $\hat{x}_t = x_t$.

- If $x_t < y_t$, then $\hat{x}_t = y_t$.

This shows that the summands of the objective function (4.1) are either 0 if $x_t - y_t \geq 0$, or $(x_t - y_t)^2$ otherwise. □

We will use Theorem 2.5 to compute 2-ND of two empirical distributions. For empirical random variables $\tilde{X}$, $\tilde{Y}$ with ordered atoms $x_1 \leq \cdots \leq x_T$, $y_1 \leq \cdots \leq y_T$, 2-ND is computed as follows:

$$\text{2-ND}(X,Y)^2 = \min_{\hat{x_1},\ldots,\hat{x_T}} \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2 \qquad (4.2)$$

$$\text{subject to } \sum_{j=1}^{t} \hat{x}_j \geq \sum_{j=1}^{t} y_j, \qquad t = 1, \ldots, T.$$

Using this, we now perform a numerical study on particular distributions.

4.1 Log-Normal Distribution

Log-normal distribution is a distribution which is in a tight relationship with normal distribution. If a random variable X is log-normally distributed, then random variable $Z = \ln(X)$ is normally distributed. The distribution is determined by two parameters, μ and σ. μ is a real number, σ is a positive real number. Its probability density function is

$$\frac{1}{x\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \cdot \mathbb{I}_{x>0}.$$

Its mean is $\eta = e^{\mu+\sigma^2/2}$ and its variance is $s^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$. Assuming positive σ, they are both always positive.

The quantile function of log-normal distribution is quite complicated:

$$F^{-1}(\alpha) = \exp(\mu + \sqrt{2\sigma^2} \cdot \text{erf}^{-1}(2\alpha - 1)),$$

where $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. We could not compute the Wasserstein distance of two log-normally distributed random variables because Theorem 2.1, which we used for it in Chapter 3, uses the quantile function. As a result, we could not compute the measure of non-dominance. We perform a numerical study to gain more understanding of it in log-normal distributions.

It was shown in *Levy* [2006] that the following rules hold for stochastic dominance in log-normal distributions hold.

Theorem 4.2. *Suppose η_X, η_Y, s_X, s_Y are the means and standard deviations of X and Y. Then*

$$X \succeq_{(1)} Y \iff \eta_X \geq \eta_Y \text{ and } s_X^2 = s_Y^2,$$

$$X \succeq_{(2)} Y \iff \eta_X \geq \eta_Y \text{ and } \frac{\eta_X}{s_X} \geq \frac{\eta_Y}{s_Y}.$$

We suppose for the purpose of the numerical study that $X \sim ln(0,1)$ and we alternate the parameters of Y. We generate 1000 numbers from each distribution. They define the empirical distributions, $\tilde{X}$, $\tilde{Y}$, that estimate the log-normal ones.

We compute 1-ND$(\tilde{X},\tilde{Y})$ and 2-ND$(\tilde{X},\tilde{Y})$ for each pair of empirical distributions. We do it 200 times for each set of parameters defining X and Y so we receive 200 values of 1-ND$(\tilde{X},\tilde{Y})$ and 2-ND$(\tilde{X},\tilde{Y})$ for each set of parameters. We then compute their mean and standard deviation. We present our results in Table 4.1. The first two columns specify the parameters defining Y, the third column describes the strongest holding stochastic dominance relationship if $X \succeq Y$, and the following columns present the mean 1-ND$(\tilde{X},\tilde{Y})$ and 2-ND$(\tilde{X},\tilde{Y})$ and their standard deviations for each set of parameters.

$\boldsymbol{\mu_Y}$	$\boldsymbol{\sigma_Y^2}$	**SD holding**	**1-ND**	**std 1-ND**	**2-ND**	**std 2-ND**
-1	0.25	-	0.003	0.001	0.003	0.001
-1	1	SSD	0.003	0.035	0.0	0.0
-1	2	SSD	0.814	0.931	0.0	0.0
-1	4	-	11.281	7.14	1.01	0.476
0	0.25	-	0.136	0.014	0.136	0.014
0	2	-	4.547	2.277	1.025	0.219
0	4	-	36.392	28.919	5.622	1.467
1	0.25	-	1.541	0.064	1.516	0.06
1	1	-	4.822	0.789	2.838	0.188
1	2	-	16.324	4.611	5.718	0.6
1	4	-	101.44	63.264	18.455	3.868
2	0.25	-	7.252	0.202	6.749	0.156
2	1	-	17.223	1.907	10.514	0.536
2	2	-	48.474	12.898	18.442	1.641
2	4	-	302.706	236.554	53.8	11.572

Table 4.1: Estimated non-dominance in log-normal distributions with reference distribution $ln(0, 1)$.

It can be seen from the results presented in Table 4.1 that 2-ND is always lower than (or equal to) 1-ND. It should be so because FSD is stricter than SSD so $\hat{X}$ has to satisfy stricter conditions in order to dominate Y with respect to FSD. As a result, it differs from X more and the measure of non-dominance is higher.

When $X \succeq_{(2)} Y$, the estimated measure of second non-dominance is 0 as we expect. The estimated measure of second non-dominance is 0.003 in the first row so it is still very close to zero but it is higher than when SSD holds. It correctly describes the fact that the rule for SSD, $F_X^{(2)}(x) \leq F_Y^{(2)}(x)$, is violated only on a very short interval consisting of very low x's.

Both 1-ND and 2-ND increase with increasing μ_Y. It is understandable because $\eta_X \geq \eta_Y$ is a necessary condition for both $X \succeq_{(1)} Y$ and $X \succeq_{(2)} Y$, and $\eta_X \geq \eta_Y \iff e^{\mu_X + \sigma_X^2/2} \geq e^{\mu_Y + \sigma_Y^2/2}$. If $\sigma_X = \sigma_Y$, it implies that $\mu_X \geq \mu_Y$ is a necessary condition for $X \succeq_{(1)} Y$ and $X \succeq_{(2)} Y$. As μ_Y increases, μ_X becomes increasingly far from being higher than μ_Y.

A similar rule holds for σ^2. With increasing σ_Y^2, both 1-ND and 2-ND increase. The only exception to this rule can be seen by comparing the first two rows of Table 4.1, where the 2-ND decreases. It is correct in this case because SSD does not hold for the parameters defining the first row.

4.2 Gamma Distribution

Gamma distribution can be seen as a generalization of exponential distribution. It uses two parameters, $k > 0$ and $\theta > 0$, and its probability density function is

$$\frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \cdot \mathbb{I}_{x>0}.$$

Gamma distribution with parameters $k = 1$ and $\theta = \frac{1}{\lambda}$ is equivalent to exponential distribution with parameter λ. It holds for $X \sim \Gamma(k, \theta)$ that $\mathbb{E}\, X = k\theta$ and $\operatorname{var} X = k\theta^2$.

It was shown in Ali [1975] and Mikulka [2011] that the following stochastic dominance rules hold for random variables with gamma distribution.

Theorem 4.3. *Let $X \sim \Gamma(k_X, \theta_X)$ and $Y \sim \Gamma(k_Y, \theta_Y)$. Then*

$$X \succeq_{(1)} Y \iff \theta_X \geq \theta_Y \text{ and } k_X \geq k_Y,$$

$$X \succeq_{(2)} Y \iff X \succeq_{(\infty)} Y \iff \frac{k_X}{k_Y} \geq \frac{\theta_Y}{\theta_X} \text{ and } k_X \geq k_Y.$$

The fact that $X \succeq_{(2)} Y$ is equivalent to $X \succeq_{(\infty)} Y$ implies that it is equivalent also to $X \succeq_{(3)} Y$ and $X \succeq_{(D)} Y$. To our best knowledge and according to Okagbue et al. [2020], there is no closed-form expression for the inverse cumulative distribution function of gamma distribution. We therefore cannot use Theorem 2.1 to receive the Wasserstein distance of two gamma distributed random variables. Hence, we are not able to derive an exact formula for the computation of the measure of non-dominance. The numerical study helps us to gain some knowledge about the values of 1-ND and 2-ND.

As in the case of the log-normal distribution, we generate 1000 numbers from the distributions of X and Y, which define the empirical distributions $\tilde{X}$and $\tilde{Y}$. We repeat it 200 times for each set of parameters. We receive 200 values of 1-ND$(\tilde{X}, \tilde{Y})$ and 2-ND$(\tilde{X}, \tilde{Y})$ for each set of parameters.

We set $X \sim \Gamma(1, 1)$ and alternate the parameters defining Y. We present the results for $\theta_Y = 1, 2, 3$ and $k_Y = 1, 2, 3, 4$ in Table 4.2. The first two columns specify the parameters defining Y, and the following columns present the average 1-ND$(\tilde{X}, \tilde{Y})$ and 2-ND$(\tilde{X}, \tilde{Y})$ and their standard deviations for each set of parameters. In case of the presented parameters $X \not\succeq_{(n)} Y$ for any n.

It can be seen in Table 4.2 that 1-ND$(\tilde{X}, \tilde{Y})$ is higher than 2-ND$(\tilde{X}, \tilde{Y})$ for all presented parameters, which is correct. Both 1-ND$(\tilde{X}, \tilde{Y})$ and 2-ND$(\tilde{X}, \tilde{Y})$ increase with increasing k_Y or θ_Y. It makes sense because as k_Y increases, $k_X = 1$ is even further from being higher than k_Y, which is a necessary condition for $X \succeq_{(n)} Y$. Similarly, $\theta_X \geq \theta_Y$ is a necessary condition for $X \succeq_{(1)} Y$, and increasing θ_Y leads to θ_X being further from satisfying it. It holds for fixed k_X and θ_X that if θ_Y increases, they become even further from satisfying $X \succeq_{(2)} Y$.

We may also notice that if the dominance is violated only by the value of k_Y ($\theta_Y = 1$), the measures non-dominance tend to be lower than when the dominance is violated only by the value of θ_Y ($k_Y = 1$).

The distributions defined by the first two rows of Table 4.2 where $k_Y = 1$ can be seen as exponential distributions with parameters $\lambda_Y = 1/2$ and $\lambda_Y = 1/3$. The reference distribution $\Gamma(1, 1)$ is also an exponential distribution with

$\boldsymbol{k}_Y$	$\boldsymbol{\theta}_Y$	**1-ND**	**std 1-ND**	**2-ND**	**std 2-ND**
1	2	1.431	0.096	1.006	0.064
1	3	2.805	0.152	1.986	0.096
2	1	1.093	0.063	0.997	0.05
2	2	3.524	0.129	3.002	0.1
2	3	5.977	0.181	5.008	0.136
3	1	2.148	0.071	2.002	0.061
3	2	5.59	0.141	5.002	0.118
3	3	9.04	0.2	7.993	0.163
4	1	3.183	0.088	3.004	0.079
4	2	7.634	0.152	7.006	0.133
4	3	12.104	0.217	11.002	0.183

Table 4.2: Estimated non-dominance in gamma distributions with reference distribution $\Gamma(1,1)$.

parameter $\lambda_X = 1$. So, it is possible to compute the measure of non-dominance in these two cases exactly based on Theorem 3.12. For $\lambda_X = 1$ and $\lambda_Y = \frac{1}{2}$, the n-ND$(X,Y)^2$ is

$$2 \cdot \left(\frac{1 - \frac{1}{2}}{1 \cdot \frac{1}{2}} \right)^2 = 2.$$

For $\lambda_X = 1$ and $\lambda_Y = \frac{1}{3}$, the n-ND$(X,Y)^2$ is

$$2 \cdot \left(\frac{1 - \frac{1}{3}}{1 \cdot \frac{1}{3}} \right)^2 = 8.$$

Their square roots are approximately 1.414 and 2.828. These values correspond quite well to the mean 1-ND$(\tilde{X},\tilde{Y})$. It seems that the approximations of gamma distributions by the empirical distributions are quite good for estimating 1-ND. Mean 2-ND$(\tilde{X},\tilde{Y})$ is substantially lower than 1-ND$(\tilde{X},\tilde{Y})$, which is incorrect in this case. It should be the same for exponentially distributed random variables. It shows the limits of estimating the measure of second non-dominance by empirical distributions. We assumed in Section 3.3 that $\hat{X}$, the variable which is as close to X as possible while dominating Y, is exponentially distributed. Such assumption is missing when comparing empirical distributions. Therefore the $\hat{X}$, which emerges here in the numerical study, can be a vector that is very unlikely to be generated from exponential distribution.

An example of it can be seen in Figure 4.1. Let $\tilde{X}$ be the empirical approximation of $\Gamma(1,1)$ and $\tilde{Y}$ be the empirical approximation of $\Gamma(1,3)$. $\hat{X}$ is the distribution that emerges from the computation of 2-ND$(\tilde{X},\tilde{Y})$. $\tilde{X}$ and $\tilde{Y}$ were generated 200 times and therefore $\hat{X}$ was computed 200 times as well. Figure 4.1 shows the distribution functions for the average $\tilde{X}$, $\tilde{Y}$ and $\hat{X}$ over those 200 generations and computations. $\hat{X}$, which emerged from computing 1-ND, was also examined. It cannot be seen in the graph because its distribution coincides with the distribution of $\Gamma(1,3)$.

We show the estimated measures of non-dominance for another set of gamma distributions. This time $X \sim \Gamma(2,1)$ so the reference variable does not correspond to any exponential distribution. We alternate the parameters defining Y:

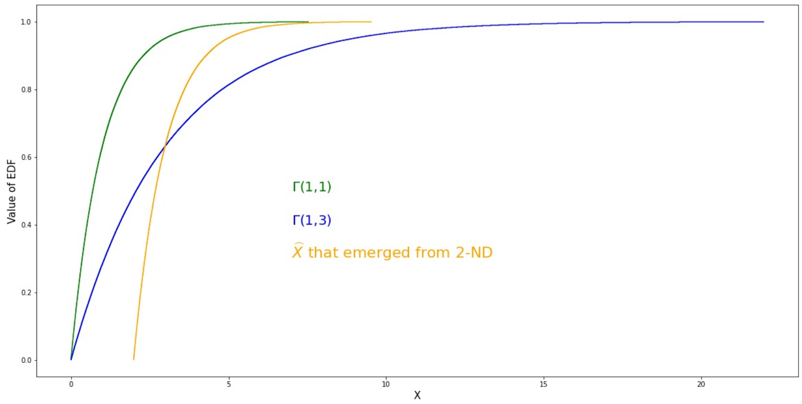

Figure 4.1: Distribution functions of the average of empirical variables.

$\theta_Y = 1, 2, 3$ and $k_Y = 1, 2, 3, 4$. The procedure of computing 1-ND$(\tilde{X}, \tilde{Y})$ and 2-ND$(\tilde{X}, \tilde{Y})$ is the same as in the previous case. The results are shown in Table 4.3. The first two columns specify the parameters defining Y, the third column describes the strongest holding stochastic dominance relationship if $X \succeq Y$, and the following columns present the average 1-ND$(\tilde{X}, \tilde{Y})$ and 2-ND$(\tilde{X}, \tilde{Y})$ and their standard deviations for each set of parameters.

k_Y	θ_Y	**SD holding**	**1-ND**	**std 1-ND**	**2-ND**	**std 2-ND**
1	1	FSD	0.002	0.008	0.0	0.0
1	2	SSD	0.575	0.119	0.033	0.049
1	3	-	1.892	0.176	1.0	0.11
2	2	-	2.454	0.134	2.005	0.099
2	3	-	4.906	0.195	4.0	0.149
3	1	-	1.062	0.074	1.001	0.063
3	2	-	4.496	0.141	3.998	0.111
3	3	-	7.929	0.188	6.974	0.156
4	1	-	2.102	0.086	2.007	0.076
4	2	-	6.551	0.164	6.014	0.141
4	3	-	10.986	0.206	9.987	0.181

Table 4.3: Estimated non-dominance in gamma distributions with reference distribution $\Gamma(2, 1)$.

Firstly, we may notice in Table 4.3 that when $X \succeq_{(1)} Y$, both 1-ND$(\tilde{X}, \tilde{Y})$ and 2-ND$(\tilde{X}, \tilde{Y})$ either equal 0, or they are very close to it, which can be attributed to the simulation error. When $X \succeq_{(2)} Y$, 2-ND$(\tilde{X}, \tilde{Y})$ is also very close to 0. Neither 1-ND, nor 2-ND is as low in cases when $X \not\succeq Y$. The empirical estimation of gamma distribution approximates whether stochastic dominance holds well in these cases.

What was noted in the description of Table 4.2 holds as well. 1-ND is always lower than 2-ND. Both 1-ND and 2-ND increase with increasing k_Y and θ_Y, and they increase more significantly with increasing θ_Y.

5. Application to Portfolio Optimization

We will now apply stochastic dominance theory to portfolio optimization, and we will show the role of the measure of non-dominance. We first present the notation and models for portfolio optimization with stochastic dominance constraints as described by Dentcheva and Ruszczynski [2006]. We proceed by incorporating the measure of non-dominance and defining the closest dominating portfolio from two perspectives, and finally, we apply the theory to actual financial data.

5.1 Portfolio Optimization with Stochastic Dominance Constraints

We assume that there are finitely many assets that we may use to construct a portfolio. We denote the number of assets n, and the weights of the portfolio $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$ or later also $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$. We assume that the weights are non-negative and they sum up to one. $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}, \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0\}$ is the set of all possible weights.

We assume that there exists a benchmark investment with random return Y, and we denote the random return of portfolio with weights $\boldsymbol{\lambda}$ as $R(\boldsymbol{\lambda})$. The general form of a portfolio optimization problem with stochastic dominance constraints was presented in Dentcheva and Ruszczynski [2006]. For f being a concave continuous function of the portfolio weights, which we wish to maximize, the form is following:

$$\begin{aligned} &\max_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda}) \\ &\text{subject to } R(\boldsymbol{\lambda}) \succeq_{(n)} Y, \\ &\boldsymbol{\lambda} \in \Lambda. \end{aligned} \tag{5.1}$$

We use past returns to estimate the distributions of returns of the n considered assets and of the benchmark investment. r_{it} denotes the return of i^{th} asset at time t, $t = 1, \dots, T$. y_t denotes the return of the benchmark investment at time t. To estimate the distribution, we consider each of these T returns equally likely - each of them happens with probability $1/T$.

We apply the above program (5.1) in two ways: we use the first-order, and the second-order stochastic dominance in the constraint. Mean return is used as the measure of return in both cases: $f(\boldsymbol{\lambda}) = \mathbb{E}\ R(\boldsymbol{\lambda})$.

The following program was formulated in Kuosmanen [2004]. The objective function is the estimated mean return, and the constraints ensure that $R(\boldsymbol{\lambda}) \succeq_{(1)} Y$.

$$
\begin{aligned}
&\max_{\boldsymbol{\lambda},\pi_{t,s}} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\lambda_i r_{it} && && (5.2)\\
&\text{subject to } \sum_{i=1}^{n}\lambda_i r_{it} \geq \sum_{s=1}^{T}\pi_{ts} y_s, && t = 1,\dots,T,\\
&\sum_{t=1}^{T}\pi_{ts} = 1, && s = 1,\dots,T,\\
&\sum_{s=1}^{T}\pi_{ts} = 1, && t = 1,\dots,T,\\
&\pi_{ts} \in \{0,1\}, && s = 1,\dots,T, \;\; t = 1,\dots,T,\\
&\boldsymbol{\lambda} \in \Lambda.
\end{aligned}
$$

We denote the optimal solution of this problem as $\boldsymbol{\lambda}^{*1}$. It represents the weights of the optimal portfolio and the optimal value of the objective function is its mean return.

We use model presented in Dentcheva and Ruszczynski [2006] for portfolio optimization with the second-order stochastic dominance constraints. They showed that the SSD relationship can be described using expected shortfall, and derived from it the following formulation of portfolio optimization problem with second-order stochastic dominance constraints:

$$
\begin{aligned}
&\max_{\boldsymbol{\lambda},s_{t,s}} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\lambda_i r_{it} && && (5.3)\\
&\text{subject to } \sum_{i=1}^{n}\lambda_i r_{is} + s_{ts} \geq y_t, && s = 1,\dots,T, \;\; t = 1,\dots,T,\\
&\frac{1}{T}\sum_{s=1}^{T} s_{ts} \leq F_Y^{(2)}(y_t), && t = 1,\dots,T,\\
&s_{ts} \geq 0, && s = 1,\dots,T, \;\; t = 1,\dots,T,\\
&\boldsymbol{\lambda} \in \Lambda.
\end{aligned}
$$

The optimal solution of this problem, $\boldsymbol{\lambda}^{*2}$, represents the weights of the optimal portfolio. The optimal value of the objective function is its expected return. We should also note that (5.3) is a linear program which allows us to solve it for much larger data than the mixed-integer linear program (5.2). But, our goal is to compare the results of these two problems so we will solve them for the same data.

The FSD constraint implemented in (5.2) is stricter than the SSD constraint implemented in (5.3), and we have seen in the first chapter that FSD implies SSD. Therefore the feasible set in the second case is larger, and it includes the feasible set of the first case. Therefore also the optimal solution of the second program leads to higher (or the same) expected returns than the optimal solution of the first program.

5.2 Closest Dominating Portfolios

It holds that $R(\boldsymbol{\lambda}^{*1}) \succeq_{(1)} Y$, $R(\boldsymbol{\lambda}^{*1}) \succeq_{(2)} Y$, $R(\boldsymbol{\lambda}^{*2}) \succeq_{(2)} Y$, but it does not usually hold that $R(\boldsymbol{\lambda}^{*2}) \succeq_{(1)} Y$ and $R(\boldsymbol{\lambda}^{*2}) \succeq_{(1)} R(\boldsymbol{\lambda}^{*1})$. Our goal is to analyze how far from it the solution $\boldsymbol{\lambda}^{*2}$ is. We will compute the corresponding measure of first non-dominance, and we introduce two other approaches to study how much must $\boldsymbol{\lambda}^{*2}$ change in order for $R(\boldsymbol{\lambda}^{*2}) \succeq_{(1)} Y$ to hold. These approaches are based on two ways in which we define the closest dominating portfolio.

Firstly, we search for a portfolio whose weights are as close as possible to $\boldsymbol{\lambda}^{*2}$ while dominating the benchmark Y with respect to FSD.

Secondly, we compute the measure of first non-dominance 1-ND$(R(\boldsymbol{\lambda}^{*2}), Y)$. $\hat{X}$ emerges from this computation. It dominates Y with respect to FSD while being as close as possible to $R(\boldsymbol{\lambda}^{*2})$ with respect to the Wasserstein distance of order 2. We can then search for such weights $\boldsymbol{\omega}$ that lead to the distribution $R(\boldsymbol{\omega})$ being as close as possible to the distribution of $\hat{X}$ with respect to the Wasserstein distance of order 2. To simplify the notation, $\boldsymbol{\lambda}^*$ denotes $\boldsymbol{\lambda}^{*2}$ in this section.

We use the Euclidean distance to measure the distance of the vectors representing the weights of portfolios. The Euclidean distance of two vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$ is defined as follows:

$$d_E(\boldsymbol{\lambda}, \boldsymbol{\omega}) = \left(\sum_{i=1}^{n} (\lambda_i - \omega_i)^2 \right)^{\frac{1}{2}}.$$

The first approach leads to the following definition.

Definition 5.1 (Closest Strongly Dominating Portfolio)**.** *Let Y be the random return of a benchmark investment and let $\boldsymbol{\lambda}^*$ be the weights of an optimal portfolio with respect to the second-order stochastic dominance. We say that portfolio with weights $\boldsymbol{\omega}^S$ is the closest strongly dominating portfolio (CSD) to $\boldsymbol{\lambda}^*$, if it is the optimal solution of the following program:*

$$\begin{aligned} &\min_{\boldsymbol{\omega}} \; d_E(\boldsymbol{\lambda}^*, \boldsymbol{\omega}) \\ &\textit{subject to } R(\boldsymbol{\omega}) \succeq_{(1)} Y, \\ &\boldsymbol{\omega} \in \Lambda. \end{aligned} \tag{5.4}$$

We can rewrite the program (5.4) in a way that can be applied to our situation. The Euclidean distance of the two vectors can be rewritten according to its definition. The condition $R(\boldsymbol{\omega}) \succeq_{(1)} Y$ was rewritten for our situation in problem (5.2).

$$
\begin{aligned}
&\min_{\boldsymbol{\omega}} \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{n}(\lambda_i^* - \omega_i)^2 && && (5.5)\\
\text{subject to } &\sum_{i=1}^{n}\omega_i r_{it} \geq \sum_{s=1}^{T}\pi_{ts}y_s, && t = 1,\dots,T,\\
&\sum_{t=1}^{T}\pi_{ts} = 1, && s = 1,\dots,T,\\
&\sum_{s=1}^{T}\pi_{ts} = 1, && t = 1,\dots,T,\\
&\pi_{ts} \in \{0,1\}, && s = 1,\dots,T,\ \ t = 1,\dots,T,\\
&\boldsymbol{\omega} \in \Lambda.
\end{aligned}
$$

The optimal solution of this program, $\boldsymbol{\omega}^S$, represents the weights of the closest strongly dominating portfolio. The optimal value of the objective function is the squared Euclidean distance of $\boldsymbol{\omega}^S$ and $\boldsymbol{\lambda}^*$.

The second approach leads to the following definition.

Definition 5.2 (Closest Weakly Dominating Portfolio)**.** *Let Y be the random return of a benchmark investment and let $\boldsymbol{\lambda}^*$ be the weights of an optimal portfolio with respect to the second-order stochastic dominance. Let $\hat{X}$ be the random variable that emerges from the computation of 1-ND$(R(\boldsymbol{\lambda}^*), Y)$. We say that a portfolio with weights $\boldsymbol{\omega}^W$ is the closest weakly dominating portfolio (CWD) to $\boldsymbol{\lambda}^*$, if it is the optimal solution of the following program.*

$$
\begin{aligned}
&\min_{\boldsymbol{\omega}} d_2(\hat{X}, R(\boldsymbol{\omega})) \qquad (5.6)\\
&\textit{subject to } \boldsymbol{\omega} \in \Lambda.
\end{aligned}
$$

We can rewrite the problem (5.6) in a way that can be applied to our situation. Although both $\hat{X}$ and $R(\boldsymbol{\omega})$ are empirically distributed random variables, we cannot use the simplification of the Wasserstein distance for empirically distributed variables from Theorem 2.3 in the objective function. It is caused by the fact that we cannot order the returns $R(\boldsymbol{\omega})$ beforehand. So instead, we compute the Wasserstein distance using the procedure for two discretely distributed random variables from Theorem 2.2. Let $\hat{x}_t$ represent the t$^{\text{th}}$ return of $\hat{X}$. Then the program can be formulated in the following way:

$$
\begin{aligned}
&\min_{\xi_{ts},\boldsymbol{\omega}} \sum_{t=1}^{T}\sum_{s=1}^{T} \xi_{ts}(\hat{x}_t - z_s)^2 && && (5.7)\\
&\text{subject to } z_s = \sum_{i=1}^{n} \omega_i \cdot r_{is}, && s = 1,\dots,T, &&\\
&\sum_{s=1}^{T} \xi_{ts} = \frac{1}{T}, && t = 1,\dots,T, &&\\
&\sum_{t=1}^{T} \xi_{ts} = \frac{1}{T}, && s = 1,\dots,T, &&\\
&\xi_{ts} \geq 0, && s = 1,\dots,T,\;\; t = 1,\dots,T. &&\\
&\boldsymbol{\omega} \in \Lambda. && &&
\end{aligned}
$$

z_s represents the s^{th} return of the portfolio with weights $\boldsymbol{\omega}$. The optimal solution of this program, $\boldsymbol{\omega}^W$, represents the weights of the closest weakly dominating portfolio. The optimal value of the objective function is the squared Wasserstein distance of order 2 of $R(\boldsymbol{\omega}^W)$ and $\hat{X}$.

5.3 Data

We use the prices of twelve assets covered by the German stock index DAX in September 2021. DAX consisted of 30 companies traded on the Frankfurt Stock Exchange. For portfolio optimization, we consider only 12 assets, which had the highest weight in the index in September 2021 and which have been traded at the stock exchange at least since October 2007. As a result, we consider the following assets: Adidas, Allianz, BASF, Bayer, Daimler, Deutsche Börse, Deutsche Post, Infineon Technologies, Munich Re, SAP, Siemens, and Volkswagen Group.

We work with the monthly returns from October 2007 to September 2021. To compute the returns, we use close prices adjusted for splits and dividends as downloaded from *finance.yahoo.com*. The return of i^{th} asset at time t, r_{it}, is computed as $\frac{p_{it}-p_{it-1}}{p_{it-1}}$ where p_{it} is the price of i^{th} asset at time t. Basic characteristics of the monthly returns of the 12 considered stocks are presented in Table 5.1. It shows their mean, minimal, and maximal return, and the standard deviation of returns based on the whole data set covering returns from October 2007 to September 2021.

To find the optimal portfolios with respect to the first-order and the second-order stochastic dominance, we must set a benchmark investment. To keep the benchmark consistent throughout the period, we use a portfolio that consists of $\frac{1}{n}$ of each of n considered assets. It means in this case that the benchmark investment is a portfolio consisting of $\frac{1}{12}$ of each of the twelve considered companies. Its mean return, minimum, maximum, and standard deviation is also presented in Table 5.1.

In order to receive more than one result for each of the programs presented in Section 5.1 and Section 5.2, we split the data regarding stock returns from the last 14 years into 7 disjoint data sets. Each of them consists of the returns in particular 2 years, hence of 24 observations. Tables presenting the basic characteristics of the

Name	Mean	Min	Max	Std
Adidas	0.015	-0.343	0.248	0.091
Allianz	0.008	-0.397	0.308	0.082
BASF	0.009	-0.228	0.303	0.08
Bayer	0.004	-0.243	0.196	0.073
Daimler	0.009	-0.267	0.423	0.101
Deutsche Borse	0.008	-0.222	0.247	0.075
Deutsche Post	0.013	-0.415	0.305	0.082
Infineon Technologies	0.021	-0.481	1.316	0.172
Munich Re	0.009	-0.202	0.163	0.057
SAP	0.01	-0.311	0.24	0.067
Siemens	0.008	-0.293	0.189	0.074
Volkswagen Group	0.013	-0.449	0.379	0.117
Benchmark	0.011	-0.27	0.259	0.065

Table 5.1: Basic characteristics calculated form monthly data.

considered assets in each of these periods separately can be found in Attachment A.1.

5.4 Results

We used Python for data preparation and for the presentation of results, and GAMS for solving the optimization problems. The scripts and the data are available in the attachment of this work. We used the CPLEX solver for the mixed-integer programming problem (5.2) as well as for the linear programming problem (5.3). We used the BARON solver for the mixed-integer non-linear programming problem (5.5), and the CONOPT solver for the nonlinear programming problem (5.7). It was necessary to set $\boldsymbol{\lambda}^{*2}$ as the starting value of $\boldsymbol{\omega}$ when solving the problem (5.7). It did not usually reach the globally optimal solution otherwise.

We computed the weights of optimal portfolios with respect to the first-order and the second-order stochastic dominance (we denote them FSD and SSD), the weights of the closest strongly dominating portfolio (CSD), and the weights of the closest weakly dominating portfolio (CWD) in each of the seven time periods. We present one table for each time period that shows them. The weights of assets that are not presented in the tables are 0 in the given period. If the weights do not add up precisely to one, it is caused by rounding errors. The tables show also the mean return of each of these portfolios. Table 5.2 shows the optimal portfolios in the first period.

It can be seen in Table 5.2 as well as in the following tables that the mean return of the optimal portfolio with respect to FSD is lower than or equal to the mean return of the optimal portfolio with respect to SSD. It corresponds to the theory because the feasible set in the first case is a subset of the feasible set in the second case. The mean return of the CSD portfolio is lower than the mean return of the FSD portfolio. It is correctly so because they must both dominate the benchmark with respect to FSD, and while the FSD portfolio is optimized to have the maximal mean return, the CSD portfolio is optimized to be as close

Name	FSD	SSD	CSD	CWD
Adidas	0.129	0	0.12	0.025
BASF	0.038	0	0.046	0
Bayer	0.002	0	0	0
Daimler	0	0	0.003	0
Deutsche Borse	0	0	0.004	0
Deutsche Post	0.2	0	0.193	0
Infineon Technologies	0.113	0.055	0.109	0.09
Munich Re	0.309	0.476	0.319	0.459
SAP	0.02	0	0.021	0
Siemens	0.055	0	0.053	0
Volkswagen Group	0.134	0.468	0.132	0.426
Mean return	0.00171	0.00517	0.00151	0.00534

Table 5.2: Optimal portfolios based on data from Oct. 2007 to Sept. 2009.

to the SSD portfolio as possible. The mean return of the CWD portfolio is similar to the mean return of the SSD portfolio. It is caused to the fact that the distribution of the optimal portfolio with respect to SSD violates first-order stochastic dominance of the benchmark only slightly. Therefore, the distribution of $\hat{X}$ is quite similar to the distribution of the SSD portfolio, and the distribution of the CWD portfolio is as similar as possible to the distribution of $\hat{X}$.

To gain a better understanding of the results, we show the empirical distribution functions (EDF) of the benchmark investment and of the optimal portfolios with respect to FSD and SSD in Figure 5.1.

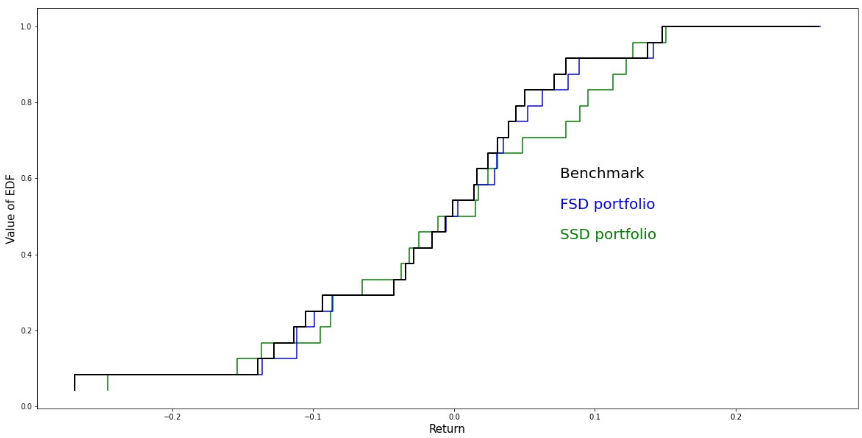

Figure 5.1: EDFs of optimal portfolios with respect to FSD and SSD based on data from Oct. 2007 to Sept. 2009.

The distribution function of the returns of the FSD portfolio is always below the distribution function of the benchmark investment. It satisfies the Alternative Definition of FSD from Theorem 1.1. The distribution function of the SSD portfolio intersects the distribution function of the benchmark investment several times. Although the distribution function of the SSD portfolio is higher than the distribution function of the benchmark investment at some points, which violates FSD, the difference is not large, and it does not happen for most of the points. This is the reason why the distribution of $\hat{X}$ is quite similar to it. It can be seen in Figure 5.2, which shows the distribution functions of the benchmark investment, $\hat{X}$ and the CWD portfolio.

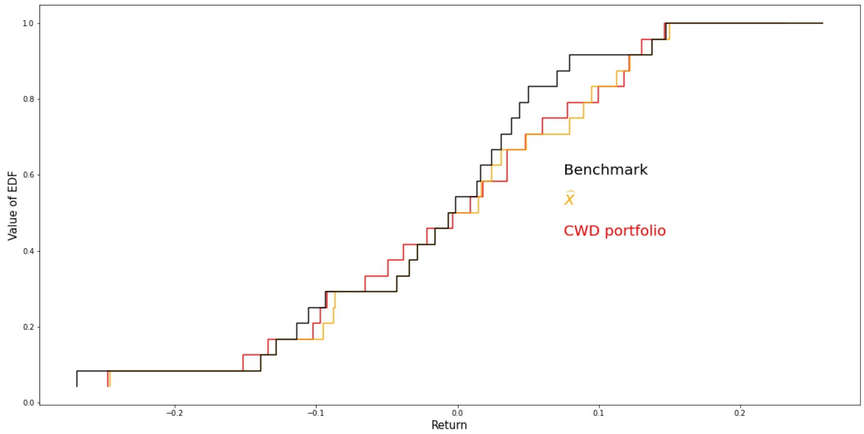

Figure 5.2: EDFs of $\hat{X}$ and the CWD portfolio based on data from Oct. 2007 to Sept. 2009.

The distribution function of $\hat{X}$ is always below the distribution function of the benchmark which is correct because $\hat{X}$ is supposed to dominate the benchmark with respect to FSD. The distribution of the CWD portfolio is supposed to be as close as possible to the distribution of $\hat{X}$ but it does not have to dominate the benchmark with respect to FSD. It can be seen in Figure 5.2 that it really does not dominate the benchmark with respect to FSD since its distribution function intersects the distribution function of the benchmark investment.

To present a complete set of graphs for this time period, we present Figure 5.3, which shows also the distribution function of returns of the CSD portfolio. A distribution function of the FSD portfolio is presented for comparison. It can be seen that they are very similar. On the other hand, the distribution of the CSD portfolio noticeably differs from the distribution of the SSD portfolio (see Figure 5.1 for comparison), even though the CSD portfolio's weights are as similar as possible to the SSD portfolio's weights. The distribution function of the benchmark investment is also shown. It can be seen that the distribution function of the CSD portfolio is always below it, which implies that the CSD portfolio dominates the benchmark with respect to FSD, as it should.

We proceed by presenting Table 5.3 which shows the optimal portfolios in the following period.

The properties of portfolios presented in Table 5.3 are similar to those of portfolios presented in Table 5.2. As before, the mean return of the CWD portfolio is similar to the mean return of the SSD portfolio. This time, however, it is slightly lower. This can also happen. The mean of $\hat{X}$ is always greater than or equal to the mean of the SSD portfolio. But, the CWD portfolio may have higher as well

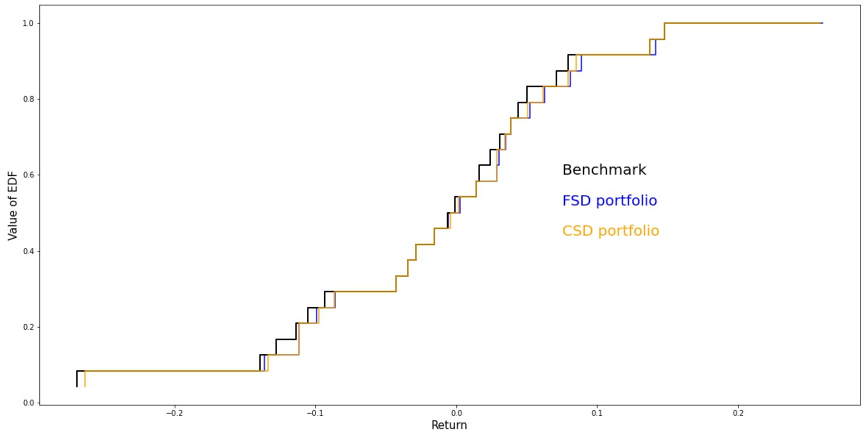

Figure 5.3: EDFs of the CSD portfolio compared to other investments based on data from Oct. 2007 to Sept. 2009.

Name	FSD	SSD	CSD	CWD
Adidas	0.045	0.081	0.067	0.083
Allianz	0	0	0.003	0.017
BASF	0.384	0.473	0.396	0.443
Daimler	0.009	0	0.013	0.01
Deutsche Borse	0.039	0	0.059	0.001
Infineon Technologies	0.155	0.17	0.161	0.173
SAP	0.277	0.229	0.257	0.222
Siemens	0	0	0	0.001
Volkswagen Group	0.09	0.048	0.044	0.051
Mean return	0.01421	0.01558	0.01362	0.01522

Table 5.3: Optimal portfolios based on data from Oct. 2009 to Sept. 2011.

as lower mean return than $\hat{X}$.

Table 5.4 shows the optimal portfolios in the third period. Deutsche Post had the highest mean return of all considered assets in the period from October 2011 to September 2013, which can be seen in Table A.3 in the attachment. Its distribution also dominates the distribution of the benchmark investment with respect to FSD in this period. Therefore, the optimal portfolio with respect to FSD consists of Deutsche Post only. Its mean return cannot be improved by adding any other asset to the portfolio. Obviously, it dominates the benchmark also with respect to SSD so the SSD optimal portfolio is the same. As a result, the CSD and the CWD portfolios are the same as well.

The optimal portfolios in the following time periods have similar properties. The following tables present their weights.

Weights as low as 0.001 can be seen in the CWD portfolios in Table 5.3 and Table 5.6. It is not a numerical error. The distribution of the CWD portfolio is very slightly closer to the distribution of $\hat{X}$ with them than without them.

It can be seen in Table 5.8 that the optimal portfolios with respect to FSD

Name	FSD	SSD	CSD	CWD
Deutsche Post		1		
Mean return		0.04451		

Table 5.4: Optimal portfolios based on data from Oct. 2011 to Sept. 2013.

Name	FSD	SSD	CSD	CWD
Adidas	0	0	0.047	0
Bayer	0.081	0	0	0.002
Deutsche Borse	0.743	0.768	0.757	0.772
Infineon Technologies	0.176	0.232	0.191	0.223
Munich Re	0	0	0.005	0.003
Mean return	0.01745	0.01764	0.01514	0.01762

Table 5.5: Optimal portfolios based on data from Oct. 2013 to Sept. 2015.

Name	FSD	SSD	CSD	CWD
Adidas	0.694	0.667	0.674	0.669
Deutsche Borse	0	0	0.005	0
Deutsche Post	0.024	0	0	0
Infineon Technologies	0.282	0.333	0.302	0.329
Munich Re	0	0	0	0.001
Volkswagen Group	0	0	0.018	0.001
Mean return	0.04357	0.04361	0.04325	0.04358

Table 5.6: Optimal portfolios based on data from Oct. 2015 to Sept. 2017.

Name	FSD	SSD	CSD	CWD
Adidas	0.273	0	0.269	0
Deutsche Borse	0.722	1	0.723	0.991
Deutsche Post	0	0	0.008	0
Munich Re	0.005	0	0	0
Volkswagen Group	0	0	0	0.009
Mean return	0.02026	0.02155	0.02012	0.02144

Table 5.7: Optimal portfolios based on data from Oct. 2017 to Sept. 2019.

Name	FSD	SSD	CSD	CWD
Deutsche Post		0.481		
Infineon Technologies		0.519		
Mean return		0.03752		

Table 5.8: Optimal portfolios based on data from Oct. 2019 to Sept. 2021.

and SSD are the same in the last considered time period. Unlike in the period from October 2011 to September 2013, they consist of more than one asset. Infineon Technologies had the highest mean return in this period, but its minimal return, −0.276, was lower than the minimal return of the benchmark (see Table A.7). Therefore, its returns cannot dominate the benchmark with respect to FSD nor with respect to SSD. Deutsche Post had the second-highest return in this period, and their combination presented in Table 5.8 dominates the benchmark investment with respect to FSD. Its minimal return, −0.185, is equal to the minimal return of the benchmark investment. It cannot be lower in order for the FSD or SSD to hold. Therefore, the SSD optimal portfolio is the same as the FSD, and so are the CSD and CWD portfolios.

In every period, the optimal portfolio with respect to SSD contains only assets which are represented in the optimal portfolio with respect to FSD. To dominate the benchmark with respect to FSD, the portfolio has to be more diversified than when the requirement is to dominate the benchmark with respect to SSD. This sometimes allows the SSD portfolio to contain fewer assets than the FSD portfolio consists of. It contains the ones with the higher mean return.

The closest strongly dominating portfolio contains all of the assets that the SSD portfolio consists of in every period. Sometimes, it also contains all of the assets that the FSD portfolio consists of as can be seen in Table 5.3. Sometimes, it contains at least some of the assets, which are in the FSD portfolio and are not in the SSD portfolio, which can be seen in Table 5.2 and Table 5.7. This could be expected because, just like the FSD portfolio, it is supposed to dominate the benchmark with respect to FSD so it is understandable that it is more similar to it. But, in all five periods when the FSD and the SSD portfolios differ, the CSD portfolio contains also assets that are not included in the FSD or SSD portfolios, which may be surprising

Based on our results, the closest weakly dominating portfolio contains all of the assets that the SSD portfolio consists of. We believe that this is a likely but not inevitable outcome. As already mentioned, the distribution of $\hat{X}$ can be very similar to the distribution of the SSD portfolio. Therefore, also the distribution of the CWD portfolio can be very similar so it tends to consist of similar assets. But, as can be seen in Table 5.3, Table 5.6, or Table 5.7, it may include completely different assets which are not included in the SSD or in the FSD portfolio. It is not as surprising this time because when constructing the CWD portfolio, we are trying to imitate a theoretical distribution of $\hat{X}$ which is unrelated to the FSD portfolio. The mean return of the CWD portfolio is higher than the mean return of the FSD portfolio in all cases. It implies that the CWD portfolio does not dominate the benchmark with respect to FSD because the FSD portfolio has the highest possible return while dominating the benchmark with respect to FSD.

The optimal portfolios with respect to FSD and SSD are different in five out of the seven considered time periods. We provide an overview of the measures of non-dominance and of the distances of weights between them in these five time periods in Table 5.9.

It can be seen in Table 5.9 that the higher the non-dominance is, the higher the distance of the weights is. It is consistent with the natural expectation that, if the portfolios are more similar, the measure of non-dominance between them is lower. It means that when the non-dominance was small, the optimal portfolio

Time Period	1-ND($\boldsymbol{\lambda}^{*2}, \boldsymbol{\lambda}^{*1}$)	$d_E(\boldsymbol{\lambda}^{*2}, \boldsymbol{\lambda}^{*1})$
2007-2009	$16.79 \cdot 10^{-3}$	$45.24 \cdot 10^{-2}$
2009-2011	$3.25 \cdot 10^{-3}$	$12.25 \cdot 10^{-2}$
2013-2015	$2.34 \cdot 10^{-3}$	$10.16 \cdot 10^{-2}$
2015-2017	$2 \cdot 10^{-3}$	$6.22 \cdot 10^{-2}$
2017-2019	$5.37 \cdot 10^{-3}$	$38.97 \cdot 10^{-2}$

Table 5.9: Comparison of the measure of non-dominance and distance between optimal portfolios with respect to SSD and FSD.

with respect to SSD did not need to be changed much in order for it to dominate FSD portfolio with respect to FSD.

Table 5.10 provides an overview of the measures of non-dominance and distances of weights between the optimal portfolios with respect to SSD and the closest strongly dominating portfolios in the five periods when they differ.

Time Period	1-ND($\boldsymbol{\lambda}^{*2}, \boldsymbol{\omega}^{S}$)	$d_E(\boldsymbol{\lambda}^{*2}, \boldsymbol{\omega}^{S})$
2007-2009	$16.55 \cdot 10^{-3}$	$44.47 \cdot 10^{-2}$
2009-2011	$2.73 \cdot 10^{-3}$	$10.26 \cdot 10^{-2}$
2013-2015	$1.81 \cdot 10^{-3}$	$4.85 \cdot 10^{-2}$
2015-2017	$1.13 \cdot 10^{-3}$	$3.65 \cdot 10^{-2}$
2017-2019	$5.12 \cdot 10^{-3}$	$38.58 \cdot 10^{-2}$

Table 5.10: Comparison of the measure of non-dominance and distance between the optimal portfolio with respect to SSD and the CSD portfolio.

The distance of weights between $\boldsymbol{\lambda}^{*2}$ and $\boldsymbol{\omega}^{S}$ (Table 5.10) is noticeably lower than the distance of weights $\boldsymbol{\lambda}^{*2}$ and $\boldsymbol{\lambda}^{*1}$ (Table 5.9) in every period. It is correct because both $\boldsymbol{\lambda}^{*1}$ and $\boldsymbol{\omega}^{S}$ dominate the benchmark with respect to FSD but $\boldsymbol{\omega}^{S}$ was found as the closest portfolio to $\boldsymbol{\lambda}^{*2}$ which satisfies that.

As in the previous table, it holds also in Table 5.10 that the higher the measure of first non-dominance is, the higher the distance of portfolio weights is. This ordering holds also when we combine the data from both tables. However, it does not mean that these measures can fully substitute one another, which can be seen from the following example. Let us focus on the third row of Table 5.9 and on the second row of Table 5.10. The Euclidean distance of weights is very similar in these two cases: $10.16 \cdot 10^{-2}$ and $10.26 \cdot 10^{-2}$. The second one is higher by approximately 1 %. However, the measures of non-dominance are $2.34 \cdot 10^{-3}$ and $2.73 \cdot 10^{-3}$. The second one is higher by 17 %. So, although it holds in this data set that the measure of non-dominance increases with increasing distance of weights, they do not increase proportionately. The measure of non-dominance brings a different type of information about the distance of portfolios so it cannot be replaced by the measure of distance.

To conclude, we have seen that the weights of the SSD portfolio can be noticeably more similar to the weights of the CSD portfolio than to the weights of the FSD portfolio. So, the FSD portfolio is quite far from being the closest dominating portfolio to the SSD portfolio. The CSD portfolio dominates the benchmark with respect to FSD by definition. On the other hand, the CWD portfolio did not dominate the benchmark with respect to FSD in any of the presented cases. The measure of non-dominance between portfolios evolves similarly to their Euclidean distance but it brings a different type of information.

Conclusion

We introduced stochastic dominance of different types in the first chapter (FSD, SSD, TSD, DARA SD, and ISD). We presented their definitions using the expected values of utility functions of compared random variables and also the alternative definitions using the distribution functions of the random variables. They were useful for deriving stochastic dominance rules for random variables with particular distributions.

We defined a measure of non-dominance in the second chapter. It quantifies how much must a random variable change to dominate another random variable with respect to stochastic dominance of a given order. In general, it is found by solving an optimization problem, but we derived its special form for empirically distributed random variables.

We computed the exact values of the measure of first and second non-dominance for uniformly, and normally distributed random variables. The measure of second non-dominance corresponds to the measure of third non-dominance in case of uniform distributions, and to the measures of any higher non-dominance in case of normal distributions. We derived a measure of non-dominance also for exponentially distributed variables. It is the same for all of the considered orders of stochastic dominance. This allows us to measure the quality of stochastic dominance approximation exactly in these distributions.

We were not able to formulate the optimization problem defining non-dominance for log-normally and gamma distributed random variables. A numerical study was used to approximate the measures of first and second non-dominance in these distributions. We approximated them by empirical distributions consisting of 1000 atoms and computed the measure of non-dominance between them. Although there were simulation errors, the results were overall in accordance with our expectations. When gamma distribution coincided with exponential distribution, we saw that the measure of first non-dominance was estimated quite well. However, estimation of the measure of second non-dominance showed weaknesses of our approach.

We concentrated on portfolio optimization in the fifth chapter and applied the theory to real-life data. We used monthly returns of twelve assets covered by the German stock index DAX. We found optimal portfolios with respect to FSD and SSD in seven different time periods, and we computed how far are they from each other. It was expressed by the measure of non-dominance as well as by the distance of their weights. We defined the closest strongly dominating and the closest weakly dominating portfolios. The closest weakly dominating portfolios did not dominate the benchmark in the presented cases, but the closest strongly dominating portfolios did. They provide a different view on how far the optimal portfolio with respect to SSD is from dominating the benchmark with respect to FSD.

In conclusion, we introduced and applied a new measure of non-dominance, which allows us to measure the quality of stochastic dominance approximation. We showed that it can be computed exactly for some distributions, and estimated for other distributions. It can be used also to measure the distance of portfolios, and to assess how far are they from dominating the benchmark. Further research could focus on searching its exact values for more distributions. One could also consider using a different measure of distance in its definition, which may allow its exact computation for more distributions such as the log-normal and gamma distribution. It could be also interesting to focus on improving the estimations of the measure of second non-dominance. Imposing more rules on the empirically distributed $\hat{X}$ in this case, may be helpful.

Bibliography

M. M. Ali. Stochastic dominance and portfolio analysis. *Journal of Financial Economics*, 2(2):205–229, 1975.

K. J. Arrow. *Aspects of the Theory of Risk-Bearing*. Yrjo Jahnsson Foundation, Helsinki, 1965.

D. Dentcheva and A. Ruszczynski. Semi-infinite probabilistic optimization: First order stochastic dominance constraints. *Optimization*, 53:583–601, 2004.

D. Dentcheva and A. Ruszczynski. Portfolio optimization with stochastic dominance constaints. *Journal of Banking and Finance*, 30(2):433–451, 2006.

K. Kozmík. Robust approaches in portfolio optimization with stochastic dominance. *Master's thesis, KPMS MFF UK*, 2019.

T. Kuosmanen. Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10):1390–1406, 2004.

H. Levy. *Stochastic Dominance: Investment Decision Making under Uncertainty*. 2nd Edition. Springer, New York, 2006.

J. Mikulka. Stochastická dominance vyšších řádů. *Master's thesis, KPMS MFF UK*, 2011.

H. Okagbue, M. O. Adamu, and T. A. Anake. Approximations for the inverse cumulative distribution function of the gamma distribution used in wireless communication. *Heliyon*, 6(11):e05523, 2020.

G. Ch. Pflug and A. Pichler. *Multistage Stochastic Optimization*. Springer, 2014.

T. Post, Y. Fang, and M. Kopa. Linear tests for decreasing absolute risk aversion stochastic dominance. *Management Science*, 61(7):1615–1629, 2015.

J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2): 122–136, 1964.

G. A. Whitmore. *Stochastic Dominance for the Class of Completely monotonic Utility functions*. Springer-Verlag, New York, 1989.

List of Figures

List of Tables

List of Abbreviations

FSD	First-order stochastic dominance
SSD	Second-order stochastic dominance
TSD	Third-order stochastic dominance
DARA SD	Decreasing absolute risk aversion stochastic dominance
ISD	Infinite-order stochastic dominance
nSD	n^{th}-order stochastic dominance
ND	Non-dominance
CSD	Closest strongly dominating
CWD	Closest weakly dominating
EDF	Empirical distribution function

A. Attachments

A.1 Data Characteristics in Each Period

Name	Mean	Min	Max	Std
Adidas	0.001	-0.343	0.248	0.127
Allianz	-0.015	-0.397	0.18	0.131
BASF	-0.001	-0.228	0.254	0.112
Bayer	-0.002	-0.167	0.125	0.071
Daimler	-0.017	-0.243	0.423	0.147
Deutsche Borse	-0.013	-0.222	0.247	0.122
Deutsche Post	-0.003	-0.415	0.305	0.147
Infineon Technologies	0.017	-0.481	1.316	0.39
Munich Re	-0.003	-0.095	0.14	0.057
SAP	-0.003	-0.267	0.151	0.088
Siemens	-0.007	-0.293	0.186	0.118
Volkswagen Group	0.012	-0.449	0.358	0.178
Benchmark	-0.003	-0.27	0.259	0.106

Table A.1: Basic characteristics of monthly data from Oct. 2007 to Sept. 2009.

Name	Mean	Min	Max	Std
Adidas	0.012	-0.15	0.244	0.104
Allianz	-0.0	-0.212	0.141	0.082
BASF	0.017	-0.214	0.137	0.085
Bayer	-0.001	-0.197	0.097	0.068
Daimler	0.005	-0.256	0.211	0.093
Deutsche Borse	-0.009	-0.197	0.112	0.083
Deutsche Post	-0.006	-0.135	0.109	0.066
Infineon Technologies	0.023	-0.205	0.283	0.121
Munich Re	-0.0	-0.118	0.106	0.054
SAP	0.008	-0.129	0.108	0.05
Siemens	0.008	-0.196	0.189	0.069
Volkswagen Group	0.018	-0.169	0.22	0.116
Benchmark	0.006	-0.17	0.113	0.063

Table A.2: Basic characteristics of monthly data from Oct. 2009 to Sept. 2011.

Name	Mean	Min	Max	Std
Adidas	0.028	-0.106	0.159	0.06
Allianz	0.027	-0.132	0.146	0.069
BASF	0.023	-0.088	0.154	0.062
Bayer	0.035	-0.05	0.117	0.044
Daimler	0.031	-0.092	0.245	0.094
Deutsche Borse	0.021	-0.188	0.133	0.067
Deutsche Post	0.045	-0.055	0.144	0.047
Infineon Technologies	0.017	-0.165	0.199	0.092
Munich Re	0.023	-0.076	0.109	0.042
SAP	0.017	-0.079	0.139	0.058
Siemens	0.017	-0.074	0.119	0.053
Volkswagen Group	0.029	-0.093	0.267	0.091
Benchmark	0.026	-0.082	0.134	0.047

Table A.3: Basic characteristics of monthly data from Oct. 2011 to Sept. 2013.

Name	Mean	Min	Max	Std
Adidas	-0.007	-0.219	0.128	0.08
Allianz	0.013	-0.065	0.093	0.045
BASF	0.003	-0.087	0.138	0.057
Bayer	0.015	-0.099	0.133	0.059
Daimler	0.009	-0.12	0.167	0.064
Deutsche Borse	0.018	-0.086	0.149	0.055
Deutsche Post	0.004	-0.108	0.066	0.046
Infineon Technologies	0.017	-0.094	0.128	0.069
Munich Re	0.011	-0.129	0.082	0.048
SAP	0.005	-0.091	0.086	0.046
Siemens	-0.001	-0.096	0.11	0.051
Volkswagen Group	-0.015	-0.423	0.137	0.106
Benchmark	0.006	-0.074	0.081	0.043

Table A.4: Basic characteristics of monthly data from Oct. 2013 to Sept. 2015.

Name	Mean	Min	Max	Std
Adidas	0.047	-0.104	0.193	0.07
Allianz	0.018	-0.091	0.136	0.052
BASF	0.017	-0.136	0.107	0.061
Bayer	0.006	-0.15	0.119	0.066
Daimler	0.01	-0.173	0.218	0.087
Deutsche Borse	0.011	-0.046	0.098	0.046
Deutsche Post	0.022	-0.14	0.113	0.053
Infineon Technologies	0.036	-0.091	0.253	0.081
Munich Re	0.009	-0.11	0.095	0.053
SAP	0.023	-0.064	0.24	0.065
Siemens	0.021	-0.084	0.145	0.06
Volkswagen Group	0.02	-0.201	0.204	0.096
Benchmark	0.02	-0.081	0.125	0.045

Table A.5: Basic characteristics of monthly data from Oct. 2015 to Sept. 2017.

Name	Mean	Min	Max	Std
Adidas	0.017	-0.075	0.163	0.071
Allianz	0.01	-0.102	0.114	0.058
BASF	-0.008	-0.185	0.132	0.068
Bayer	-0.017	-0.18	0.154	0.081
Daimler	-0.008	-0.204	0.126	0.077
Deutsche Borse	0.022	-0.07	0.108	0.042
Deutsche Post	-0.003	-0.15	0.141	0.074
Infineon Technologies	-0.005	-0.235	0.189	0.088
Munich Re	0.016	-0.033	0.089	0.036
SAP	0.009	-0.107	0.112	0.057
Siemens	-0.004	-0.112	0.113	0.057
Volkswagen Group	0.01	-0.1	0.142	0.075
Benchmark	0.003	-0.081	0.089	0.045

Table A.6: Basic characteristics of monthly data from Oct. 2017 to Sept. 2019.

Name	Mean	Min	Max	Std
Adidas	0.007	-0.199	0.185	0.095
Allianz	0.004	-0.195	0.308	0.097
BASF	0.012	-0.187	0.303	0.092
Bayer	-0.005	-0.243	0.196	0.095
Daimler	0.03	-0.267	0.271	0.113
Deutsche Borse	0.004	-0.157	0.134	0.071
Deutsche Post	0.034	-0.141	0.22	0.076
Infineon Technologies	0.041	-0.276	0.262	0.103
Munich Re	0.008	-0.202	0.163	0.089
SAP	0.011	-0.311	0.128	0.088
Siemens	0.023	-0.165	0.164	0.081
Volkswagen Group	0.018	-0.281	0.379	0.125
Benchmark	0.016	-0.185	0.177	0.075

Table A.7: Basic characteristics of monthly data from Oct. 2019 to Sept. 2021.