UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
DEPARTMENT OF MATHEMATICS
AND INFORMATICS

mr Sanja Lončar

Negative Selection - An Absolute Measure of Arbitrary Algorithmic Order Execution

-PhD thesis-

Novi Sad, 2017

A man is born to work, to suffer and to fight;
he who doesn't, must perish.
Nikola Tesla

Abstract

Algorithmic trading is an automated process of order execution on electronic stock markets. It can be applied to a broad range of financial instruments, and it is characterized by a significant investors' control over the execution of his/her orders, with the principal goal of finding the right balance between costs and risk of not (fully) executing an order. As the measurement of execution performance gives information whether best execution is achieved, a significant number of different benchmarks is used in practice. The most frequently used are price benchmarks, where some of them are determined before trading (Pre-trade benchmarks), some during the trading day (Intraday benchmarks), and some are determined after the trade (Post-trade benchmarks). The two most dominant are VWAP and Arrival Price, which is along with other pre-trade price benchmarks known as the Implementation Shortfall (IS).

We introduce Negative Selection as a posteriori measure of the execution algorithm performance. It is based on the concept of Optimal Placement, which represents the ideal order that could be executed in a given time window, where the notion of ideal means that it is an order with the best execution price considering market conditions during the time window. Negative Selection is defined as a difference between vectors of optimal and executed orders, with vectors defined as a quantity of shares at specified price positions in the order book. It is equal to zero when the order is optimally executed; negative if the order is not (completely) filled, and positive if the order is executed but at an unfavorable price.

Negative Selection is based on the idea to offer a new, alternative performance measure, which will enable us to find the optimal trajectories and construct optimal execution of an order.

The first chapter of the thesis includes a list of notation and an overview of definitions and theorems that will be used further in the thesis. Chapters 2 and 3 follow with a theoretical overview of concepts related to market microstructure, basic information regarding benchmarks, and theoretical background of algorithmic trading. Original results are presented in chapters 4 and 5. Chapter 4 includes a construction of optimal placement, definition and properties of Negative Selection. The results regarding the properties of a Negative Selection are given in [35]. Chapter 5 contains the theoretical background for stochastic optimization, a model of the optimal execution formulated as a stochastic optimization problem with regard to Negative Se-

lection, as well as original work on nonmonotone line search method [31], while numerical results are in the last, 6th chapter.

Apstrakt

Algoritamsko trgovanje je automatizovani proces izvršavanja naloga na elektronskim berzama. Može se primeniti na širok spektar finansijskih instrumenata kojima se trguje na berzi i karakteriše ga značajna kontrola investitora nad izvršavanjem njegovih naloga, pri čemu se teži nalaženju pravog balansa između troška i rizika u vezi sa izvršenjem naloga. S ozirom da se merenjem performasi izvršenja naloga određuje da li je postignuto najbolje izvršenje, u praksi postoji značajan broj različitih pokazatelja. Najčešće su to pokazatelji cena, neki od njih se određuju pre trgovanja (eng. Pre-trade), neki u toku trgovanja (eng. Intraday), a neki nakon trgovanja (eng. Post-trade). Dva najdominantnija pokazatelja cena su VWAP i Arrival Price koji je zajedno sa ostalim "pre-trade" pokazateljima cena poznat kao Implementation shortfall (IS).

Pojam negative selekcije se uvodi kao "post-trade" mera performansi algoritama izvršenja, polazeći od pojma optimalnog naloga, koji predstavlja idealni nalog koji se mogao izvršiti u datom vremenskom intervalu, pri čemu se pod pojmom "idealni" podrazumeva nalog kojim se postiže najbolja cena u tržišnim uslovima koji su vladali u toku tog vremenskog intervala. Negativna selekcija se definiše kao razlika vektora optimalnog i izvršenog naloga, pri čemu su vektori naloga definisani kao količine akcija na odgovarajućim pozicijama cena knjige naloga. Ona je jednaka nuli kada je nalog optimalno izvršen; negativna, ako nalog nije (u potpunosti) izvršen, a pozitivna ako je nalog izvršen, ali po nepovoljnoj ceni.

Uvođenje mere negativne selekcije zasnovano je na ideji da se ponudi nova, alternativna, mera performansi i da se u odnosu na nju nađe optimalna trajektorija i konstruiše optimalno izvršenje naloga.

U prvom poglavlju teze dati su lista notacija kao i pregled definicija i teorema neophodnih za izlaganje materije. Poglavlja 2 i 3 bave se teorijskim pregledom pojmova i literature u vezi sa mikrostrukturom tržista, pokazateljima trgovanja i algoritamskim trgovanjem. Originalni rezultati su predstavljeni u 4. i 5. poglavlju. Poglavlje 4 sadrži konstrukciju optimalnog naloga, definiciju i osobine negativne selekcije. Teorijski i paraktični rezultati u vezi sa osobinama negativna selekcije dati su u [35]. Poglavlje 5 sadrži teorijske osnove stohastičke optimizacije, definiciju modela za optimalno izvršenje, kao i originalni rad u vezi sa metodom nemonotonog linijskog pretraživanja [31], dok 6. poglavlje sadži empirijske rezultate.

Acknowledgement

Firstly, I would like to thank my family and friends for their continuous support throughout writing this thesis.

My sincere thanks go to my advisor Professor Dr. Nataša Krejić for the support of my Ph.D. study and related research.

I would also like to thank Professor Dr. Zorana Lužanin and Professor Dr. Danijela Rajter - Ćirić for their insightful comments and encouragement.

· ∘ ◎ ⊙ ◎ ∘ ·

Acknowledgement to Dr. Miles Kumaresan

The topic of this research was proposed by Dr. Miles Kumaresan who acted as an industrial supervisor for the thesis. Dr. Kumaresan provided guidance in all sophisticated properties of markets and algorithmic trading that were needed for the obtained results. His insightful comments and generous help were essential for the work presented in the thesis.

Sanja Lončar
Novi Sad, June 6, 2017

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Algorithmic Trading is the ubiquitous way of trading at electronic stock exchanges. The advantage of such a trading is a significant control over the execution. It is a system of rules based on knowledge of trading, quantitative analysis, and programming with the primary concern to decide where and how to trade.

The idea is to find the right balance between risk and cost i.e. to solve trader's dilemma [27] which describes a tradeoff between market impact and timing risk. To solve this and make an optimal execution we must decide whether to trade aggressively or passively in the prevailing market conditions. For this purpose, there is a wide variety of orders, some even behave like algorithms, but most common are market and limit orders. In essence, they have opposed behavior regarding the aggressive and passive trading. Market orders are perceived as aggressive, while limit orders are viewed as passive. However, one must bear in mind that limit orders with price limit at the best bid for buying (or best ask for selling) order are also regarded as aggressive. The only thing that protects execution from unfavorable price, in the case when market order "walks the book," is the inbuilt price limit. The market order guarantees execution, it has a minimal timing risk but can produce large market impact. On the contrary, the limit order does not guarantee that order will be filled. Therefore, the significant timing risk is included, while the market impact is minimal because of the price limit of the order.

A performance measurement plays a great role in trading and execution, and for this purpose, there is a significant number of different benchmarks. Price benchmarks are the most frequently used in practice and are usually categorized by the time when they are determined, so there are pre-trade, intraday, and post-trade benchmarks. The two most dominant benchmarks are VWAP, and Arrival Price, which is along with other pre-trade price benchmarks known as the Implementation Shortfall (IS).

VWAP - the intraday benchmark was introduced in 1988 by Berkowitz, Logue, and Noser [7], as the alternative to Open-High-Low-Close benchmark in measuring market impact cost. It owes its popularity to the simplicity of its calculation, intuitiveness and the fact that it gives a realistic portray of market conditions [3, 10, 20, 21, 38, 41]. Cook [10] emphasizes two major drawbacks of the VWAP benchmark: the first one is that it could be gamed, and the other one is that it discourages investors to give an extra effort to achieve the best execution. Instead, they execute order throughout the day, to avoid slippage to benchmark. Freyre-Sanders, Guobuzaite, and Byrne [14] suggest that VWAP good benchmark only for trades which will make a small market impact.

Implementation Shortfall (IS) was introduced by Perold [44] and it uses the pre-trade benchmark arrival price as a proxy for the decision price in IS concept. The advantage od using the pre-trade benchmark is that it cannot be gamed, but it does not always give proper information of market conditions. Implementation Shortfall, as a difference between the performance of the theoretical and real portfolio, is regarded as one of the most reliable trading cost measurement. But as Cook states [10] having a low value of IS does not mean that execution was good, neither a high value of IS is an indicator of bad execution.

With all this in mind, the idea of this research was to introduce Negative Selection, as a new, alternative performance measure. We start with the notion of Optimal Placement, which represents the ideal order that could be executed in a given time window, where the notion of ideal means that this an order with the best execution price considering market conditions during the time window. Negative Selection is defined as the difference between vectors of optimal and executed orders, with vectors defined as quantities of shares at specified price positions in the order book. It is equal to zero when the order is optimally executed; negative if the order is not filled, and positive if the order is executed but at the unfavorable price.

The term "Negative Selection" is used by some authors as "Adverse Selection." Saraiya and Mittal [48] describes Adverse selection (Negative Selection) "an interaction between an uninformed trader and an informed trader" in Dark pools, where for uninformed trader exist the possibility of filling his/her order at an unfavorable price. In the same context, Self [50] states

”Negative Selection can be described as only ever receiving an execution when it would have been better not to trade.” However, here, the term is related to execution process and the execution algorithm’s performance with regard to price movements. If the price comes in our direction, one does not want to be filled too early. Therefore it is better to be more passive and try to execute at a more favorable price, which means the order being ”selected” will have sub-optimal execution.

The thesis is organized as follows. In Chapter 1 is given a list of notation and an overview of definitions and theorems. Chapters 2 and 3 follow with a theoretical overview of concepts related to market microstructure, basic information regarding benchmarks, and theoretical background of algorithmic trading. Original results are presented in chapters 4 and 5. Chapter 4 includes a construction of optimal placement, definition and properties of Negative Selection. The results regarding the properties of a Negative Selection are given in [35]. Chapter 5 contains the theoretical background for stochastic optimization, a model of the optimal execution formulated as a stochastic optimization problem with regard to Negative Selection, as well as original work on nonmonotone line search method [31]. Chapter 6 contains numerical results, obtained using simulator developed in Matlab and MySQL.

1.1 List of Notations

$\mathcal{N}$	Negative Selection value for simple order
$\mathcal{N}(\mathcal{S})$	Negative Selection vector for complex order $\mathcal{S}$
$x = [x_1, \ldots, x_n]^T$	real n-dimensional (column) vector
$A = [a_{ij}]_{m \times n}$	an $m \times n$ matrix with elements a_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$
$U(c)$	an upper triangular matrix with diagonal and all above elements equal to c
$L(c)$	a lower triangular matrix with diagonal and all lower elements equal to c
$\mathbf{1}_{m \times n}$	an $n \times m$ matrix with all elements equal to 1 i.e. $a_{ij} = 1$, $i = 1, \ldots, m$, $j = 1, \ldots, n$
$\mathbf{0}_{m \times n}$	an $n \times m$ matrix with all elements equal to 0 i.e. $a_{ij} = 0$, $i = 1, \ldots, m$, $j = 1, \ldots, n$
E	the identity matrix
$det(A)$	determinant of a matrix A
$rank(A)$	rank of a matrix A
$\mathbb{E}(X)$	mathematical expectation of a random variable X
$\mathbb{D}(X), \sigma^2(X)$	variance of a random variable X
$\sigma(X)$	standard deviation of a random variable X
$\|\| \quad \|\|_1$	l_1 norm
$\|\| \quad \|\|_2$	l_2 norm
$\nabla f(x)$	gradient of a function $f(x)$
$\nabla^2 f(x)$	Hessian of a function $f(x)$

1.2 List of Abbreviations

BB	Black Box
bps	Basis Points
i.i.d.	independently and identically distributed
IS	Implementation Shortfall
NS	Negative Selection
VWAP	Volume Weighted Average Price
w.p.1	with probability 1

1.3 Background Material

1.3.1 Linear Algebra and Analysis

For $x = [x_1, \ldots, x_n]^T$ we define the following vector norms.

Definition 1.3.1. *(1-norm or Sum-norm)*

$$||x||_1 = \sum_{i=1}^{n} |x_i|$$

Definition 1.3.2. *(2-norm or Eucledean norm)*

$$||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

Definition 1.3.3. *(∞ -norm or Sup-norm)*

$$||x||_\infty = \max_{1 \le i \le n} |x_i|$$

Definition 1.3.4. *(Matrix norm) Let $|| \quad ||$ be real-valued function on $\mathbb{R}^{n\times n}$, i.e. $|| \quad || : \mathbb{R}^{n\times n} \to \mathbb{R}$, with following properties:*

- $||A|| \ge 0$, *for all* $A \in \mathbb{R}^{n\times n}$ *and*

$$||A|| = 0 \text{ if and only if } A = 0.$$

- $||\alpha A|| = |\alpha| ||A||$, *for all* $A \in \mathbb{R}^{n\times n}$ *and* $\alpha \in \mathbb{R}$.
- $||A + B|| \le ||A|| + ||B||$, *for all* $A, B \in \mathbb{R}^{n\times n}$.
- $||AB|| \le ||A|| ||B||$, *for all* $A, B \in \mathbb{R}^{n\times n}$.

Definition 1.3.5. *Matrix norm $|| \quad ||$ defined on $\mathbb{R}^{n\times n}$ is induced by (or subordinate to) vector norm $|| \quad ||_v$ on $\mathbb{R}^n$ if and only if*

$$||A|| = \sup_{x \ne 0} \frac{||Ax||_v}{||x||_v}$$

These matrix norms are called operators or natural norms.

The following matrix norms are induced by 1-norm, 2-norm and ∞ -norm, respectively.

Definition 1.3.6. *(1-norm or Maximum absolute column sum norm)*

$$||A||_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |x_{ij}|.$$

Definition 1.3.7. *(2-norm or Eucledean-norm)*

$$||A||_2 = \sqrt{\rho(A^T A)},$$

where $\rho(A)$ is spectral radius of A, i.e. if λ_i $i = 1, \ldots, n$ are eigenvalues of matrix A, then

$$\rho(A) = \max_{i\leq i\leq n} |\lambda_i|.$$

Definition 1.3.8. *(∞ -norm or Maximum absolute row sum norm)*

$$||A||_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |x_{ij}|.$$

Let $S \subset I\!R^p$. The set of functions $f : S \to I\!R^q$ which are continuous on S is denoted by $C(S)$, and for $k \in I\!N$, and $C^k(S)$ denotes set of functions that have continuous kth derivatives.

Definition 1.3.9. *Let $S \subset I\!R^p$. The function $f : S \to I\!R^q$ is Lipschitz continuous on the set S if there exists constant L such that for every $x, y \in S$*

$$||f(x) - f(y)|| \leq L||x - y||$$

1.3.2 Probability Theory

Let Ω be the non-empty set of all logically possible outcomes of an experiment. Elements of the set Ω are called the *elementary events* or *states* and are denoted by ω.

Any subset of Ω is called *random event* or just *event*. The set Ω is called *sure* event, and empty set $\varnothing$ is *impossible* event. For any event A, we define its *complement* $\overline{A} = \Omega \setminus A$.

Definition 1.3.10. *Let $\mathcal{F} = \{A | A \subseteq \Omega\}$ be a collection of subsets of Ω, for which the following three properties hold:*

(1) $\Omega \in \mathcal{F}$

(2) $A \in \mathcal{F} \implies \overline{A} \in \mathcal{F}$

(3) $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$

Then $\mathcal{F}$ is called σ-algebra (σ-field) on Ω.

Definition 1.3.11. *Let Ω be a set of the elementary events and $\mathcal{F}$ σ-algebra on Ω. A map $P : \mathcal{F} \mapsto [0, 1]$ such that the following conditions are satisfied*

(1) $P(\Omega) = 1,$

(2) $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}, A_i \cap A_j = \varnothing, i, j \in \mathbb{N}, i \neq j \implies P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$

is called probability on the space $(\Omega, \mathcal{F})$, and triple $(\Omega, \mathcal{F}, P)$ is called the probability space.

Theorem 1.3.1. *Properties of probability*

(1) $P(\varnothing)=0$.

(2) If $A_1, \dots, A_n \in \mathcal{F}$ and $A_i \cap A_j = \varnothing$, $i, j = 1, \dots, n$, $i \neq j$, then

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

(3) For $A \in \mathcal{F}$, $P(\overline{A}) = 1 - P(A)$.

(4) For $A_1, \ldots, A_n \in \mathcal{F}$,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i,j=1\\ i<j}}^{n} P(A_i \cap A_j) + \cdots + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i).$$

Definition 1.3.12. *Let $(\Omega, \mathcal{F}, P)$ be the probability space and an event $A \in \mathcal{F}$ such that $P(A) > 0$. A map $P(\cdot|A) : \mathcal{F} \to I\!R$ defined in the following way*

$$P(B|A) = \frac{P(A \cap B)}{(A)},$$

is called conditional probability. $P(B|A)$ is the conditional probability of B given A.

Definition 1.3.13. *Let $(\Omega, \mathcal{F}, P)$ be the probability space. Two events $A, B \in \mathcal{F}$ are called independent if and only if*

$$P(A \cap B) = P(A) \cdot P(B).$$

Definition 1.3.14. *Let $(\Omega, \mathcal{F}, P)$ be the probability space. An arbitrary collection of events $A_i \in \mathcal{F}$, $i \in I$ are called independent if and only if for each finite set of distinct indices $i_1, i_2, \ldots, i_k \in I$ holds*

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \ldots \cdot P(A_{i_k}).$$

Let us consider topological space $(\mathbb{R}, \tau)$. The smallest σ-algebra generated by τ is called Borel σ-algebra on $\mathbb{R}$ and it is denoted by $\mathcal{B}$, i.e.

$$\mathcal{B} = \bigcap_{\widetilde{\mathcal{B}} \in \pi} \widetilde{\mathcal{B}},$$

where π is collection of σ-algebras, such that $\tau \subseteq \widetilde{\mathcal{B}}$. Elements of Borel σ-algebra are called Borel sets.

Definition 1.3.15. *Let $(\Omega, \mathcal{F}, P)$ be the probability space and $\mathcal{B}$ Borel σ-algebra on $\mathbb{R}$. A map $X : \Omega \to \mathbb{R}$ is called random variable if for every $B \in \mathcal{B}$ holds*

$$X^{-1}(B) = \{\omega | X(\omega) \in B\} \in \mathcal{F}.$$

Then, we say that X is $\mathcal{F}$-measurable.

Theorem 1.3.2. *For a probability space $(\Omega, \mathcal{F}, P)$ and random variable X, triple $(\mathbb{R}, \mathcal{B}, P_X)$, where*

$$P_X(B) = P(X^{-1}(B)),\ B \in \mathcal{B}$$

is probability space, which is called the probability space induced by random variable X.

Definition 1.3.16. *For random variable X on the probability space $(\Omega, \mathcal{F}, P)$ the cumulative distribution function (CDF) of X, denoted by F_X is defined as:*

$$F_X(x) = P_X((-\infty, x)) = P(\{\omega | X(\omega) < x\}) = P(X < x).$$

Theorem 1.3.3. *Let $(\Omega, \mathcal{F}, P)$ be the probability space and X random variable. The cumulative distribution function of X has the following properties:*

(1) F_X is left-continuous nondecreasing function.

(2) $\lim\limits_{x \to -\infty} F_X(x) = 0$.

(3) $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Definition 1.3.17. *Let $(\Omega, \mathcal{F}, P)$ be the probability space and $\mathcal{B}$ Borel σ-algebra on $\mathbb{R}$. Random variable X is said to be discrete if and only if there exists finite or countably infinite set $\mathbb{R}_X \subset \mathbb{R}$ such that $P(X \in \mathbb{R}_X) = 1$.*
If $\mathbb{R}_X = \{x_1, x_2, \ldots, x_k, \ldots\}$ then

$$\forall B \in \mathcal{B} \quad P(X \in B) = \sum_{x_k \in B} P(X = x_k).$$

Probability distribution function (PDF) for discrete random variable X is given by

$$X : \begin{pmatrix} x_1 & x_2 & \ldots & x_k & \ldots \\ p_1 & p_2 & \ldots & p_k & \ldots \end{pmatrix} \tag{1.1}$$

where $p_k = P(X = x_k)$.

Definition 1.3.18. *Let $(\Omega, \mathcal{F}, P)$ be the probability space and $\mathcal{B}$ Borel σ-algebra on $\mathbb{R}$. Random variable X is said to be absolutely continuous if and only if there is an integrable nonnegative function φ_X such that*

$$\forall B \in \mathcal{B} \quad P(X \in B) = \int_B \varphi_X(x) dx.$$

φ_X is called density function of X. The cumulative distribution function is then defined by

$$F_X = P(X < x) = \int_{-\infty}^{x} \varphi_X(s)ds.$$

For absolutely continuous random variable X and every interval $(a, b] \subset \mathbb{R}$ there holds

$$P_X((a, b]) = P(a < X \leq b) = \int_a^b \varphi_X(s)ds = F_X(b) - F_X(a).$$

Density function φ_X is piecewise continuous and for an arbitrary point x, such that φ_X is continuous in x, holds

$$\varphi_X(x) = \frac{dF_X(x)}{dx}$$

and for every Borel set B holds

$$\int_B dF_X(s) = \int_B \varphi_X(s)ds.$$

Independence of random variables will be defined using definition of independent events:

Definition 1.3.19. *An arbitrary collection of random variables $X_1, X_2, \ldots X_n$ are called independent if and only if the events $X_1^{-1}(S_1), X_2^{-1}(S_2), \ldots X_n^{-1}(S_n)$ are independent for each $S_i \in \mathcal{B}$, $i = 1, 2, \ldots, n$*

Definition 1.3.20. *Let X be a discrete random variable with probability distribution function given by (1.1). Mathematical expectation of X, denoted by $\mathbb{E}(X)$ is defined by*

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k p_k.$$

It exists if and only if $\sum_{k=1}^{\infty} |x_k| p_k < \infty$.

Definition 1.3.21. *Let X be an absolutely continuous random variable with density function φ. Mathematical expectation of X, denoted by $\mathbb{E}(X)$ is defined by*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\varphi(x)dx.$$

It exists if and only if $\int_{-\infty}^{\infty} |x|\varphi(x)dx < \infty$.

Theorem 1.3.4. *Properties of mathematical expectation*

(1) $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

(2) $\mathbb{E}(c) = c$ *where c=const.*

(3) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$, *where* $\alpha, \beta \in \mathbb{R}$.

(4) $\mathbb{E}(X - \mathbb{E}(X)) = 0$.

(5) *If for each* $\omega \in \Omega$, $X(\omega) \geq 0$ *then* $\mathbb{E}(X) \geq 0$.

(6) *If for each* $\omega \in \Omega$, $X(\omega) \geq Y(\omega)$ *then* $\mathbb{E}(X) \geq E(Y)$.

(7) *If* X *and* Y *are two independent random variables such that* $\mathbb{E}(X) < \infty$ *and* $\mathbb{E}(Y) < \infty$ *then*

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Theorem 1.3.5. *(Fundamental theorem of mathematical expectation)*
Let X *be a random variable and* $f : \mathbb{R} \to \mathbb{R}$ *Borel function, then the following holds*

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} f(x)dF_X(x)$$

Definition 1.3.22. *Variance of random variable X, denoted by* $\mathbb{D}(X)$ *or* σ_X^2, *is defined as*

$$\sigma_X^2 = \mathbb{D}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Theorem 1.3.6. *Properties of variance*

(1) $\mathbb{D}(X) \geq 0$.

(2) $\mathbb{D}(X) = 0$ *if and only if* $X = const$ *almost surely.*

(3) $\mathbb{D}(\alpha X) = \alpha^2\mathbb{D}(X)$, *where* $\alpha \in \mathbb{R}$.

(4) $\mathbb{D}(X + \alpha) = \mathbb{D}(X)$, *where* $\alpha \in \mathbb{R}$.

(5) If X and Y are two independent random variables such that

$$\mathbb{D}(X) < \infty \text{ and } \mathbb{D}(Y) < \infty$$

then

$$\mathbb{D}(X+Y) = \mathbb{D}(X) + \mathbb{D}(Y).$$

Definition 1.3.23. *Standard deviation of random variable X is defined as*

$$\sigma_X = \sqrt{\mathbb{D}(X)}.$$

Definition 1.3.24. *Let X and Y be two random variables. If there exist $\mathbb{E}(X)$, and $\mathbb{E}(Y)$, then the covariance of two random variables X and Y, denoted by $cov(X,Y)$ or $\sigma_{X,Y}$ is defined as*

$$\sigma_{X,Y} = cov(X,Y) = \mathbb{E}((X - \mathbb{E}(X) \cdot (Y - \mathbb{E}(Y)) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Definition 1.3.25. *A sequence of random variables $\{X_n\}_{n\in I\!N}$ converge in probability to random variable X (denoted by $X_n \xrightarrow{P} X$), if for every $\epsilon > 0$*

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

Definition 1.3.26. *A sequence of random variables $\{X_n\}_{n\in I\!N}$ converge almost surely (or with probability one) to random variable X (denoted by $X_n \xrightarrow{a.s.} X$), if*

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

Definition 1.3.27. *A sequence of random variables $\{X_n\}_{n\in I\!N}$ converge in mean-square to random variable X (denoted by $X_n \xrightarrow{L^2} X$), if*

$$\mathbb{E}(X_n^2) < \infty, n \in I\!N \text{ and } \lim_{n\to\infty} \mathbb{E}((X_n - X)^2) = 0.$$

Let $\{X_n\}_{n\in I\!N}$ be sequence of independent random variables in probability space $(\Omega, \mathcal{F}, P)$. Let $\overline{X}_n$ be the sample mean of the first n terms of the sequence,

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n.$$

We consider a sequence $\{\overline{X}_n - \mathbb{E}(\overline{X}_n)\}_{n\in I\!N}$ and state a set of conditions that are sufficient to guarantee its convergence to zero. For the convergence is in probability, we have Weak Laws of Large Numbers, and for the almost sure convergence, Strong Laws of Large Numbers.

Theorem 1.3.7. *If there exists a constant C such that for all $n \in I\!N$, $\mathcal{D}(X_n) \leq C$ then for a sequence $\{\overline{X}_n - \mathbb{E}(\overline{X}_n)\}_{n\in I\!N}$ holds a Weak Law of Large Numbers, i.e.*

$$\overline{X}_n - \mathbb{E}(\overline{X}_n) \xrightarrow{P} 0, n \to \infty$$

Theorem 1.3.8. *If is the sequence $\{X_n\}_{n\in I\!N}$ of independently and identically distributed random variables with finite mathematical expectation $\mathbb{E}(X_n) = \mu < \infty$, then then for a sequence $\{\overline{X}_n - \mathbb{E}(\overline{X}_n)\}_{n\in I\!N}$ holds a Weak Law of Large Numbers.*

1.3.3 Constrained Optimization

In general, optimization problem can be written as

$$\begin{aligned} &\min_{x\in\mathbb{R}^p} f(x) \\ &s.t.\ h_i(x) = 0,\ i \in \mathcal{E} \\ &\quad\ \ h_i(x) \geq 0,\ i \in \mathcal{I}, \end{aligned} \tag{1.2}$$

where $\mathcal{E}, \mathcal{I}$ are sets of indices, $x \in I\!R^p$, $f : I\!R^p \to I\!R$, $h_i : I\!R^p \to I\!R$, $i \in \mathcal{E} \cup \mathcal{I}$. x is the vector of *variables*, $f = f(x)$ is the objective function, and the functions $h_i = h_i(x),\ i \in \mathcal{E}$ are equality constraints, and $h_i = h_i(x),\ i \in \mathcal{I}$ are inequality constraints.

We consider only minimization problem because maximization of a function f subject to some constraints is equivalent to minimization of the function $-f$ under the same conditions.

The distinction between optimization problems can be made regarding whether the problem has constraints or not, or according to properties (linear, nonlinear, convex, stochastic, etc.) of the objective function f and constraints h_i. In the former case, we differ constrained and unconstrained optimization. And in the latter case, there are nonlinear optimization problems, convex optimization, etc. One extensively used class of these problems, where all constraints and objective function are linear, is linear programming problem.

We begin with convexity - one of fundamental concepts in optimization. The term *convex* is used both to sets and functions.

Definition 1.3.28. *A set $S \in \mathbb{R}^p$ is a* convex set *if the straight line segment connecting any two points in S lies entirely inside S.*

$$(\forall\ x, y \in S)\ (\forall\ \alpha \in [0,1])\ (\alpha x + (1-\alpha)y \in S)$$

Definition 1.3.29. *A function f is a* convex function *if its domain S is a convex set and for any two points x and y in domain, the graph of f lies below the strait line connecting $(x, f(x))$ and $(y, f(y))$ in the space $\mathbb{R}^{p+1}$.*

$$(\forall\ x, y \in S)\ (\forall\ \alpha \in [0,1])\ \big(f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)\big)$$

A function f is said to be concave *if the function $-f$ is convex.*

To introduce basic concepts regarding constraint optimization, we define feasible set $\mathcal{S}$, consisting of all points that satisfy constraints in (1.2):

$$\mathcal{S} = \{x|\ h_i(x) = 0,\ i \in \mathcal{E};\quad h_i(x) \geq 0,\ i \in \mathcal{I}\}.$$

Now, problem (1.2) can be rewritten as:

$$\min_{x \in S}\ f(x) \tag{1.3}$$

Definition 1.3.30. *A vector x^* is a local solution of problem (1.3) if $x^* \in \mathcal{S}$ and there is a neighborhood $\mathcal{N}$ of x^* such that*

$$f(x) \geq f(x^*) \text{ for } x \in \mathcal{N} \cap \mathcal{S}.$$

Definition 1.3.31. *A vector x^* is a strict (strong) local solution of problem (1.3) if $x^* \in \mathcal{S}$ and there is a neighborhood $\mathcal{N}$ of x^* such that*

$$f(x) > f(x^*) \text{ for } x \in \mathcal{N} \cap \mathcal{S} \text{ with } x \neq x^*.$$

Inequality constraint $h_i(x)$ is said to be *active* at $x \in \mathcal{S}$ if $h_i(x) = 0$, and *inactive* if $h_i(x) > 0$.

FIRST ORDER OPTIMALITY CONDITIONS

For the constrained optimization problem (1.2) the Lagrangian $\mathcal{L} : \mathbb{R}^p \times \mathbb{R}^{|\mathcal{E}|+|\mathcal{I}|}$ is defined as

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x). \tag{1.4}$$

At any feasible x we define active set $\mathcal{A}(x)$, as the union of indices of all equality and active inequality constraints.

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I}|\ h_i(x) = 0\}. \tag{1.5}$$

Definition 1.3.32. *Given the point x^* and an active set $\mathcal{A}(x^*)$ defined by (1.5), we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients $\{\nabla h_i(x^*), i \in \mathcal{A}(x^*)\}$ is linearly independent.*

Theorem 1.3.9. *(First-Order Necessary Conditions)*
Suppose that x^* is a local solution of (1.2)and that LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with comonents λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{1.6}$$

$$h_i(x^*) = 0, \text{ for all } i \in \mathcal{E} \tag{1.7}$$

$$h_i(x^*) \geq 0, \text{ for all } i \in \mathcal{I} \tag{1.8}$$

$$\lambda_i^* \geq 0, \text{ for all } i \in \mathcal{I} \tag{1.9}$$

$$\lambda_i^* h_i(x^*) = 0, \text{ for all } i \in \mathcal{E} \cup \mathcal{I} \tag{1.10}$$

The conditions (1.6-1.10) are called *Karush-Kuhn-Tucker condition*, or shortly *KKT conditions*.

For each inactive inequality constraint h_i at x^*, $i \notin \mathcal{A}(x^*)$ holds $h_i(x^*) > 0$. The complementary conditions (1.10) then imply that $\lambda_i = 0$ for all $i \notin \mathcal{A}(x^*)$, which allows us to rewrite (1.6) as:

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = f(x) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla h_i(x^*) = 0. \tag{1.11}$$

Definition 1.3.33. *(Strict Complementarity)*
Given a local solution x^* of (1.2) and a vector λ^* satisfying (1.6-1.10) we say that the strict complementarity condition holds if exactly one of λ_i^* and $h_i(x^*)$ is zero for each index $i \in \mathcal{I}$. In other words, we have that

$$\lambda_i^* > 0, \quad i \in \mathcal{I} \cap \mathcal{A}(x^*).$$

Now, we define the tangent cone F_1 to the feasible set at x^*, when constraint the qualifications are satisfied.

Definition 1.3.34. *Given a point x^* and active constraint $\mathcal{A}(x^*)$ defined by (1.5) the set F_1 is defined by*

$$F_1 = \{\alpha w | \alpha > 0, w^T \nabla h_i(x^*) = 0 \ \forall i \in \mathcal{E}, w^T \nabla h_i(x^*) \geq 0 \ \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}\}$$

SECOND ORDER OPTIMALITY CONDITIONS

The first-order conditions i.e. the KKT conditions, give information about first derivatives of f, active constraints and their relationship at x^*. And the second-order conditions deal with the Lagrangian function in the directions $w \in F_1$ for which $w^T \nabla f(x^*) = 0$.

To continue with the second-order conditions we will assume that the objective function f, and constraints h_i, $i \in \mathcal{E} \cup \mathcal{I}$ are twice continuously differentiable on $S \subset I\!R^p$, i.e. $f, h_i \in C^2(S)$, $i \in \mathcal{E} \cup \mathcal{I}$.

Now, for a given Lagrange multiplier λ^* satisfying the KKT conditions, we are able to define a subset of F_1, denoted by $F_2(\lambda^*)$ in following way:

$$F_2(\lambda^*) = \{w \in F_1 | w^T \nabla h_i(x^*) = 0, \text{ all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0\}. \quad (1.12)$$

Equivalently, $w \in F_2(\lambda^*)$ if and only if the following three conditions are satisfied:

$$\begin{cases} w^T \nabla h_i(x^*) = 0, & \text{for all } i \in \mathcal{E}, \\ w^T \nabla h_i(x^*) = 0, & \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0, \\ w^T \nabla h_i(x^*) w \geq 0, & \text{for all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* = 0. \end{cases} \quad (1.13)$$

The following theorem gives us the necessary condition using the second derivative:

Theorem 1.3.10. *(Second-Order Necessary Conditions)*
Suppose that x^* is a local solution of (1.2) and that the LICQ condition is satisfied. Let λ^* be the Lagrange multiplier vector such that the KKT conditions are satisfied, and let $F_2(\lambda^*)$ be defined as (1.12). Then

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0 \textit{ for all } w \in F_2(\lambda^*). \quad (1.14)$$

Contrary to the necessary conditions, where we assumed that x^* is a local solution, and then were able to infer what are properties of the objective function f and constraints h_i, now we impose conditions on f and h_i to ensure that x^* is a local solution.

Theorem 1.3.11. *(Second-Order Sufficient Conditions)*
Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is Lagrange multiplier vector λ^* such that KKT conditions are satisfied. Suppose also that

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0 \textit{ for all } w \in F_2(\lambda^*),\ w \neq 0. \quad (1.15)$$

Then x^* is a strict local solution for (1.2).

RATES OF CONVERGENCE

Rate of convergence is an important measure of performance of an optimization algorithm. We now define different types of convergence for a sequence $\{x^k\}_{k=0}^{\infty}$, where $x^k \in \mathbb{R}^n$ and $\lim_{k\to\infty} x^k = x^*$. We say that convergence is:

Q-linear if there is constant $r \in (0, 1)$ such that

$$\frac{||x^{k+1} - x^*||}{||x^k - x^*||} \leq r,$$

for all k sufficiently large. The quantity r is called the asymptotic rate of convergence.

Q-superlinear if the following holds

$$\lim_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = 0.$$

Q-quadratic if

$$\frac{||x^{k+1} - x^*||}{||x^k - x^*||^2} \leq M,$$

for all k sufficiently large, where $M < \infty$ is a positive constant.

R-linear if there is a sequence of nonnegative scalars $\{\nu_k\}_{k=0}^{\infty}$, such that

$$||x^k - x^*|| \leq \nu_k, \text{ for all } k$$

and $\{\nu_k\}$ converges Q-linearly to zero.

R-superlinear if there is a sequence of nonnegative scalars $\{\nu_k\}_{k=0}^{\infty}$, such that

$$||x^k - x^*|| \leq \nu_k, \text{ for all } k$$

and $\{\nu_k\}$ converges Q-superlinearly to zero.

R-quadratic if there is a sequence of nonnegative scalars $\{\nu_k\}_{k=0}^{\infty}$, such that

$$||x^k - x^*|| \leq \nu_k, \text{ for all } k$$

and $\{\nu_k\}$ converges Q-quadratically to zero.

1.3.4 Linear Programming

Linear programming problems are optimization problems such that all constraints and objective function are linear. We now define terminology associated with linear programming.

Definition 1.3.35. *General form of linear programming problem (LPG) is a special case of optimization problem (1.2), where objective function $f(x)$, equality and inequality constraints are linear functions. It is defined as:*

$$\begin{aligned} \min f(x) &= c^T x \\ s.t.\ Ax &= b \\ Cx &\geq d \\ x_j &\geq 0, j \in P, P \subseteq \{1, 2, \ldots, n\} \end{aligned} \tag{1.16}$$

where $x = [x_1, x_2, \ldots, x_n]^T \in I\!R^n$, $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $d \in \mathbb{R}^p$.

Definition 1.3.36. *The linear programming problem has the standard form (LPS) if it is defined in following way:*

$$\begin{aligned} \min f(x) &= c^T x \\ s.t.\ Ax &= b \\ x &\geq 0 \end{aligned} \tag{1.17}$$

where $x = [x_1, x_2, \ldots, x_n]^T \in I\!R^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

Before defining basic solution, we first define the feasible set of the LPS problem (1.17):

$$U_S = \{x | x \in \mathbb{R}^n, Ax = b, x \geq 0\} \tag{1.18}$$

Definition 1.3.37. *We consider an LPS problem, where $rank(A) = r = m$. A feasible solution $x^* \in U_S$ is basic (feasible) solution of problem (1.17) if there exist r-linearly independent columns of matrix A A_{i_j}, $j = 1, \ldots, r$ of a matrix A such that $\{i_1, \ldots, i_r\} \subset \{1, 2, \ldots, n\}$ and*

$$\sum_{j=1}^{r} x_{i_j} A_{i_j} = b,$$

where

$$A_k = \begin{bmatrix} a_{1\,k} \\ a_{2\,k} \\ \vdots \\ a_{r\,k} \end{bmatrix}$$

The components $x^*_{i_1}, \ldots, x^*_{i_r}$ *of vector* x^* *are called basic variables, the other components are called nonbasic and are equal to zero.*

Definition 1.3.38. *If one or more of the basic variables has value zero, that solution is said to be a degenerate basic solution. Otherwise, it is called non-degenerate basic solution.*

Definition 1.3.39. *The column vectors* $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$*, are basic vectors for basic feasible solution* x^**. The matrix* $B = [A_{i_1}, A_{i_2}, \ldots, A_{i_r}]$ *is a basic matrix for basic feasible solution* x^**.*

Definition 1.3.40. *A linear programming problem (1.17) is said to be degenerate (singular) if it has at least one degenerate basic solution. Otherwise, it is non-degenerate (regular) linear problem.*

Definition 1.3.41. *Let* S *be a convex set. The point* $x \in S$ *is an extreme point (vertex) of the set* S *if it cannot be expressed in the form*

$$x = \alpha y + (1 - \alpha) z,$$

where $0 < \alpha < 1$*,* $y, z \in S$ *and* $y, z \neq x$*.*

Definition 1.3.42. *Closed bounded convex set with a finite number of extreme points is called convex polyhedron.*

Theorem 1.3.12. *(Representation of convex polyhedron)*
Let K *be convex polyhedron with extreme points* $v_1, v_2, \ldots, v_k$*. For an arbitrary* $x \in K$ *holds*

$$x = \sum_{i=1}^{k} \alpha_i v_i$$

where

$$\sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \geq 0,\ i = 1, \ldots, k\ ,$$

i.e. convex polyhedron is convex combination of its extreme points.

Theorem 1.3.13. *The LPG problem (1.16) has no solution if the feasible set S is empty, or the objective function is unbounded on non-empty feasible set, i.e. there exists a sequence $\{x^k\} \subset S$ such that*

$$\lim_{k \to \infty} f(x^k) = -\infty.$$

Otherwise, the problem has a solution and then can occur exactly one of the following two cases:

(1) The problem has a unique solution, that is an extreme point of S.

(2) The problem has an infinite number of solutions, which are convex combination of extreme points, which are solutions of the problem.

Theorem 1.3.14. *A point x is an extreme point of the set U_S defined as (1.18) if and only if it is a basic feasible solution of the LPS problem (1.17).*

To introduce term *reduced cost*, we consider the LPS problem (1.17) with $rank(A) = r = m$ and the feasible set U_S defined as (1.18) without degenerate extreme points. Let x be an extreme point with base $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$. Without loss of generality we assume that first r columns of a matrix A are basic vectors, hence it can be written in form

$$A = [B \quad N],$$

where B is a basic matrix, and N is a nonbasic matrix. Now, the extreme point x can be written as

$$x = \begin{bmatrix} x_B \\ x_N \end{bmatrix},$$

where x_B is the vector of basic variables and x_N the vector of nonbasic variables.

From the previous notation, it is obvious that $x_B = B^{-1}b$ and for each feasible point $y \in U_S$, using the notation as for x (y_B the vector of basic variables, y_N the vector of nonbasic variables) holds

$$y_B = x_B - B^{-1}Ny_N.$$

Now, the objective function has the following value in y:

$$
\begin{aligned}
f(y) &= c^T y \\
&= c_B^T y_B + c_N^T y_N \\
&= c_B^T x_B - (c_B^T B^{-1} N - c_N^T) y_N \\
&= f(x) - (c_B^T B^{-1} N - c_N^T) y_N .
\end{aligned}
$$

Here, c_B and c_N are coefficients of basic and nonbasic variables in c, respectively. With the matrix N written in form $N = [A_{r+1}, \ldots, A_n]$ and the previous result, there holds

$$
f(y) = f(x) - \sum_{i=r+1}^{n} \big(c_B^T B^{-1} A_i - (c_N)_i\big)(y_N)_i = f(x) - \sum_{i=r+1}^{n} \Delta_i (y_N)_i
$$

The coefficients Δ_i are called reduced costs of extreme point x. It can be proved that, if there exists $j \in \{1, \ldots n\}$ such that $\Delta_j > 0$, then the extreme point x is not solution of the LPS problem.

Chapter 2

Market Microstructure

Market microstructure has a vital importance for finance practitioners, as it is one of the fields of financial research with rapid growth in recent twenty years. Stoll [53] writes: "[It] deals with the purest form of financial intermediation - the trading of financial asset, such as a stock or bond."

The term "market microstructure" was introduced by Mark Garman [16] in 1976 where he describes "the 'temporal microstructure' or moment-to-moment trading activities in asset market." O'Hara [42] defines it as "the study of process and outcomes of exchanging assets under explicit trading rules," while Madhavan [36] gives the following definition: "Studies of market microstructure analyze the process by which investors' latent demands are translated into executed trades."

Market microstructure theory consists of three key areas [21]:

- Market structure and design,
- Trading mechanism research,
- Transition cost measurement and analysis.

By many academic studies, market structure and design have a significant influence on quality and speed of price discovery, liquidity as well as, on the total cost of trading. On the other hand, the focus of trading mechanism research area is mainly in price formation, price discovery, and trade execution. The area of transaction cost measurement and analysis covers usage of spreads and benchmarks as cost measurement and analysis of other cost components like volatility risk, market impact, opportunity cost, etc. Some of the key features of market microstructure will be described in the following sections.

2.1 Types of Markets

Financial markets are venues that accommodate requirements for both investors and issuers of financial instruments. One of the key characteristics of market structure and design is a market type, which could influence overall transaction cost, and consequently, the profitability of a trade. Markets are usually classified by their mechanism and frequency of trading.

2.1.1 Trading mechanism

By trading mechanism, markets can be separated into three groups:

- quote-driven (dealer market),
- order-driven (auction market),
- hybrid or mix of previous two.

In an entirely quote-driven market, investors cannot transact directly with each other but only with a dealer, who quotes ask and bid prices at which he/she is willing to buy and sell particular quantity. In an order-driven market investors trade without dealer intermediation, and it allows equal participation of all the investors which is actualized through placing and matching their orders in order book by specific rules.

The difference between quote-driven and order-driven markets manifests through levels of visibility in terms of orders and bid and ask prices. In quote-driven market one can see only market maker's two-way quote (to buy given quantity at ask price, or to sell at bid price), which guarantees execution at that price, for set size (see Table 2.1):

Bid size	Bid price	Ask price	Ask size
1000	101.0	101.5	2500

Table 2.1: Example of market maker's two-way quote. (Primer dvosmernih kotacija)

This way market maker is the one who provides liquidity to the market and takes on the risk of a less favorable position, but in return, the difference in bid and sell price enables him/her to earn a profit.

A significant advantage of the order-driven market is its transparency, because of visibility of prices and corresponding order sizes of both buyers and sellers, which are organized and displayed in an order book. The other benefit is that they provide visible liquidity and more options for placing an order (at any given price and size). The drawback of this market is that the execution of an order is not guaranteed.

Markets that combine properties of order-driven and quote-driven markets are called hybrid markets, an example of such market is NASDAQ [21].

2.1.2 Trading frequency

By frequency of trading, markets can be also separated in three groups:

- Continuous trading,
- Periodic trading,
- Request-driven trading.

In continuous trading, orders can be traded at any point in time and are executed as soon as they are received. Periodic, or scheduled trading or call auctions are organized at specified times in the day. Request-driven trading involves requesting a quote from a market maker.

Continuous trading is preferable if one needs an immediacy of execution, but the drawback is possible price volatility. Periodic trading is often used for less liquid assets (when continuous trading cannot be sustained). Also, it can have an effect on reducing price volatility.

Some markets made the transition from periodic to continuous trading and some use combination of the two. The trend is "continuous trading based on hybrid mechanisms, often with additional call auctions to help start and/or close the market." [21].

2.2 Order Book

In a pure order-driven market, which is the majority of the world's financial markets, order book has a central role in the trading process. Roşu [46] gives examples of pure order-driven markets: Euronext, Hong Kong, Tokyo, Toronto, and different ECN (Island, Instinet). He also lists examples of hybrid exchanges like NYSE, Nasdaq, and London Stock Exchange.

Hasbrouck [19], states that *consolidated limit order book* (CLOB) is used in most Asian and European markets.

The term CLOB is used when all trading for a security takes place in a single order book. The same acronym is used for the term *centralized limit order book*, and both represent the same thing. CLOB is also referred to as a double-sided auction.

The order is a building block of an order book, and it represents a trade instruction. For a given financial instrument, an order is determined by its

- direction (buy or sell),
- price,
- size and
- time when it is submitted.

Nowadays the most prevalent priority mechanism is *price-time* also known as FIFO (first in first out). This means that price priority is primary, and orders with better price are given a higher priority. Secondary priority is based on time of order submission. Consequently, previously listed properties allow explicit determining order priority in a limit order book.

There is also *pro-rata* priority mechanism, but it is commonly used in futures markets.

Before specifying some properties of limit order book (LOB), we continue with its definition:

Connor, Goldberg, and Korajczyk [9] define it in the following way: "The menu of outstanding limit orders is called the limit order book." While Narang [41] gives a definition with additional intuitive insight into its mechanism:

> "The collection of all available bids and offers (all of which are passive orders) for a given security is known as the limit order book, which can be thought of as a queue of limit orders to buy and sell."

For a given limit order book, both price and size of an order depend directly on *tick size* and *lot size*, collectively referred to as *resolution parameters.*

The *tick size* of a limit order book is the smallest permissible price increment. The consequence of mandatory tick size is a discrete set of prices and its direct effect on bid-ask spread. Harris [18] noted, since tick size is a cost of getting priority, with a too small tick size, one can cheaply step in front of a queue, just by slightly improving his order price, a phenomenon often called "pennying." In this case, time priority becomes meaningless.

The *lot size* is the smallest quantity of the financial instrument that can be traded within the limit order book. All submitted orders must have a size that is multiple of a lot size. A lot size can vary from one to hundred, or even thousand and more shares. Because of the practice to break large market orders in smaller blocks, a lot size directly influences traders execution strategy.

The *depth* of a limit order book at a specified price is the total number of shares of all the active orders at that price. It is often expressed as multiple of lot size and usually influenced by tick size. Larger tick size gives incentive for providing liquidity.

The *best bid price* at the time t denoted by $b(t)$ is the highest price of all buy orders at the time t.

The *best ask price* at the time t denoted by $a(t)$ is the lowest price of all sell orders at the time t.

The *mid price* at the time t denoted by $m(t)$ is arithmetic mean of best ask and best bid price at time t:

$$m(t) = \frac{a(t) + b(t)}{2}.$$

The *bid-ask spread* at the time t denoted by $s(t)$ is a difference between best ask and best bid price at time t:

$$s(t) = a(t) - b(t).$$

The *proportional bid-ask spread* at the time t is defined in the following way:

$$\frac{a(t) - b(t)}{m(t)}.$$

It is relative measure of spread and usually it is given in basis points (bps).

Liquidity plays a great role in trading, and it represents a cost of converting financial instrument to cash and vice versa. The notion was first introduced by Demsetz [11] using term "immediacy." Zubulake and Lee [58] define it as "the amount of a security that is available on the bid/buy and offer/sell of a market, as well as the depth of both buyers and sellers."

2.3 Order Types

Each market allows different order types, and the two most important ones are *market* and *limit* orders.

Market orders are instructions to trade a given size immediately at the best available market price. They are liquidity takers and are exposed to the risk of unfavorable execution price.

Limit orders are instructions to trade given size at the specified price or better. They provide liquidity and are exposed to the risk of not being executed.

It is worth noting that there is another order type, which is, in essence, a hybrid of market and limit order, and it is called *Marketable Limit Order*. Kissell and Glantz [26] explain:

> "This order will either be executed in the market at the specified price or better, or be cancelled if there are no existing orders at that price or better in the market."

A trader can apply additional conditions to order, to achieve better control of execution, these are fill instructions, duration instructions, etc. Conditions regarding fill instructions include:

- Immediate-or-cancel: is an order for which the part that cannot be immediately executed at given price will be cancelled, i.e. order can be partially filled.

- Fill or kill: is a limit order that must be completely filled immediately or is automatically cancelled.

- All-or-none: is a limit order that must be completely filled, but not immediately, and expires at the end of the day.

- Minimum-volume: is an order with the condition that some specified quantity must be filled. For example at Euronext, a minimum-volume order will be cancelled if it cannot fill the required minimum quantity.
- Must-be-filled: is an order which must be completely filled. Johnson [21] notes that " [they] are generally associated with trading to fulfill expiring futures or option contracts."

Order's "lifetime" lasts from its submission to limit order book until it is cancelled or completely filled. Duration instructions specify additional conditions to alter its lifetime, for example, "Day Order" or "good for day" (GFD) represent order that will be cancelled at the end of the day. Other instructions of this type include: "good 'til date" (GTD), "good after time/date" (GAT), etc.

Many exchanges support and create new types of orders, which are actually a combination of market and a limit orders. Besides hybrid orders, these are [21]:

- Conditional orders (stop-loss orders, trailing stop orders, contingent orders, tick sensitive orders)
- Hidden and iceberg orders
- Discretional orders (not-held, pegged orders)
- Routed orders (pass-through orders)

Some of them are even behaving like algorithms, and are collectively referred as *dynamic order types*, the examples are stop orders and pegging orders.

2.4 Aggressive and Passive Trading

To achieve the best execution, which by Kissell and Glantz [26] depends on price, size, and time factors, a trader must find a balance between passive and aggressive trading illustrated by Figure 2.1. He/she must solve trader's dilemma formulated by Kissell [27] in the following way:

> "Trading too aggressive will lead to higher impact cost but trading too passively will lead to higher risk and may result in even more costly trades."

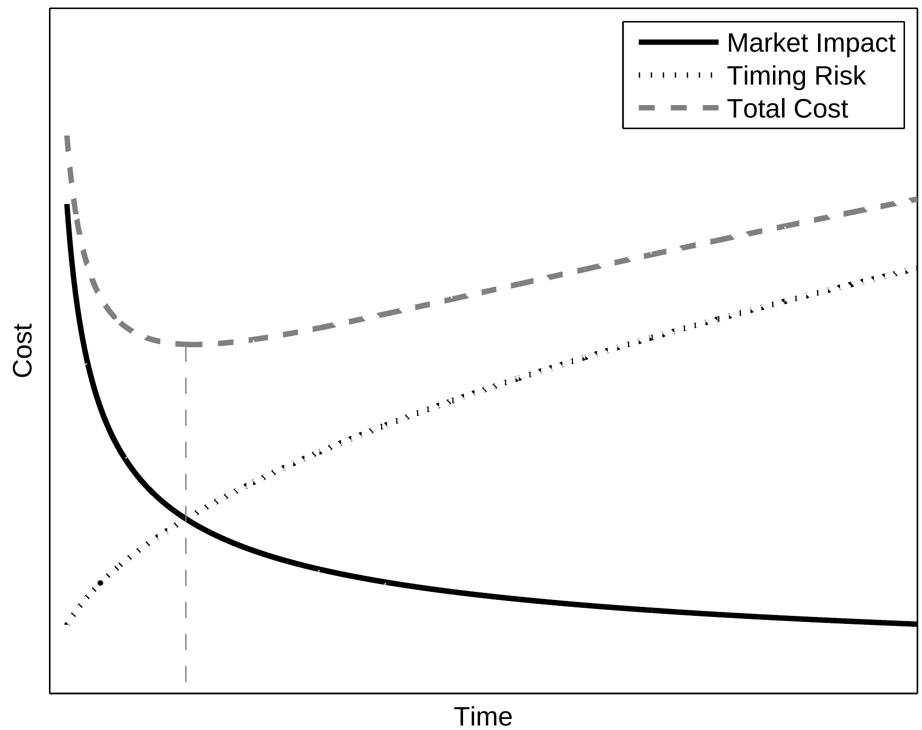

Figure 2.1: Trade-off between market impact and timing risk. (Odnos između tržnog impakta i vremenskog rizika)

Narang [41] notes that a difference between passive and aggressive trading is reflected in "how immediately a trader wants to do a trade." With that in mind, it can be said that there is a trade-off between cost and risk: aggressive trading style leads to a higher cost and a lower risk, while passive trading style is associated with a lower cost and a higher risk. In general, using market orders is considered aggressive, and using limit orders passive trading.

Using market orders guarantees fast execution, but with the price that cannot be controlled. For example, if a size of market buy order does not exceed quantity at best ask, the order will be completely filled at this price. On the other hand, if order's size exceeds quantity at best ask, then it will "walk the book" and take available liquidity until it is filled. With a market order, there is always a cost of crossing the spread.

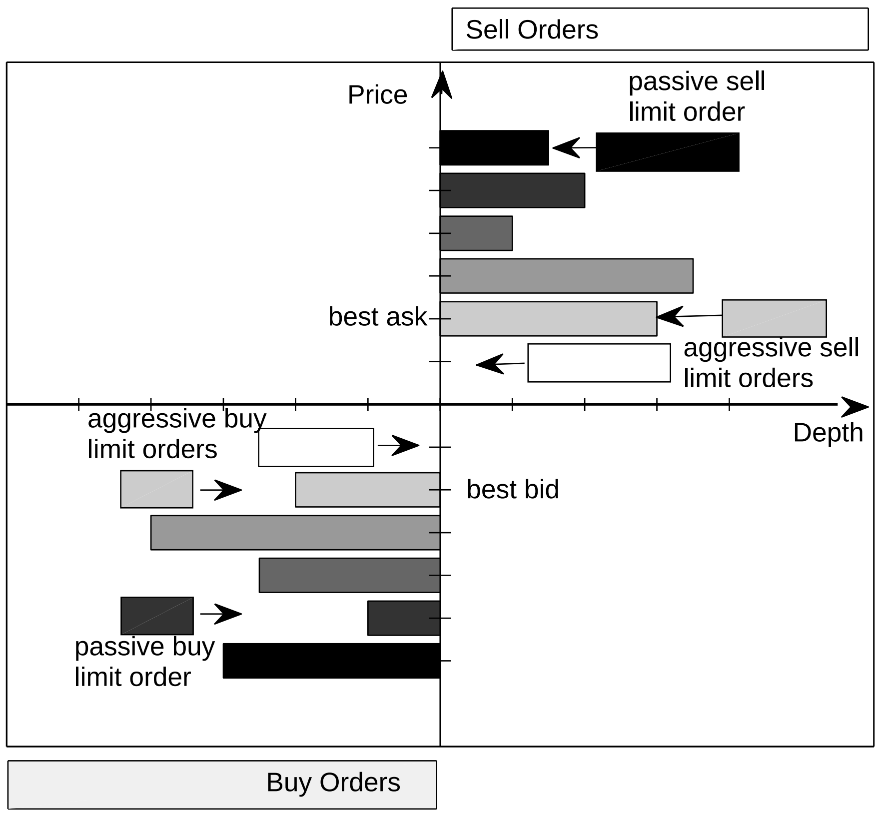

Figure 2.2: Aggressive and passive limit orders in the limit order book. Agresivni i pasivni limit nalozi u knjizi naloga

Limit orders are opposite of market orders. They have an inbuilt price limit but do not guarantee execution. However, using limit orders can also be aggressive: for example placing an order at the best bid or best ask price (also known as *joining*) or even adding it to limit order book, so that it will create a new best bid or best ask (known as *improving*). Orders submitted at price levels that are far lower than best bid (or higher than best ask) are considered passive.

Figure 2.2 illustrates aggressive and passive orders, both on buy and sell side. White and light gray rectangles represent aggressive limit orders that are *improving* and *joining*, respectively. Passive limit orders in the limit order book are in black and dark gray color.

A style of execution depends on many factors, Narang [41] explains that momentum strategies are prone to aggressive style, while mean reversion strategies tend to be passive. But, the deciding factor is a signal strength.

There is psychological research in the field of behavioral finance. Madhavan [36] states that "traders tend to overestimate the precision of their information." Further, some traders tend to overreact, while the other overreact to new information, which leads to a different understanding of signal strength.

2.5 Transaction costs

As the transaction cost are included in each trade, they play a major role in trading. In 1988 Perold introduced the measure of total transaction cost, called *Implementation Shortfall*. In his paper [44], he proposes to run a theoretical (paper) portfolio alongside the real one and calculate implementation shortfall as the difference between the performance of the theoretical and real portfolio. Further, he separates the two basic components of Implementation Shortfall:

- *Execution cost* - represents the cost of transacting like taxes and commissions, and also include market impact
- *Opportunity cost* - represents "the cost of not transacting" i.e. measures performance of theoretical orders that were not executed in reality.

Narang [41] emphasizes the importance of estimation of transaction costs. He explains that with underestimation of transaction costs trader makes "too many trades that have insufficient benefit," and with overestimation, the trader makes fewer trades "which usually results in holding positions too long." Therefore, proper estimation of transaction cost is crucial for overall trading performance. Pre-trade and post-trade analysis enables identification and modeling of transaction costs and therefore taking appropriate action for reducing the overall costs.

Almgren et al [4] classifies costs into two main categories: direct and indirect costs. Other authors, like Kendall [22] and Johnson [21] use different terminology for the same classification, they differentiate *explicit* and *implicit costs*.

Explicit or direct costs can be easily measured and usually account for a small part of the total cost. They are fees, commission, and taxes.

Implicit or indirect costs are costs which are not directly observable and therefore difficult to measure. They make up the largest part of the total costs. Cook [10] notes "As these costs are highly variable, they offer a greater potential for cost management." Implicit costs are the spread cost, delay cost, market impact, price trend, timing risk, opportunity cost, etc.

SPREAD COST

Spread cost is a difference between best ask and best bid price, it is variable and easily measured at any point of time. Kissel and Glantz [26] point that "Spreads represent a round-trip cost of transacting; however, this is only true for small orders." The high spread cost is characteristic for less liquid financial instruments, and aggressive trading style. The spread cost for trading Q shares with n executions of size Q_i at time t_i is defined as

$$\sum_{i=1}^{n} Q_i \cdot \frac{s(t_i)}{2},$$

where $s(t)$ is bid-ask spread at time t.

DELAY COST

Delay cost is defined by Wagner and Glass [54] as "the change in a stocks price that occurs once the manager makes a decision to buy or sell a stock, but before releasing it to a specific broker." It is mathematically expressed as

$$Q \cdot (P_0 - P_d),$$

where P_d and P_0 are mid prices at the time of manager's decision, and at the time when the order (of the size Q) is released to the broker, respectively. The delay cost is more present at volatile financial instruments, and also when the price is trending in the opposite direction from our order.

MARKET IMPACT

Market impact, notes Johnson [21], "represent price change caused by specific trade or order." Kissell and Glantz [26] give the following definition:

> "It is the difference between the stock's price trajectory with the

order and what the price trajectory would have been had the order not been submitted to the market."

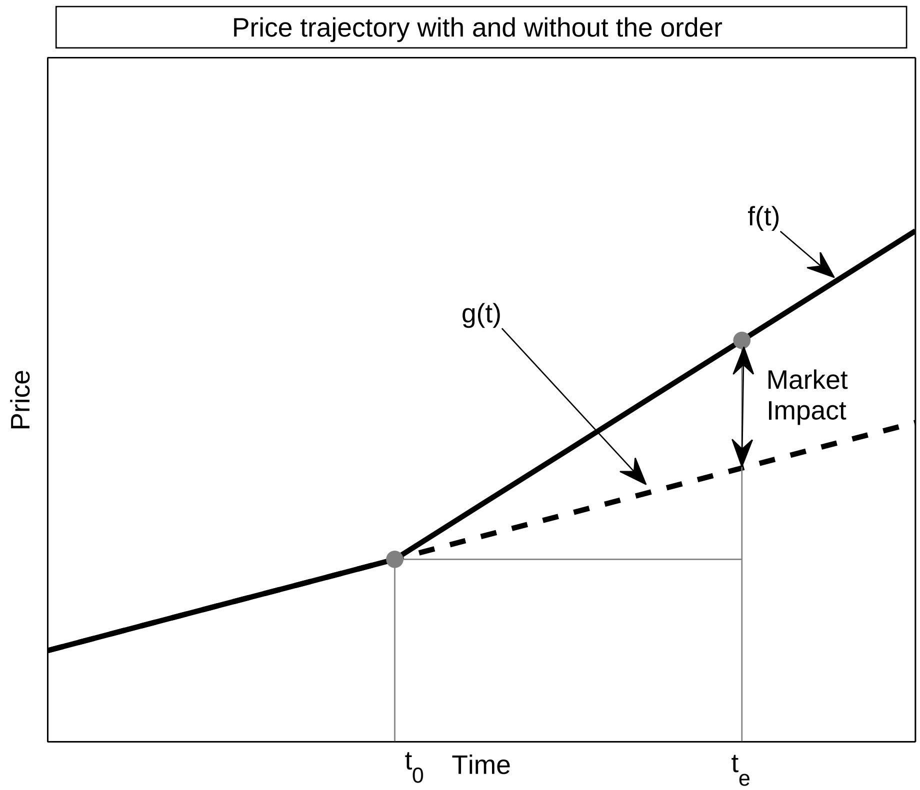

Figure 2.3: Illustration of Market Impact for buying order in rising market. Ilustracija tržisnog impakta kupovnog naloga pri rastućem trendu cene.

The authors further give mathematical expression of market impact at time t as

$$\kappa(t) = f(t) - g(t)$$

where function f describes price trajectory of the stock with the order, and g describes price trajectory of the stock without the order being released to market.

Figure 2.3 illustrates the total cost as the difference between price when the order was released $f(t_0)$ and price of execution $f(t_e)$ i.e. $f(t_e) - f(t_0)$

and the market impact cost that is $f(t_e) - g(t_e)$. Also, Kissel [24] states that market impact is a consequence of the liquidity demands or the information contained in the order, like its size, urgency of trading or leakage of information. The last especially concerns asset managers regarding execution algorithm information leakage to trading predators, Sofianos and Xiang in their article "Do Algorithmic Executions Leak Information?" [43] present their results on the matter.

It is obvious that accurate measurement and estimate of market impact is difficult, one approximate for market impact for trading Q shares with n executions of size Q_i and price P_i at time t_i [21] is

$$\begin{cases} \sum_{i=1}^{n} Qi \cdot (P_i - a(t_i)), & \text{for buy order;} \\ \sum_{i=1}^{n} Qi \cdot (P_i - b(t_i)), & \text{for sell order.} \end{cases}$$

Here, $a(t)$ and $b(t)$ represent best ask and best bid at time t, respectively.

It consists of two components: *temporary* and *permanent* market impact. Generally, the temporary impact is a consequence of demanding liquidity and the information content of the order is a cause for permanent impact.

PRICE TREND

Price trend also known as price appreciation, drift, momentum, volatility cost or short-term alpha. Kissell [24] describes it as "the cost (savings) associated with buying stock in a rising (falling) market or selling (buying) stock in a falling (rising) market." He further explains that this is not directly observable cost, and depends on price trend and implementation strategy.

The price trend cost for trading Q shares with the price trend function $\widetilde{P}(t)$ and n executions of size Q_i at time t_i is defined as

$$\sum_{i=1}^{n} Q_i \cdot (\widetilde{P}(t_i) - m(t_0)),$$

where $m(t_0)$ is a mid price at time t_0 when the order was received by a broker.

TIMING RISK

Timing risk of an asset describes the volatility of its price and liquidity. Kissell and Glantz [26] define

> "Timing risk is the associated uncertainty in trading cost estimates due to price volatility and liquidity risk. Price volatility affects the price appreciation estimate, and liquidity risk affects the market impact estimate."

With the assumption of independence of volume and price movement, Kissell and Glantz [26] compute price volatility $\sigma(\mu(x_k))$ and liquidity risk $\sigma(\kappa(x_k))$ separately, and formulate timing risk of strategy x_k as

$$\Re(\Phi) = \sqrt{\sigma^2(\mu(x_k)) + \sigma^2(\kappa(x_k))}$$

Johnson [21] gives a way for measuring timing risk as

$$\sum_{i=1}^{n} Q_i \cdot (m(t_i) - \widetilde{P}(t_i)),$$

where $\widetilde{P}(t)$ is the price trend function and there are n executions of size Q_i and mid price $m(t_i)$ at time t_i.

OPPORTUNITY COST

Opportunity cost is defined by Kissell [25] as "a measure of the forgone profit or avoided loss of not being able to transact the entire order." It is defined as

$$(Q - \sum Q_i) \cdot (P_n - P_0),$$

where P_0 is an arrival price and P_n price at the end of a period. The opportunity cost is present as a consequence of price movement in adverse direction or insufficient liquidity. High opportunity cost is a sign of passive trading.

Chapter 3

Benchmarks and Algorithmic Trading

Each transaction includes costs, and their correct quantification has a crucial role in formulation of investment strategy. Madhavan [37] describes investment process as a cycle (Figure 3.1) consisting of development of investment strategy and its implementation, which is realized through the following stages: portfolio formation, pre-trade analysis, execution and post-trade analysis.

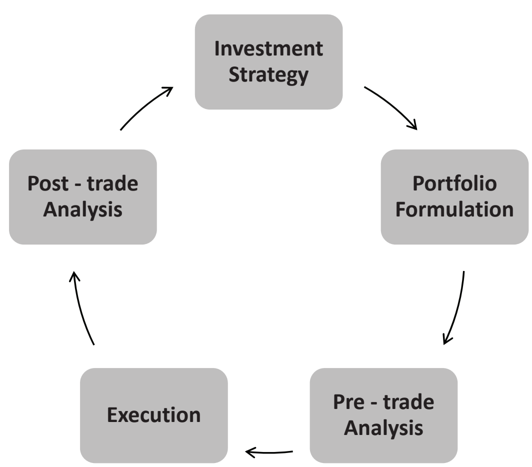

Figure 3.1: Investment cycle. Source: Madhavan [37]. Investicioni ciklus. Izvor: [37]

Pre-trade analysis focuses on (historical) data, which Baker and Filbeck [5] classify into two groups: "fundamentals of security" and "trade-related factors" that are:

- Prices - ranges (day's high and low), trends, momentum, market prices, last traded price etc.
- Liquidity - average daily volume (ADV), volume profile, trading stability which can be quantified using coefficient of variation (see [25])

$$CV = \frac{\sigma(ADV)}{ADV}$$

- Risk - volatility, beta, risk exposure.
- Cost estimates - explicit costs, and components of implicit costs like market impact, timing risk, opportunity cost etc.

The post-trade analysis measures transaction costs and execution performance. Its main purpose is an improvement of future performance. The key part in the identification of sources of under- and over-performance is a choice of proper benchmark, which is used for comparison with average execution price of a trade.

Essentially, measuring performance for an individual asset, as performance in units of money, is described in (3.1)

$$\sum Q_j P_j - Q P_b \tag{3.1}$$

where Q_j is amount of shares executed in period j at price P_j and Q is total amount to execute, and P_b is benchmark price. In this case, negative performance is more favorable than positive. Also performance can be expressed in units of money per share (3.2), which is average price of execution (P_{avg})

$$\frac{\sum Q_j P_j - Q P_b}{\sum Q_j} \tag{3.2}$$

or in basis points (3.3)

$$\frac{P_{avg} - P_b}{P_b} \cdot 10^4. \tag{3.3}$$

Cook [10] insists on following properties of benchmark: it should be weighted average for any order that cannot be completed in one trade. Further, it should be transparent and achievable. Finally, it should not be influenced by trader's own trading nor susceptible to gaming.

3.1 Types of Benchmark

The most common way in classifying price benchmarks is based on the time when they are determined (See [5, 13, 21, 22]), in this case, we distinguish between pre-trade, intraday, and post-trade benchmarks. The hybrid benchmarks are a combination of some already known benchmarks. And there are also non-price benchmarks, like *relative performance measure* (RPM) proposed by Kissel and Glantz [26].

3.1.1 Pre-Trade benchmarks

Pre-Trade benchmarks are also known as implementation shortfall [22] benchmarks. They are easy to determine and immediately available, but they do not always reflect market conditions. The advantages of pre-trade benchmarks are that they cannot be influenced by trader's own trading and they could be applied to any order size. They are:

- *Previous close* represents closing price of the asset on the previous day.
- *Opening price* is the opening price for the same day.
- *Decision price* is in close relationship with Perold's implementation shortfall [44], as Baker and Filbeck [5] note, "it is the price at which the choice to invest was actually made."
- *Arrival price* represent price at time when order could be traded.

First two benchmarks are directly observable, while the latter two are not publicly available and often not recorded by investors, which leads to inaccuracy in the measurement of transaction costs.

3.1.2 Intraday benchmarks

Intraday benchmarks are also known as average, or across-day benchmarks, They try to give the more accurate picture of prevailing market conditions, unlike pre- and post-trade benchmarks, they are recalculated during the trading day. Kendal [22] notes that "[they] are composed of prices that occur during a trading session".

- *Open-High-Low-Close* (OHLC) is average of four prices: open, high, low and close, and as a consequence of its definition (3.4), in case of extreme values of those four prices, value of this benchmark can be distorted. Cook [10] emphasize that "[OHCL] is the most arbitrary of this class of benchmarks, an as such, most difficult to consistently achieve"

$$OHCL = \frac{P_{open} + P_{high} + P_{low} + P_{close}}{4} \tag{3.4}$$

- *Time Weighted Average Price* (TWAP) is average of all trade prices for a given period of time. If n_i is frequency of trades at price P_i, then its definition is qiven with (3.5). Drawback of TWAP is obvious in case of large number of small trades with extreme prices.

$$TWAP = \frac{\sum n_i P_i}{\sum n_i} \tag{3.5}$$

- *Volume Weighted Average Price* (VWAP) defined as (3.6),where V_i is a number of shares transacted at price P_i during the trading day. It gives more accurate information about intraday market conditions. With VWAP, small trades with extreme prices do not have crucial influence on benchmark, but the trades with the largest volume.

$$VWAP = \frac{\sum V_i P_i}{\sum V_i} \tag{3.6}$$

- *Participation Weighted Average Price* (PWAP), as Gomes and Waelbroeck [17] explain, "[is] calculated as VWAP for time period starting an order arrival until the time that is required to complete the order at selected participation rate". Cook [10] claims that participation rate "lies in the 10% range", while many traders assert 20-25%.

3.1.3 Post-Trade benchmarks

Post-Trade benchmarks are determined at the end of a trade, or at the end of a trading day. As markets are usually more active at the end of the day, the closing price may not be a good representative of conditions during the day. Additionally, post-trade benchmark could give incorrect information of execution performance, for example: if the order was filled during some part of a day, unfavorable price trend would make execution look good. Madhavan [38] states that this benchmark supports trading at the close, which can lead to significant hidden costs.

The two most common post-trade benchmarks are:

- *Close* is closing price for the day.
- *Future close* is closing price for the next day.

3.1.4 Hybrid benchmarks

Hybrids of pre-, intra- and post-trade benchmarks are also used in practise. Kissel and Glantz [26] gave an example of composite 30-40-30 benchmark, computed in following way:

$$30\% P_{open} + 40\% OHLC + 30\% P_{close}$$

where OHCL is calculated with different weights for each price:

$$OHCL = \frac{40\% P_{open} + 10\% P_{high} + 10\% P_{low} + 40\% P_{close}}{4}$$

3.1.5 Relative Performance Measure

Kissel and Glantz [26] proposed *relative performance measure* (RPM) as an alternative to price benchmarks. In the essence of this performance metric is measuring the percentage of all market activity, that traded less favorable then execution price. It is computed for market volume and number of trades in following way

$$RPM_{volume} = \frac{\sum\limits_{\substack{i=1 \\ d=1, P<P_{avg}}}^{N} Q(i,P,d) + \sum\limits_{\substack{i=1 \\ d=-1, P>P_{avg}}}^{N} Q(i,P,d)}{Q}$$

$$RPM_{trades} = \frac{N_{buy}(P < P_{avg}) + N_{sell}(P > P_{avg})}{N},$$

where Q denotes the total market volume, N is the total number of trades, P_{avg} is an execution price and $Q(i,P,d)$ represents quantity traded at ith trade at a price P with a direction d which takes value 1 for buy, and -1

for sell orders. Furthermore, $N_{buy}(P < P_{avg})$ is a number of buy trades for which $P < P_{avg}$, and $N_{sell}(P > P_{avg})$ is a number of buy trades for which $P > P_{avg}$.

Authors, additionally proposed qualitative representation for average RPM, which is calculated as

$$RPM = \frac{RPM_{trades} + RPM_{volume}}{2},$$

in the form of qualitative label, defined as follows:

$$\text{Qualitative label} = \begin{cases} Excellent & 80\% < RPM \leq 100\% \\ Good & 60\% < RPM \leq 80\% \\ Average & 40\% \leq RPM \leq 60\% \\ Fair & 20\% \leq RPM < 40\% \\ Poor & 0\% \leq RPM < 20\% \end{cases}$$

3.2 Volume Weighted Average Price

Berkowitz, Logue and Noser [7] introduced *Volume Weighted Average Price* (VWAP) in 1988, as the alternative to OHCL in measuring market impact cost. The idea behind it was to provide better, unbiased estimates of market impact. For an individual asset, it is defined by (3.6).

The VWAP benchmark owes its popularity to simplicity of its calculation, intuitiveness and the fact that it gives a good indication of market behavior during given time interval.

Freyre-Sanders, Guobuzaite and Byrne [14] suggest that VWAP is "best used for smaller trades that have little or no impact on existing market prices." Because in the case of a large order or illiquid financial instrument, there is a chance for influencing the benchmark. The authors explain additional limitations of VWAP benchmark: first, it does not take into account difficulty of a trade and second, in given period of time for which VWAP is calculated might be included prices that have no significance for analyzed trade.

As Kissell, Glantz and Malamut [26] noted, VWAP does not allow benchmark comparison between different assets, or between different days for the same asset. For example, let us have benchmark comparison of two assets during two different days, as depicted in Table 3.1.

	Asset X	**Asset Y**
Day 1	20 bps	40 bps
Day 2	50 bps	40 bps

Table 3.1: Example of benchmark comparison for different assets and days. Primer poređenja pokazatelja za različite finansijske instrumente i dane.

By just looking at the data for Day 1 we could draw the conclusion that X performed better than Y, but the smaller difference for X could be the consequence of order size or market conditions. And again, looking at same asset at different days, we cannot conclude that X was better at Day 1, nor the Y had the same performance on these two days. Even with the same order size, market condition during Day 1 and Day 2 might be completely different.

Johnson [21] notes that even order of size over 30% of ADV influences the benchmark. Kissell, Glantz, and Malamut [26] explain that this situation illustrates the case when the difference between average price and benchmark does not give proper information about the quality of a trade.

3.3 Arrival Price and Implementation Shortfall

Arrival price benchmark is, as Almgren [3] defines it, "the quoted market price in effect at the time the order was released to the trading desk." It is in close relationship with total cost measure *Implementation shortfall* (IS) introduced by Perold [44] in 1988. With VWAP, it is one of the most popular benchmarks used in practice.

Kissel [25] suggests that this arrival price is the benchmark of choice for managers, whose decisions for buying and selling are based on company's long-term growth expectations. Further, he adds

> "[It] is also an appropriate benchmark price for situations where a market event triggers the portfolio manager or trader to release an order to the market."

For the reason that it is a pre-trade benchmark it is not susceptible to gaming, but as such, it does not always give true information of market conditions.

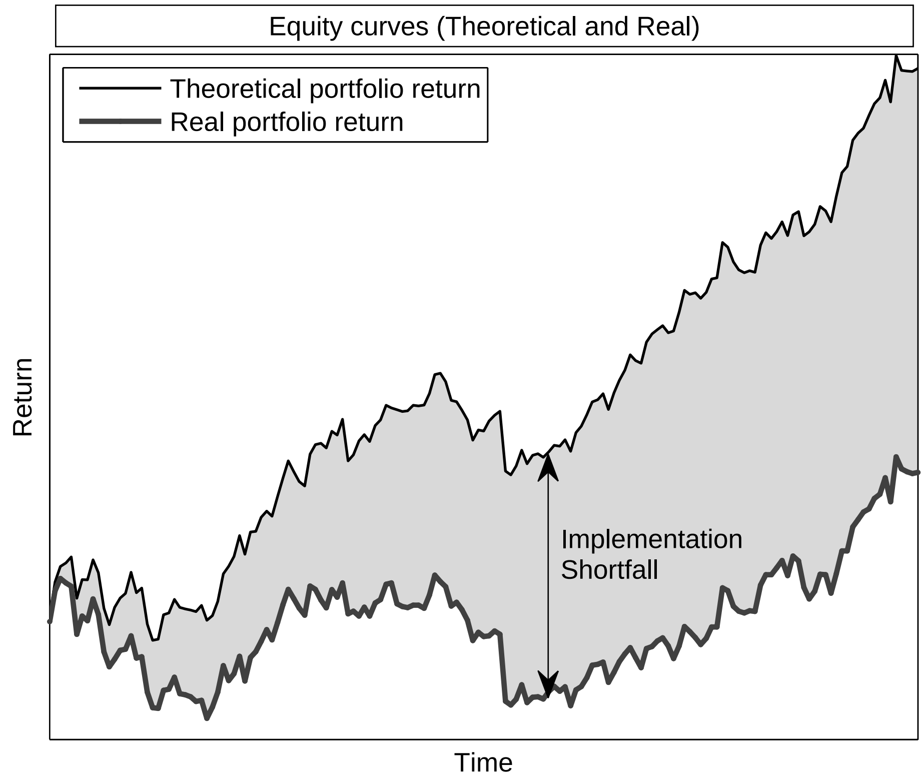

Figure 3.2: Comparison of returns between theoretical and real portfolio. Poređenje prinosa teorijskog i realnog portfolia

It is often used as a decision price in Perold's [44] *Implementation shortfall*. In his paper, he recommends parallel running of a theoretical (also known as paper or ideal) and a real portfolio, and Implementation Shortfall then represents the difference between the performance of the theoretical and real portfolio (Figure 3.2), which can be mathematically expressed as (3.7). All calculation of theoretical portfolio performance is done using mid price, which enables measurement of spread cost included in real portfolio costs.

$$IS = R_{theoretical} - R_{real} \tag{3.7}$$

With Q representing a total number of shares intended to trade, and $\sum Q_i$ representing executed shares, unexecuted part is given by $Q - \sum Q_i$, which makes possible expressing IS by (3.8), i.e. as a sum of execution and opportunity cost, respectively.

$$\sum Q_i P_i - \sum Q_i P_d + (Q - \sum Q_i) \cdot (P_n - P_d) \tag{3.8}$$

Here P_d is a *decision price* and P_n price at the end of a period.

Wagner and Glass [54] introduced delay cost (3.9) as "the change in a stock's price that occurs once the manager makes a decision to buy or sell a stock, but before releasing it to a specific broker."

$$Q(P_0 - P_d) \tag{3.9}$$

Authors showed that the delay costs are part of transaction cost, and their result was included in *Expanded implementation shortfall* (3.10) formulated by Kissell and Glantz [26], where P_0 is arrival price.

$$Q(P_0 - P_d) + \sum Q_i P_i - \sum Q_i P_0 + (Q - \sum Q_i) \cdot (P_n - P_0) \tag{3.10}$$

Even with the fact that IS is regarded as the most reliable measure of total transaction cost, it cannot be used as a measure of the quality of trader's performance. Minimization of IS is a trader's objective, but as Cook [10] explains "low shortfall number does not necessarily indicate a good result, nor a large shortfall number a poor result."

3.4 Algorithmic Trading

Algorithmic trading refers to the algorithmic order execution. Narang [41] defines it as "the use of computer software to manage and work an investors buy and sell orders in electronic markets." As the vast majority of financial instruments can be traded electronically, it is clear that algorithmic trading has become the crucial tool for the electronic trading. The term is sometimes used for quantitative and black-box trading, but as Schmidt [49] states, algorithmic trading only "focuses on making decisions where and how to trade." Smart order routing deals with the question "where to trade" in markets that have multiple liquidity pools. We here describe algorithms that give an answer to question "how to trade?".

The idea behind algorithmic trading is to formulate and use mathematical models, that react to changes in a market environment, to achieve specified goals for the execution while minimizing execution costs. The goals in most cases are meeting the benchmark or minimizing the total transaction costs. One of the classifications of execution algorithms is based on previously mentioned goals. Therefore we distinguish two major groups of algorithms:

- Benchmark-driven
- Cost-driven

Johnson [21] gives classification based on their underlying mechanism, he differs the following types of algorithms:

- Impact-driven
- Cost-driven
- Opportunistic

Moreover, he further explains their fundamental goals: the first group tries to minimize total market impact, the second aims to reduce total transaction costs, while the last group is made of dynamic algorithms designed to utilize favorable market conditions.

3.4.1 Benchmark-driven algorithms

Benchmark-driven algorithms aim to minimize slippage to chosen benchmark. Besides the fact that all the execution algorithms try to minimize costs related to the completion of a buy or sell order, these group of algorithms, Schmidt [49] explains "are based on some simple measures of market dynamics rather than on explicit optimization protocols." The most widely known algorithms from this group are Time Weighted Average Price (TWAP), Volume Weighted Average Price (VWAP) and Percent of Volume (POV).

All three of previously mentioned algorithms rely on the mechanism of splitting large orders into smaller parts, also known as *child orders* or *atomic orders*, with the intent to minimize overall market impact. With this aspect, they are also classified as impact-driven algorithms.

TWAP algorithm has a goal of meeting the TWAP benchmark, defined by (3.5), which is an average trade price for a time period when the order is submitted until it is completely filled. The most simple TWAP algorithm for a given order size Q that will trade over time period T minutes splits a large order into for example n parts, therefore we have n child orders each with size $\frac{Q}{n}$, trading in so called *waves* every $\Delta t = \frac{T}{n}$ minutes. It is obvious that this slicing of an order does not take into account market volume and price.

The drawback of this algorithm is predictability, so there is a risk of information leakage. One of the possible solutions for the problem is to randomize size of each child order and even to randomize trading waves.

VWAP algorithm is one of the most famous execution algorithms. It aims to minimize slippage to VWAP benchmark, defined by (3.6). Here trader specifies the time intervals, and then the algorithm splits the order into child orders with size proportional to period trading order. It is evident that one cannot know the future volume profile, so algorithm relies on historical, or combination of historical and real-time data. Kissell and Glantz [26] suggest VWAP execution strategy: First, they divide given trading day into n intervals, with volume weighted average price $\overline{P}_i$ and volume traded in that interval $\widetilde{V}_i$, then VWAP can be rewritten as

$$VWAP = \sum_{i=1}^{n} \frac{\widetilde{V}_i}{\sum_{j=1}^{n} \widetilde{V}_j} \overline{P}_i.$$

They state that strategy y for creating child orders minimizes the expected difference from the VWAP benchmark, if y_i are equal to previously defined weights.

POV algorithm is based on the idea to trade predetermined percent of volume p, known as *participation rate*. So the child order with size q_k in time interval k with the total trading volume Q_k is calculated so that the following holds

$$p = \frac{q_k}{Q_k + q_k},$$

So, then we have

$$q_k = \frac{p}{1-p} \, Q_k.$$

With this approach algorithm, child orders have a lower market impact. Leshik [33] explains that this strategy gives certain cover, especially for large orders, by ”hiding” the order from the rest of the market. The biggest problems with POV algorithm are that it does not guarantee completion of an order, applicability on trades with illiquid asset and predictability.

3.4.2 Cost-driven algorithms

Cost-driven algorithms are also called Implementation Shortfall algorithms, and their goal is to minimize total transaction costs. All of them try to reduce cost by finding a balance between timing risk and market impact.

Implementation shortfall algorithm is, in essence, same as arrival price algorithm. Both are based on Perold's [44] implementation shortfall, but as Glantz and Kissell [25] note, they differ in "real-time adaptation tactics". In most cases, *arrival price* benchmark is used as decision price. The objective of the implementation shortfall algorithm is to achieve average execution price which minimizes the Implementation Shortfall.

Adaptive Shortfall is an Implementation Shortfall algorithm with the additional property of dynamical adaptation to market conditions.

Market on close relay on post-trade benchmark *close*. The idea behind this algorithm is to find an optimal start time for trading, so that, it is not too early, nor too late because in the first case the order is exposed to timing risk and in the second, to market impact.

Chapter 4

Negative Selection

In Automated Order Execution, a computer program based on some trading strategy creates and submits orders with the primary goal, to achieve benchmark specified by a client. There are many different approaches for measuring execution performance, and each of them with some advantages and some flaws, as it is truly difficult to define one standard performance measure that would fulfill all the requirements for a valid benchmark.

The two most popular benchmarks VWAP and Implementation Shortfall (IS) represent the standard in the financial industry and are the subject of many academic studies [8, 10, 14, 32]. The majority of Benchmark-driven algorithms are developed to minimize slippage to these two benchmarks. As already mentioned, some problems accompany measuring slippage on VWAP and IS. They manifest in the loss of objectiveness and giving actual information of toughness of market conditions. In the case of VWAP, by its definition, trading with a large order or illiquid financial instrument can significantly distort the slippage. On the other hand, IS, with Arrival Price as the reference price, cannot be influenced by our own trading, but it is insensitive to market conditions and does not capture the nature of absolute slippage. Let us consider an example of executing a buy limit order placed at the best bid. In both cases, in rising and falling market, VWAP will decrease with the size of a large order even for the suboptimal execution. With IS, there is a different situation, in rising market, it will be high, which reflects the market conditions. But in falling market it will be constant representing the difference between the best bid price and Arrival Price, i.e. it does not indicate that we got filled too early, and could have done better in these market conditions. Therefore there is a need for an absolute performance measure. The performance measure, which we introduce here, has an objective to provide an alternative way of measuring the performance of execution algorithms. It takes a posteriori view of market conditions, which allows us to determine what would have been the optimal order placement if we knew in advance the complete market information during the trading window. Then, the difference between the optimal trading position and actual execution represents the performance measure, which takes into account all traded quantities with considered time window.

Optimal placement for a given quantity Q and time window $[0, T]$ is defined as a solution of the Linear Programming problem, where unknowns are quantities at specified price levels, which add up to Q and would achieve the best price under prevailing market conditions during $[0, T]$.

Negative Selection (NS) of an order is the distance between the vector of Optimal Placement and actual order. The measure differentiate between filled and unfilled orders because unfilled orders have negative and filled orders have nonnegative NS. It is also capable of showing the toughness of market conditions during the given time window. For example, buy order at bid1 will have positive NS in falling market, because the order is filled at an unfavorable price, while in rising market it could be optimal with zero NS or negative because the order was not (completely) filled.

We continue with notation and assumptions necessary for the definition of Negative Selection.

4.1 Notation and Assumptions

To define Negative Selection, we first assume that we have to buy Q shares either by placing a market order or taking a passive position at some of the bid levels. For the opposite case, selling Q shares, the definition is completely symmetric. We consider a market governed by the limit order book implying that the orders are placed in queues by price and arrival time priority. Next, let us assume that the buy order of the size Q has to be executed within the time window $[0, T]$. At $t = 0$ the following information is available.

- The price vector
$$\mathcal{P} = [P_0, ..., P_k]^T,$$
where P_0 is market price (the best ask price at $t = 0$), and $P_1, ..., P_k$ are bid prices at corresponding bid levels. Clearly, $P_k < P_{k-1} < \ldots <$

$P_1 < P_0$.

- Volume ahead

$$\mathcal{V} = [V_0, ..., V_k]^T,$$

represents the sizes of the existing orders in the corresponding bid queues at $t = 0$. We will assume that $V_0 = 0$, so a market order with price P_0 is immediately traded.

- The gain coefficients are defined as

$$\mathcal{G} = [g_0, ..., g_k]^T, \text{ where } g_j = \frac{P_0 - P_j}{P_0} \text{ for } j = 0, \ldots, k. \tag{4.1}$$

Clearly, $0 = g_0 < g_1 < ... < g_k$.

We consider a simple order determined by its quantity Q placed at the end of the existing queue, at some of the available bid levels or as a market order. We are assuming that the quantity Q is small enough so that it can be traded as a simple order. For technical reasons, such order will be represented by *order vector*

$$\mathcal{Q} = [Q_0, ..., Q_k]^T. \tag{4.2}$$

If the simple order of quantity Q is placed at some price level P_m, $m = 0, \ldots, k$, then the components Q_j, $j = 0, \ldots, k$ of its order vector $\mathcal{Q}$ are defined in the following way:

$$Q_j = \begin{cases} \text{Q}, & j = m; \\ 0, & j \neq m. \end{cases} \tag{4.3}$$

At the end of the time window $[0, T]$ the following information is available.

- The traded quantity at each price level during the time window $(0, T]$ is represented by

$$\mathcal{T} = [T_0, ..., T_k]^T.$$

We assume that $T_0 \geq Q$, i.e. there is enough liquidity at the price level P_0, so that the market order can be filled at P_0.

- Available quantity

$$\mathcal{A} = [A_0, ..., A_k]^T, \; A_j = \max\{T_j - V_j, 0\}.$$

The assumption $T_0 \geq Q$ guaranties that of indices

$$I_L = \{j | A_j > 0, j = 0, \ldots, k\} \tag{4.4}$$

is a nonempty set and we denote $l = \max I_L$. We also define set of indices I_H with cumulative sums of available quantity exceeds Q

$$I_H = \{j | \sum_{i=j}^{l} A_i \geq Q, j = 0, \ldots, l\} s \tag{4.5}$$

Again, the assumption $T_0 \geq Q$ guaranties that I_H is a nonempty set with its maximal element $h = \max I_H$. The price P_h is the lowest price level that allows for the order to be completely filled with respect to price-time priority.

4.2 Optimal Placement

The optimal placement can be defined at the time $t = T$, that is, only when the complete information about market conditions in considered time window is available. It represents an ideal order placed at time $t = 0$ that would be executed and achieve the best (the lowest buy) price during $[0, T]$. If we denote optimal placement as a vector

$$\mathcal{O} = [O_0, ..., O_k]^T,$$

where O_0 represents quantity at market, and O_i is quantity placed at corresponding bid price P_i, $i = 1, \ldots, k$. Since the objective is to buy Q shares at the lowest possible price, the optimal placement is a solution of the following Linear Programming Problem.

$$\min_{O_0,\ldots,O_k} \sum_{i=0}^{k} P_i O_i \tag{4.6}$$

$$s.t. \sum_{i=0}^{k} O_i = Q \tag{4.7}$$

$$\sum_{i=j}^{l} O_i \leq \sum_{i=j}^{l} A_i, j = 1, \ldots, k \tag{4.8}$$

$$O_j \geq 0, j = 0, \ldots, k. \tag{4.9}$$

The objective of achieving the lowest average execution price for buying Q shares $\frac{1}{Q}\sum_{i=0}^{k} P_i O_i$, is described trough the objective function (4.6) and first constraint (4.7). The objective function (4.6) represents the total cost for buying Q shares, and the first constraint (4.7) that we want to buy Q shares. The second constraint (4.8) specifies that we can buy only the available quantities at each price levels with respect to the execution price-queue priority in the order book, i.e. it cannot be stated simply as $O_i \leq A_i$. The following toy example shows that filling the available quantities from below does not yield the smallest price for the total order of Q shares. Let us consider data shown in Table 4.1 and order with $Q = 100$ shares is submitted to order book at bid2, i.e. at price $P_2 = 98$. If the constraint is just $O_i \leq A_i$, then the optimal placement would be 40 shares at market price ($P_0 = 100$) and 60 shares at bid3 ($P_3 = 97$) which yields average execution price of 98.2. On the other hand, by using constraint (4.8) price-time priority is incorporated which leads to optimal placement with just 5 shares at market price ($P_0 = 100$), 95 shares at bid2 ($P_2 = 98$) and average execution price 98.1.

level	$\mathcal{P}$	$\mathcal{V}$	$\mathcal{T}$
0	100	0	100
1	99	200	200
2	98	80	80
3	97	35	95

Table 4.1: Snapshot of data in a given time window that is necessary to determine an optimal placement. Prikaz podataka u vremenskom intervalu neophodnih za određivanje optimalnog naloga.

For the purpose of displaying Mangasarian's results [39] on uniqueness of solution in linear programming, which will be used to prove the uniqueness of optimal placement, we consider the problem of the following form:

$$\begin{aligned} \min_{x}\ & c^T x \\ s.t.\ & Ax = b \\ & Cx \geq d \end{aligned} \tag{4.10}$$

where $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $d \in \mathbb{R}^p$.

Theorem 4.2.1. *[39] A solution x^* of the linear programming problem (4.10) is unique if and only if it remains a solution to all the linear programs obtained from (4.10) by arbitrary but sufficiently small perturbation of its cost vector c, or equivalently for each $q \in \mathbb{R}^n$ there exist a positive real number ϵ such that x^* remains a solution of the perturbed linear program (4.11)*

$$\begin{aligned} \min_x \; & (c + \epsilon q)^T x \\ s.t. \; & Ax = b \\ & Cx \geq d. \end{aligned} \tag{4.11}$$

Theorem 4.2.2. *The vector $\mathcal{O}$ with components O_j, $j = 0, \ldots, k$ defined by*

$$O_j = 0, \qquad j = 0, \ldots, h-1 \tag{4.12}$$

$$O_j = Q - \sum_{i=h+1}^{l} A_i, \qquad j = h \tag{4.13}$$

$$O_j = A_j, \qquad j = h+1, \ldots, l \tag{4.14}$$

$$O_j = 0, \qquad j = l+1, \ldots, k, \tag{4.15}$$

is the unique solution of (4.6)-(4.9).

Proof. As there is no liquidity at levels $l+1, \ldots, k$, i.e., $A_j = 0$ for $j = l+1, \ldots k$ and because of constrains (4.8)-(4.9) it is clear that all feasible solutions satisfy $O_j = 0$ for $j = l+1, \ldots k$, thus (4.15) holds. So, instead of the problem (4.6)-(4.9) the following problem will be considered

$$\min_{O_0, \ldots, O_l} \sum_{i=0}^{l} P_i O_i \tag{4.16}$$

$$s.t. \sum_{i=0}^{l} O_i = Q \tag{4.17}$$

$$\sum_{i=j}^{l} O_i \leq \sum_{i=j}^{l} A_i, j = 1, \ldots, l \tag{4.18}$$

$$O_j \geq 0, j = 0, \ldots, l \tag{4.19}$$

In order to solve the problem(4.16)-(4.19), inequality constraints (4.18) will be transformed into equality constraints, by adding nonnegative variable d_j to left side of each inequality in (4.18), i.e. we consider LPS problem (4.20)

$$\begin{aligned}
\min_{O_0,\ldots,O_l} \; & \sum_{i=0}^{l} P_i O_i \\
s.t. \; & \sum_{i=0}^{l} O_i = Q \\
& \sum_{i=j}^{l} O_i = d_j + \sum_{i=j}^{l} A_i, j = 1, \ldots, l \\
& O_j \geq 0, j = 0, \ldots, l \\
& d_j \geq 0, j = 1, \ldots, l
\end{aligned} \tag{4.20}$$

or in the short form:

$$\begin{aligned}
\min_x \; & c^T x \\
s.t. \; & Mx = b \\
& x \geq 0
\end{aligned} \tag{4.21}$$

where $c = [c_0, ..., c_{2l}]^T$, $x = [x_0, ..., x_{2l}]^T$ and $b = [b_0, ..., b_l]^T$ are defined by

$$c_j = \begin{cases} P_j, \; j = 0, \ldots, l \\ 0, \; j = l+1, \ldots, 2l \end{cases}$$

$$x_j = \begin{cases} O_j, \; j = 0, \ldots, l \\ d_{j-l}, \; j = l+1, \ldots, 2l \end{cases}$$

$$b_j = \begin{cases} Q, \; j = 0 \\ \sum_{i=j}^{l} A_i, \; j = 1, \ldots, l \end{cases}$$

and matrix

$$M = \left[\begin{array}{lll} 1 & \mathbf{1}_{1\times l} & \mathbf{0}_{1\times l} \\ \mathbf{0}_{l\times 1} & U(1)_{l\times l} & E_{l\times l} \end{array} \right].$$

Here, $U(c)$ is an upper triangular matrix with diagonal and all above elements equal to c, $\mathbf{0}$ is a matrix with all zero elements, and $\mathbf{1}$ is a matrix with all elements equals to 1, E is an identity matrix. The following vector is a basic solution of problem (4.21)

$$v = [0, 0, \ldots, 0, Q - \sum_{j=h+1}^{l} A_j, A_{h+1}, \ldots, A_{l-1}, A_l,$$
$$(\sum_{j=1}^{l} A_j) - Q, \ldots, (\sum_{j=h-1}^{l} A_j) - Q, (\sum_{j=h}^{l} A_j) - Q, 0, \ldots, 0, 0]^T, \quad (4.22)$$

with a basis matrix

$$B = \begin{bmatrix} 1 & \mathbf{1}_{1\times(l-h)} & \mathbf{0}_{1\times h} \\ \mathbf{1}_{h\times 1} & \mathbf{1}_{h\times(l-h)} & E_{h\times h} \\ \mathbf{0}_{(l-h)\times 1} & U(1)_{(l-h)\times(l-h)} & \mathbf{0}_{(l-h)\times h} \end{bmatrix}$$

and a non-basis matrix

$$N = \begin{bmatrix} U(1)_{h\times h} & \mathbf{0}_{h\times(l-h)} \\ \mathbf{0}_{1\times h} & \mathbf{0}_{1\times(l-h)} \\ \mathbf{0}_{(l-h)\times h} & E_{(l-h)\times(l-h)} \end{bmatrix}.$$

As inverse of B is

$$B^{-1} = \begin{bmatrix} 1 & \mathbf{0}_{1\times h} & [-1, \mathbf{0}_{1\times(l-h-1)}] \\ \mathbf{0}_{(l-h)\times 1} & \mathbf{0}_{(l-h)\times h} & U(1)^{-1}_{(l-h)\times(l-h)} \\ -\mathbf{1}_{h\times 1} & E_{h\times h} & \mathbf{0}_{h\times(l-h)} \end{bmatrix}$$

and

$$B^{-1}N = \begin{bmatrix} \mathbf{1}_{1\times h} & [-1, \mathbf{0}_{1\times(l-h-1)}] \\ \mathbf{0}_{(l-h)\times h} & U(1)^{-1}_{(l-h)\times(l-h)} \\ L(-1)_{h\times h} & \mathbf{0}_{h\times(l-h)} \end{bmatrix},$$

where $L(c)$ is a lower triangular matrix with diagonal and all lower elements equal to c, the Simplex table for the basic solution v

	O_0	O_1	$\dots$	O_{h-1}	d_{h+1}	d_{h+2}	$\dots$	d_l	v_B
O_h	1	1	$\dots$	1	-1	0	$\dots$	0	$Q-\sum_{j=h+1}^{l} A_j$
O_{h+1}	0	0	$\dots$	0	1	-1	$\dots$	0	A_{h+1}
O_{h+2}	0	0	$\dots$	0	0	1	$\dots$	0	A_{h+2}
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
O_{l-1}	0	0	$\dots$	0	0	0	$\dots$	-1	A_{l-1}
O_l	0	0	$\dots$	0	0	0	$\dots$	1	A_l
d_1	-1	0	$\dots$	0	0	0	$\dots$	0	$(\sum_{j=1}^{l} A_j)-Q$
d_2	-1	-1	$\dots$	0	0	0	$\dots$	0	$(\sum_{j=2}^{l} A_j)-Q$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
d_h	-1	-1	$\dots$	-1	0	0	$\dots$	0	$(\sum_{j=h}^{l} A_j)-Q$
	Δ_{O_0}	Δ_{O_1}	$\dots$	$\Delta_{O_{h-1}}$	$\Delta_{d_{h+1}}$	$\Delta_{d_{h+2}}$	$\dots$	Δ_{d_l}	

All the reduced costs are negative: $\Delta_{O_j} = P_h - P_j < 0,\ j = 0, \dots, h-1$ and $\Delta_{d_j} = P_j - P_{j-1} < 0,\ j = h+1, \dots, l$. Now it could be concluded that a basic solution v is optimal solution of problem (4.21). Further, the vector

$$[0, 0, \dots, 0, Q - \sum_{h+1}^{l} A_j, A_{h+1}, \dots, A_{l-1}, A_l]^T$$

is optimal solution of (4.16)-(4.19) and optimal solution of (4.6)-(4.9) is indeed given by (4.12)-(4.15).
Uniqueness will be proved using Theorem 4.2.1. Let $q \in \mathbb{R}^{l+h+1}$ be an arbitrary vector, and we will show that there is positive number $\epsilon > 0$ such that vector v, defined in (4.22), remains a solution of perturbed problem

$$\begin{aligned} &\min_x \ (c+\epsilon q)^T x \\ &s.t.\ Mx = b \\ &\quad x \geq 0. \end{aligned} \tag{4.23}$$

The vector v is a basic solution of the problem (4.23), and reduced costs for the perturbed problem are

$$\Delta_{O_j} = P_h + \epsilon q_h - (P_j + \epsilon q_j) = P_h - Pj + \epsilon(q_h - q_j),\ j = 0, \dots, h-1$$

$$\Delta_{d_j} = (P_j + \epsilon q_j) - (P_{j-1} + \epsilon q_{j-1}) = (P_j - P_{j-1}) + \epsilon(q_j - q_{j-1}),\ j = h+1, \dots, l.$$

By choosing ϵ as minimum of set

$$\{\frac{P_j - P_h}{2(q_h - q_j)} | q_h > q_j, j = 0, \dots, h-1\} \cup \{\frac{P_{j-1} - P_j}{2(q_j - q_{j-1})} | q_j > q_{j-1}, j = h+1, \dots, l\}$$

we get

$$\Delta_{O_j} < \frac{1}{2}(P_h - Pj) < 0,\ j = 0, \dots, h-1$$

$$\Delta_{d_j} =< \frac{1}{2}(P_j - P_{j-1}) < 0,\ j = h+1, \dots, l.$$

This means that v is an optimal solution of (4.23). □

4.3 Negative Selection

Definition 4.3.1. *For an order with the size Q at the price level P_m and execution time window $[0, T]$, Negative Selection is defined as*

$$\mathcal{N} = (\mathcal{O} - \mathcal{Q})^T \mathcal{G}, \tag{4.24}$$

where $\mathcal{O}$ is the optimal placement vector, $\mathcal{Q}$ is the order vector and $\mathcal{G}$ is the vector of gain coefficients defined by (4.1).

We continue with properties of Negative Selection that make it a well defined performance measure.

Lemma 4.3.1. *Negative Selection of optimally placed order is zero.*

Proof. If the order is optimally placed then $\mathcal{O} = \mathcal{Q}$, therefore

$$\mathcal{N} = (\mathcal{O} - \mathcal{Q})^T \mathcal{G} = (\mathcal{O} - \mathcal{O})^T \mathcal{G} = 0.$$

□

Lemma 4.3.2. *Negative Selection of filled order is nonnegative.*

Proof. If the order is filled at a level m then $m \in I_H$ and $m \leq h$ holds. Furthermore,

$$\mathcal{N} = (\mathcal{O} - \mathcal{Q})^T \mathcal{G} = \sum_{i=0}^{k} O_i g_i - Q g_m \geq Q(g_h - g_m) \geq 0.$$

□

Lemma 4.3.3. *Negative Selection of (partially) unfilled order is negative.*

Proof. If the order is unfilled at level m it is clear that $m \geq l$. Suppose that holds $m < l$ this means that $P_m > P_l$ and as $A_l > 0$ it is clear that there was some trading at level l. For the price to go down to P_l, all orders placed at higher price levels would be filled, which means that our order at P_m would be filled, and this contradicts the assumption of being unfilled.

If the order is unfilled at level m it is clear that $m \geq l$, because if $m < l$ holds, this means that $P_m > P_l$ and as $A_l > 0$ it is clear that there was some trading at level l. For the price to go down to P_l, all orders placed at higher price levels would be filled, which means that our order at P_m would be filled, and this contradicts the assumption of being unfilled.

$$\mathcal{N} = (\mathcal{O} - \mathcal{Q})^T \mathcal{G} < \sum_{i=0}^{k} O_i g_l - Q g_m = Q(g_l - g_m) \leq 0.$$

□

Lemma 4.3.4. *Consider two orders with the same size Q placed at two price levels P_m and P_{m+1} with $P_m > P_{m+1}$. If $\mathcal{N}_m$ and $\mathcal{N}_{m+1}$ are their Negative Selections respectively then $\mathcal{N}_m > \mathcal{N}_{m+1}$.*

Proof. Order vectors of orders of size Q at prices P_m and P_{m+1}, are denoted by $\mathcal{Q}_m$ and $\mathcal{Q}_{m+1}$, respectively. For quantity Q exists an optimal placement $(\mathcal{O})$, then

$$\mathcal{N}_m - \mathcal{N}_{m+1} = (\mathcal{O} - \mathcal{Q}_m)^T \mathcal{G} - (\mathcal{O} - \mathcal{Q}_{m+1})^T \mathcal{G} = Q(g_{m+1} - g_m) > 0.$$

□

Lemma 4.3.5. *Consider two different order of the sizes $Q_1 > Q_2$ placed at the same price level P_m. If $\mathcal{N}_{Q_1}$ and $\mathcal{N}_{Q_2}$ are their Negative Selections, respectively, then we have two cases:*

(i) If the larger order is filled then $\mathcal{N}_{Q_1} \geq \mathcal{N}_{Q_2}$.

(ii) If the larger order is unfilled then $\mathcal{N}_{Q_1} < \mathcal{N}_{Q_2}$.

Proof. Because of the optimal placement definition both orders "share" same starting point from the lowest price level P_l where available quantity is positive. i.e. both orders have the same set I_L defined in (4.4), with maximal element l. For both order sizes Q_1 and Q_2, we define sets I_{H_1} and I_{H_2} using definition (4.5), respectively. Let $h_1 = max(I_{H_1})$, and $h_2 = max(I_{H_2})$. It is quite clear, because of difference in sizes and optimal placement definition, that $h_2 \geq h_1$.

The order vector for the order of size Q_1 is denoted by $\mathcal{Q}_1$ and its optimal placement is given by

$$\mathcal{O}_{Q_1} = [O_0^1, ..., O_k^1]^T.$$

Similarly, we denote the order vector for the order of size Q_2, by $\mathcal{Q}_2$ and its optimal placement by

$$\mathcal{O}_{Q_2} = [O_0^2, ..., O_k^2]^T.$$

Because of optimal placement definition $O_i^1 = O_i^2, i = h_2 + 1, \ldots, l$

$$Q_1 - Q_2 = \sum_{i=h_1}^{h_2-1} O_i^1 + O_{h_2}^1 - O_{h_2}^2$$

(i) If both orders are filled that means that $m \leq h_1 \leq h_2 \leq l$. We will consider two cases:

$(i.1.)$ If $A_l \geq Q_1$ and $m = l$, then it holds $m = h_1 = h_2 = l$

$$\mathcal{N}_{Q_1} - \mathcal{N}_{Q_2} = (\mathcal{O}_{Q_1} - \mathcal{Q}_1)^T \mathcal{G} - (\mathcal{O}_{Q_2} - \mathcal{Q}_2)^T \mathcal{G} = 0.$$

$(i.2.)$ If $A_l < Q_1$ then $m \leq h_1 < h_2 = l$ or $m \leq h_1 \leq h_2 < l$

$$\begin{aligned}\mathcal{N}_{Q_1} - \mathcal{N}_{Q_2} &= \sum_{i=h_1}^{h_2-1} O_i^1 g_i + (O_{h_2}^1 - O_{h_2}^2) g_{h_2} - (Q_1 - Q_2) g_m \\ &> \sum_{i=h_1}^{h_2-1} O_i^1 g_{h_1} + (O_{h_2}^1 - O_{h_2}^2) g_{h_1} - (Q_1 - Q_2) g_m \\ &= (Q_1 - Q_2)(g_{h_1} - g_m) \geq 0\end{aligned}$$

(ii) If order with size Q_1 is unfilled, then there are two possibilities:
$(ii.1.)$ Order with size Q_2 is filled, then

$$\mathcal{N}_{Q_2} \geq 0 > \mathcal{N}_{Q_1}.$$

$(ii.2.)$ Order with size Q_2 is unfilled ($m \geq l$)

$$\begin{aligned}\mathcal{N}_{Q_1} - \mathcal{N}_{Q_2} &= \sum_{i=h_1}^{h_2-1} O_i^1 g_i + (O_{h_2}^1 - O_{h_2}^2) g_{h_2} - (Q_1 - Q_2) g_m \\ &< \sum_{i=h_1}^{h_2-1} O_i^1 g_l + (O_{h_2}^1 - O_{h_2}^2) g_l - (Q_1 - Q_2) g_m \\ &= (Q_1 - Q_2)(g_l - g_m) \leq 0\end{aligned}$$

□

4.4 Negative Selection of a Complex Order

With every order one is faced with the dilemma whether to trade aggressively or passively. Aggressive trading guarantees that buy order will be filled in rising market, but in falling market it will cause unnecessary costs. On the other hand placing buy order passively will be beneficial in falling market, but in rising market, a passive buy order will be (partially) unfilled. One way to get the best of both trading behaviors while trying to adapt to future market conditions and therefore achieve maximum profit is to split the order into multiple price levels. Thus, we construct a complex order, which can be regarded as trading strategy for splitting an order for buying Q shares, denoted by

$$\mathcal{S} = [Q_0, ..., Q_k]^T, \tag{4.25}$$

where $Q_i,\ i = 0, \ldots, k$ represent a nonnegative quantity placed at a price level P_i, whereby $\sum_{i=0}^{k} Q_i = Q$ is satisfied. Each quantity Q_i in (4.25) has its corresponding order vector $\mathcal{Q}_i$ defined in (4.2). Therefore, we are in position to calculate its Negative Selection, which we denote by $\mathcal{N}_i$. Thus, the Negative Selection for the complex order (4.25) is defined in the following way

$$\boldsymbol{\mathcal{N}}(\mathcal{S}) = [\mathcal{N}_0, ..., \mathcal{N}_k]^T. \tag{4.26}$$

Unlike the case when determining the optimal placement of a simple order, there is an inevitable interaction between optimal placements for orders in the strategy, so the following algorithm will be used for determination of Negative Selection for orders in strategy. When determining starting level for the algorithm, denoted by s, we distinguish two cases:

1) $Q_l \leq A_l$ Order at level l is filled, orders at levels $l+1, \ldots, k$ are completely unfilled, then $s = l$.

2) $Q_l > A_l$ Order at level l is partially filled with quantity is A_l (A_l is positive number by definition), and orders at levels $l+1, \ldots, k$ are completely unfilled, then $s = l - 1$.

To cover both cases in algorithm let us define vector $\alpha = [\alpha_0, \ldots, \alpha_k]^T$

$$\alpha_i = \begin{cases} A_{s+1},\ i = s+1 \\ 0,\ i \neq s+1 \end{cases}$$

Remark 4.4.1. *The purpose of a vector α is to "make a reservation" of a quantity for the order that is partially filled.*

Algorithm 1 consist of two phases, in the first phase, is calculated Negative Selection for all filed orders. It starts at level s and finishes at 0. In the second phase, we determine Negative Selection for unfilled orders, going from $s+1$ to k. The idea behind the algorithm is to take into account all our positions in strategy, that way it constructs the optimal placement $\mathcal{O}_i$ taking into account the corresponding available quantity $\mathcal{A}_i$.

Algorithm 1: $\textsc{StrategyNegSel}(\mathcal{S}, \mathcal{A}, s, \alpha)$

---

comment: Filled orders

$\mathcal{A}_s := \mathcal{A} - \alpha$
$\mathcal{O}_s := \textsc{OptimalPlacement}(\mathcal{A}_s, Q_s)$
$\mathcal{N}_s := \mathcal{O}_s - \mathcal{Q}_s$

for $i := s - 1$ **to** 0

$$\textbf{do} \begin{cases} \mathcal{A}_i := \mathcal{A}_{i+1} - \mathcal{O}_{i+1} \\ \mathcal{O}_i := \textsc{OptimalPlacement}(\mathcal{A}_i, Q_i) \\ \mathcal{N}_i := \mathcal{O}_i - \mathcal{Q}_i \end{cases}$$

comment: Unfilled orders

$\mathcal{A}_{s+1} := \mathcal{A}_0 - \mathcal{O}_0 + \alpha$
$\mathcal{O}_{s+1} := \textsc{OptimalPlacement}(\mathcal{A}_{s+1}, Q_{s+1})$
$\mathcal{N}_{s+1} := \mathcal{O}_{s+1} - \mathcal{Q}_{s+1}$

for $i := s + 2$ **to** k

$$\textbf{do} \begin{cases} \mathcal{A}_i := \mathcal{A}_{i-1} - \mathcal{O}_{i-1} \\ \mathcal{O}_i := \textsc{OptimalPlacement}(\mathcal{A}_i, Q_i) \\ \mathcal{N}_i := \mathcal{O}_i - \mathcal{Q}_i \end{cases}$$

return $(\mathcal{N}(\mathcal{S}))$

Remark 4.4.2. *When $Q_i = 0$, its optimal placement is by definition zero vector.*

The Algoritam is well defined, in sense that sum of all optimal placement vectors $\mathcal{O}_i$, $i = 0, 1, \ldots, k$ is an optimal placement vector for overall quantity Q. First we will prove the statement for just two orders.

Lemma 4.4.1. *Let $\mathcal{A} = [A_0, A_1, \ldots, A_l, 0, \ldots, 0]^T$ be available quantity. For two orders with size Q_1 and Q_2, following equality is satisfied*

$$\mathcal{O}_{Q_1} + \mathcal{O}_{Q_2} = \mathcal{O}_{Q_1+Q_2}. \tag{4.27}$$

Where $\mathcal{O}_{Q_1}$ and $\mathcal{O}_{Q_2}$ are optimal placements obtained by Algorithm 1, and $\mathcal{O}_{Q_1+Q_2}$ is optimal placement for quantity $Q_1 + Q_2$.

Proof. To be able to use Algorithm 1, two orders will be represented in the form of strategy $\mathcal{S}$, in which all orders have zero quantity except orders at levels k_1 and k_2, where they have quantities Q_1 and Q_2, respectively. Because of *time priority* in order execution and the fact that available quantity is determined for moment $t = 0$, a case where both orders are put at same price level will be regarded as impossible. Now, without loss of generality, it could be assumed that $k_1 < k_2$. We consider three cases.

1) Both orders are filled: It is evident that when both orders are filled $s = l$, and α is zero vector. Then $\mathcal{A}_s = \mathcal{A}$, and because of Remark 4.4.2, $\mathcal{A}_{k_2} = \mathcal{A}$. For the quantity Q_2 we define set I_{H_2} using (4.5). and denote its maximal element by h_2. Then by Theorem 4.2.2

$$\mathcal{O}_{Q_2} = [0, \dots, 0, Q_2 - \sum_{j=h_2+1}^{l} A_j, A_{h_2+1}, \dots, A_l, 0, \dots, 0]^T$$
$$\mathcal{A}_{k_2-1} = \mathcal{A}_{k_2} - \mathcal{O}_{Q_2}.$$

Again Remark 4.4.2 gives $\mathcal{A}_{k_1} = \mathcal{A}_{k_2-1}$ so

$$\mathcal{A}_{k_1} = [A_0, A_1, \dots, (\sum_{j=h_2}^{l} A_j) - Q_2, 0, \dots, 0]^T.$$

Now, we can define I_{H_1} and optimal placement for Q_1.

$$I_{H_1} = \{j | (\sum_{i=h_2}^{l} A_i) - Q_2 + \sum_{i=j}^{h_2-1} A_i \geq Q_1, j = 0, \dots, h_2\}.$$

I_{H_1} could be rewritten as

$$I_{H_1} = \{j | \sum_{i=j}^{l} A_i \geq Q_2 + Q_1, j = 0, \dots, h_2\}$$
$$h_1 = \max(I_{H_1})$$

$$\begin{aligned}\mathcal{O}_{Q_1} = [0, \ldots, 0,& \\ & Q_1 + Q_2 - \sum_{j=h_1+1}^{l} A_j, A_{h_1+1}, \ldots, \\ & A_{h_2-1}, (\sum_{j=h_2}^{l} A_j) - Q_2, 0, \ldots, 0]^T\end{aligned}$$

Sum of optimal placements of these orders is

$$\begin{aligned}&\mathcal{O}_{Q_2} + \mathcal{O}_{Q_1} = \\ &\quad = [0, \ldots, 0, Q_2 - \sum_{j=h_2+1}^{l} A_j, A_{h_2+1}, \ldots, A_l, 0, \ldots, 0]^T + \\ &\quad + [0, \ldots, 0, Q_1 + Q_2 - \sum_{j=h_1+1}^{l} A_j, \\ &\quad A_{h_1+1}, \ldots, A_{h_2-1}, (\sum_{j=h_2}^{l} A_j) - Q_2, 0, \ldots, 0]^T \\ &\quad = [0, \ldots, 0, Q_1 + Q_2 - \sum_{j=h_1+1}^{l} A_j, A_{h_1+1}, \ldots, A_l, 0, \ldots, 0]^T.\end{aligned}$$

Now, optimal placement for order with size $Q_1 + Q_2$ is defined by Theorem 4.2.2

$$\mathcal{O}_{Q_1+Q_2} = [0, \ldots, 0, Q_1 + Q_2 - \sum_{j=h+1}^{l} A_j, A_{h_1+1}, \ldots, A_l, 0, \ldots, 0]^T,$$

where h is maximum element of set I_H defined for $Q_2 + Q_1$ using (4.5). To prove (4.27), it has to be proved that $h = h_1$ holds. The fact $I_{H_1} \subseteq I_H$ gives $h_1 \leq h$.

As $\sum_{j=h}^{l} A_j \geq Q_2 + Q_1 > Q_1$ holds, then $h \in I_{H_1}$, which leads to $h_1 \geq h$. This means that $h_1 = h$, and (4.27) is true for this case.

2) One order is filled, and the other one is (partially) unfilled: Here we distinguish two cases, the first, when order is partially unfilled, and the second case when it is completely unfilled.

(2.1) Unfilled order is partially unfilled: $s = l-1$ then $\mathcal{A}_s = \mathcal{A} - \alpha$, and because of Remark (4.4.2)

$$\mathcal{A}_{k_1} = [A_0, A_1, \dots, A_{l-1}, 0, \dots, 0]$$

$$\begin{aligned}
&I_{H_1} = \{j | \sum_{i=j}^{l-1} A_j \geq Q_1, j = 0, \dots, l-1\} \\
&h_1 = \max(I_{H_1}) \\
&\mathcal{O}_{Q_1} = [0, \dots, 0, Q_1 - \sum_{j=h_1+1}^{l-1} A_j, A_{h_1+1}, \dots, A_{l-1}, 0, \dots, 0]^T \\
&\mathcal{A}_{k_1-1} = \mathcal{A}_{k_1} - \mathcal{O}_{Q_1}
\end{aligned}$$

As order with quantity Q_2 is partially filled it is clear that $k_2 = l = s+1$ then available quantity is

$$\begin{aligned}
\mathcal{A}_{s+1} &= \mathcal{A}_0 - 0 + \alpha \\
&= [A_0, A_1, \dots, A_{h_1-1}, \sum_{j=h_1}^{l-1} A_j - Q_1, 0 \dots, 0, A_l, 0, \dots, 0]
\end{aligned}$$

Also, because of being partially filled, it means $Q_2 > A_l$, so none of indices $j = h_1+1, \dots, l$ are in the set for which is satisfied $\sum_{i=j}^{l} A_i \geq Q_2$, so

$$\begin{aligned}
I_{H_2} &= \{j | A_l + \sum_{i=h_1}^{l-1} A_i - Q_1 + \sum_{i=j}^{h_1-1} A_i \geq Q_2, j = 0, \dots, h_1\} \\
&= \{j | \sum_{i=j}^{l} A_i \geq Q_2 + Q_1, j = 0, \dots, h_1\}.
\end{aligned}$$

If h_2 is maximum element of I_{H_2}, then optimal placement is

$$\begin{aligned}
\mathcal{O}_{Q_1} = [0, \dots, 0, Q_1 + Q_2 - \sum_{h_2+1}^{l} A_j, \dots, A_{h_1-1}, \\
\sum_{j=h_1}^{l-1} A_j - Q_1, 0, \dots, 0, A_l, 0 \dots, 0]^T
\end{aligned}$$

Now,

$$\mathcal{O}_{Q_1} + \mathcal{O}_{Q_2} = [0, \dots, 0, Q_1 + Q_2 - \sum_{h_2+1}^{l} A_j, A_{h_2+1} \dots, A_l, 0 \dots, 0]^T$$

Similarly to previous case it could be proved that

$$h_2 = \max\{j | \sum_{i=j}^{l} A_i \geq Q_2 + Q_1, j = 0, \dots, l\}$$

which means $\mathcal{O}_{Q_1} + \mathcal{O}_{Q_2} = \mathcal{O}_{Q_1+Q_2}$.

(2.2) Unfilled order is completely unfilled: Then is $s = l$ and available quantity is

$$\mathcal{A}_{k_1} = [A_0, A_1, \dots, A_l, 0, \dots, 0]$$

Now optimal placement for quantity Q_1 and available quantity for order with quantity Q_2 could be defined

$$I_{H_1} = \{j | \sum_{i=j}^{l} A_i \geq Q_1, j = 0, \dots, l\}$$

$$h_1 = \max(I_{H_1})$$

$$\mathcal{O}_{Q_1} = [0, \dots, 0, Q_1 - \sum_{j=h_1+1}^{l} A_j, A_{h_1+1}, \dots, A_l, 0, \dots, 0]^T$$

$$\mathcal{A}_{k_1-1} = [A_0, \dots, A_{h_1-1}, \sum_{j=h_1}^{l} A_j - Q_1, 0, \dots, 0]^T$$

And then optimal placement for Q_2 could be defined

$$I_{H_2} = \{j | \sum_{i=j}^{l} A_i \geq Q_2 + Q_1, j = 0, \dots, h_1\}$$

$$h_2 = \max(I_{H_2})$$

$$\mathcal{O}_{Q_2} = [0, \dots, 0, Q_1 + Q_2 - \sum_{j=h_2+1}^{l} A_j, \dots, A_{h_1-1}, \sum_{j=h_1}^{l} A_j - Q_1, 0 \dots, 0]^T$$

Like in previous case because

$$h_2 = \max\{j | \sum_{i=j}^{l} A_i \geq Q_2 + Q_1, j = 0, \ldots, h_1\}$$

and (4.27) holds.

3) Both orders are unfilled: In both cases $s = l$ and $s = l - 1$,

$$\mathcal{A}_s = \mathcal{A} - \alpha$$

because all orders at levels $j \leq s$ are with quantity zero $\mathcal{A}_0 = \mathcal{A}_s$ then

$$\mathcal{A}_{s+1} = (\mathcal{A} - \alpha) + \alpha = \mathcal{A}$$

Then, simply by following proof for case where one order is filled, and other is completely unfilled, one gets (4.27). □

Corollary 4.4.1. *Let $\mathcal{O}_Q$ be optimal placement for quantity Q, and available quantity $\mathcal{A}$. If $\mathcal{O}_i$ is optimal placements for orders Q_i, $i = 0, \ldots, k$ in (4.25) obtained by Algorithm 1, then*

$$\sum_{i=0}^{k} \mathcal{O}_i = \mathcal{O}_Q. \tag{4.28}$$

Proof. The proof consist of successive application of result (4.27), starting from index s following the path like in Algorithm 1. □

Chapter 5

Stochastic Optimization

Many optimization problems that arise in fields such as finance, engineering, medicine, machine learning involve uncertainty as part of objective function, constraints or both. It can manifest itself trough noisy measurement, i.e. it consists of an ideal part (exact value) and error. In some cases, the decision depends on one or more parameters whose values are unknown at the moment of the decision but will be known in the future. For example, in Markowitz model investor wants to invest his capital in some number of stocks, he has to make a decision about weights of each stock in his portfolio to maximize overall return and minimize risk, while his decision depends on random returns.

The uncertainty and its formalization bring difficulty, both in modeling and optimization of the problem under consideration, and as a result, there is a vast number of approaches to formulating and solving optimization problems involving uncertainty. The stochastic optimization problem is usually formulated as

$$\min_{x \in \Theta} f(x) := \mathbb{E}(g(x, \omega)) \tag{5.1}$$

where Θ is nonempty subset of $I\!R^p$ and $g(x, \omega)$ is a function of decision variable x, and random vector ω with support $\Omega \subset I\!R^d$ and probability distribution P. The mathematical expectation $\mathbb{E}$ is defined with respect to ω in the probability space $(\Omega, \mathcal{F}, P)$.

An example of such method is the problem (5.2), where we have to decide what would be the optimal execution with respect to the Negative Selection.

$$\begin{aligned} &\min_{Q_0,\dots,Q_k} E(||\boldsymbol{\mathcal{N}}(\mathcal{S})||_1) \\ &\text{subject to} \sum_{j=0}^{k} Q_j = Q \\ &\qquad Q_j \geq 0, j = 0, \dots, k. \end{aligned} \tag{5.2}$$

Where $\mathcal{S}$ and $\boldsymbol{\mathcal{N}}(\mathcal{S})$ are defined by (4.25) and (4.26) respectively. Given the fact that for single order NS represents the distance from optimal placement, the idea is to consider norm-1 of NS vector $\boldsymbol{\mathcal{N}}(\mathcal{S})$ and to make it as small as possible for a given quantity of shares Q.

Clearly, it is stochastic value, as we face the uncertainty of the quality of order execution under prevalent market conditions, i.e. the uncertainty of the achieved execution price for both, market and limit orders, as well as the uncertainty of execution of a limit order. It is indubitably difficult to analytically formulate the relation between market conditions and resulting Negative Selection values. Instead, we deal with large amounts of data as the input for the simulator to run, and obtained simulations enable us to analyze our model and perform the optimization. In problem (5.2) we consider constrained optimization, as we impose the constraint that the sum of decision variables must be equal to a given quantity, and their value must also be nonnegative.

It is clear that the function $f(x)$, defined in (5.1), does not include uncertainty, as it is a mathematical expectation over random variable ω. However, application of methods for deterministic problems can be difficult, as the analytical form of the function f is seldom available, and consequently, evaluation of gradient is often intractable.

Here, we consider two approaches for solving stochastic optimization problems of the form (5.1): Stochastic Approximation (SA) and Sample Average Approximation (SAA). SA is a very developed field with an abundance of literature covering theory and applications. It allows noisy inputs, but the efficiency of classical methods depends on the choice of the gain sequence. On the other hand, SAA is very general, does not require convexity of the objective function. It often needs a large sample of data, which can be computationally expensive.

5.1 Stochastic Approximation

Stochastic Approximation (SA) was introduced 1951 by Robbins and Monro [45] for solving root-finding problems with noisy measurements. Afterward, gradient-free version of the algorithm (5.4) was formulated by Kiefer and Wolfowitz [23]. Robbins-Monro (RM) and Kiefer-Wolfowitz (KW) algorithms address the one-dimensional unconstrained case and are often regarded as classical methods, as they are the foundation for the development of SA. In case of solving $\nabla f = 0$, RM algorithm is defined as

$$x_{k+1} = x_k - a_k \widehat{\nabla} f(x_k). \tag{5.3}$$

The nonnegative scalar a_k is the step size, which is also called *gain*, *stochastic gradient* $\widehat{\nabla} f(x_k)$ is an estimate of the gradient $\nabla f(x_k)$. Clearly, the formulation of the algorithm is inspired by steepest descent in deterministic optimization and is also called stochastic gradient descent. The KW algorithm uses finite difference as an estimate for stochastic gradient

$$x_{k+1} = x_k - a_k \frac{g(x_k + c_k, \omega_k) - g(x_k + c_k, \omega_k)}{2c_k}. \tag{5.4}$$

Under certain conditions (see [15]), RM converges in mean square to its solution, while KW converges in probability. A practical performance of both classical methods depends on the choice of sequences $\{a_k\}$ and $\{c_k\}$. Sacks [47] proved that under appropriate assumptions, RM algorithm converges asymptotically at a rate of $O(k^{-\frac{1}{2}})$. In his book [6], Banks suggests $a_k = \frac{a}{k}, c_k = \frac{c}{k^{1/6}}$, for all k and some positive scalars a and c. Then, under certain conditions, KW algorithm (5.4) converges asymptotically at a rate of $O(k^{-\frac{1}{3}})$. When one-sided finite difference is an estimate for stochastic gradient, algorithm can reach asymptotic rate of convergence of $O(k^{-\frac{1}{4}})$.

Now we consider root finding problem $h(x) = 0$ for a function $h : I\!R^p \to I\!R^p$, whereby in each iteration k only noisy measurements $h_k(x)$ of $h(x)$ are available, i.e.

$$h_k(x) = h(x) + \epsilon_k(x). \tag{5.5}$$

Then the SA algorithm is defined by

$$x_{k+1} = \Pi_\Theta(x_k - a_k h_k(x_k)), \tag{5.6}$$

where $\Pi_\Theta(x)$ denotes projection of x into the feasible region Θ and it is required only in the framework for constrained optimization. The following theorem states its convergence property.

Theorem 5.1.1. *[52] Consider the unconstrained algorithm 5.6. Suppose that x^* is a unique solution of $h(x) = 0$ and that the following conditions hold*

1. *$\{a_k\}$ is a sequence of positive constants such that $\sum_{k=1}^{\infty} a_k = \infty$ and $\sum_{k=1}^{\infty} a_k^2 < \infty$.*
2. *For some symmetric, positive definite matrix B and every $0 < \eta < 1$*
$$\inf_{\eta<||x-x^*||<1/\eta} (x - x^*)^T B \nabla f(x) > 0.$$
3. *$\mathbb{E}(\epsilon_k(x)) = 0$ for all x and k*
4. *$||h(x)||^2 + \mathbb{E}(||\epsilon_k(x)||^2) \leq c(1 + ||x||^2)$ for all x and k and some $c > 0$.*

Then $lim_{k\to\infty} x_k \to x^*$ w.p.1.

The first condition, i.e. restriction on gain sequence a_k, regulates the speed of step size convergence to zero. Obviously, condition $\sum_{k=1}^{\infty} a_k^2 < \infty$ ensures $a_k \to 0$, while $\sum_{k=1}^{\infty} a_k$ slows down its convergence, preventing the algorithm to converge to early with poor estimate of x^*. In practice, a_k has the following form
$$a_k = \frac{a}{(k + A)^{\alpha}}, \tag{5.7}$$
where $a > 0$, $A \geq 0$ and $\alpha \in (0.5, 1]$.

In second condition, common choice for matrix B is an identity matrix. The third condition implies that $h_k(x)$ is unbiased estimator of $h(x)$ Because, the following holds
$$\mathbb{E}(||h_k(x)||^2) = ||h(x)||^2 + \mathbb{E}(||\epsilon_k(x)||^2),$$
then we see that the fourth condition puts a bound on growth of $\mathbb{E}(||h_k(x)||^2)$.

Many realistic optimization problems come in the form of "Black Box", in which case analytic form of the function is unavailable, and also the gradient cannot be obtained. We consider objective function $f(x)$, and its noisy measurement
$$\widehat{f}(x) = f(x) + \xi(x).$$

In this case, like in KW algorithm, derivatives are approximated using evaluations of the objective function. The FDSA algorithm is defined by

(5.6), where $h_k(x_k)$ represents the finite difference approximation of gradient. Here, there are two possibilities for finite difference approximation. The first is two-sided (central) FD approximation where jth component of $h_k(x_k)$ is

$$\frac{\widehat{f}(x_k + c_k e_j) - \widehat{f}(x_k - c_k e_j)}{2c_k} \tag{5.8}$$

where e_j denotes the jth coordinate vector in $\mathbb{R}^p$. The second possibility is one-sided finite difference approximation, where jth component of $h_k(x_k)$ is defined as

$$\frac{\widehat{f}(x_k + c_k e_j) - \widehat{f}(x_k)}{c_k}. \tag{5.9}$$

In first case there is $2p$ evaluation of the objective function, while in second $p+1$, but as we can see in case of classical methods in first case there is a better convergence rate. Like in previous case practical performance depends on the choice of sequences $\{a_k\}$ and $\{c_k\}$, so in the following result on convergence of FDSA, there are restrictions placed on gain sequences.

Theorem 5.1.2. *[52] Consider the unconstrained algorithm (5.6) with stochastic gradient defined by (5.8). Suppose that the following conditions hold*

1. *Let $\{a_k\}$ and $\{c_k\}$ be positive tuning sequences satisfying the conditions*

$$c_k \to 0,\ \sum_{k=1}^{\infty} a_k = \infty,\ \sum_{k=1}^{\infty} a_k c_k < \infty, \sum_{k=1}^{\infty} a_k^2 c_k^{-2} < \infty.$$

2. *There is a unique minimum x^* such that for every $\eta > 0$,*

$$\inf_{|||x-x^*||>\eta} ||h(x)|| > 0 \text{ and } \inf_{|||x-x^*||>\eta} ||f(x) - f(x^*)|| > 0$$

3. *For all i and k, $\mathbb{E}(\xi_k(x_k + c_k e_i) - \xi_k(x_k - c_k e_i)|\mathcal{I}_k) = 0$ w.p.1 and $\mathbb{E}((\xi_k(x_k \pm c_k e_i))^2|\mathcal{I}_k) \leq C$ w.p.1 for some $C >$ that independent of k and i, where $\mathcal{I}_k = \{x_0, \ldots, x_k\}$ contains information on previous iterations.*

4. *The Hessian matrix $\nabla^2 f(x)$ exists for all x and it is uniformly bounded in norm for all $x \in I\!R^p$*

Then $lim_{k\to\infty} x_k \to x^*$ w.p.1.

For the proper practical performance of algorithm the two sequences $\{a_k\}$ and $\{c_k\}$ need some tuning, this is usually done on a small scale version of the full problem. Spall suggests [52]

$$a_k = \frac{a}{(k+1+A)^\alpha}, c_k = \frac{c}{(k+1)^\gamma}, \tag{5.10}$$

with $\alpha = 0.602$ and $\gamma = 0.101$. One way of tuning the sequences is trough trail and error. However, sometimes one can apply "semiautomatic" method for choosing the a, A, c, α and γ, that is in full details described in [52]. Proper choice of gain sequences is essential because if a_k is too small relative to a gradient, iterations slowly progress towards to optimum. On the other hand, large a_k relative to a gradient may cause iterations to extremely oscillate without approaching to optimum. One approach to reduce the sensitivity is introduction of an adaptive step-size rule. For example Kestens's rule allows step size to decrease only if there is a directional change in iterations.

Many researchers have worked on the enhancement of classical methods. We already mentioned Kesten's rule, and there is also Averaging Iterates variation of SA. The idea behind Averaging Iterates lies in choosing the biggest step size a_k so that iterates oscillate around optimum. Then instead of taking the last iteration x_N as output, optimum is estimated by taking average of all N iteration,i.e. $\frac{1}{N}\sum_{k=1}^{N} x_k$. Another variation is called "sliding window", which takes the average over last m iterates.

Other variations of SA are Varying Bonds, Simultaneous perturbation stochastic approximation (SPSA), for the comprehensive overview of variation of classical algorithms and new advances, see [6, 15].

5.2 Sample Average Approximation

Sample Average Approximation for solving the problem (5.1) can be outlined in the following way. First, we choose a random sample $\omega_1, \omega_2, \ldots, \omega_N$ that is independently and identically distributed (i.i.d.) with the same distribution as ω, then consider the following estimation of the problem (5.1).

$$\min_{x\in\Theta} f_N(x) := \frac{1}{N}\sum_{i=1}^{N} g(x, \omega_i). \tag{5.11}$$

Clearly, for a fixed sample $f_N(\cdot)$ is a deterministic, and we are in a position to apply deterministic optimization methods to solve the problem (5.11). At the end we take the optimizer of (5.11), denoted by x_N^* to be estimator of the solution of (5.1).

Now, we approach to SAA more formally. We will assume that the objective function $f(x)$ of problem (5.1) is well defined and finite valued for all $x \in \Theta$ $\omega_1, \omega_2, \ldots, \omega_N$ that is independently and identically distributed (i.i.d.) with the same distribution as ω.

As f_N depends on random sample $\omega_1, \omega_2, \ldots, \omega_N$, it is a random function.As the sample is i.i.d., by the Strong Law of Large Numbers (SLLN) for every $x \in \Theta$

$$\lim_{N\to\infty} f_N(x) = f(x) \text{ w.p.1.}$$

As this is pointwise convergence, we continue with definition of uniform convergence w.p.1., we say that f_N converges to f w.p.1 uniformly on S if

$$\lim_{N\to\infty} \sup_{x\in\Theta} |f_N(x) - f(x)|.$$

Before stating the condition for uniform uniform convergence w.p.1., we need the following: We say that $g(x, \omega)$ is dominated by integrable function, if there exists a nonnegative measurable function $H(\omega)$, such that $\mathbb{E}(H(\omega)) < \infty$ and $P(g(x, \omega) \leq H(\omega)) = 1$.

Theorem 5.2.1. *[51] Let S be nonempty compact subset of $I\!R^p$, and suppose that*

i) for any $x \in S$ the function $g(\cdot, \omega)$ is continuous at x for almost every ω

ii) the sample $\omega_1, \ldots, \omega_N$ is i.i.d., and

iii) $g(x, \omega), x \in S$ is dominated by integrable function.

Then $f(x) = \mathbb{E}(g(x, \omega))$ is finite valued and continuous on S, and f_N converges to f w.p.1 uniformly on S.

Let X^* and X_N^* be sets of optimal solution for original optimization problem (5.1), and sample average approximation (SAA) (5.11), respectively i.e.

$$f^* = f(x^*) = \min_{x\in\Theta} \ f(x), \forall \ x^* \in X^*$$

$$f_N^* = f_N(x_N^*) = \min_{x\in\Theta} \ f_N(x), \forall \ x_N^* \in X_N^*$$

Theorem 5.2.2. *[51] Suppose that $f_N(x)$ converges to $f(x)$ w.p.1, as $N \to \infty$, uniformly on Θ. Then f_N^* converges to f^* w.p.1 as $N \to \infty$.*

Before stating the theorem on convergence results, we need a definition of the deviation between two sets $A, B \subset I\!R^n$ we denote by

$$Dev(A, B) = \sup_{x \in A} d(x, B),$$

the *deviation* of set A from set B, whereby $d(x, B)$ denotes the Euclidean distance of point x, form set B,i.e. $d(x, B) = \inf_{x' \in B} ||x - x'||$.

Theorem 5.2.3. *[51] Suppose that there exists a compact set $S \subset I\!R^p$ such that $\varnothing \neq X^* \subset S$ and the following holds:*

i) the function f is finite valued and continuous on S,

ii) $f_N(x)$ converges to $f(x)$ w.p.1, as $N \to \infty$ uniformly in $x \in S$,

iii) for N large enough the set $X_N^* \neq \varnothing$ and $X_N^* \subset S$.

Then $f_N^* \to f^*$ and $Dev(X_N^*, X^*) \to 0$ w.p.1 as $N \to \infty$.

$Dev(X_N^*, X^*) \to 0$ w.p.1, guarantees that the distance of the optimal solution x_N^* of SAA problem (5.11) from set of solutions of original problem converges to zero w.p.1, i.e. $d(x_N^*, X^*) \to 0$, w.p.1. In special case, when $X^* = \{x^*\}$, the sequence x_N^* converges to solution w.p.1 as $N \to \infty$.

If the function $g(x, \omega)$ is convex for any $\omega \in \Omega$ then $f(x)$ in original problem (5.1) is convex, and also f_N, are convex functions. This property with theory of epiconvergence, leads to relaxation of Theorem 5.2.3 in [51].

We now consider case when set of constraints is finite, and for some $\epsilon \geq 0$ we define sets of ϵ-optimal solutions

$$X^\epsilon = \{x \in \Omega | f(x) \leq f^* + \epsilon\}$$

$$X_N^\epsilon = \{x \in \Omega | f_N(x) \leq f_N^* + \epsilon\}$$

of original (5.1) and SAA (5.11) problem, respectively. Clearly, for $\epsilon = 0$, X^ϵ and X_N^ϵ coincide with X^* and X_N^*, respectively.

It can be shown that for all $\epsilon \geq 0$, $P(X_N^\epsilon \subset X^\epsilon) \to 1$, as $N \to \infty$ [28], which basically means that for large enough N, ϵ-optimal solution of SAA is ϵ-optimal solution of the original problem.

To state the next theorem, we need the following definitions. For some random variable with mean $\mu = \mathbb{E}(Y)$, its moment-generating function is $M(t) = \mathbb{E}(e^{tY})$, and conjugate function $I(z) = \sup_{t\in\mathbb{R}}\{tz - \Lambda(t)\}$ of logarithmic moment-generating function $\Lambda(t) = \log M(t)$, is called (large deviation (LD)) rate function of Y.

Let us also, consider the mapping $u : \Omega \setminus X^{\epsilon} \to \Omega$ such that

$$f(u(x)) \leq f - \epsilon^*, \forall x \in \Omega \setminus X^{\epsilon},$$

for some $\epsilon^* \geq \epsilon$, and LD rate function of $g(u(x), \omega) - g(x, \omega)$ denoted by $I_x(\cdot)$.

Assumption A1. *For every $x \in \Omega \setminus X^{\epsilon}$, the moment generating function of the random variable $Y(x, \omega) = F(u(x), \omega) - F(x, \omega)$ is finite valued in a neighborhood of t=0.*

Theorem 5.2.4. *[51] Let ϵ and δ be nonegative numbers such that $\delta \leq$. Then*

$$1 - P(X_N^{\delta} \subset X^{\epsilon}) \leq |X| e^{-N\gamma(\delta,\epsilon)},$$

where

$$\gamma(\delta, \epsilon) = \min_{x\in\Omega\setminus X^{\epsilon}} I_x(-\delta).$$

Moreover, if $\delta < \epsilon^*$ and Assumption A1, holds, then $\gamma(\delta, \epsilon) > 0$.

The lower bound for $I_x(-\delta)$ is

$$I_x(-\delta) \geq \frac{(\epsilon^* - \delta)^2}{2\sigma_x^2},$$

where $\sigma_x^2 = Var(Y(x, \omega))$.

Theorem 5.2.5. *[51] Suppose that there is a constant $\sigma > 0$ such that for any $x \in \Omega \setminus X^{\epsilon}$ the momentum-generating function $M_x(t)$ of the random variable $Y(x, \omega) - \mathbb{E}(Y(x, \omega))$ satisfies*

$$M_x(t) \leq e^{\sigma^2 t^2/2}.$$

Then for $\epsilon > 0$, $0 \leq \delta < \epsilon$, and $\alpha \in (0, 1)$ and for the sample size

$$N \geq \frac{2\sigma^2}{\epsilon - \delta^2} \ln(\frac{|\Omega|}{\alpha})$$

it follows that

$$P(X_N^{\delta} \subset X^{\epsilon}) \geq 1 - \alpha.$$

Lower bound of N gives us information of the problem complexity, as we can see that it linearly depends on variance and that even exponential increase in size $|\Omega|$, produces the linear increase of N. However, it is regarded as "conservative for practical estimates" [28], as it can lead to computational inefficiency.

In essence, the sample size, represents a tradeoff between precision and cost, as large sample size provides better approximation but causes higher computation costs and vice versa. In general, a large sample size is needed to obtain estimates of reasonable accuracy. This fact causes considerable computational effort in solving (5.1) as the computation of the objective function, as well as its derivatives, becomes very costly. The general approach is to consider a sequence of approximations (5.11) with an increasing sample size, i.e., with a different sample size in each iteration and lower the cost of the overall optimization procedure. The dominant way of sample size scheduling is an increasing sample size sequence that results in smaller computational costs than working with a large sample from the beginning. There are main approaches in the sample size scheduling - a predetermined sample size schedule, for example, [56] or an adaptive sample size schedule, [30, 55, 57]. An overview of different sample size scheduling is presented in [29].

Next, we preset results from original work from Krejić and Lončar [31], a nonmonotone line search method for solving unconstrained optimization problems using SAA is presented, the convergence is obtained in the sense of zero upper density.

5.3 Nonmonotone Line Search Method

The problem that we consider is an unconstrained problem of the form

$$\min_{x \in \mathbb{R}^p} f(x), \tag{5.12}$$

where the objective function f is given as

$$f(x) = \mathbb{E}(g(x, \omega)). \tag{5.13}$$

The mathematical expectation $\mathbb{E}$ is defined with respect to ω in the probability space $(\Omega, \mathcal{F}, P)$. It is assumed that the function $g : \mathbb{R}^p \times \Omega \to \mathbb{R}$ is known analytically. However, the analytical form of the function f is seldom available and needs to be estimated in some way, and for this purpose, we use the Sample Average Approximation defined as

$$G(x,w) = \frac{1}{n}\sum_{j=1}^{n} g(x,\omega_j), \tag{5.14}$$

where $\omega = \{w_1, \ldots, w_n\}$ is random sample of size n.

A tradeoff between precision and cost is represented by the sample size n because large samples cause higher computation costs but also provide more precision, while smaller samples reduce costs at the expense of approximation quality. There are many different ways of choosing the sequence $\{n(i)\}$ of sample sizes at each iteration, but the prevailing way of sample size scheduling is an increasing sample size sequence, and in this way taking advantage of having smaller computational costs at the beginning, when the iterations are not close to the solution. Two main approaches can be distinguished in the sample size scheduling - a predetermined sample size schedule [56] or an adaptive sample size schedule [30, 55, 57].

The monotone line search method for (5.12)-(5.13) with a predetermined sample size sequence is defined and considered for problems of type (5.12) in [56]. The method is based on a decrease determined by the Armijo rule in each iteration, for the approximate objective function defined with the current sample in the iteration, with search direction determined as an approximate negative gradient. The method converges with zero upper density.

Here we consider the nonmonotone line search rule due to Li, Fukushima [34] which is applied in many papers, for deterministic and stochastic problems, for example, see [1, 30].

The main contribution of the following result is a generalization of the results presented in [56] in the following sense. We define a nonmonotone line search strategy that allows us to take an arbitrary search direction, which needs to approach the negative gradient only in the limit and prove the convergence of the proposed algorithm in terms of zero upper density, as in [56]. We also present a set of initial testing results that confirm the theoretical results and provide empirical evidence for the proposed algorithm.

Preliminaries

First, we give outline of the results of Wardy [56] that will allow us to propose a nonmonotone line search method and prove its convergence. We start with the definition of upper density convergence.

Definition 5.3.1. *[56] Let K be a set of integers. The upper density of K, denoted by $ud(K)$ is the quantity*

$$ud(K) = \limsup_{i\to\infty} \frac{|K \cap [1,i]|}{i}\ , \tag{5.15}$$

where $|S|$ denotes cardinality of set S, and for integers i and j, $j \geq i$

$$[i,j] := \{i, i+1, \ldots, j\}.$$

The convergence in upper density in formulated in terms of optimality functions. The function $\theta : I\!R^p \to I\!R^+$ is an optimality function if $\theta(x) = 0$ if and only if x satisfies the optimality conditions.

Definition 5.3.2. *[56] An algorithm which generates sequences $x_1, x_2, \ldots$ in $\mathbb{R}^p$ is said to converge with zero upper density (ud) on compact sets if with probability 1, if $\{x_i\}$ is a bounded sequence, then there exists a set of integers J, such that $ud(J) = 0$ and $\theta(x_i) \underset{i \notin J}{\to} 0$.*

We will prove that the nonmonotone line search method we propose here converges in upper density as in [56]. To do that, we need to assume the following.

Assumption A2. *[56]*

$$\textit{If } x_i \to x,\ x_i \in I\!R^p, i = 1,2,3,..\ \textit{then } \theta(x) = 0 \textit{ if and only if } \theta(x_i) \to 0. \tag{5.16}$$

The optimality function we consider is the norm of the gradient of the objective function i.e.

$$\theta(x) = ||\nabla f(x)||$$

and thus the assumption above is satisfied.

An algorithm which generates sequence $\{x_i\}_{i \in I\!N}$, converges with *zero upper density* on a compact set if the sequence is bounded and there exists w.p.1 a set J with $ud(J) = 0$, such that the any accumulation point of subsequence $\{x_i\}_{i \in I\!N \setminus J}$ satisfies the optimality conditions.

Let us now recall the notation needed for formulation of conditions for convergence with zero upper density on compact sets, [56]. For every compact set $\Gamma \subset I\!R^p$, $r \geq 0$, $s \geq 0$ and integer i, the following events are defined:

- $E_i(\Gamma, r)$ is the event that $x_i \in \Gamma$ and $\theta(x_i) \geq r$.
- $G_i(\Gamma, s)$ is the event that $x_i \in \Gamma$ and $f(x_{i+1}) - f(x_i) \geq -s$.
- $H_i(\Gamma, s)$ is the event that $x_i \in \Gamma$ and $f(x_{i+1}) - f(x_i) \geq s$.

Here, $\mathcal{F}_i$ is the σ-algebra generated by all the information leading to the construction of x_i.

The following two conditions together constitute a sufficient condition for the convergence in zero upper density if f is continuous function and the iterations are generated by a line search with a random sample of predetermined size at each iteration. Let C_i be an arbitrary event from $\mathcal{F}_i$.

Condition 1. *[56] For every compact set $\Gamma \subset I\!R^p$ and $r > 0$, there exists $s > 0$ such that, for every $\epsilon > 0$, there exists an integer I such that for every $i \geq I$ and event $C_i \in \mathcal{F}_i$*

$$P(G_i(\Gamma, s)|E_i(\Gamma, r), C_i) < \epsilon \tag{5.17}$$

Condition 2. *[56] For every compact set $\Gamma \subset I\!R^p$, $s > 0$ and $\epsilon > 0$, there exists an integer I such that for every $i \geq I$ and event $C_i \in \mathcal{F}_i$*

$$P(H_i(\Gamma, s)|C_i) < \epsilon \tag{5.18}$$

We consider the prototype algorithm proposed in [56]. For integer sequence $n(i)$ determined a priori, the algorithm has the following structure:

Algorithm Prototype. *Data, $x_0 \in I\!R^p$.*

Step 0. *Set $i = 0$.*

Step 1. *Randomly draw $n(i)$ sample points $\omega^i := \{\omega_{i,1}, \omega_{i,2}, \ldots, \omega_{i,n(i)}\} \in \Omega$.*

Step 2. *Use ω^i to compute an approximation to $f(x_i)$, compute a descent direction h_i, and the next point x_{i+1}.*

Step 3. *Set $i = i + 1$ and go to Step 1.*

Theorem 5.3.1. *[56] If the Conditions 1 - 2 are satisfied and if f is continuous, then the Algorithm Prototype converges with zero upper density on compact sets.*

The following two assumptions characterise the problem we consider more closely.

Assumption A3. *The objective function f has the form (5.13), and $g(\cdot, \omega) \in C^2(\mathbb{R}^p)$.*

Assumption A4. *For every compact set $\Gamma \subset \mathbb{R}^p$, there exists $K > 0$ such that, for every $x \in \Gamma$ and $\omega \in \Omega$,*

$$|g(x,\omega)| + ||\frac{\partial g}{\partial x}(x,\omega)^T|| + ||\frac{\partial^2 g}{\partial x^2}(x,\omega)|| \leq K, \tag{5.19}$$

where $||\cdot||$ denotes vector norm, or induced matrix norm, depending on context.

The consequence of A4 is that f is continuously differentiable and ∇f is Lipschitz continuous on compact sets, so

$$\nabla f(x) = E\left(\frac{\partial g}{\partial x}(x,\omega)^T\right). \tag{5.20}$$

This fact justifies the choice of $||\nabla f(x)||$ as the optimality function i.e. $\theta(x) = ||\nabla f(x)||$. Clearly, the condition (5.16) holds.

The Nonmonotone Line Search Method

The modification that we introduce in the algorithm presented in [56] is that we use a general search direction satisfying (5.23), and nonmonotone Armijo rule, instead of monotone Armijo-type line search with a negative gradient as the search direction. The nonmonotonicity is defined by a sequence $\{\epsilon_i\}_{i\in\mathbb{N}}$ such that

$$\epsilon_i > 0, \ \sum_{i=0}^{\infty} \epsilon_i < \infty. \tag{5.21}$$

Algorithm 2. *Input: $x_0 \in \mathbb{R}^p$, $\{n(i)\}_{i\in\mathbb{N}}$, $\{\epsilon_i\}_{i\in\mathbb{N}}$, $\alpha \in (0,1)$, $\beta \in (0,1)$*

Step 0. *Set $i = 0$.*

Step 1. *Randomly draw $n(i)$ sample points $\omega^i := \{\omega_{i,1}, \omega_{i,2}, \ldots, \omega_{i,n(i)}\} \in \Omega$.*

Step 2. *Choose a search direction h_i.*

Step 3. *Set $k(i)$ to be the smallest integer k satisfying*

$$G(x_i - \beta^k h_i, \omega^i) - G(x_i, \omega^i) \leq -\alpha\beta^k ||h_i||^2 + \epsilon_i. \tag{5.22}$$

Set $x_{i+1} = x_i - \beta^{k(i)} h_i$, $i = i + 1$ and go to Step 1.

In Step 3 our goal is to find the step size that satisfies the nonmonotone Armijo condition, i.e. find the appropriate $k(i)$ that satisfies (5.22). Notice that Algorithm 2 is well defined for an arbitrary search direction as $\epsilon_i > 0$ so for any h_i there exists $k(i)$ large enough such that (5.22) holds and Step 3 finishes with a finite $k(i)$.

Theorem 5.3.2. *Assume that A3-A4 hold. If the search directions h_i in Step 2 of Algorithm are chosen such that*

$$\lim_{i\to\infty} \frac{||\nabla G(x_i, \omega^i) - h_i||}{\epsilon_i} = 0, \tag{5.23}$$

where $G(x_i, \omega^i) := \frac{1}{n(i)} \sum_{j=1}^{n(i)} g(x_i, \omega_{i,j})$ and $\nabla G(x_i, \omega^i) := \frac{\partial G}{\partial x}(x_i, \omega^i)^T$, then Algorithm 2 converges with zero upper density on compact sets to a stationary point of (5.12).

Proof. To prove the statement we need to show that Conditions 1 and 2 hold. Then the statement follows by Theorem 5.3.1. Let $\Gamma \subset I\!R^p$ be a compact set. First, we show that the sequence $||h_i||$ is bounded from above. Due to (5.23), there exists a constant K_0 such that $||h_i - \nabla G(x_i, \omega^i)|| \leq K_0$. Also, (5.19) guaranties that there exists $K_1 > 0$ such that $||\nabla G(x_i, \omega^i)|| \leq K_1$. So, for $M = 2\max\{K_0, K_1\}$, we have

$$||h_i|| \leq ||h_i - \nabla G(x_i, \omega^i)|| + ||\nabla G(x_i, \omega^i)|| \leq M. \tag{5.24}$$

Therefore, $||h_i||$ is bounded from above. Let us prove now that

$$\lim_{i\to\infty} \frac{|h_i^T h_i - \nabla G(x_i, \omega^i)^T h_i|}{\epsilon_i} = 0. \tag{5.25}$$

Given that

$$0 < |h_i^T h_i - \nabla G(x_i, \omega^i)^T h_i| \leq ||h_i - \nabla G(x_i, \omega^i)|| \cdot ||h_i|| \tag{5.26}$$

and that $\|h_i\|$ is bounded, the limit (5.23) implies that (5.25) holds.

Let us now prove that for an arbitrary compact set $\Gamma \subset I\!R$ there exists an integer $\overline{k}$ such that for every $x_i \in \Gamma$ we have . Let $x_i \in \Gamma$ and $\lambda \geq 0$. By the Mean value theorem we have

$$G(x_i - \lambda h_i, \omega^i) - G(x_i, \omega^i) = -\lambda \frac{\partial G}{\partial x}(x_i, \omega^i) h_i + \lambda^2 \int_0^1 (1-s) \langle \frac{\partial^2 G}{\partial x^2}(x_i - s\lambda h_i, \omega^i) h_i, h_i \rangle ds \quad (5.27)$$

By (5.25), there exists an integer i_0 such that for every $i \geq i_0$

$$-\lambda \frac{\partial G}{\partial x}(x_i, \omega^i) h_i \leq -\lambda ||h_i||^2 + \epsilon_i. \quad (5.28)$$

By CauchySchwarz's inequality, continuity of $\frac{\partial^2 G}{\partial x^2}(\cdot, \omega^i)$ and boundness of $||h_i||$ we obtain

$$|\lambda^2 \int_0^1 (1-s) \langle \frac{\partial^2 G}{\partial x^2}(x_i - s\lambda h_i, \omega^i) h_i, h_i \rangle ds| \leq \lambda^2 K ||h_i||^2 \quad (5.29)$$

Now, (5.27)-(5.29) implies

$$G(x_i - \lambda h_i, \omega^i) - G(x_i, \omega^i) \leq -\lambda(1 - \lambda K)||h_i||^2 + \epsilon_i \quad (5.30)$$

Substituting $\lambda = \beta^k$ in the above inequality, we get that (5.22) is satisfied if $\beta^k \leq (1-\alpha)/K$ holds, i.e for all $k \geq \frac{\log((1-\alpha)/K)}{\log(\beta)}$. Therefore, there exists $\overline{k}$ such that $k(i) \leq \overline{k}$.

Let us consider Condition 1. Let $\Gamma \subset I\!R^p$ be a compact set. Take $r > 0$ and $s = \frac{1}{2}\alpha\beta^{\overline{k}} r^2$ and $\epsilon > 0$. We can choose $\delta \in (0, r)$ such that

$$\alpha\beta^{\overline{k}}(r - \delta)^2 \geq s.$$

As $\sum_{i=0}^{\infty} \epsilon_i < \infty$, there exists an integer i_1 such that for every $i \geq i_1$ we have $\epsilon_i \leq \delta$.

Let $A(i)$ be the event: $x_i \in \Gamma$, and

$$||\nabla f(x_i) - \nabla G(x_i, \omega^i)|| < \frac{\delta}{2}, \ |f(x_i) - G(x_i, \omega^i)| < \frac{\delta}{2}, \ |f(x_{i+1}) - G(x_{i+1}, \omega^i)| < \frac{\delta}{2}.$$

By the Weak Law of Large Numbers there exists an integer i_2 such that for every $i \geq i_2$

$$P(A(i)|C_i, x_i \in \Gamma) \geq 1 - \epsilon.$$

With $I = \max\{i_0, i_1, i_2\}$, for all $i \geq I$, if $A(i)$ is satisfied and $||\nabla f(x_i)|| \geq r$ then

$$\begin{aligned}||\nabla f(x_i) - h_i)|| &= ||\nabla f(x_i) - G(x_i, \omega^i) + G(x_i, \omega^i) - h_i|| \\ &\leq ||\nabla f(x_i) - G(x_i, \omega^i)|| + ||G(x_i, \omega^i) - h_i|| \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta,\end{aligned}$$

and

$$\begin{aligned}||h_i|| &= ||\nabla f(x_i) - (\nabla f(x_i) - h_i)| \geq \big| ||\nabla f(x_i)|| - ||(\nabla f(x_i) - h_i)|| \big| \\ &\geq ||\nabla f(x_i)|| - ||(\nabla f(x_i) - h_i)|| \geq r - \delta.\end{aligned}$$

Then the following holds

$$\begin{aligned}f(x_{i+1}) - f(x_i) &= f(x_{i+1}) - G(x_{i+1}, \omega^i) \\ &\quad - \big(f(x_i) - G(x_i, \omega^i)\big) + G(x_{i+1}, \omega^i) - G(x_i, \omega^i) \\ &\leq \delta - \alpha\beta^{\overline{k}}||h_i||^2 + \epsilon_i \leq 2\delta - \alpha\beta^{\overline{k}}(r - \delta)^2 \leq -s.\end{aligned}$$

The above inequalities imply that under $E_i(\Gamma, r)$ and C_i, $A(i)$ implies $\overline{G_i(\Gamma, s)}$. Therefore, $G_i(\Gamma, s)$ implies $\overline{A(i)}$ consequently conditional probability of $G_i(\Gamma, s)$ is less or equal than conditional probability of $\overline{A(i)}$. As

$$P(\overline{A(i)}|C_i, x_i \in \Gamma) \leq \epsilon$$

we conclude that

$$P(G_i(\Gamma, s)|E_i(\Gamma, r), C_i) < \epsilon$$

i.e., Condition 1 is fulfilled.

To prove Condition 2 we consider again a compact set $\Gamma \subset I\!R$, $s > 0$ and $\epsilon > 0$. As $\sum_{i=0}^{\infty} \epsilon_i < \infty$, we can take an integer i_0 such that for every $i \geq i_0$ there holds

$$\epsilon_i \leq \frac{s}{3}.$$

As f is Lipschitz continuous on Γ and (5.24) holds, for $x_{i+1} = x_i - \beta^{k(i)} h_i$ there exist constants $L > 0$, and $M > 0$ such that

$$|f(x_{i+1}) - f(x_i)| \leq LM\beta^{k(i)}.$$

Thus, there exists an integer $\overline{k}$, such that if $k(i) \geq \overline{k}$, then

$$f(x_{i+1}) - f(x_i) \leq s. \tag{5.31}$$

Now, we consider the case $k(i) \leq \overline{k}$. Let $B(i)$ be the event

$$x_i \in \Gamma,\ k(i) \leq \overline{k},\ |f(x_i) - G(x_i, \omega^i)| < \frac{s}{3},\ |f(x_{i+1}) - G(x_{i+1}, \omega^i)| < \frac{s}{3}.$$

If the event $B(i)$ is realized, then

$$\begin{aligned} f(x_{i+1}) - f(x_i) &= f(x_{i+1}) - G(x_{i+1}, \omega^i) \\ &\quad - \big(f(x_i) - G(x_i, \omega^i)\big) + G(x_{i+1}, \omega^i) - G(x_i, \omega^i) \\ &\leq \frac{2s}{3} - \alpha\beta^{k(i)}||h_i||^2 + \epsilon_i \leq s. \end{aligned}$$

Again, by the Weak law of large number, there exists an integer i_1, such that for all $i \geq i_1$

$$P(B(i), k(i) \leq \overline{k} | C_i, x_i \in \Gamma) \geq 1 - \epsilon.$$

Taking $I = \max\{i_0, i_1\}$, we have that for all $i \geq I$ and $C_i \in \mathcal{F}_i$

$$P(\overline{B(i)}, k(i) \leq \overline{k} | C_i, x_i \in \Gamma) \leq \epsilon. \tag{5.32}$$

Now, (5.32), (5.31) and (5.18) imply that Condition 2 is fulfilled. As Conditions 1 - 2 are satisfied, the statement follows by Theorem 2.1 in [56]. □

Numerical Results

The numerical result that we present here confirm theoretical results and demonstrate efficiency of the proposed approach. We consider the following four test examples, defined as

$$g(x, \omega) = \phi(\omega x), \quad \omega : \mathcal{N}(1, \sigma^2),$$

where $\phi : I\!\!R^p \to I\!\!R$. The testing is done for two variance levels $\sigma^2 = 0.1$ and $\sigma^2 = 1$, using test functions ϕ taken from [2] and [40]:

AP Aluffi-Pentini's Problem, $p = 2$

$$g(x, \omega) = 0.25(\omega x_1)^4 - 0.5(\omega x_1)^2 + 0.1(\omega x_1) + 0.5(\omega x_2)^2.$$

EXP Exponential Problem $p = 10$

$$g(x,\omega) = \exp(-0.5\sum_{i=1}^{10}(\omega x_i)^2).$$

SAL Salomon Problem $p = 10$

$$g(x,\omega) = 1 - \cos(2\pi||\omega x||) + 0.1||\omega x||, \text{ where } ||\omega x|| = \sqrt{\sum_{i=1}^{10}(\omega x_i)^2}.$$

SPH Sphere function or first function of De Jongs $p = 10$

$$g(x,\omega) = \sum_{i=1}^{10}(\omega x_i)^2.$$

Theoretical results are obtained for the case $n \to \infty$. However, practical implementation is possible only with finite sample size. Let $n_{\max}$ denote the maximal sample size allowed and we fixed $n_{\max} = 100$ for the first two problems, $n_{\max} = 1300$ for the third problem and $n_{\max} = 200$ for the last problem. The choice of $n_{\max}$ is highly non-trivial but we will not discuss it here as our aim is only to illustrate the potential advantages of nonmonotone line search rule.

The algorithm is implemented and tested against classical Armijo monotone line search rule ($\epsilon_i = 0$ in Algorithm 2) for two search directions, the first one being the negative gradient while the second direction is the finite difference approximation of the negative gradient $\nabla_\xi G(x_i, \omega^i)$, defined in [52]. The jth component is defined as

$$\frac{G(x_i + \xi e_j, \omega^i) - G(x_i - \xi e_j, \omega^i)}{2\xi},$$

where e_j denotes the jth coordinate vector in $\mathbb{R}^p$ and $\xi = 10^{-4}$. The sequence $\{\epsilon_i\}$ is defined as $\epsilon_i = 2^{-i}, i = 1, 2, \ldots$. Therefore, we have implemented four different methods.

- NM1 Nonmonotone line search with the negative gradient search direction, $h_i = \nabla G(x_i, \omega^i)$

- NM2 Nonmonotone line search with the finite difference approximation of the negative gradient. $h_i = \nabla_\xi G(x_i, \omega^i)$
- M1 Monotone (Armijo) line search with the the negative gradient search direction, $h_i = \nabla G(x_i, \omega^i)$
- M2 Monotone (Armijo) line search with the finite difference approximation of the negative gradient. $h_i = \nabla_\xi G(x_i, \omega^i)$

The sample size in each iteration is defined as

$$n(i+1) = \min\{\lceil 1.1n(i) \rceil, n_{max}\},$$

with the initial value $n(0) = 3$ and a new sample of the size $n(i)$ is generated in ith iteration. The algorithmic parameters are the same for all problems, the starting point is $x_0 = 10 \cdot [1, 1, \ldots, 1]^T$, $\alpha = 10^{-4}$ and backtracking is performed with $\beta = 0.5$. Maximal allowed number of of backtracking steps is 5. The stopping criteria is satisfied in x_i if the norm of the gradient or its approximation is smaller than 10^{-2} and $n(i) = n_{max}$. The number of function evaluations is used as the algorithm performance measure. Thus, for NM1 and M1, each gradient calculation is counted as p function evaluation, while for NM2 and M2 we used the two-sided approximation of gradient, so each gradient calculation is counted as $2p$ function evaluation. The method is stopped if the maximal allowed number of function evaluation is exhausted, with the maximal number set to 10^7.

In the testing process, we generated 5 independent samples for each variance levels and all problems are tested using the same collection of samples.

The results are presented at Figure 5.1, using the performance profile graph [12], where the cost function $t_{p,m}$ is defined as the number of function evaluations for solving problem p with method m, and performance profile for set of test problems S_p and set of methods S_m is

$$\rho(\tau) = \frac{1}{n_p} |\{p \in S_p | r_{p,m} \leq \tau\}|,$$

where n_p is number of test problems and

$$r_{p,m} = \frac{t_{p,m}}{\min\{t_{p,m} | m \in S_m\}}$$

is performance ratio. The graph demonstrate that the nonmonotone line search outperforms the classical Armijo line search at the considered test collection for both search directions. As expected, negative gradient performs better than the finite difference approximation of the negative gradient but nevertheless works reasonable well, which is an important property for problems where the function is calculated using a black box and the exact gradient of g is not available.

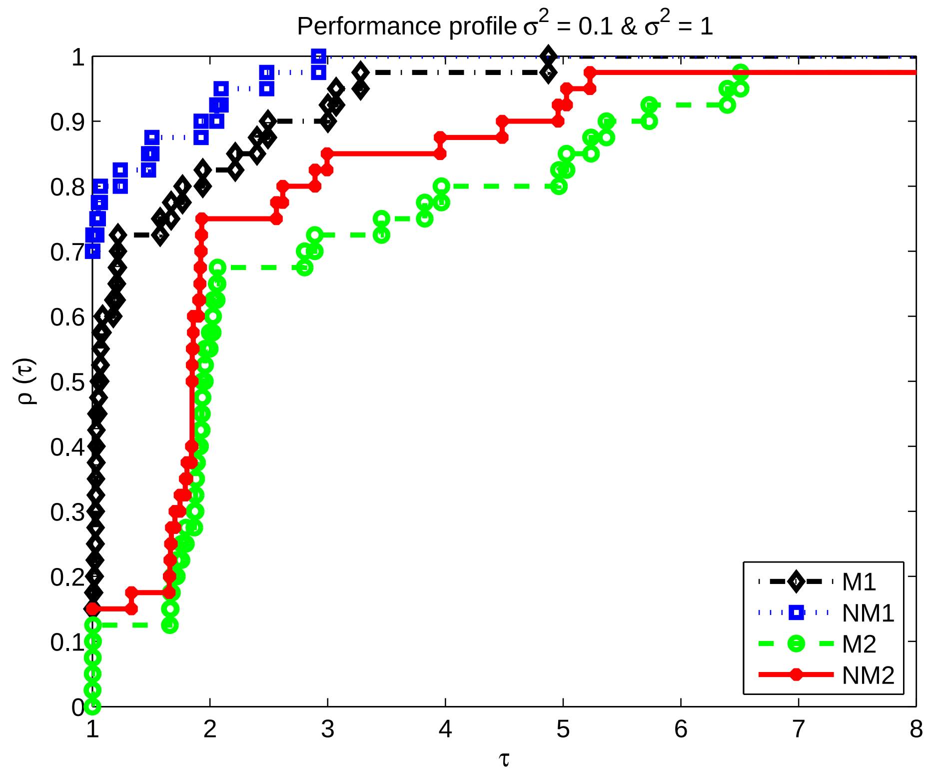

Figure 5.1: Performance profile for methods M1, NM1, M2, NM2 and two variance levels 0.1 and 1.

Chapter 6

Empirical results

In this chapter, we demonstrate properties of Negative Selection and test the execution model using real trade data four stocks from London Stock Exchange and Euronext, specifically Vodafone Group (VOD.L), AstraZeneca (AZN.L), Barclays PLC (BARC.L), and Sanofi SA (SASY.PA), which will be denoted by VOD, AZN, BARC and SASY.

Negative Selection satisfies important properties of any performance measure. It can distinguish between filled and partially filled orders clearly. It makes a distinction between orders filled at the different price level as well as orders of various sizes at the same price level. Furthermore, it possesses a continuity, in the sense that small shifts in the order size or price should yield negligible changes in the performance measure. Moreover, probably the most important feature is that it is capable of reflecting the toughness of market condition at a particular time window and thus providing means for objective judgment of the quality of execution. Thus, the additional potential of NS lies in its possibility of application in the process of testing a trading strategy on relevant historical data i.e. backtesting. To demonstrate this feature, we consider a simple example, in which we compare the behavior of NS, VWAP and IS benchmarks in both falling and rising markets. We place an order at bid1 until filled or the time of 10 minutes expires. If the order is not completely filled within 10 minutes, the residual is filled by crossing the spread at the end of given time window. We tested a sequence of orders with increasing sizes, from 0 to 35% of average traded quantity in the selected time window. The 10 minutes windows are chosen randomly, and the relevant trajectories for AZN are shown at Figure 6.1 and Figure 6.2. The price trajectories are shown at the left-hand side while the right-hand side shows the slippages with respect to all three benchmarks at both Figures. The horizontal axis shows the traded amount in thousands. The average traded quantity for AZN is 50000 shares in 10 minutes, so the simulations are performed for orders of size 1 to 17500 shares with the step size of 500 shares.

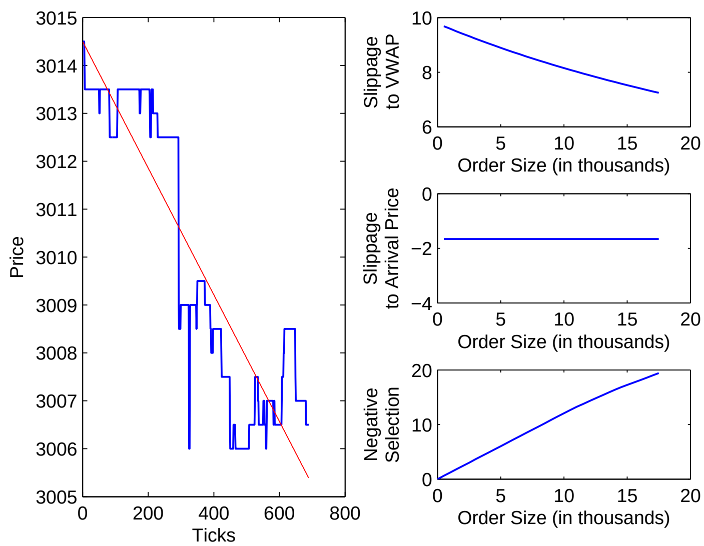

Figure 6.1: VWAP, Arrival Price and Negative Selection for falling market

Figure 6.1 shows the case of falling market. The slippage to VWAP is positive and decreasing with the increase of order quantity. Being positive gives the right information of our execution, but the reduction of slippage provides false information. The decreasing slippage implies that the execution strategy is good, although it is quite clear that in the failing market one should have placed orders at lower price levels. This decrease in the slippage is a consequence of the already mentioned VWAP flaw - the slippage is declining due to the impact of large traded quantity. With IS, the situation is different: the slippage is negative and constant. Its value is the difference of Arrival Price and the bid1 price at the beginning of the time window. Here, the negative sign of slippage gives false information on the execution performance as a consequence of insensitivity of Arrival Price to the market conditions in the trading time window.

The rising market is shown in Figure 6.2. In this case, an order placed

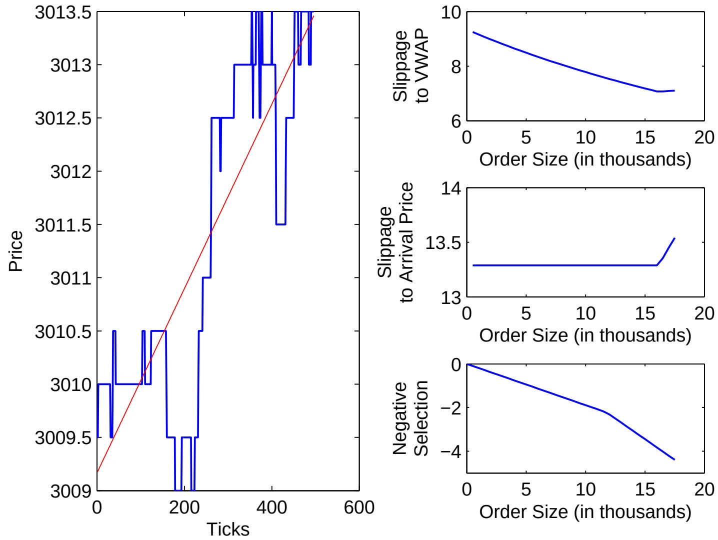

Figure 6.2: VWAP, Arrival Price and Negative Selection for rising market

at bid1 can be thought of as being passive. If the order is not filled, it will result in crossing the spread at the end of the time window and paying a higher price. The slippage to VWAP is positive because the order is filled at a price higher than the benchmark. However, again, we see the decrease in the slippage with the increase of the order size, giving the false impression that the execution strategy is improving with the order size. The slippage to Arrival Price is high and positive. It is constant while there is enough liquidity at $t = T$, but when the order size increases enough - above the quantity available at ask1, the order starts to "walk the book, " and the slippage to Arrival Price starts to rise. Whereas, NS is negative and increasing with the order size. Thus the information we get is correct - the execution strategy should have been more aggressive.

6.1 Data and Simulations

For the propose of demonstrating NS properties, and also testing the execution model, we developed a simulator in MATLAB and MySQL, while optimization is done using MATLAB function *fmincon*. The simulator is built in such way that mimics real-time trading functionality with additional support for statistical analysis of all inputs and outputs of trading simulations with the possibility to keep track of open position i.e. all positions are closed with the opposite operation (buy/sell). Its functionality is based on level-2 tick data provided by Reuters, which gives a view at order book and its dynamics during trading hours. The data consist of 5 levels of depth, with information about:

- date and time of event taking place
- bid and ask prices for all five levels
- volume at each bid and ask level
- number of orders at bid and ask levels
- trade data: indicator, price, size and time of a trade.

We consider real trade tick data for VOD, AZN, BARC and SASY from the beginning of January to the end of August 2006. For every day during this period, the simulator used trading data from 8:15 to 16:25. With statistical properties of the whole sample of data, shown in Table 6.1 we can classify VOD as a super-liquid, and AZN, BARC, and SASY as liquid stocks.

Stock	Average Spread [bps]	Average Daily Volume	Standard deviation of Price	Average Daily Trades
VOD	21.65	189,321,705.10	6.01	3373.19
AZN	6.42	3,951,778.23	226.39	3278.09
BARC	10.36	14,843,907.84	30.14	3086.18
SASY	8.60	1,923,534.09	2.51	2482.29

Table 6.1: Properties of stock data. Svojstva podataka o finansijskim instrumentima

To be able to simulate trading, we consider a sequence of orders generated by Black Box (BB) trading strategy with inventory. It is a momentum strategy generating signals using a mathematical model with constant parameters selected as follows:

- execution time window width $T = 10$ minutes,
- cancel threshold of $45bps$,
- approximate average volume traded in 10-minutes window.

The tested order sizes vary from 1% and 5% of average traded volume in the time windowwhich is approximately $40,000$ to $200,000$ shares for Vodafone, 500 to $2,500$ shares for AstraZeneca, $3,000$ to $15,000$ for Barclays PLC, 420 to $2,100$ shares for Sanofi SA, respectively.

The cancel threshold coupled with time window width create a criterion for canceling the order. An active order is canceled if either time expires, i.e. during that time there was not enough liquidity to be completely filled or the cancellation threshold is reached, i.e. price moves in an adverse direction. Therefore, the order is either (completely) filled or (partially) unfilled. In the former case, we define the *fill time* (T_{fill}) as the time elapsed from submission until filling the order, in the latter, the *cancel time* (T_{cancel}) as the time elapsed from submission until canceling the order. Clearly, both T_{cancel} and T_{fill} are bounded above by execution time window width.

In addition to the Black Box trading strategy, we also consider the so-called Default strategy which is formulated as the alternation of buy and sell signal every 10 minute. When producing signals, it does not take into account the actual market conditions. However, like BB, it obeys rules regarding the possible cancellation of an order. As the quality of signals, regarding profitability, is quite random, the purpose of Default strategy is to give us the baseline for market conditions during the observed period. The Default Strategy, in fact, reflects the toughness of the market as it landscapes the data.

The properties of Default and Black Box are presented in Table 6.2. The dollar sign represents monetary units, i.e., British Pound for Vodafone, AstraZeneca and Barclays PLC, and Euro for Sanofi SA and bid/ask1 denotes that buy order was placed at bid1, and sell order at the ask1 price level. Analogously, we introduce notation bid/ask2 and bid/ask3. Clearly, the BB strategy has short-term alpha and can generate profit. The Default strategy is evidently losing money.

Ticker:		VOD			AZN				BARC				SASY			
		D	BB1	BB2	D	BB1	BB2	BB3	D	BB1	BB2	BB3	D	BB1	BB2	BB3
100% at bid/ ask1	Total number of Triggers	8329	637	704	8219	1424	1043	1081	8337	1019	954	735	7505	860	601	438
	Profitable days [%]	5.36	64.49	60.16	3.61	57.76	58.97	59.35	2.38	57.67	59.87	61.22	8.19	59.01	57.05	51.24
	Profit [in 000]	-68230	12942	10035	-18628	1552	1045	2483	-29055	1996	2289	2096	-384	53	40	33
	Slippage to VWAP [bps]	1.84	-7.36	-3.18	1.84	-1.25	-1.26	-2.33	2.27	-0.75	-0.51	-2.18	2.14	-1.79	-1.93	-3.58
	Win[$]/Los[$]	-0.35	1.55	1.58	-0.39	1.19	1.14	1.31	-0.35	1.23	1.3	1.26	-0.46	1.38	1.35	1.31
	Profitable trades [%]	25.73	54.1	59.84	35.69	56.11	53.07	52.84	32.63	52.97	53.41	52.48	34.92	55.13	54.68	53.59
100% at bid/ ask2	Total number of Triggers	8329	637	704	8219	1424	1043	1081	8337	1019	954	735	7505	860	601	438
	Profitable days [%]	31.21	51.11	56.76	12.65	60.51	54.86	59.15	16.67	64.44	59.23	63.39	19.88	68.28	66.4	59.13
	Profit [in 000]	-15183	536	2131	-11856	1988	296	2589	-12404	1903	2531	1828	-191	79	73	70
	Slippage to VWAP [bps]	0.78	-0.48	0.27	1.25	-1.36	-0.91	-2.8	1.12	-0.92	-1.09	-1.49	1.3	-2.92	-3.34	-7.23
	Win[$]/Los[$]	-0.39	1.08	1.65	-0.43	1.34	1.05	1.45	-0.46	1.48	1.75	1.42	-0.55	2.05	2.14	1.95
	Profitable trades [%]	33.89	46.15	58.93	40.52	56.81	53.69	56.72	38.67	59.2	60.28	56.77	42.91	64.95	64.76	60.22

Table 6.2: Properties of Default (D) and Black Box (BB) trading strategy. Svojstava Default i Black Box strategije trgovanja

6.2 Review of Negative Selection Properties

Tables 6.3 - 6.6 contain the simulation results for all four stocks and all BB strategies using the whole data set. We tested two order sizes 1% and 5% of average traded quantity in 10 minutes intervals, across all five bid/ask positions. Here, bid/ask1 denotes placing an order at the price level bid1 for a buy signal, and at ask1 for a sell signal. Price levels bid/ask2, bid/ask3, ... bid/ask5 are analogously defined. The mean values across the whole data set, for each stock are, are given in Tables 6.3 - 6.6. It is evident that theoretical properties stated in Lemma 4.3.4 and Lemma 4.3.5 are empirically confirmed. For fixed order size, comparison of NS across different price levels shows that the more passive the placement of the order is, the more negative is the value of NS. When comparing NS by size, for two orders placed at the same price level, we see that when the larger order is filled, then the NS of smaller order is lower than NS of the larger order, which captures market impact caused by larger order. Furthermore, when the larger order is unfilled, then its value of NS is always more negative than the value of NS for the smaller order. When there is not enough liquidity for the larger order to be completely filled, but for the smaller is enough to be executed, NS of the smaller order is positive and negative for the larger order, this is illustrated in Table 6.5 for BB3 trading strategy. Tables 6.3 - 6.6 depict additional feature of NS: the mean values of NS vary considerably between all considered stocks, which also captures the behavior of the particular stock in prevailing market conditions. For example, VOD is the most liquid stock with the widest spread (21.65bps) in this data sample - trading takes place at bid1 and ask1 for an extended period followed by a shift in price to the next price level or a few price levels above / below and repeats the bid/ask bouncing. Unlike VOD, AZN is the least liquid stock with the smallest spread (6.42bps). AZN price tends to trade in a narrow price channel e.g. going from bid1 to bid3 and then bounce back and repeated the process, with a different price trajectory. This behavior justifies the fact that VOD has the best performance, regarding profits and NS, at bid/ask1 and for AZN we see the same at level bid/ask2. BARC exhibits similar behavior as VOD, while SASY is more like AZN.

More detailed results, which also include mean values of T_{fill} and T_{cancel} for the first three bid/ask levels are given in Table 6.7 and Table 6.8 for 5% and 1% of average traded quantity in 10 minutes intervals, respectively.

Figures 6.3 - 6.6 illustrate in more detail Lemma 4.3.4 for all four stocks and all BB trading strategies. NS is calculated without an actual trading mechanism, but for each signal is considered order with fixed size. Figures depict orders of 5% of the average traded quantity placed at bid/ask1, bid/ask2,..., bid/ask5 for some random sample of triggers. Apparently, the measure shows the sensitivity to changes in price levels. Furthermore, passive behavior is always accompanied with more negative values of NS if the price went in an adverse direction for considered window. On the other hand, aggressive trading, when the price went in our direction, is "punished" with high positive values of NS.

Figures 6.7 - 6.10 portray relationship between order sizes fixed price levels, described in Lemma 4.3.5. Again, we used all four stocks and all BB trading strategies. We tested two order sizes 1% and 5% of average traded quantity in 10 minutes intervals at bid/ask1 position for some random sample of triggers. In cases when the larger order is filled, then NS of larger order dominates over NS value of smaller, in this way incorporating the impact caused by our trading. We interpret this information as being "punished" for aggressive trading a large order when passive behavior would get a more favorable price, i.e. overall costs are greater for suboptimal large order than of a suboptimal smaller order. When the larger order is unfilled, NS value of smaller dominates NS value of larger. In this situation, there are two scenarios for the smaller order. The first scenario, when there is enough liquidity for smaller order to be filled, which is characterized by nonnegative NS value of smaller and negative of the larger order. The second, when the smaller order is also (partially) unfilled, in this situation both orders have negative NS, but for the larger order, it is more negative. Clearly, for both order sizes the strategy was too passive, but for the larger order, there is an additional cost for passiveness, as filling the larger order, when the price is going away from us, requires more aggressive behavior.

Figures 6.11 - 6.14 represent the relative distribution of Negative Selection for all four stocks and all BB trading strategies. The scatter plot of Cancel and fill time and NS for the same sequence of orders is depicted on Figures 6.15 - 6.18. Cancel time corresponds to points with negative NS, while fill time is on the nonnegative side of the axis, which is a consequence of Lemma 4.3.1- Lemma 4.3.3, because filled orders have nonnegative NS, while unfilled orders have negative NS. The grouping of cancel time data at the 10th minute is caused by the limitation of execution time window width to $T = 10$ minutes.

Ticker:		VOD		
		Q	BB1	BB2
Mean Negative Selection for order with 100% at	bid/ask1	1%	5.73	-28.56
		5%	17.00	-152.12
	bid/ask2	1%	-79.98	-113.86
		5%	-409.92	-579.29
	bid/ask3	1%	-164.15	-197.90
		5%	-831.74	-998.74
	bid/ask4	1%	-248.26	-282.01
		5%	-1251.97	-1420.06
	bid/ask5	1%	-332.61	-366.11
		5%	-1673.73	-1840.51

Table 6.3: VOD: Comparison of Negative Selection by order size. VOD: Poređenje negativne selekcije po veličini naloga.

Ticker:		AZN			
		Q	BB1	BB2	BB3
Mean Negative Selection for order with 100% at	bid/ask1	1%	0.23	0.26	0.36
		5%	0.96	1.12	1.53
	bid/ask2	1%	0.02	0.05	0.13
		5%	-0.10	0.06	0.40
	bid/ask3	1%	-0.21	-0.19	-0.13
		5%	-1.25	-1.15	-0.88
	bid/ask4	1%	-0.48	-0.46	-0.43
		5%	-2.57	-2.49	-2.36
	bid/ask5	1%	-0.81	-0.80	-0.78
		5%	-4.23	-4.16	-4.13

Table 6.4: AZN: Comparison of Negative Selection by order size. AZN: Poređenje negativne selekcije po veličini naloga.

Ticker:			BARC		
		Q	BB1	BB2	BB3
Mean Negative Selection for order with 100% at	bid/ask1	1%	0.52	0.38	1.30
		5%	1.02	-0.16	4.62
	bid/ask2	1%	-1.90	-2.15	-1.15
		5%	-11.12	-12.75	-7.74
	bid/ask3	1%	-4.49	-4.73	-3.77
		5%	-24.11	-25.67	-20.72
	bid/ask4	1%	-7.20	-7.37	-6.51
		5%	-37.61	-38.84	-34.51
	bid/ask5	1%	-10.02	-10.16	-9.31
		5%	-51.76	-52.79	-48.54

Table 6.5: BARC: Comparison of Negative Selection by order size. BARC: Poređenje negativne selekcije po veličini naloga.

Ticker:			SASY		
		Q	BB1	BB2	BB3
Mean Negative Selection for order with 100% at	bid/ask1	1%	0.25	0.30	0.55
		5%	1.07	1.34	2.47
	bid/ask2	1%	-0.04	0.02	0.26
		5%	-0.38	-0.10	1.03
	bid/ask3	1%	-0.34	-0.28	-0.05
		5%	-1.85	-1.55	-0.51
	bid/ask4	1%	-0.64	-0.58	-0.36
		5%	-3.35	-3.05	-2.01
	bid/ask5	1%	-0.94	-0.89	-0.69
		5%	-4.87	-4.61	-3.66

Table 6.6: SASY: Comparison of Negative Selection by order size. SASY: Poređenje negativne selekcije po veličini naloga.

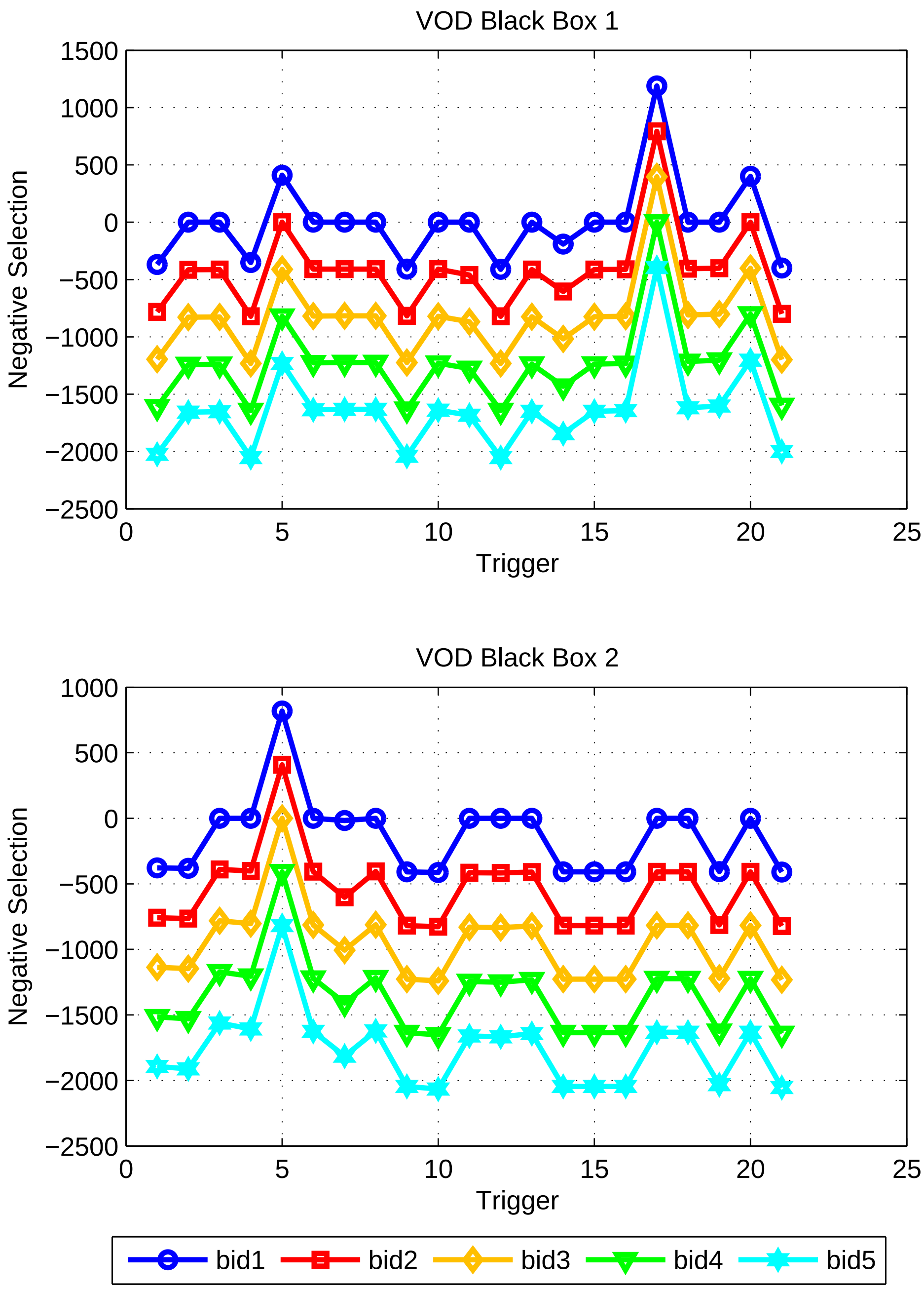

Figure 6.3: VOD: Comparison of Negative Selection by price level. VOD: Poređenje negativne selekcije po nivoima cena.

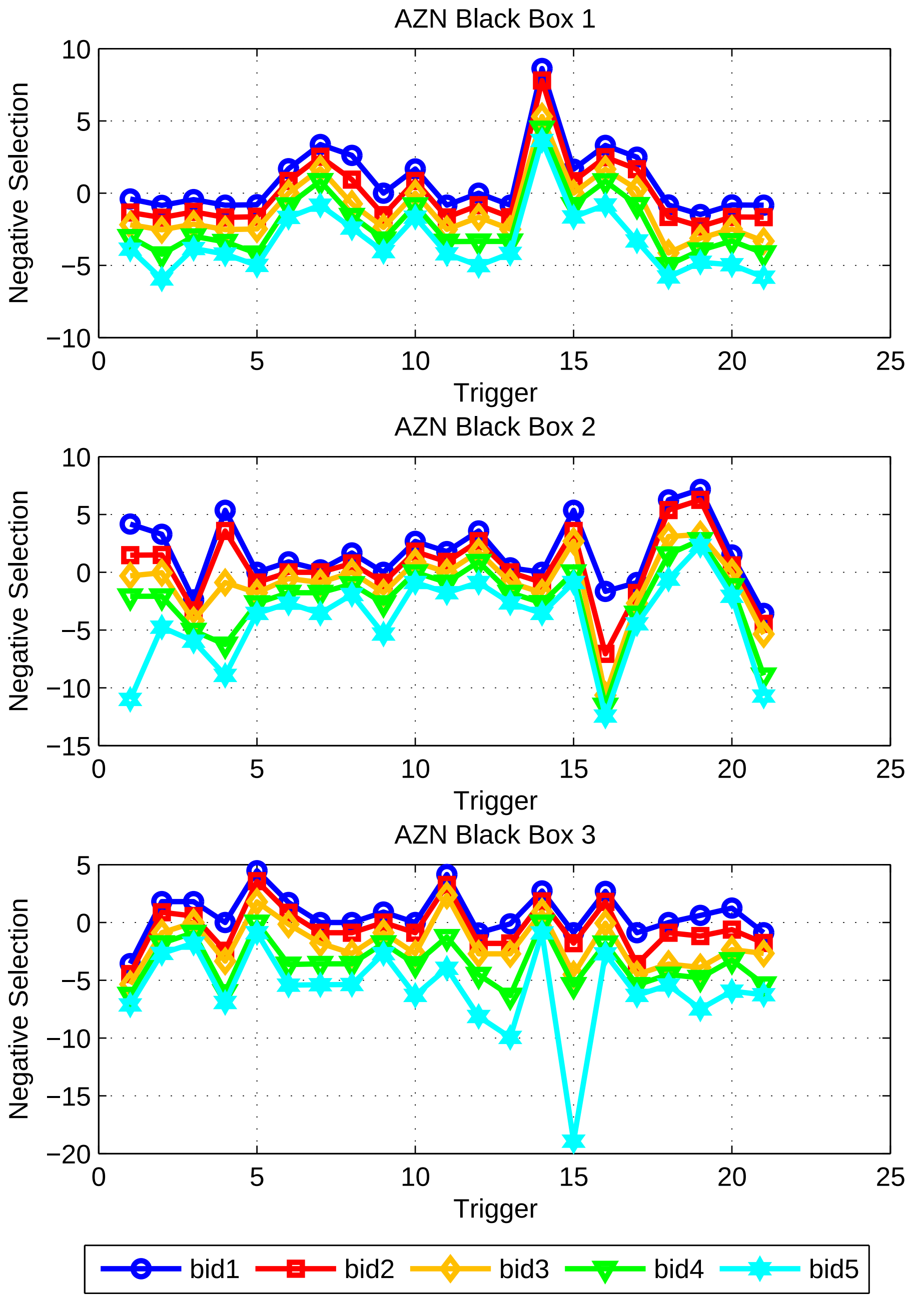

Figure 6.4: AZN: Comparison of Negative Selection by price level. AZN: Poređenje negativne selekcije po nivoima cena.

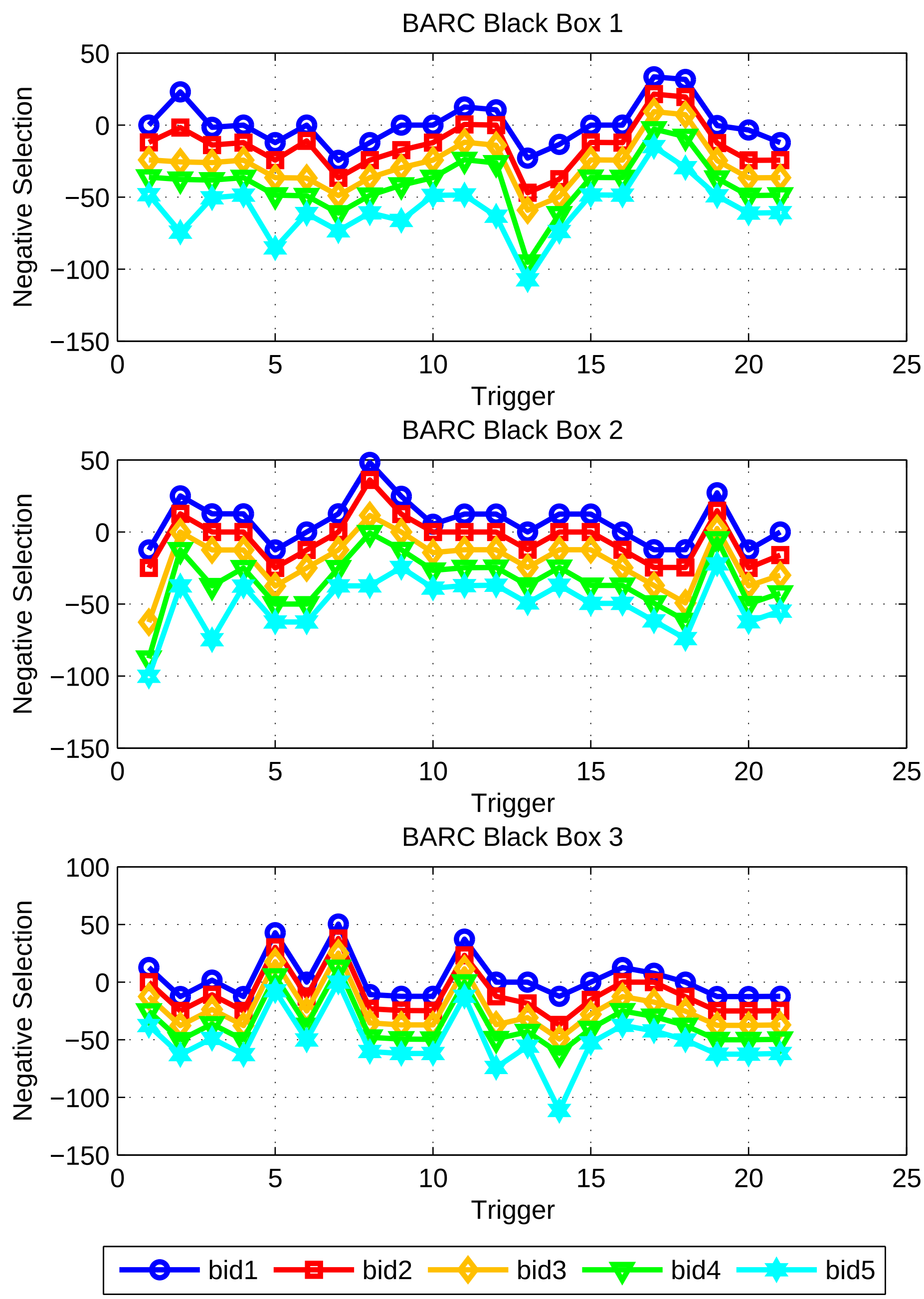

Figure 6.5: BARC: Comparison of Negative Selection by price level. BARC: Poređenje negativne selekcije po nivoima cena.

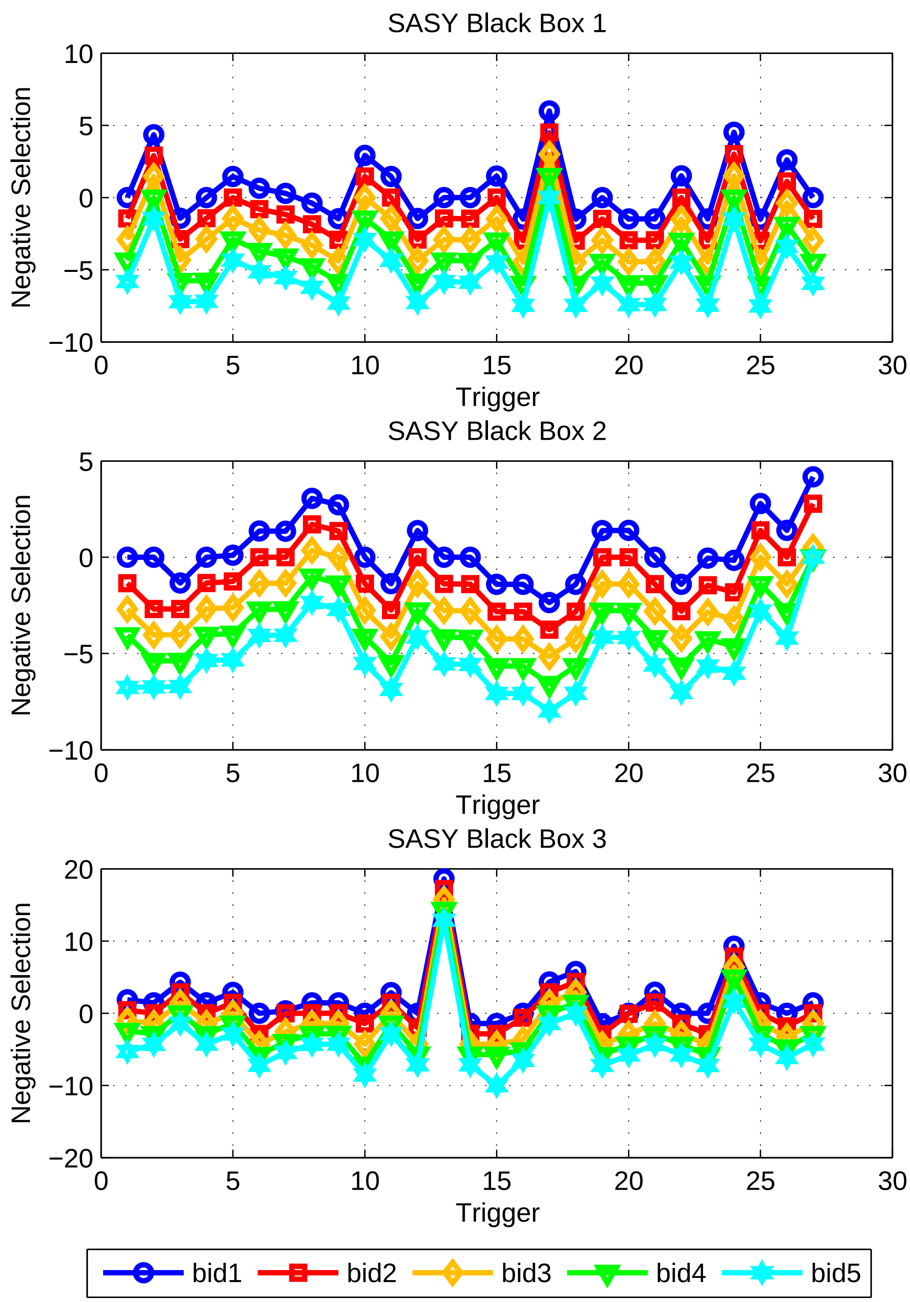

Figure 6.6: SASY: Comparison of Negative Selection by price level. SASY: Poređenje negativne selekcije po nivoima cena.

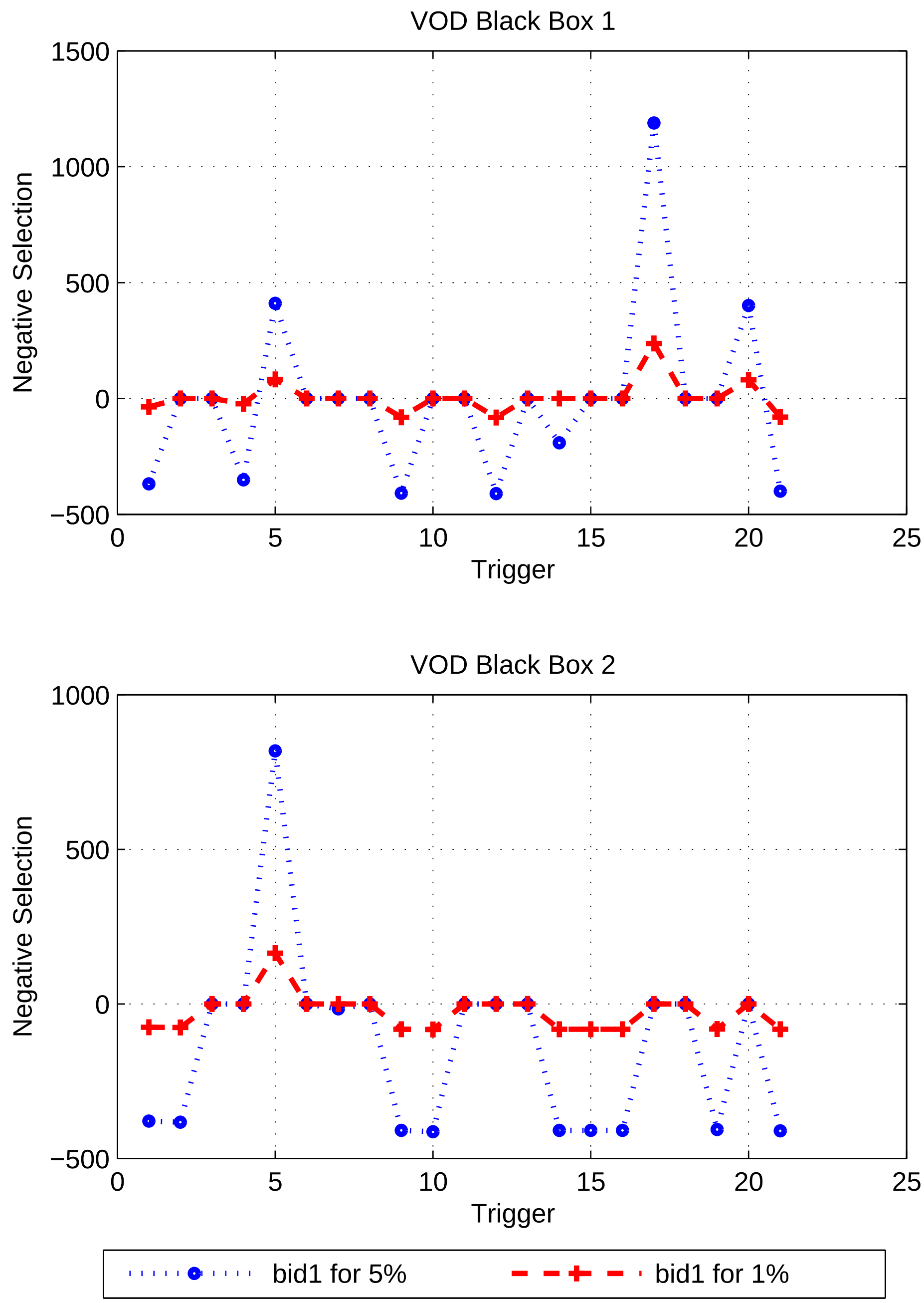

Figure 6.7: VOD: Comparison of Negative Selection by size of an order for bid/ask1. VOD: Poređenje negativne selekcije po nivoima cena.

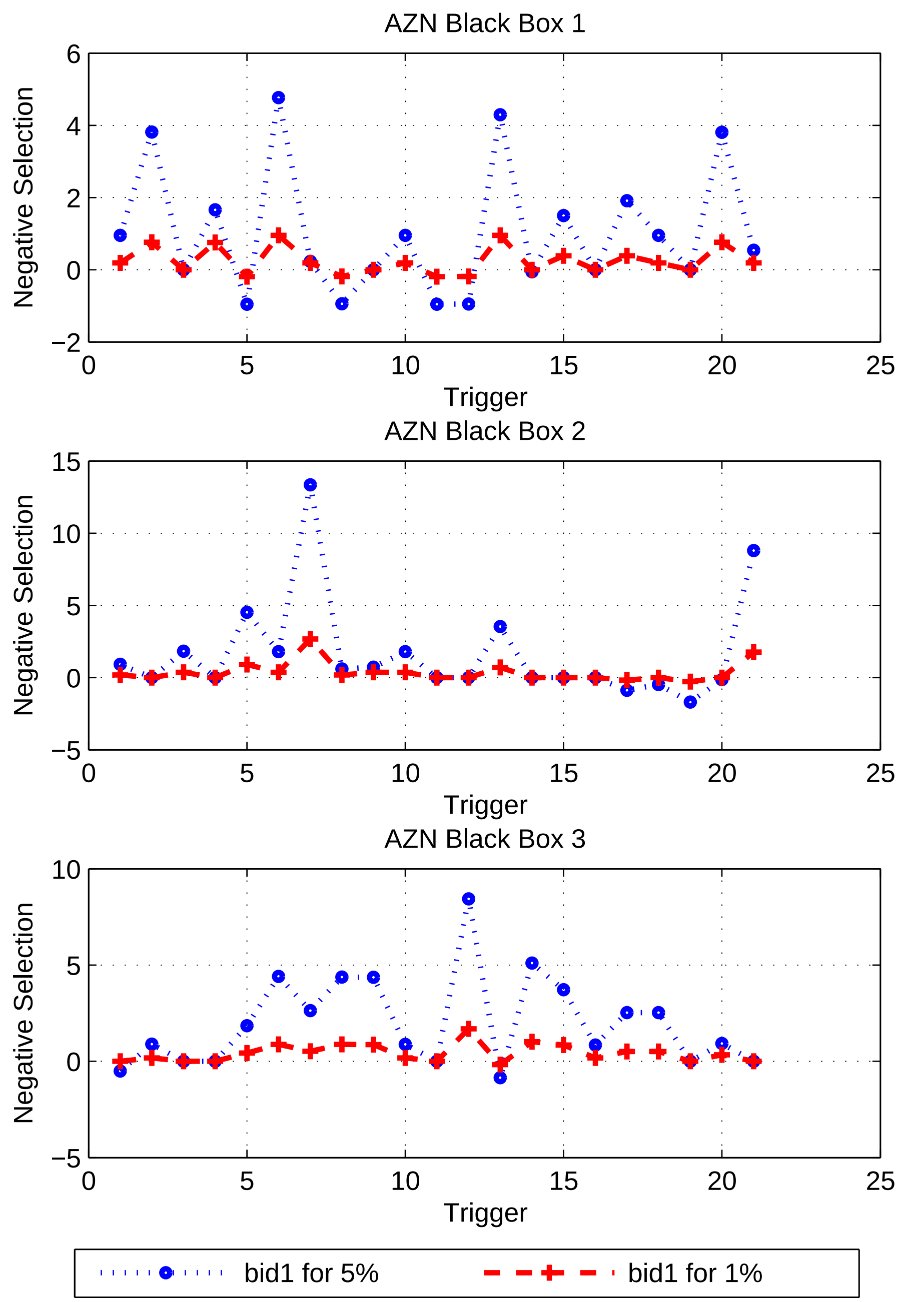

Figure 6.8: AZN: Comparison of Negative Selection by size of an order for bid/ask1. AZN: Poređenje negativne selekcije po nivoima cena.

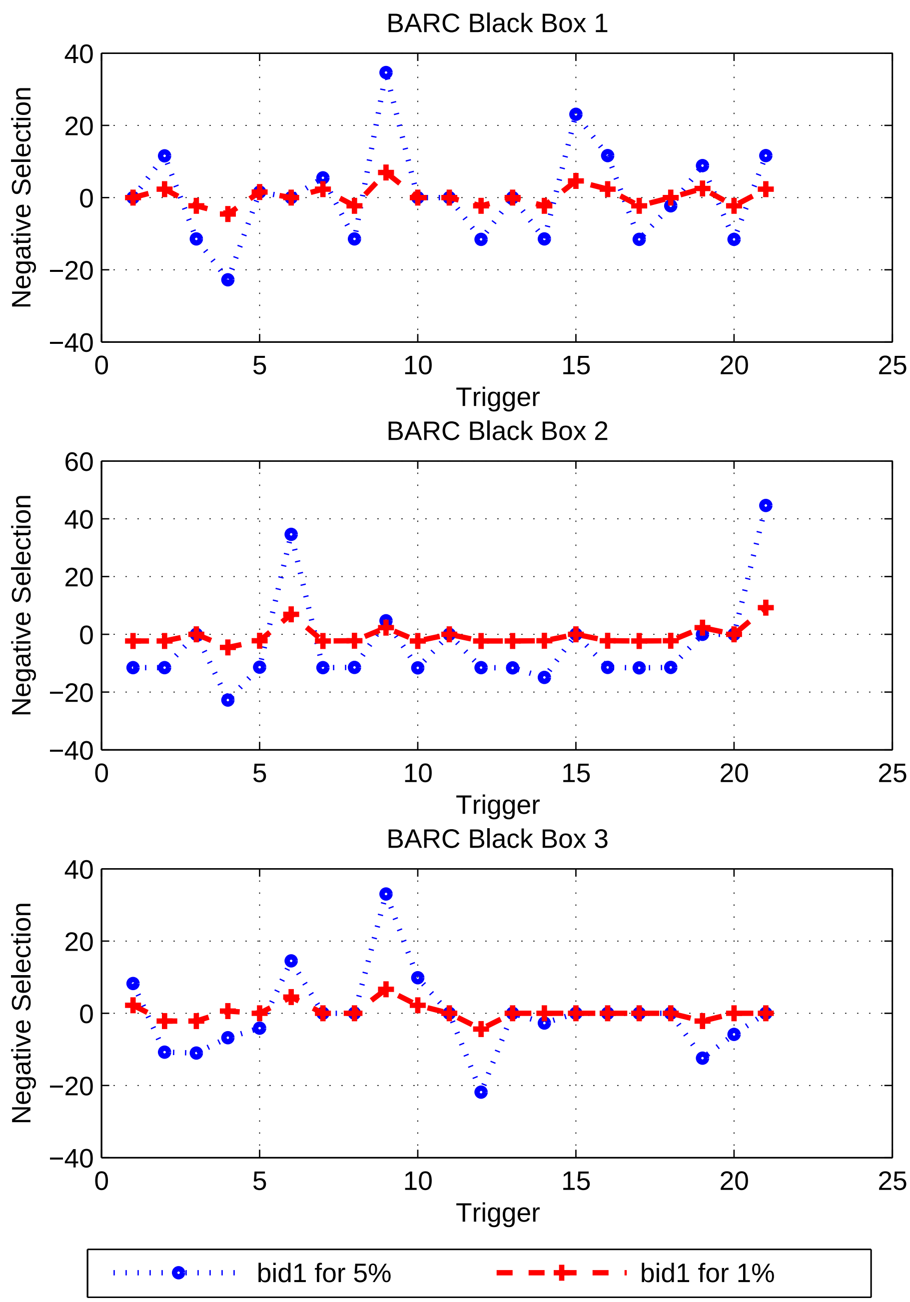

Figure 6.9: BARC: Comparison of Negative Selection by size of an order for bid/ask1. BARC: Poređenje negativne selekcije po nivoima cena.

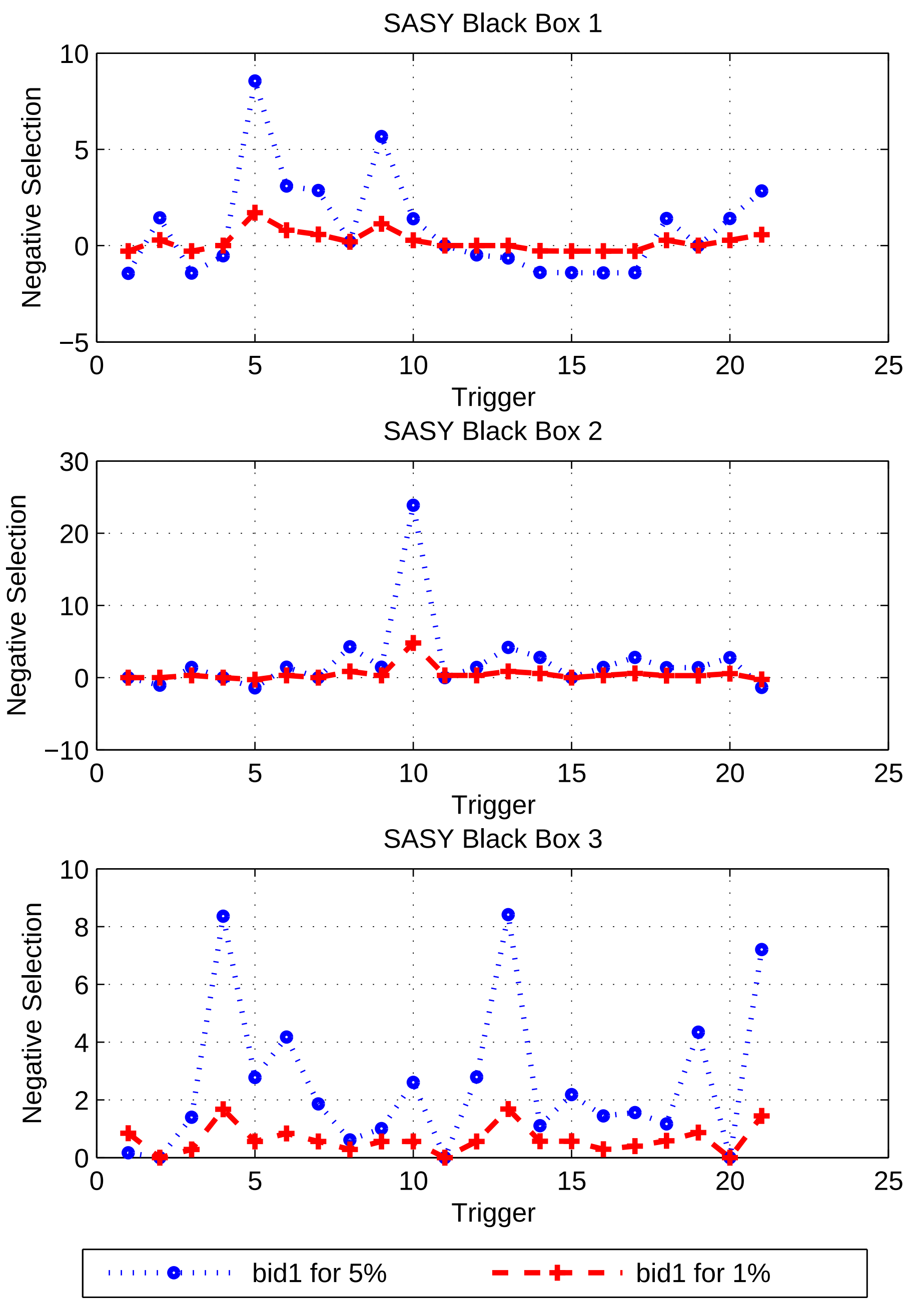

Figure 6.10: SASY: Comparison of Negative Selection by size of an order for bid/ask1. SASY: Poređenje negativne selekcije po nivoima cena.

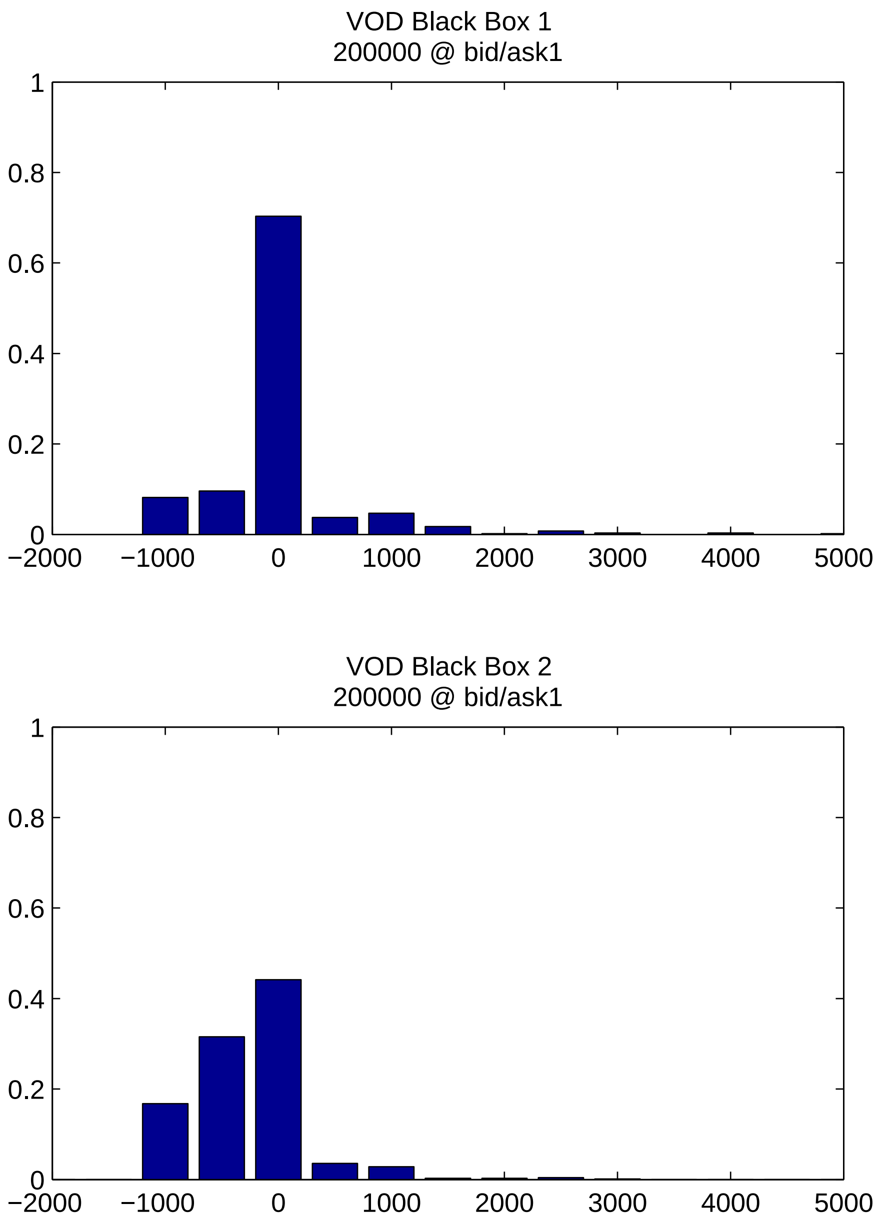

Figure 6.11: VOD: Relative frequency histogram of Negative Selection. VOD: Histogram relativne frekvencije Negativen Selekcije.

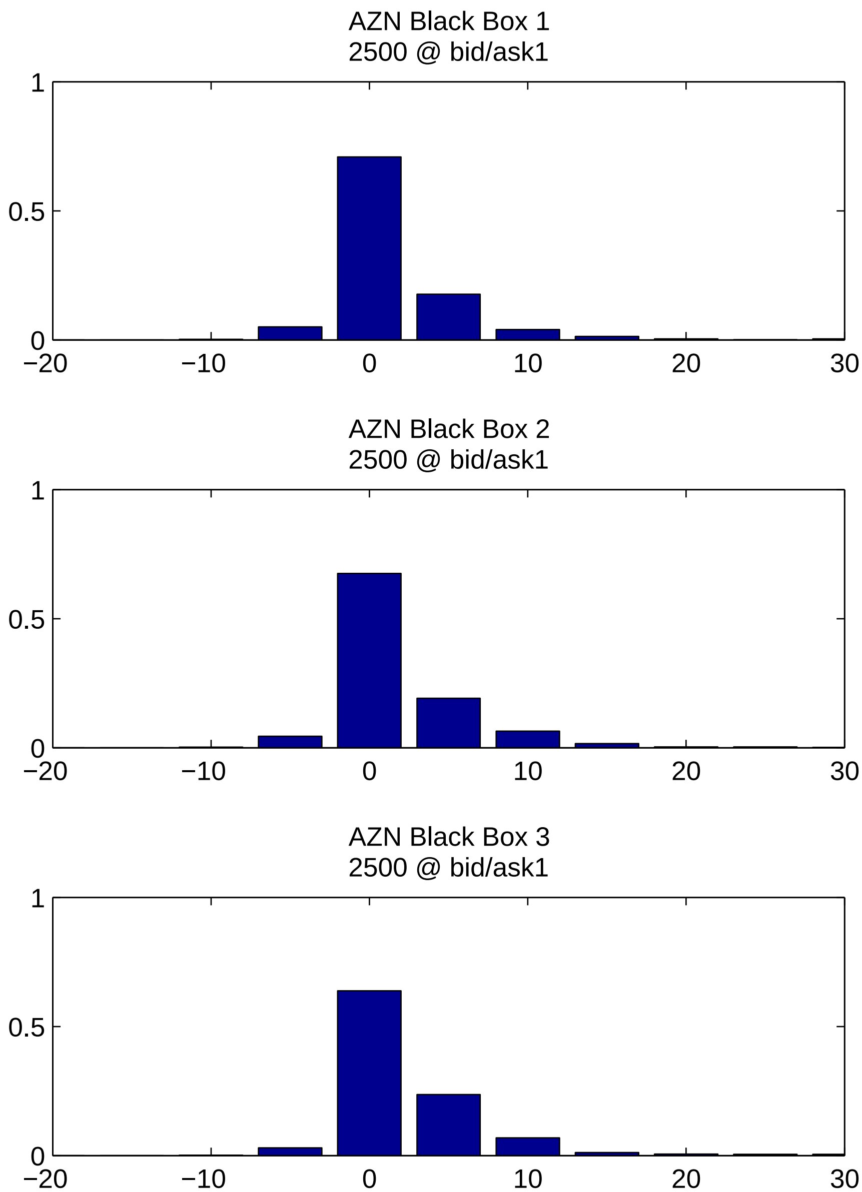

Figure 6.12: AZN: Relative frequency histogram of Negative Selection. AZN: Histogram relativne frekvencije Negativen Selekcije.

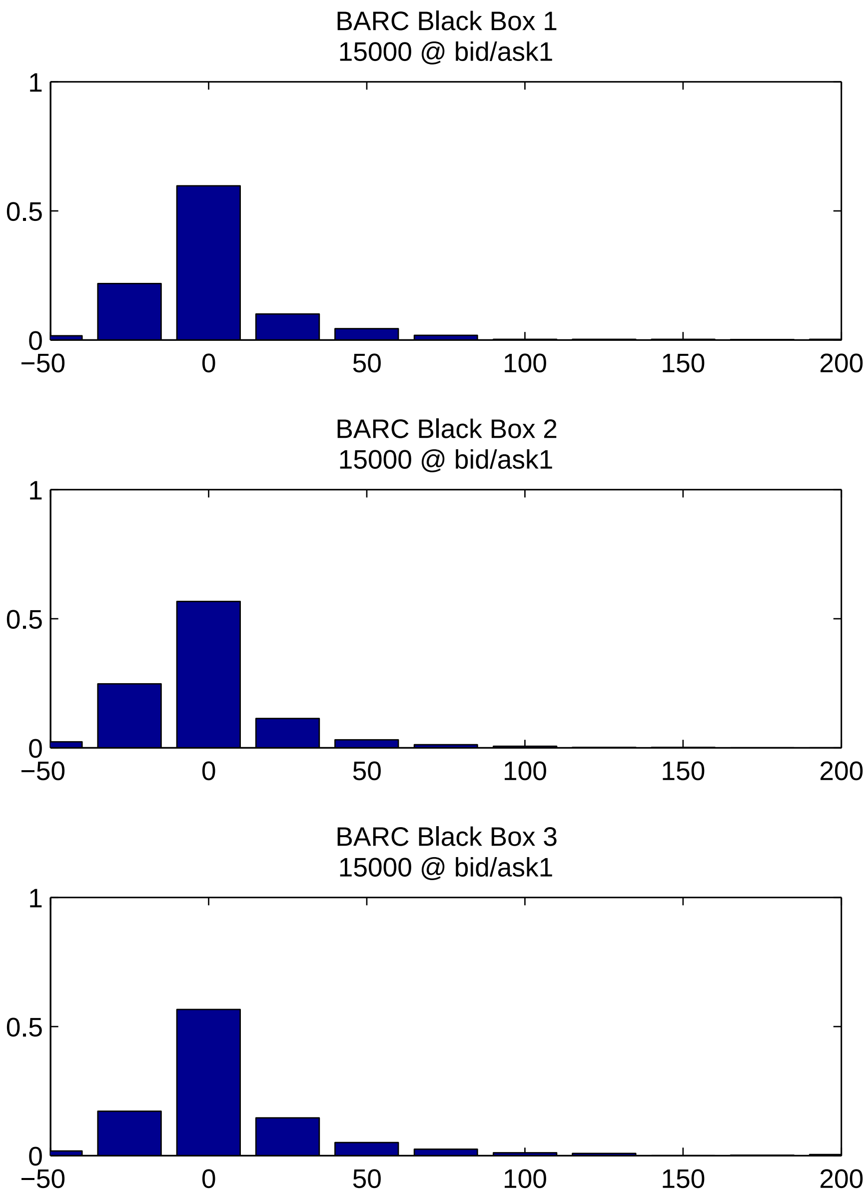

Figure 6.13: BARC: Realtive frequency histogram of Negative Selection. BARC: Histogram relativne frekvencije Negativen Selekcije.

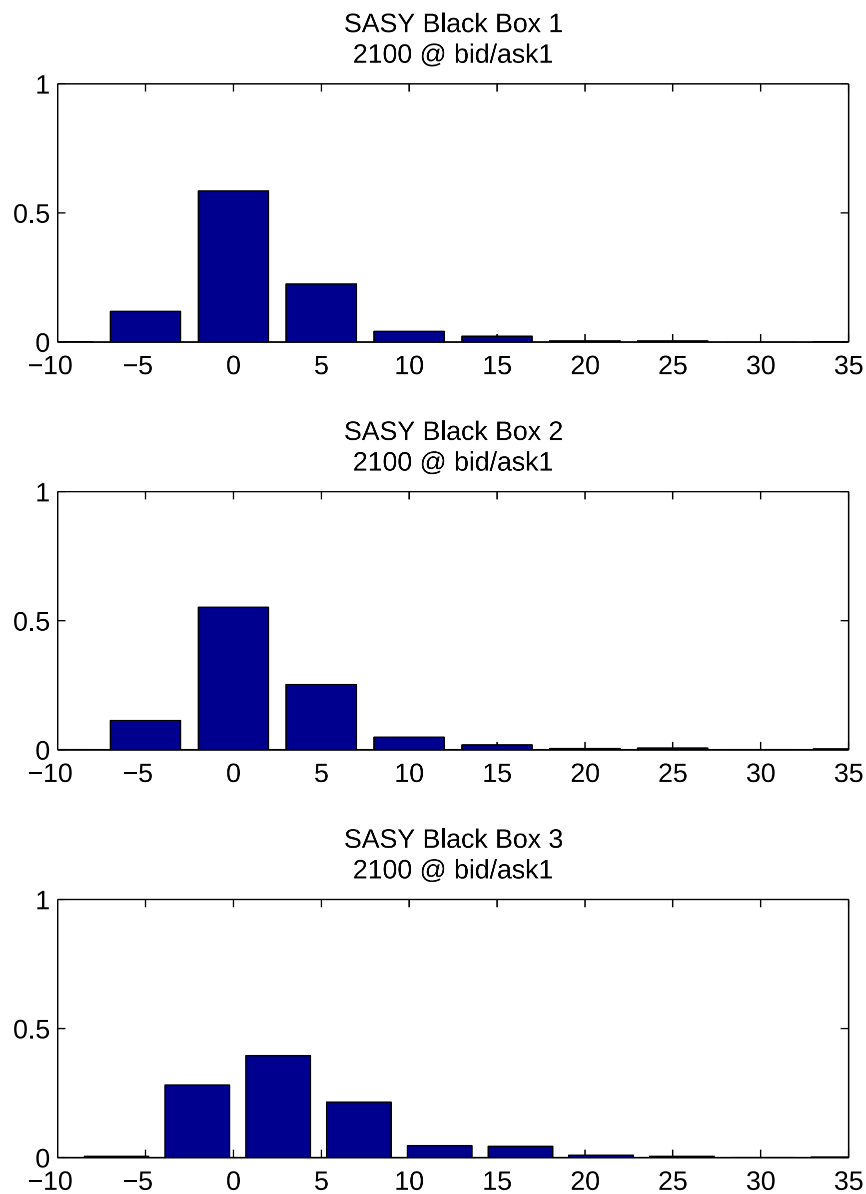

Figure 6.14: SASY: Relative frequency histogram of Negative Selection. SASY: Histogram relativne frekvencije Negativen Selekcije.

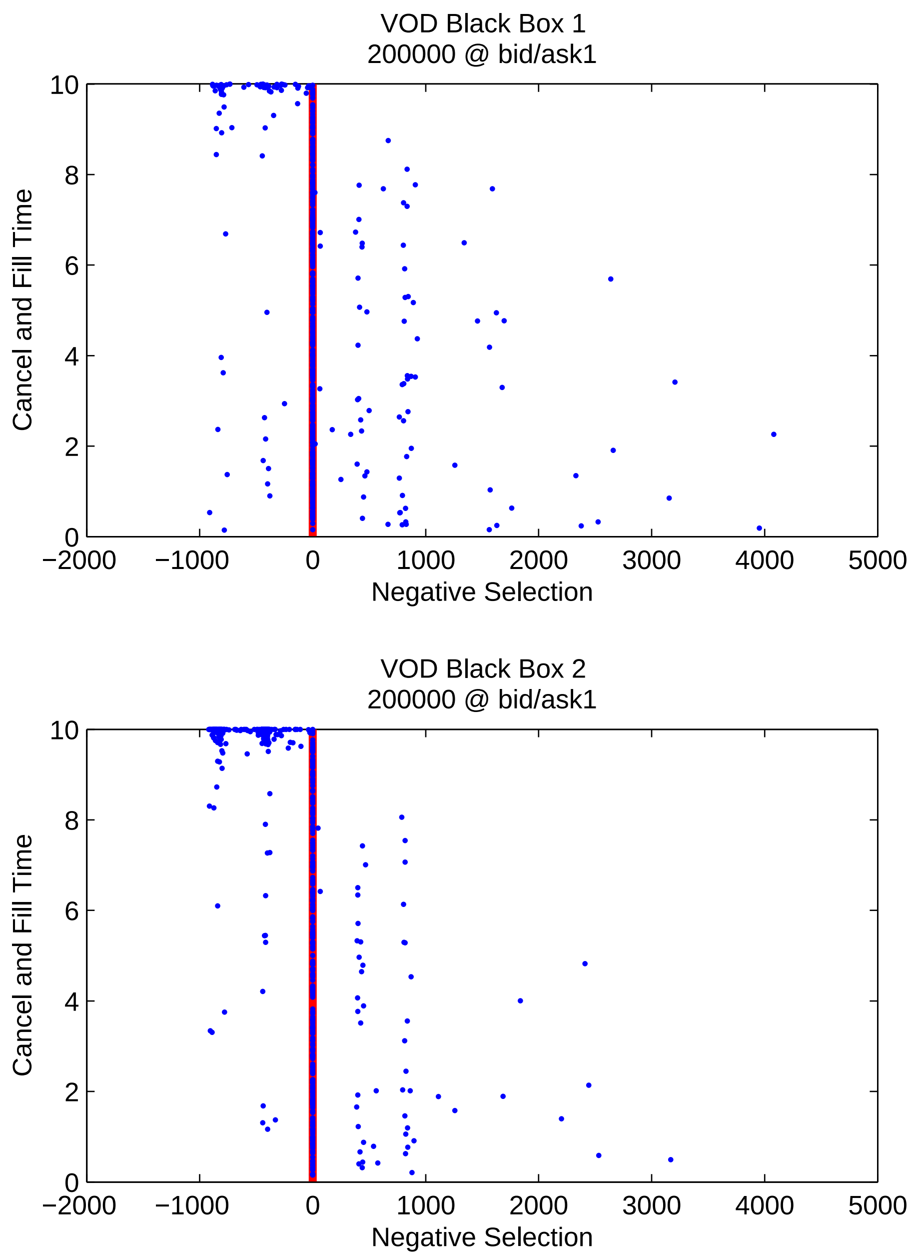

Figure 6.15: VOD: The scatter plot of Cancel and Fill time and NS. VOD: Dijagram rasipanja za Cancel i Fill vremena i NS.

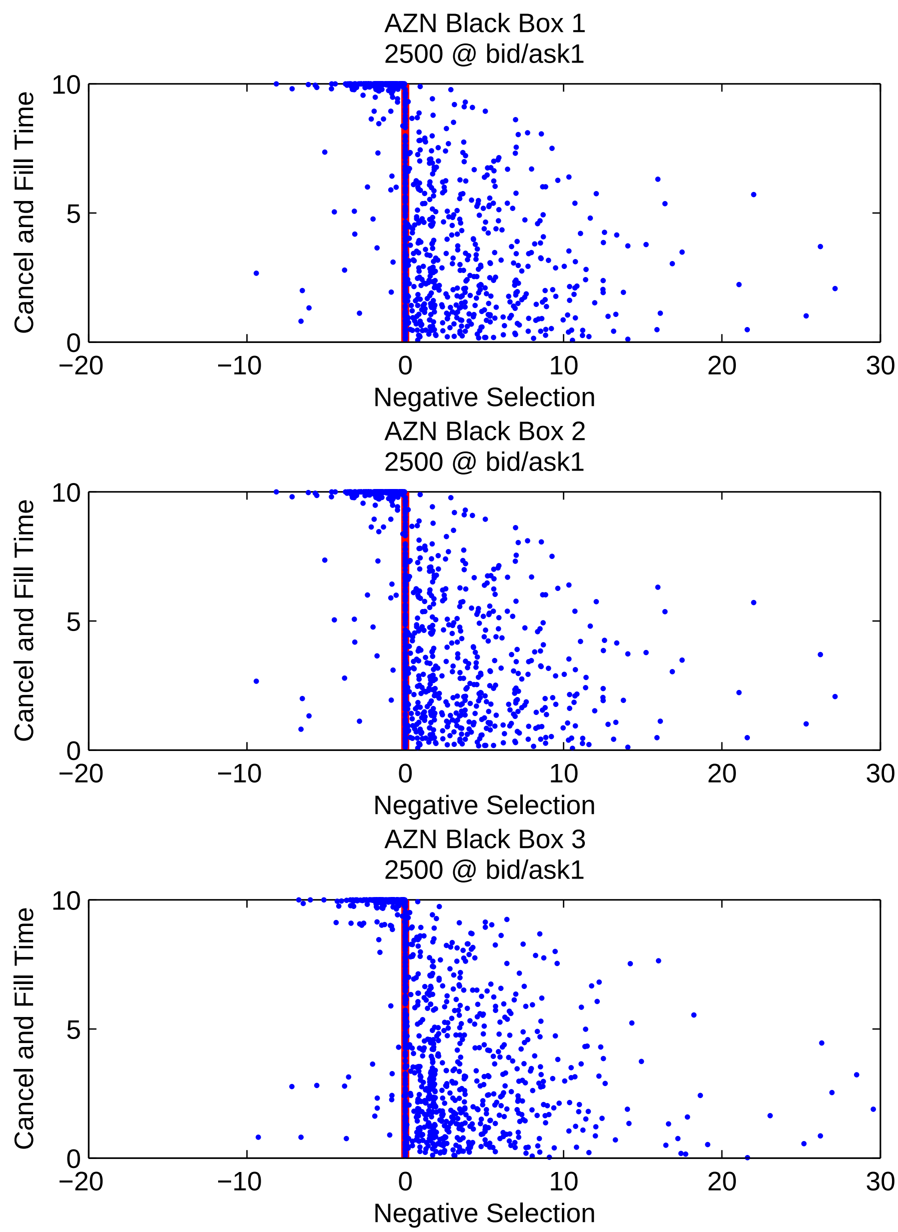

Figure 6.16: AZN: The scatter plot of Cancel and Fill time and NS. AZN: Dijagram rasipanja za Cancel i Fill vremena i NS.

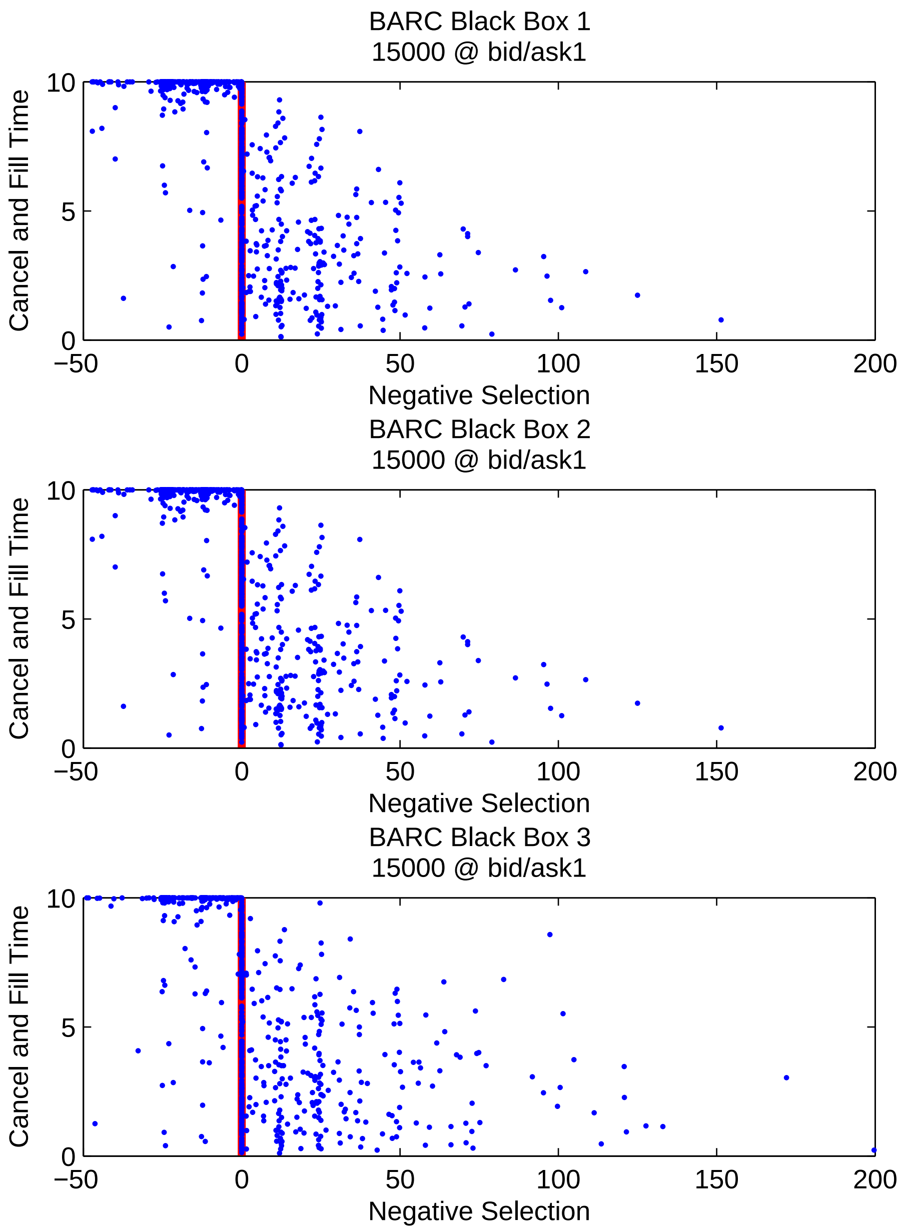

Figure 6.17: BARC: The scatter plot of Cancel and Fill time and NS. BARC: Dijagram rasipanja za Cancel i Fill vremena i NS.

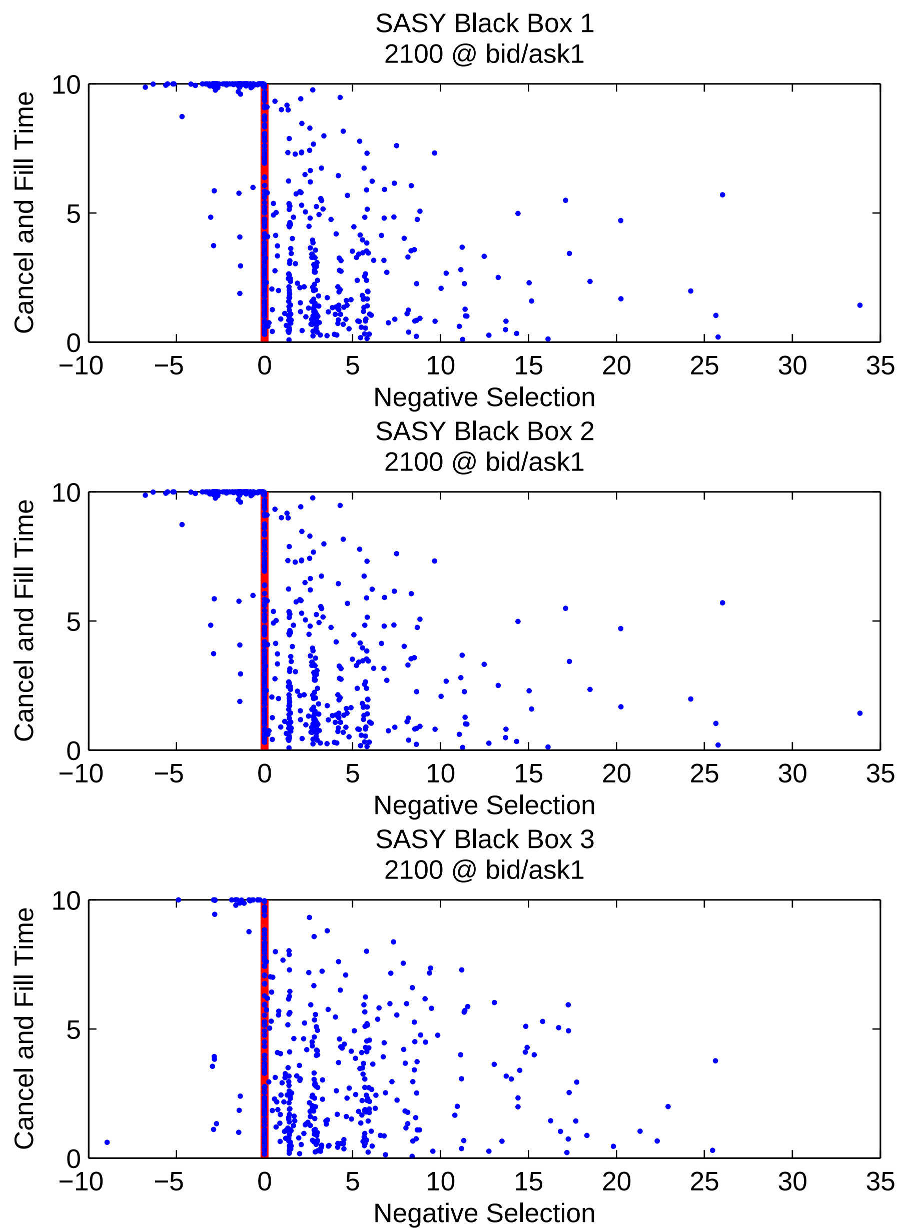

Figure 6.18: SASY: The scatter plot of Cancel and Fill time and NS. SASY: Dijagram rasipanja za Cancel i Fill vremena i NS.

Ticker:		VOD		AZN			BARC			SASY		
		BB1	BB2	BB1	BB2	BB3	BB1	BB2	BB3	BB1	BB2	BB3
100% at bid/ ask1	Mean (NS)	25.23	-215.35	1.40	1.62	2.38	0.48	-1.76	5.90	1.44	1.84	3.82
	St.Dev. (NS)	605.60	468.14	4.94	4.30	4.76	25.00	22.15	32.08	4.31	4.89	5.13
	Coeff. Of Variation (NS)	24.00	-2.17	3.53	2.65	2.00	52.48	-12.61	5.44	3.00	2.66	1.34
	Median (NS)	0.00	-23.34	0.00	0.00	1.12	0.00	0.00	0.00	0.00	0.00	2.61
	Mean(T_{fill})	4.37	4.39	3.74	3.63	3.86	4.33	4.32	4.20	3.68	3.49	3.36
	Mean(T_{cancel})	9.03	9.68	9.68	9.61	9.20	9.59	9.69	9.31	9.67	9.70	7.57
100% at bid/ ask2	Mean (NS)	-450.57	-609.94	-0.20	-0.04	0.49	-14.14	-16.10	-10.93	-0.63	-0.26	1.33
	St.Dev. (NS)	428.19	370.43	4.72	3.86	4.21	22.90	20.69	28.32	3.95	4.47	4.36
	Coeff. Of Variation (NS)	-0.95	-0.61	-23.44	-102.22	8.63	-1.62	-1.29	-2.59	-6.27	-17.12	3.28
	Median (NS)	-428.27	-754.94	-0.78	-0.76	0.00	-14.40	-20.17	-12.49	-1.35	-0.74	0.00
	Mean(T_{fill})	5.99	5.35	4.69	4.63	4.76	5.33	5.62	5.27	4.61	4.43	4.48
	Mean(T_{cancel})	9.56	9.80	9.76	9.70	9.49	9.71	9.77	9.54	9.75	9.76	9.00
100% at bid/ ask3	Mean (NS)	-863.71	-1011.38	-1.60	-1.55	-1.16	-26.60	-28.30	-24.52	-2.26	-2.09	-0.82
	St.Dev. (NS)	406.38	355.80	4.56	3.76	4.09	21.21	20.16	27.18	3.43	3.82	3.95
	Coeff. Of Variation (NS)	-0.47	-0.35	-2.85	-2.42	-3.52	-0.80	-0.71	-1.11	-1.52	-1.82	-4.78
	Median (NS)	-847.46	-1002.02	-1.86	-1.79	-1.52	-25.86	-31.09	-25.00	-2.78	-2.75	-1.38
	Mean(T_{fill})	5.62	5.90	5.43	5.33	5.36	5.62	6.20	5.96	5.23	5.21	5.17
	Mean(T_{cancel})	9.56	9.80	9.79	9.72	9.54	9.74	9.79	9.59	9.76	9.78	9.37

Table 6.7: Negative Selection statistics and Fill/Cancel average time for Black Box (BB) trading strategies and 5% of average traded quantity. Svojstva Negativna Selekcija i Fill i Cancel vremena

Ticker:		VOD		AZN			BARC			SASY		
		BB1	BB2	BB1	BB2	BB3	BB1	BB2	BB3	BB1	BB2	BB3
100% at bid/ ask1	Mean (NS)	11.01	-40.40	0.39	0.45	0.63	0.83	0.46	2.13	0.39	0.49	0.94
	St.Dev. (NS)	121.72	96.11	1.03	0.93	1.04	5.37	5.04	7.01	0.92	1.05	1.07
	Coeff. Of Variation (NS)	11.05	-2.38	2.63	2.07	1.65	6.45	10.98	3.30	2.36	2.14	1.14
	Median (NS)	0.00	0.00	0.02	0.15	0.35	0.00	0.00	0.00	0.00	0.27	0.58
	Mean(T_{fill})	3.97	4.02	3.17	3.14	3.31	3.46	3.78	3.45	3.08	2.89	2.81
	Mean(T_{cancel})	8.71	9.68	9.67	9.50	8.96	9.58	9.65	9.24	9.63	9.64	5.26
100% at bid/ ask2	Mean (NS)	-88.15	-119.62	0.04	0.08	0.22	-2.37	-2.73	-1.59	-0.06	0.04	0.42
	St.Dev. (NS)	85.29	75.06	0.97	0.81	0.94	4.74	4.40	5.96	0.84	0.95	0.94
	Coeff. Of Variation (NS)	-0.97	-0.63	26.01	10.24	4.37	-2.00	-1.62	-3.75	-14.91	24.54	2.23
	Median (NS)	-84.93	-95.46	0.00	0.00	0.00	-2.48	-2.53	-2.41	-0.16	0.00	0.00
	Mean(T_{fill})	5.83	5.16	4.39	4.37	4.43	5.15	5.37	5.09	4.25	4.15	3.98
	Mean(T_{cancel})	9.56	9.79	9.75	9.67	9.41	9.69	9.76	9.51	9.75	9.77	8.79
100% at bid/ ask3	Mean (NS)	-170.47	-200.32	-0.26	-0.24	-0.15	-4.96	-5.25	-4.46	-0.41	-0.36	-0.07
	St.Dev. (NS)	81.01	71.12	0.94	0.78	0.90	4.42	4.19	5.74	0.71	0.80	0.80
	Coeff. Of Variation (NS)	-0.48	-0.36	-3.60	-3.19	-6.21	-0.89	-0.80	-1.29	-1.74	-2.26	-11.44
	Median (NS)	-168.07	-181.00	-0.34	-0.34	-0.18	-4.95	-5.02	-4.90	-0.55	-0.54	0.00
	Mean(T_{fill})	5.65	5.78	5.22	5.12	5.23	5.61	6.04	5.74	5.04	5.10	4.97
	Mean(T_{cancel})	9.56	9.80	9.79	9.71	9.52	9.74	9.79	9.58	9.76	9.77	9.31

Table 6.8: Negative Selection statistics and Fill/Cancel average time for Black Box (BB) trading strategies and 1% of average traded quantity. Svojstva Negativna Selekcija i Fill i Cancel vremena

6.3 Review of Optimization Results

In previous section, we started with a simple order for which we defined Negative Selection 4.24, as the distance between the vector of Optimal Placement and actual order. When placing an order, one faces the dilemma of either being aggressive and cross the spread to buy at the prevailing asking price or take the chance of a better price by placing a more passive limit order. For example, in a rising market, a passive buy order at Bid1 will remain unfilled which would lead to chasing the market to get filled, resulting in a larger slippage than crossing the spread. While in a sideways market, one is likely to save the spread cost by being passive. In the case of a falling market, a buyer is considered too aggressive if the entire order is placed at Bid1 since one would achieve a better average price by having placed it at an even more passive price level. However, in the latter case, the probability of fill decreases significantly with more passive orders. Therefore, there is a need to split the orders into multiple price levels.

To calculate NS of a complex order, we use Algorithm 1. As NS of complex order is a vector of NS values of corresponding placements at different price levels, we still have all the information of execution performance for each level, taking into account our own trading. From practical reason, as we need sensitivity to small entries, we use norm-1 of NS vector.

By solving optimization problem defined in (5.2), we aim to have execution as close as possible to optimal. For this purpose, the data was divided in in-sample consisting of 80% and out-sample containing 20% of overall data. The optimization was done using in-sample data, then the profitability of the obtained result was compared to the profitability of 15 strategies across the out-sample data. We tested two order sizes 1% and 5% of average traded quantity in 10 minutes, for all four stock and all BB trading strategies.

For VOD.L in the in-sample optimization, the optimal is $(0, 100, 0, 0, 0, 0)$ i.e. placing 100% of shares at bid/ask1 gives us the minimal expected norm of NS for all BB trading strategies and both order sizes. Here, 1% and 5% of average traded quantity in 10 minutes intervals is approximate 40 000 and 2000 000. From Table 6.9 and Table 6.10 we can see that profitability of this placement is the highest in the out-sample data, again for all BB strategies, and both order sizes. For BB1 and 1%-size by far the best overall all possible options for splitting an order, while for 5% size profit for 100% at bid/ask2 is positive but considerably smaller than profit for 100% at bid/ask1.

BARC.L behaves similarly to VOD.L regarding optimization results, i.e.,

$(0, 100, 0, 0, 0, 0)$ is optimal for both order sizes, 1% and 5% of average traded quantity in 10 minutes intervals (3000 and 15000 shares, respectively) and all BB strategies. The difference is that BARC.L allows more passive placements, i.e., bid/ask3, which will not lose money, but the profit is still poor in comparison with placing everything on bid/ask1 as we can see in Table 6.11 and Table 6.12.

For AZN.L optimization results in in-sample are not in line with profits in out-sample for BB1 and BB2. Specifically, in the case of BB1 and order size of 1% of average traded quantity in 10 minutes, i.e. 500 shares, the optimal solution is $(0, 50, 50, 0, 0, 0)$, but the highest profit is achieved by placing 100% of shares at bid/ask3 level (Table 6.13). For BB2 and same order size, again the optimal solution is $(0, 50, 50, 0, 0, 0)$, i.e. placing 50% at bid/ask1 and 50% at bid/ask2, but the highest profit is achieved by placing 70% at bid/ask1 and 30% at bid/ask2. For trading strategy BB3 and order size like in previous cases, the optimal solution $(0, 50, 50, 0, 0, 0)$ is the most profitable. When we consider order size 5% of average traded quantity in 10 minutes, i.e. 2500 shares, Table 6.14 shows profits in out sample and results are as follows. For BB1 optimal solution is $(0, 70, 30, 0, 0, 0)$, its profit in out-sample is slightly smaller than the highest profit obtained with 65% at bid/ask1 and 35% at bid/ask2. BB2 has the optimal solution $(0, 40, 60, 0, 0, 0)$, and its profit is quite smaller than the highest profit achieved with 85% at bid/ask1 and 15% at bid/ask2. Results for BB3 are completely in line with profits in out-sample, i.e. $(0, 40, 60, 0, 0, 0)$ is the optimal solution, and its profit is the highest.

In the case of SASY.PA, optimal solution is the same for all BB trading strategies and both order sizes and it is $(0, 0, 100, 0, 0, 0)$ i.e placing entire quantity at bid/ask2. Table 6.15 and Table 6.16) show that only case of discrepancy between optimal solution and out-sample profits is for BB3 and order size of 420 shares. Here the profit is the highest for even more passive placement, i.e. 100% at bid/ask3. However, if we compare this with Table 6.8, we see that for entire data sample the smallest absolute value of mean (NS) for BB3 is exactly at 100% at bid/ask3.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2 profit
$(100, 0, 0, 0, 0, 0)$	-2,532,291.72	-1,719,285.29
$(0, 100, 0, 0, 0, 0)$	110,176.20	549,994.97
$(0, 0, 100, 0, 0, 0)$	-157,021.05	80,740.03
$(0, 0, 0, 100, 0, 0)$	-176,557.21	-124,057.25
$(0, 0, 0, 0, 100, 0)$	-107,293.59	-120,871.65
$(0, 0, 0, 0, 0, 100)$	-82,934.01	-13,714.82
$(0, 10, 90, 0, 0, 0)$	-107,968.99	155,246.77
$(0, 90, 10, 0, 0, 0)$	92,666.91	534,298.50
$(0, 85, 15, 0, 0, 0)$	86,071.64	524,548.71
$(0, 50, 50, 0, 0, 0)$	80,716.03	433,271.12
$(0, 60, 40, 0, 0, 0)$	75,457.23	463,477.06
$(0, 70, 30, 0, 0, 0)$	74,956.39	490,830.55
$(0, 40, 60, 0, 0, 0)$	33,931.69	366,652.51
$(0, 65, 35, 0, 0, 0)$	74,545.03	477,474.86
$(0, 55, 45, 0, 0, 0)$	77,556.47	448,769.17

Table 6.9: Total profit in out-sample for 40000 VOD.L shares. Ukupni profit van uzorka za 40000 VOD.L akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2 profit
$(100, 0, 0, 0, 0, 0)$	-24,207,417.84	-31,239,497.88
$(0, 100, 0, 0, 0, 0)$	8,461,780.77	6,804,664.53
$(0, 0, 100, 0, 0, 0)$	839,537.25	944,565.52
$(0, 0, 0, 100, 0, 0)$	-53,014.54	-391,963.48
$(0, 0, 0, 0, 100, 0)$	-836,255.99	-549,903.04
$(0, 0, 0, 0, 0, 100)$	-414,670.03	-64,118.85
$(0, 10, 90, 0, 0, 0)$	1,531,332.33	1,643,492.52
$(0, 90, 10, 0, 0, 0)$	7,715,836.43	6,272,520.23
$(0, 85, 15, 0, 0, 0)$	7,343,022.41	6,014,265.35
$(0, 50, 50, 0, 0, 0)$	4,572,432.56	4,637,729.58
$(0, 60, 40, 0, 0, 0)$	5,375,880.30	5,048,748.78
$(0, 70, 30, 0, 0, 0)$	6,216,059.59	5,429,418.41
$(0, 40, 60, 0, 0, 0)$	3,746,833.55	3,910,130.43
$(0, 65, 35, 0, 0, 0)$	5,789,839.83	5,248,665.01
$(0, 55, 45, 0, 0, 0)$	4,964,342.73	4,847,878.97

Table 6.10: Total profit in out-sample for 200,000 VOD.L shares. Ukupni profit van uzorka za 200,000 VOD.L akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2: profit	BB3: profit
$(100, 0, 0, 0, 0, 0)$	-114,819.59	-303,638.69	-90,760.22
$(0, 100, 0, 0, 0, 0)$	385,899.72	431,572.27	528,809.10
$(0, 0, 100, 0, 0, 0)$	54,988.04	108,387.49	196,735.86
$(0, 0, 0, 100, 0, 0)$	59,867.76	78,092.47	109,482.47
$(0, 0, 0, 0, 100, 0)$	-638.52	-47,318.91	-9,738.50
$(0, 0, 0, 0, 0, 100)$	-13,692.20	-19,469.16	-21,912.12
$(0, 10, 90, 0, 0, 0)$	76,044.58	128,456.93	225,262.34
$(0, 90, 10, 0, 0, 0)$	342,696.21	384,484.96	490,200.79
$(0, 85, 15, 0, 0, 0)$	320,890.38	361,570.64	470,986.57
$(0, 50, 50, 0, 0, 0)$	170,116.10	218,759.01	338,250.85
$(0, 60, 40, 0, 0, 0)$	211,442.69	258,120.61	374,986.83
$(0, 70, 30, 0, 0, 0)$	254,050.68	297,147.81	412,727.38
$(0, 40, 60, 0, 0, 0)$	145,898.65	196,397.50	309,560.95
$(0, 65, 35, 0, 0, 0)$	232,531.96	277,466.12	393,780.84
$(0, 55, 45, 0, 0, 0)$	190,865.13	238,473.56	356,478.41

Table 6.11: Total profit in out-sample for 3,000 BARC.L shares.Ukupni profit van uzorka za 3,000 BARC.L akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2: profit	BB3: profit
$(100, 0, 0, 0, 0, 0)$	-585,901.98	-1,531,673.10	-464,587.83
$(0, 100, 0, 0, 0, 0)$	1,365,225.87	1,761,834.88	2,065,003.56
$(0, 0, 100, 0, 0, 0)$	181,611.83	528,368.95	1,010,274.91
$(0, 0, 0, 100, 0, 0)$	206,898.32	215,181.44	375,582.43
$(0, 0, 0, 0, 100, 0)$	59,517.46	-219,920.52	-11,666.05
$(0, 0, 0, 0, 0, 100)$	-85,974.32	-102,570.37	-116,073.02
$(0, 10, 90, 0, 0, 0)$	310,201.91	639,589.38	1,144,384.90
$(0, 90, 10, 0, 0, 0)$	1,288,511.05	1,623,324.73	1,981,759.72
$(0, 85, 15, 0, 0, 0)$	1,241,985.95	1,549,282.20	193,4452.86
$(0, 50, 50, 0, 0, 0)$	753,657.52	1,005,771.73	1,575,178.44
$(0, 60, 40, 0, 0, 0)$	928,898.50	1,151,755.66	1,669,452.95
$(0, 70, 30, 0, 0, 0)$	1,067,419.84	1,304,759.76	1,7782,65.29
$(0, 40, 60, 0, 0, 0)$	641,011.38	922,782.65	1,486,112.64
$(0, 65, 35, 0, 0, 0)$	1,004,912.18	1,225,053.47	1,724,479.88
$(0, 55, 45, 0, 0, 0)$	847,179.12	1,079,047.51	1,615,261.02

Table 6.12: Total profit in out-sample for 15,000 BARC.L shares. Ukupni profit van uzorka za 15,000 BARC.L akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2: profit	BB3: profit
$(100, 0, 0, 0, 0, 0)$	-186,588.42	-190,744.77	-254,832.96
$(0, 100, 0, 0, 0, 0)$	143,300.42	50,188.67	74,734.56
$(0, 0, 100, 0, 0, 0)$	78,540.67	16,176.66	105,230.09
$(0, 0, 0, 100, 0, 0)$	161,861.71	31,375.71	55,918.88
$(0, 0, 0, 0, 100, 0)$	53,219.82	42,234.19	70,870.58
$(0, 0, 0, 0, 0, 100)$	16,989.73	9,972.14	12,251.22
$(0, 10, 90, 0, 0, 0)$	86,005.87	23,620.34	106,842.27
$(0, 90, 10, 0, 0, 0)$	139,981.42	53,439.99	80,970.70
$(0, 85, 15, 0, 0, 0)$	138,144.09	54,280.05	84,338.25
$(0, 50, 50, 0, 0, 0)$	116,844.56	51,022.03	108,943.55
$(0, 60, 40, 0, 0, 0)$	124,189.48	52,834.35	102,501.34
$(0, 70, 30, 0, 0, 0)$	130,727.48	54,647.77	95,526.30
$(0, 40, 60, 0, 0, 0)$	108,552.00	44,326.50	108,506.67
$(0, 65, 35, 0, 0, 0)$	127,541.71	53,754.63	99,024.19
$(0, 55, 45, 0, 0, 0)$	120,682.10	51,976.49	105,826.13

Table 6.13: Total profit in out-sample for 500 AZN.L shares. Ukupni profit van uzorka za 500 AZN.L akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2: profit	BB3: profit
$(100, 0, 0, 0, 0, 0)$	-938,053.97	-957,295.45	-1,278,139.51
$(0, 100, 0, 0, 0, 0)$	551,574.09	242,032.24	338,915.65
$(0, 0, 100, 0, 0, 0)$	405,813.97	-32,760.28	411,341.20
$(0, 0, 0, 100, 0, 0)$	581,127.97	92,333.10	183,953.36
$(0, 0, 0, 0, 100, 0)$	226,235.27	210,074.82	133,570.87
$(0, 0, 0, 0, 0, 100)$	69,742.73	47,664.30	-4,346.71
$(0, 10, 90, 0, 0, 0)$	442,110.05	22,330.39	427,698.67
$(0, 90, 10, 0, 0, 0)$	575,201.53	259,314.60	332,646.30
$(0, 85, 15, 0, 0, 0)$	586,164.18	264,871.33	332,082.51
$(0, 50, 50, 0, 0, 0)$	575,822.56	215,474.38	457,459.46
$(0, 60, 40, 0, 0, 0)$	595,333.02	240,399.06	411,005.13
$(0, 70, 30, 0, 0, 0)$	596,216.00	257,920.59	351,925.08
$(0, 40, 60, 0, 0, 0)$	536,353.72	180,979.45	469,826.25
$(0, 65, 35, 0, 0, 0)$	601,537.23	249,651.44	386,942.52
$(0, 55, 45, 0, 0, 0)$	586,567.26	229,232.22	434,127.64

Table 6.14: Total profit in out-sample for 2,500 AZN.L shares.Ukupni profit van uzorka za 2,500 AZN.L akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2: profit	BB3: profit
$(100, 0, 0, 0, 0, 0)$	-2,334.72	-4,428.36	-2,733.08
$(0, 100, 0, 0, 0, 0)$	3,485.83	2,020.42	3,101.85
$(0, 0, 100, 0, 0, 0)$	5,352.82	4,691.98	4,998.70
$(0, 0, 0, 100, 0, 0)$	2,991.87	4,273.81	5,298.23
$(0, 0, 0, 0, 100, 0)$	1,929.52	1,718.23	1,325.56
$(0, 0, 0, 0, 0, 100)$	123.68	639.97	434.98
$(0, 10, 90, 0, 0, 0)$	5,198.75	4,525.59	4,885.63
$(0, 90, 10, 0, 0, 0)$	3,716.13	2,398.28	3,452.47
$(0, 85, 15, 0, 0, 0)$	3,837.68	2,578.73	3,594.56
$(0, 50, 50, 0, 0, 0)$	4,533.21	3,794.42	4,468.57
$(0, 60, 40, 0, 0, 0)$	4,407.42	3,453.53	4,247.78
$(0, 70, 30, 0, 0, 0)$	4,193.22	3,108.55	3,986.58
$(0, 40, 60, 0, 0, 0)$	4,701.29	3,978.95	4,579.83
$(0, 65, 35, 0, 0, 0)$	4,308.24	3,281.30	4,119.39
$(0, 55, 45, 0, 0, 0)$	4,471.08	3,624.11	4,357.39

Table 6.15: Total profit in out-sample for 420 SASY.PA shares. Ukupni profit van uzorka za 420 SASY.PA akcija.

$(x_0, x_1, x_2, x_3, x_4, x_5)$	BB1:profit	BB2: profit	BB3: profit
$(100, 0, 0, 0, 0, 0)$	-13,357.61	-25,078.28	-16,246.47
$(0, 100, 0, 0, 0, 0)$	15,441.77	10,485.07	12,666.99
$(0, 0, 100, 0, 0, 0)$	22,172.02	18,145.93	22,179.49
$(0, 0, 0, 100, 0, 0)$	14,229.30	17,577.81	19,904.48
$(0, 0, 0, 0, 100, 0)$	8,573.98	6,800.46	4,759.27
$(0, 0, 0, 0, 0, 100)$	489.48	3,187.62	1,983.91
$(0, 10, 90, 0, 0, 0)$	22,106.79	17,993.59	21,984.56
$(0, 90, 10, 0, 0, 0)$	16,683.25	11,885.91	14,792.90
$(0, 85, 15, 0, 0, 0)$	17,288.98	12,575.65	15,666.88
$(0, 50, 50, 0, 0, 0)$	20,511.35	17,012.81	20,837.66
$(0, 60, 40, 0, 0, 0)$	19,749.19	15,655.42	19,387.99
$(0, 70, 30, 0, 0, 0)$	18,935.04	14,355.52	17,979.82
$(0, 40, 60, 0, 0, 0)$	20,916.60	17,367.42	21,152.75
$(0, 65, 35, 0, 0, 0)$	19,340.23	14,965.01	18,666.90
$(0, 55, 45, 0, 0, 0)$	20,139.34	16,375.76	20,105.81

Table 6.16: Total profit in out-sample for 2,100 SASY.PA shares. Ukupni profit van uzorka za 2,100 SASY.PA akcija.

Bibliography

[1] Ahookhosh, M. and Ghaderi, S. On efficiency of nonmonotone armijo-type line searches. *Applied Mathematical Modelling*, 43:170–190, 2014.

[2] Ali, M. M., Khompatraporn, C., and Zabinsky, Z. B. A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems. *Journal of Global Optimization*, 31(4):635–672, 2005.

[3] Almgren, R. Execution costs. In *Encyclopedia of Quantitative Finance*. Wiley Online Library, 2008.

[4] Almgren, R., Thum, C., Hauptmann, E., and Li, H. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.

[5] Baker, H. K. and Filbeck, G. *Portfolio Theory and Management*. Oxford University Press, 2013.

[6] Banks, J. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley & Sons, 1998.

[7] Berkowitz, S. A., Logue, D. E., and Noser, E. A. The total cost of transactions on the NYSE. *The Journal of Finance*, 43(1):97–112, 1988.

[8] Bertsimas, D. and Lo, A. W. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[9] Connor, G., Goldberg, L. R., and Korajczyk, R. A. *Portfolio Risk Analysis*. Princeton University Press, 2010.

[10] Cook, G. C. Trading benchmark choice and transition management performance attribution. *Journal of Investing*, 20(2):143 – 154, 2011.

[11] Demsetz, H. The cost of transacting. *The quarterly journal of economics*, 82(1):33–53, 1968.

[12] Dolan, E. D. and Moré, J. J. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[13] Fabozzi, F. J. and Markowitz, H. M. *The Theory and Practice of Investment Management*, volume 118. John Wiley & Sons, 2002.

[14] Freyre-Sanders, A., Guobuzaite, R., and Byrne, K. A review of trading cost models: Reducing transaction costs. *Journal of Investing*, 13(3):93–115, 2004.

[15] Fu, M. et al. *Handbook of Simulation Optimization*, volume 216. Springer, 2015.

[16] Garman, M. B. Market microstructure. *Journal of Financial Economics*, 3(3):257–275, 1976.

[17] Gomes, C. and Waelbroeck, H. Transaction cost analysis to optimize trading strategies. *The Journal of Trading*, 5(4):29–38, 2010.

[18] Harris, L. Stock price clustering and discreteness. *Review of Financial Studies*, 4(3):389–415, 1991.

[19] Hasbrouck, J. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, 2006.

[20] Hu, G. VWAP cost excluding own trades. *The Journal of Trading*, 2(1):30–34, 2007.

[21] Johnson, B. *Algorithmic Trading & DMA: An introduction to direct access trading strategies*. 4Myeloma Press, London, 2010.

[22] Kendall, K. *Electronic and Algorithmic Trading Technology: The Complete Guide*. Academic Press, 2007.

[23] Kiefer, J., Wolfowitz, J., et al. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[24] Kissell, R. The expanded implementation shortfall: Understanding transaction cost components. *The Journal of Trading*, 1(3):6–16, 2006.

[25] Kissell, R. *The Science of Algorithmic Trading and Portfolio Management*. Academic Press, 2013.

[26] Kissell, R., Glantz, M., Malamut, R., Association, A. M., et al. *Optimal Trading Strategies: Quantitative Approaches for Managing Market Impact and Trading Risk*. Amacom, 2003.

[27] Kissell, R. and Malamut, R. Understanding the profit and loss distribution of trading algorithms. *Trading*, 2005(1):41–49, 2005.

[28] Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[29] Krejić, N. and Jerinkić, N. K. Stochastic gradient methods for unconstrained optimization. *Pesquisa Operacional*, 34(3):373–393, 2014.

[30] Krejić, N. and Jerinkić, N. K. Nonmonotone line search methods with variable sample size. *Numerical Algorithms*, 68(4):711–739, 2015.

[31] Krejić, N. and Lončar, S. A noonmonotone line search for stochastic optimization problems. Filomat(to appear), University of Novi Sad, Faculty of Sciences, Department of Mathematics and Informatics `http://people.dmi.uns.ac.rs/~natasa/Natasa%20Krejic%20-%20Curriculum%20Vitae.htm`.

[32] Kritzman, M., Myrgren, S., and Page, S. Implementation shortfall. *Journal of Portfolio Management*, 33(1):25 – 30, 2006.

[33] Leshik, E. A. and Cralle, J. An introduction to algorithmic trading. *Padstow: John Wiley and Sons*, 2011.

[34] Li, D.-H. and Fukushima, M. A derivative-free line search and global convergence of broyden-like method for nonlinear equations. *Optimization Methods and Software*, 13(3):181–201, 2000.

[35] Lončar, S., Kumaresan, M., and Krejić, N. Negative selection - a new performance measure for automated order execution. Technical report,

University of Novi Sad, Faculty of Sciences, Department of Mathematics and Informatics `http://people.dmi.uns.ac.rs/~natasa/Natasa%20Krejic%20-%20Curriculum%20Vitae.htm`.

[36] Madhavan, A. Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258, 2000.

[37] Madhavan, A. N. Implementation of hedge fund strategies. *Special Issues*, 2002(1):74–80, 2002.

[38] Madhavan, A. N. VWAP strategies. *Trading*, 2002(1):32–39, 2002.

[39] Mangasarian, O. L. Uniqueness of solution in linear programming. *Linear Algebra and its Applications*, 25:151–162, 1979.

[40] Molga, M. and Smutnicki, C. Test functions for optimization needs. *http://new.zsd.iiar.pwr.wroc.pl/files/docs/functions.pdf*, 2005.

[41] Narang, R. K. *Inside the Black Box: A Simple Guide to Quantitative and High Frequency Trading*, volume 884. John Wiley & Sons, 2013.

[42] O'Hara, M. *Market Microstructure Theory*, volume 108. Blackwell Cambridge, MA, 1995.

[43] O'Hara, M. and de Prado, M. L. *High Frequency Trading*. Risk Books, London, 2013.

[44] Perold, A. F. The implementation shortfall: Paper versus reality. *The Journal of Portfolio Management*, 14(3):4–9, 1988.

[45] Robbins, H. and Monro, S. A stochastic approximation method. *The Annals of Mathematical Statistics*, 400 − 407, 1951.

[46] Roşu, I. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, 2009.

[47] Sacks, J. Asymptotic distribution of stochastic approximation procedures. *The Annals of Mathematical Statistics*, 29(2):373–405, 1958.

[48] Saraiya, N. and Mittal, H. Understanding and avoiding adverse selection in dark pools. *Investment Technology Group, November*, 2009.

[49] Schmidt, A. B. *Financial Markets and Trading: an Introduction to Market Microstructure and Trading Strategies*, volume 637. John Wiley & Sons, 2011.

[50] Self, O. Anti-gaming and anti-negative selection strategies for optimal execution. In *Algorithmic Trading & SOR Handbook 3rd edition*, 133–140. The Trade, 2009.

[51] Shapiro, A., Dentcheva, D., et al. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9. SIAM, 2009.

[52] Spall, J. C. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, 2005.

[53] Stoll, H. R. Market microstructure. *Handbook of the Economics of Finance*, 1:553–604, 2003.

[54] Wagner, W. H. and Glass, S. What every plan sponsor needs to know about transaction costs. *Trading*, 2001(1):20–35, 2001.

[55] Wardi, Y. A stochastic steepest-descent algorithm. *Journal of Optimization Theory and Applications*, 59(2):307–323, 1988.

[56] Wardi, Y. Stochastic algorithms with Armijo stepsizes for minimization of functions. *Journal of Optimization Theory and Applications*, 64(2):399–417, 1990.

[57] Yan, D. and Mukai, H. Optimization algorithm with probabilistic estimation. *Journal of Optimization Theory and Applications*, 79(2):345–371, 1993.

[58] Zubulake, P. and Lee, S. *The High Frequency Game Changer: How Automated Trading Strategies Have Revolutionized the Markets*, volume 486. John Wiley & Sons, 2011.

Biography

I was born in 1976 in Novi Sad, where I finished primary school "Svetozar Marković Toza" and the high school "Gimnazija Jovan Jovanović Zmaj." I enrolled study program "Professor in Mathematics" at Faculty of Sciences, University of Novi Sad and graduated in 2002 with the average grade 9.76. In 2010 I obtained M.Sc. in Mathematics, at University of Novi Sad.

From November to December 2002 I worked as a mathematics teacher at primary school "Ivo Lola Ribar," Novi Sad. In the period between February 2004 and September 2006, I worked in "Novosadska banka"/"Erste Bank" as IT Application Analyst/Designer. My job included maintenance and development of applications for domestic and foreign currency savings, mostly using COBOL programming language, JCL, REXX, and some VB6. From November 2006 to October 2007 I worked as a mathematics teacher at primary school "Jovan Dučić" Petrovaradin. From October 2007 I am employed at Novi Sad Business School Higher Education Institution for Applied Studies, where I work as a lecturer since 2012. I held courses on Mathematics, Information Technology Application, Principles of E-business, Quantitative Methods in Business Decision Making and Discrete Mathematical Structures.

Novi Sad, June 6, 2017 Sanja Lončar

UNIVERSITY OF NOVI SAD
FACULTY OF NATURAL SCIENCES AND MATHEMATICS
KEY WORDS DOCUMENTATION

Accession number:
ANO
Identification number:
INO
Document type: Monograph type
DT
Type of record: Printed text
TR
Contents Code: PhD thesis
CC
Author: Sanja Lončar
AU
Mentor: Prof. Dr. Nataša Krejić
MN
Title: Negative Selection - An Absolute Measure of Arbitrary Algorithmic Order Execution
XI
Language of text: English
LT
Language of abstract: English/Serbian
LA
Country of publication: Republic of Serbia
CP
Locality of publication: Vojvodina
LP
Publication year: 2017
PY
Publisher: Author's reprint
PU
Publ. place: Novi Sad, Faculty of Natural Sciences and Mathematics, Trg Dositeja Obradoviča 4
PP
Physical description: (6,153,58,19,5,19,0)

(Chapters, pages, literature, tables, pictures, graphics, appendices)
PD
Scientific field: Mathematics
SF
Scientific discipline: Numerical mathematics
SD
Key words:
KW
Holding data: library of the Department of Mathematics and Informatics, Novi Sad
HD
Note:
N
Abstract:
Algorithmic trading is an automated process of order execution on electronic stock markets. It can be applied to a broad range of financial instruments, and it is characterized by a significant investors' control over the execution of his/her orders, with the principal goal of finding the right balance between costs and risk of not (fully) executing an order. As the measurement of execution performance gives information whether best execution is achieved, a significant number of different benchmarks is used in practice. The most frequently used are price benchmarks, where some of them are determined before trading (Pre-trade benchmarks), some during the trading day (Intraday benchmarks), and some are determined after the trade (Post-trade benchmarks). The two most dominant are VWAP and Arrival Price, which is along with other pre-trade price benchmarks known as the Implementation Shortfall (IS).

We introduce Negative Selection as a posteriori measure of the execution algorithm performance. It is based on the concept of Optimal Placement, which represents the ideal order that could be executed in a given time window, where the notion of ideal means that it is an order with the best execution price considering market conditions during the time window. Negative Selection is defined as a difference between vectors of optimal and executed orders, with vectors defined as a quantity of shares at specified price positions in the order book. It is equal to zero when the order is optimally executed; negative if the order is not (completely) filled, and positive if the order is executed but at an unfavorable price.

Negative Selection is based on the idea to offer a new, alternative per-

formance measure, which will enable us to find the optimal trajectories and construct optimal execution of an order.

The first chapter of the thesis includes a list of notation and an overview of definitions and theorems that will be used further in the thesis. Chapters 2 and 3 follow with a theoretical overview of concepts related to market microstructure, basic information regarding benchmarks, and theoretical background of algorithmic trading. Original results are presented in chapters 4 and 5. Chapter 4 includes a construction of optimal placement, definition and properties of Negative Selection. The results regarding the properties of a Negative Selection are given in [35]. Chapter 5 contains the theoretical background for stochastic optimization, a model of the optimal execution formulated as a stochastic optimization problem with regard to Negative Selection, as well as original work on nonmonotone line search method [31], while numerical results are in the last, 6th chapter.
AB
Accepted by the Scientific Board on: June 25, 2015
ASB
Defended:
DE
Thesis defend board:
President: Zorana Lužanin, PhD, Full Professor, Faculty of Sciences, University of Novi Sad
Member: Nataša Krejić, PhD, Full Professor, Faculty of Sciences, University of Novi Sad
Member: Danijela Rajter Ćirić, PhD, Full Professor, Faculty of Sciences, University of Novi Sad
Member: Branko Urošević, PhD, Full Professor, Faculty of Economics,University of Belgrade

DB

UNIVERZITET U NOVOM SADU
PRIRODNO - MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR
Identifikacioni broj:
IBR
Tip dokumentacije: Monografska dokumentacija
TD
Tip zapisa: Tekstualni štampani materijal
TZ
Vrsta rada: Monografska dokumentacija
VR
Autor: Sanja Lončar
AU
Mentor: Prof. dr Nataša Krejić
MN
Naslov rada: Negativna selekcija - Apsolutna mera algoritamskog izvršenja proizvoljnog naloga
MR
Jezik publikacije: Engleski
JP
Jezik izvoda: engleski/srpski
JI
Zemlja publikovanja: Republika Srbija
ZP
Uže geografsko područje: Vojvodina
UGP
Godina: 2017.
GO
Izdava v c: Autorski reprint
IZ
Mesto i adresa: Novi Sad, Departman za matematiku i informatiku, PMF, Trg Dositeja Obradovića 4
MA
Fizički opis rada: (6,153,58,19,5,19,0)
(Broj poglavlja, strana, citata, tabela, slika, grafika, priloga)

FO
Naučna oblast: Matematika
NO
Naučna disciplina: Numerička matematika
ND
Ključne reči:
PO
UDK:
Čuva se: u biblioteci Departmana za matematiku i informatiku, Novi Sad
ČU
Važna napomena:
VN
Izvod:
Algoritamsko trgovanje je automatizovani proces izvršavanja naloga na elektronskim berzama. Može se primeniti na širok spektar finansijskih instrumenata kojima se trguje na berzi i karakteriše ga značajna kontrola investitora nad izvršavanjem njegovih naloga, pri čemu se teži nalaženju pravog balansa između troška i rizika u vezi sa izvršenjem naloga. S ozirom da se merenjem performasi izvršenja naloga određuje da li je postignuto najbolje izvršenje, u praksi postoji značajan broj različitih pokazatelja. Najčešće su to pokazatelji cena, neki od njih se određuju pre trgovanja (eng. Pre-trade), neki u toku trgovanja (eng. Intraday), a neki nakon trgovanja (eng. Post-trade). Dva najdominantnija pokazatelja cena su VWAP i Arrival Price koji je zajedno sa ostalim "pre-trade" pokazateljima cena poznat kao Implementation shortfall (IS).

Pojam negative selekcije se uvodi kao "post-trade" mera performansi algoritama izvršenja, polazeći od pojma optimalnog naloga, koji predstavlja idealni nalog koji se mogao izvršiti u datom vremenskom intervalu, pri čemu se pod pojmom "idealni" podrazumeva nalog kojim se postiže najbolja cena u tržišnim uslovima koji su vladali u toku tog vremenskog intervala. Negativna selekcija se definiše kao razlika vektora optimalnog i izvršenog naloga, pri čemu su vektori naloga definisani kao količine akcija na odgovarajućim pozicijama cena knjige naloga. Ona je jednaka nuli kada je nalog optimalno izvršen; negativna, ako nalog nije (u potpunosti) izvršen, a pozitivna ako je nalog izvršen, ali po nepovoljnoj ceni.

Uvođenje mere negativne selekcije zasnovano je na ideji da se ponudi nova, alternativna, mera performansi i da se u odnosu na nju nađe optimalna trajektorija i konstruiše optimalno izvršenje naloga.

U prvom poglavlju teze dati su lista notacija kao i pregled definicija i teorema neophodnih za izlaganje materije. Poglavlja 2 i 3 bave se teorijskim pregledom pojmova i literature u vezi sa mikrostrukturom tržista, pokazateljima trgovanja i algoritamskim trgovanjem. Originalni rezultati su predstavljeni u 4. i 5. poglavlju. Poglavlje 4 sadrži konstrukciju optimalnog naloga, definiciju i osobine negativne selekcije. Teorijski i paraktični rezultati u vezi sa osobinama negativna selekcije dati su u [35]. Poglavlje 5 sadrži teorijske osnove stohastičke optimizacije, definiciju modela za optimalno izvršenje, kao i originalni rad u vezi sa metodom nemonotonog linijskog pretraživanja [31], dok 6. poglavlje sadži empirijske rezultate.
IZ
Datum prihvatanja teme od strane NN veća: 25.06.2015.
DP
Datum odbrane:
DO
Članovi komisije:
Predsednik: dr Zorana Lužanin, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu
Član: dr Nataša Krejić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu
Član: dr Danijela Rajter Ćirić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu
Član: dr Branko Urošević, redovni profesor, Ekonomski fakultet, Univerzitet u Beogradu
KO