NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Nikolay A. Andreev

ON LINEARITY OF TRANSACTION COSTS IN ORDER DRIVEN MARKET

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: Financial Economics
WP BRP 38/FE/2014

Nikolay A. Andreev[1]

ON LINEARITY OF TRANSACTION COSTS IN ORDER DRIVEN MARKET[2]

We research the properties of implicit transaction costs function for general-shaped limit order book. Equivalent conditions for linearity of the function are presented in terms of market liquidity. We also present a suitable functional form of implicit costs for order-driven market on the Moscow Interbank Currency Exchange (MICEX), based on high-frequency trading data. The proposed form meets the definition of costs and implied properties while corresponding to the real form of order distribution.

JEL Classification: C60, G11, G17.

Key words: transaction costs, limit order book, trading volume, market microstructure, market liquidity.

The results of the research have been revised and published in

Andreev, N. A. (2020). On the Convexity of the Instantaneous Impact Cost Function. Journal of Mathematical Sciences, 248(1), 116-122.

[1] National Research University Higher School of Economics. Laboratory for Financial Engineering and Risk Management. Researcher; E-mail: nandreev@hse.ru

[2] Study was implemented in the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2014.

Introduction

The development of modern concept of transaction costs has begun in the middle of the XX century [Coase, 1937], [Marschak, 1950]; for an overview of the history of the concept see [Klaes, 2000]. Transaction costs are commonly decomposed into implicit and explicit part [Keim and Madhavan, 1998]. Explicit costs are the direct costs of performing a deal on the market which include taxes and brokerage commission, usually a fixed fee per deal or a percent of deal volume, while percent might vary over time. Broker can also charge a percent of the deal value, i.e. costs are proportional to $P \cdot V$, where P is the asset price calculated according to brokerage services agreement, V denotes deal's volume. In practice, commission formula can vary based on deal's volume or value. Implicit costs, on the other hand, depend solely on market microstructure and liquidity, and have drawn much attention since the Long Term Capital Management (LTCM) hedge-fund's severe losses followed by collapse, due to high transaction costs. The LTCM crisis threatened to create substantial losses for many Wall Street lenders thus triggering a chain reaction in numerous markets, which made the Federal Reserve Bank of New York organize a bailout of the fund [Jorion, 2000].

Many key works on portfolio selection and management theory assume that transaction function is either fixed per deal or directly proportional to deal volume[3]. While these assumptions are justified for explicit costs in many cases [Keim and Madhavan, 1998], they must be justified or, better, proved for implicit costs. Otherwise, linearity can only be considered an approximation. Constant approximation, despite the simplicity, allows to introduce a fee per single deal and to regularize portfolio management problem by making strategies with high or infinite trade intensity (number of trades per unit of time) suboptimal. Linear approximation allows to incorporate explicit costs and impact of a deal size on implicit costs value. Absence of zero-degree linear term is a crucial assumption in many classical portfolio management frameworks, which allows to obtain closed-form analytic solution to the problem. Direct proportionality assumes missing or negligible fixed fee when compared to other costs. Works with non-linear parametrization include [Constantinides, 1979], [Constantinides, 1986] for convex cost function; [Vath et al., 2007] which introduces price-dependent implicit costs $C(V, \tilde{P}) = V \cdot I(V, \tilde{P})$ and permanent price impact $I(V, \tilde{P}) = \tilde{P}e^{\lambda V}$. The explicit form of impulse strategy for convex transaction costs, including fixed fee, is

[3]See., for example, [Magill and Constantinides, 1976], [Davis and Norman, 1990], [Shreve and Soner, 1994], [Framstad et al., 2001], [Ø ksendal and Sulem, 2010].

obtained in 2007 in [Ma et al., 2013]. [Андреев et al., 2011] considers polynomial form of implicit costs with auto-regressive coefficients when solving the optimal liquidation problem.

We present results of theoretical and empirical research of implicit costs function in order-driven market. For general-form distribution of limit orders in terms of cumulative volume as a function of depth, we introduce a definition of implicit costs and market liquidity. We show that linearity of implicit costs is equivalent to absolute market liquidity, while not identically zero only if agent's estimate of fair price differs from best quoted price. This draws to the conclusion that linearity of implicit costs does not match any real distribution of orders thus can lead to severe underestimation of losses due to large deals.

Based on empirical study of real distributions of orders at MICEX stock exchange for the period of 2006-2007, we introduce a parametric model of implicit costs which we believe to be more adequate to the data observed. We also demonstrate, for a specific management policy, how one can obtain an upper bound of deal volume, for which implicit costs can be approximated by a linear function.

Order-driven market

In order-driven (ODM) market, market maker's functions are delegated to automatic mechanism of order matching. We study a simplified model of the market which demonstrates its necessary aspects. At any time participants can place one of two types of orders: limit or market. Market orders trigger immediate trade (or sequence of trades) at available best prices while limit orders mean just the intent to trade at any price, up to the limit specified by the order.

Definition 1. *Market order is an order to perform immediate deal, which is characterized by volume v and direction ζ: $\zeta = 1$ denotes buying, $\zeta = -1$ denotes selling.*

Definition 2. *Limit order is specified by all the parameters of market order and limit execution price P. Limit order can be interpreted as an intent to trade volume V at a price not worse than P.*

Each order also has identification number in the automatic matching system (time stamp). All participants have access to currently available set of limit orders sorted by limit price, also known as limit order book. At any point participants can perform the following actions:

1. Placement of a limit order. The order is added to the book and remains there until withdrawn by the owner, liquidated due to a deal or cancelled by the system at the end of trading period. Execution of a limit order can be partial; execution is performed immediately after an order of suitable price and opposite direction (market or limit) appears in the book.

2. Placement of a market order. The order is not added to the book but triggers its immediate change. Market order can lead to a sequence of deals with available limit orders of opposite direction, so that cumulative traded volume equals market order's. In case of partial execution, the rest of market order is cancelled.

3. Cancellation of owned limit order. Removes the order from the book.

At the end of trading period, all unmatched limit orders are automatically cancelled by the system. Execution of orders is not random, it starts with best price orders, continues with second-best ones and so on. In case of several orders with the same limit price, execution starts in order of detection by the system: smaller time stamp means higher priority.

ODM market has the following characteristics related to costs research:

Best prices: at any moment t there are best bid and best offer price, $P_{b,t}$ and $P_{a,t}$ correspondingly[4].

Market price is a numeric characteristic of an asset reflecting its fair value. Market price might differ for buy and sell deals or differ from best prices. It can also coincide for both types. Market value of a deal is a product of deal volume and its market price. Note that market price is not directly observable on the market, though its estimates are often postulated as market price. It is participant's estimate of the asset's fair price thus subjective in nature. Participants might have different estimates of market price, thus different estimates of implicit costs (see below); however, heterogeneity does not deny equilibrium and semi-strong efficiency of the market [Kyle, 1985], [Glosten and Milgrom, 1985].

Trading volume for a specified price interval is a cumulative volume of available orders with limit price inside the interval.

[4] We do not consider degenerate case of empty order book or one of its sides. In that case implicit transaction costs should be considered infinite.

Depth is a modulus of distance from best price. Each limit order has specific depth, while best orders have zero depth. Buy/Sell volume at depth D is cumulative trading volume of buy/sell limit orders with depth less or equal to D.

It is shown below that implicit costs do not depend on characteristics of specific orders inside the book, what really matters is distribution of trading volume as a function of depth for each book side. This leads to the study of ODM market and limit order book in terms of distribution functions. Academic works in this fields usually assume absolutely continuous distribution [Obizhaeva and Wang, 2013], [Alfonsi et al., 2010]. Such distributions do not match any real order book; however, empirical distribution can sometimes be fitted by simple absolute continuous function [Bouchaud et al., 2002]. This leads to the definition of generalized ODM-market.

By $F^+(t, D)$, we denote cumulative buy volume at time t and depth D, i. e. volume available for buying at prices within D distance from the best price; by analogy, $F^+(t, D)$ is cumulative sell volume. As in [Predoiu et al., 2011], consider general form of distributions of volume with $F^+(t, D), F^-(t, D)$ being càdlàg, non-negative non-decreasing functions of D for each t. For absolute continuous distributions, $q^+(t, p)$ and $q^-(t, p)$ denote corresponding densities.

Definition 3. *Generalized ODM market* $(\mathbb{R}_+, \bar{\mathbb{R}}_+, \mathbb{T})$ *is a single-asset order-driven market, where prices take values from* $\mathbb{R}_+$*, order volumes – from* $\bar{\mathbb{R}}_+ = \mathbb{R}_+ \cup +\infty$*, time values are from* $\mathbb{T}$*, and* $F^+(t, D), F^-(t, D)$ *are càdlàg, non-negative non-decreasing functions of* D *for each* t*.*

Definition 4. *Buy (Sell) side of the limit order book in generalized ODM market* $(\mathbb{R}_+, \bar{\mathbb{R}}_+, \mathbb{T})$ *has absolute liquidity level* V *at time* t *if*

$$F^+(t, 0) \geq V \quad (F^-(t, 0) \geq V).$$

Absolute liquidity level V means that trades up to volume V can be performed with no implicit costs.

Definition 5. *Buy (Sell) side of the limit order book in generalized ODM market* $(\mathbb{R}_+, \bar{\mathbb{R}}_+, \mathbb{T})$

is absolutely liquid at time t if

$$F^+(t,0) = +\infty \quad (F^-(t,0) = +\infty).$$

Absolute liquidity means that trades of any volume can be performed with no implicit costs. Limit order book is absolutely liquid at time t when its sides are absolutely liquid at t.

Definition 6. *Limit order book is absolutely liquid (has absolute liquidity level V) during interval $\mathcal{T} \subseteq \mathbb{T}$ if it is absolutely liquid (has absolute liquidity level V) for each $t \in \mathcal{T}$.*

It is clear from the definitions that absolute liquidity is unachievable in a real market due to required infinite volume at best positions. This, in turn, requires infinite volume of one of the best orders, or infinite number of orders which is impossible in reality. However, rather active market can be considered absolutely liquid by the investor if maximum volume of her deals is expected to be smaller then cumulative volume of best orders. This is true for modern markets, where best volumes are so big that a single participant cannot match it with a single deal, and after execution of the triggered trade sequence, the book is instantly replenished (see example in [Biais et al., 1995]). Nevertheless, even such markets have limited level of absolute liquidity, which can become noticeable during financial crises like Black Monday in 1987 or Flash Crash in 2010.

Definition 7. *Price response function is denoted by*

$$\rho(t,V) = \begin{cases} \inf\Big\{D \geq 0\colon F^-(t-,D) > |V|\Big\}, & V \geq 0, \\ \inf\Big\{D \geq 0\colon F^+(t-,D) > |V|\Big\}, & V < 0, \end{cases}$$

where $\inf\{\varnothing\} = +\infty$. $\rho(t,V)$ is the value of jump of best price due to execution of market order of volume V at time t. $V > 0$ means buy, $V < 0$ means sell.

Market value of a deal of volume V can be expressed as

$$C^*(t,V;P_a^*,P_b^*) = \begin{cases} P_a^* V, & V \geq 0, \\ P_b^* V, & V < 0, \end{cases} \tag{1}$$

where P_a^* and P_b^* are market buy and offer prices at time $t-$ correspondingly, with $P_{a,t}^* \leq P_{a,t}$, $P_{b,t}^* \geq P_{b,t}$. Real value of a deal can differ from market value and equals $C_r(t,V)$.

Definition 8. *Implicit transaction costs from a deal with volume V is denoted by*

$$C_I(t,V;P_a^*,P_b^*) = C_r(t,V) - C^*(t,V;P_a^*,P_b^*),$$

where

$$C_r(t,V) = \begin{cases} \int\limits_0^{\rho(t,V)} (P_{a,t-} + D)\, dF^-(t-,D-) + \\ \quad + \Big(V - F^-\left(t-,\rho(t,V)-\right)\Big)\left(P_{a,t-} + \rho(V)\right), & V \geq 0, \\ \int\limits_0^{\rho(V)} (P_{b,t-} - D)\, dF^+(t-,D-) + \\ \quad + \Big(V - F^+\left(t-,\rho(t,V)\right)\Big)\left(P_{b,t-} - \rho(V)\right), & V < 0. \end{cases}$$

Definition 9. *Explicit transaction costs $C_E(t,V)$ from a deal with volume V is a sum of taxes and brokerage commission for performing the deal on the market.*

We assume that for each t $C_E(t,V)$ is non-negative, non-decreasing convex function of V. Reason for postulating properties in the definition follows from economic interpretation and will be explained below. In this work, study of explicit costs is not performed because they are usually known to participants beforehand and rarely change. Besides, the shape of $C_E(t,V)$ is completely defined by brokerage services agreement and does not require modelling. The center of the research is implicit costs function because its shape depends solely on market microstructure and estimates of the participant; underestimation of implicit costs can lead to substantial unexpected losses on markets with low liquidity in case of sufficiently large deals.

Definition 10. *Total transaction costs (also referred to as total costs or transaction costs) on generalized ODM market $(\mathbb{R}_+, \bar{\mathbb{R}}_+, \mathbb{T})$ at time t is*

$$C(t,V;P_a^*,P_b^*) = C_E(t,V) + C_I(t,V;P_a^*,P_b^*).$$

Properties of transaction costs function on a generalized ODM market

Hereinafter, we consider only buy market orders and sell side of the book. But the results hold by analogy for the opposite side as well. We also omit the case of infinite costs and volumes. Consider total costs function

$$C(t,V;P_a^*,P_b^*) = C_E(t,V) + C_I(t,V;P_a^*,P_b^*)$$

for $V > 0$ on generalized ODM market $(\mathbb{R}_+, \mathbb{R}_+, \mathbb{T})$. Let $F(t,D)$ be the distribution function of volume. Denote $C(t,V;P_{a,t},P_{b,t}) = C(t,V)$ henceforth for simplicity of notation. It follows from Definition 8 that

$$C(t,V;P_{a,t}^*,P_{b,t}^*) = \begin{cases} C(t,V) + (P_{a,t} - P_{a,t}^*)V, & V \geq 0, \\ C(t,V) + (P_{b,t}^* - P_{b,t})V, & V < 0. \end{cases} \tag{2}$$

To simplify notation, dependence on t and $P_{b,t}^*$ is henceforth omitted because only sell side snapshot is of interest.

Lemma 1. *For any absolute continuous distributions $F^-(D), F^+(D)$, it is true for any V, $\rho(V) < +\infty$, that*

$$C_I(V) = \begin{cases} \int\limits_0^{\rho(V)} Dq^-(P_a + D)\,dD, & V \geq 0, \\ \\ -\int\limits_0^{\rho(V)} Dq^+(P_b - D)\,dD, & V < 0. \end{cases} \tag{3}$$

Proof. 1. For absolute continuous $F^-(D), F^+(D)$ we have:

$$F^+(D) = \int\limits_{\max\{P_b - D, 0\}}^{+\infty} q^+(p)\,dp = \int\limits_{\max\{P_b - D, 0\}}^{P_b} q^+(p)\,dp = \Big\{p = P_b - y\Big\} =$$

$$= -\int\limits_{\min\{P_b, D\}}^{0} q^+(P_b - y)\,dy = \int\limits_0^{\min\{P_b, D\}} q^+(P_b - y)\,dy. \tag{4}$$

$$F^-(D) = \int\limits_0^{P_a+D} q^-(p)\,dp = \int\limits_{P_a}^{P_a+D} q^-(p)\,dp = \Big\{p = P_a + y\Big\} =$$

$$= \int\limits_0^{D} q^-(P_a + y)\,dy. \tag{5}$$

2. If $\rho(V) < +\infty$ then, by definition of response function, we have:

$$\begin{cases} F^+(\rho(V)) = V, & V < 0, \\ F^-(\rho(V)) = V, & V \geq 0 \end{cases} \tag{6}$$

3. Substituting (6) into implicit costs definition, (4) and (5) yield:

   for $V \geq 0$,

$$C_I(V) = \int\limits_0^{\rho(V)} (P_a + D)\,dF^-(D) - P_a V =$$

$$= \int\limits_0^{\rho(V)} Dq^-(P_a + D)\,dD + P_a \int\limits_0^{\rho(V)} q^-(P_a + D)\,dD - P_a V =$$

$$= \int\limits_0^{\rho(V)} Dq^-(P_a + D)\,dD + P_a F^-(\rho(V)) - P_a V = \int\limits_0^{\rho(V)} Dq^-(P_a + D)\,dD;$$

   for $V < 0$,

$$C_I(V) = \int\limits_0^{\rho(V)} (P_b - D)\,dF^+(D) - P_b V =$$

$$= -\int\limits_0^{\rho(V)} Dq^+(P_b - D)\,dD + P_b \int\limits_0^{\rho(V)} q^+(P_b - D)\,dD - P_b V =$$

$$= -\int\limits_0^{\rho(V)} Dq^+(P_b - D)\,dD + P_b F^+(\rho(V)) - P_b V = -\int\limits_0^{\rho(V)} Dq^+(P_b - D)\,dD.$$

$\square$

Transaction costs function must have the following properties according to its economic interpretation:

1. Cost of buying is not less than market value of the deal.

2. Deal of larger volume incurs larger costs. This property becomes obvious since deal with volume $V_2 > V_1$ can be decomposed into two consecutive deals with volumes V_1 and $V_2 - V_1$, with implicit costs from both parts being non-negative due to property 1.

3. In generalized ODM market $(\mathbb{R}_+, \mathbb{R}_+, \mathbb{T})$ infinitesimal increase of a deal volume incurs infinitesimal increase of costs, i. e. $C(t, V)$ is continuous in V.

4. Unit costs $\frac{C(t,V)-C(t,0)}{V}$ are increasing in V. This follows from the automatic system's attempt to match limits order in order of limit price which leads to a gradual increase in costs per unit when worse orders are matched, thus increasing unit costs for the whole deal.

Consistency of introduced definition of implicit costs with aforementioned properties is proven in

Theorem 1. *If for any distribution $F(D)$*

$$C(V) = C_E(V) + \int_0^{\rho(V)} (P_a + D)\, dF(D-) + \\ + \Big(V - F\left(\rho(V)-\right)\Big)\left(P_a + \rho(V)\right) - P_a V, \quad (7)$$

$$\rho(V) = \inf\Big\{D \geq 0\colon F(D) > V\Big\}, \quad V \geq 0,$$

then

1. $C(V) \geq 0\ \forall V \geq 0$;

2. $C(V_1) \leq C(V_2)$ *if* $V_1 \leq V_2$;

3. $C(V)$ *is convex in* V;

4. $\frac{C(V_1)-C(0)}{V_1} \leq \frac{C(V_2)-C(0)}{V_2}$ *if* $V_1 \leq V_2$.

Proof. Hereinafter, the following statement is used ([Revuz and Yor, 1999, proposal 4.5]): let $f(D), g(D)$ be functions of bounded variation on $\mathbb{R}_+$. Then

$$\int_0^D f(s-)dg(s) + \int_0^D g(s)df(s) = f(D)g(D) - f(0)g(0).$$

In particular, for $f(D) \equiv 1$, $g(D) = F(D-)$ we have $\int\limits_0^{\rho(V)} dF(D-) = F(\rho(V)-) - F(0-) = F(\rho(V)-)$.

1) To prove the first statement, we use definition of $\rho(V)$, along with càdlàg and non-decreasing properties of $F(D)$, to see that $F(\rho(V)-) \leq F(\rho(V)) \leq V$ for all $V > 0$. Then

$$C(V) \geq \int\limits_0^{\rho(V)} (P_a + D)\, dF(D-) + V\rho(V) - F\left(\rho(V)-\right)\left(P_a + \rho(V)\right) \geq$$

$$\geq P_a \int\limits_0^{\rho(V)} dF(D-) + V\rho(V) - F\left(\rho(V)-\right) P_a - F\left(\rho(V)-\right)\rho(V) =$$

$$= P_a F(\rho(V)-) + V\rho(V) - F\left(\rho(V)-\right) P_a - F\left(\rho(V)-\right)\rho(V) \geq$$

$$\geq \left(V - F(\rho(V)-)\right)\rho(V) \geq 0.$$

2) The second statement is proven similarly. Denote $\rho(V_1) = \rho_1$, $\rho(V_2) = \rho_2$. Since $\Big\{D \geq 0\colon F(D) > V_1\Big\} \supseteq \Big\{D \geq 0\colon F(D) > V_2\Big\}$, $\rho_1 \leq \rho_2$ and we get

$$C(V_2) - C(V_1) = C_E(V_2) - C_E(V_1) + \int\limits_{\rho_1}^{\rho_2} (P_a + D)\, dF(D-) + V_2\rho_2 - V_1\rho_1 -$$

$$- F(\rho_2-)(P_a + \rho_2) + F(\rho_1-)(P_a + \rho_1) \geq$$

$$\geq (P_a + \rho_1)(F(\rho_2-) - F(\rho_1-)) + V_2\rho_2 - V_1\rho_1 - F(\rho_2-)(P_a + \rho_2) + F(\rho_1-)(P_a + \rho_1) \geq$$

$$\geq -F(\rho_2-)(\rho_2 - \rho_1) + V_2\rho_2 - V_1\rho_1 = -F(\rho_2-)(\rho_2 - \rho_1) + V_2(\rho_2 - \rho_1) + \rho_1(V_2 - V_1) =$$

$$= (V_2 - F(\rho_2-))(\rho_2 - \rho_1) + \rho_1(V_2 - V_1) \geq 0.$$

3) Continuity is derived form convexity of $C(V)$. For proof of convexity of implicit costs for general càdlàg distribution, see [Predoiu et al., 2011]. Here, we demonstrate the property for absolute continuous distribution with density $q(p) > 0$ where $p \geq P_a$, $\rho \in \mathcal{C}^1$, $C_I \in \mathcal{C}^2$. By definition,

$$\int\limits_0^{\rho(V)} q(P_a + D) dD \equiv V \Longrightarrow \rho'(V) q(P_a + \rho) \equiv 1.$$

Lemma 1 implies

$$C_I(V) = \int\limits_0^{\rho(V)} Dq(P_a + D)dD \Longrightarrow$$

$$\Longrightarrow C_I'(V) = \rho'(V)\rho(V)q(P_a + \rho) = \rho(V) \Longrightarrow$$

$$\Longrightarrow C_I''(V) = \rho'(V) = \frac{1}{q(P_a + \rho)} > 0.$$

Since $C_E(V)$ is postulated convex, $C(V)$ is also convex.

4) Since the case of $V_2 = 0$ is trivial, assume $V_2 > 0$. Denote $\alpha = \frac{V_1}{V_2} \leq 1$. $(1-\alpha)\cdot 0 + \alpha \cdot V_2 = V_1$, thus convexity of $C(V)$ implies $C(V_1) \leq (1-\alpha)C(0) + \alpha C(V_2)$. Then

$$\frac{C(V_2) - C(0)}{V_2} - \frac{C(V_1) - C(0)}{V_1} = \frac{V_1 C(V_2) - V_2 C(V_1) - V_1 C(0) + V_2 C(0)}{V_1 V_2} \geq$$

$$\geq \frac{V_1 C(V_2) - V_2((1-\alpha)C(0) + \alpha C(V_2)) - V_1 C(0) + V_2 C(0)}{V_1 V_2} =$$

$$= \frac{(V_1 - \alpha V_2)C(V_2) + (\alpha V_2 - V_1)C(0)}{V_1 V_2} = 0.$$

□

Corollary 1. *Theorem 1 holds for $C(V; P_a^*)$ for any $P_a^* \leq P_a$.*

Proof. Since the statements hold for $C(V; P_a)$, then (2) implies 1) и 2) due to $P_a^* \leq P_a$; 3) holds since $C(V; P_a^*)$ is a sum of convex and linear function, hence, convex; 4) follows from $\frac{C(V;P_a^*)}{V} = \frac{C(V;P_a)}{V} + \text{const}$. □

Note that properties in Theorem 1 follow from the definition for C_I, while postulated for C_E (see Definition 9), since study of explicit costs is out the scope of the work due to aforementioned reasons. However, explicit costs should satisfy the same properties due to economic sense.

Results below shed light on the possibility of popular shapes $C(V) \equiv \text{const}$ and $C(V) = aV + C_f$.

Theorem 2. *For any distribution $F(D)$, the following statements are equivalent:*

1. *$C_I(V) \equiv 0$, where $0 \leq V < V^*$;*

2. *sell side of the limit order book has absolute liquidity level V^*.*

Proof. According to definition, sell side with absolute liquidity level V^* means $F(0) \geq V^*$.

$2 \Rightarrow 1$ is obtained directly from (7) since $F(0-) = 0$, $\rho(V) = 0$ for all $V < V^*$.

$1 \Rightarrow 2$ is proven by contradiction. Let $F(0) = V_0 < V^*$. Consider two possible cases:

1. $\lim\limits_{D \to +\infty} F(D) = \bar{V} < V^*$. Then $\rho(\frac{\bar{V}+V^*}{2}) = +\infty$ and $C(\frac{\bar{V}+V^*}{2}) = +\infty$, which contradicts the main assumption since $\frac{\bar{V}+V^*}{2} < V^*$.

2. $\lim\limits_{D \to +\infty} F(D) \geq V^*$. Denote $V' = \frac{V_0+V^*}{2} > V_0$. Since $F(D)$ is right-continuous, there exists $\varepsilon > 0$: $F(D) < V'$ for $0 \leq D < \varepsilon$. Then $\rho' = \rho(V') \geq \varepsilon > 0$ and

$$C_I(V') = \int\limits_0^{\rho'} (P_a + D) dF(D-) + V'\rho' - F(\rho'-)(P_a + \rho') =$$

$$= \int\limits_0^{\rho'} DdF(D-) + (V'\rho' - F(\rho'-))\rho' \geq$$

$$\geq \max\Big\{ \int\limits_0^{\rho'} DdF(D-), (V'\rho' - F(\rho'-))\rho' \Big\}.$$

If $F(D) \equiv V_0$ for all $D < \rho'$, then $V' - F(\rho'-) = V' - V_0 > 0$. Otherwise, $\int\limits_0^{\rho'} DdF(D-) > 0$ since $F(D-)$ is non-decreasing thus cannot have a discontinuity at zero. Hence, $C_I(V') > 0$ which leads to contradiction.

□

Lemma 2. *Assume* $0 < D_1 < D_2$ *such that* $0 < F(D_1) = V_1 < F(D_2) = V_2$. *Then* $C_I(V)$ *cannot be a linear function of* V *in the interval* $0 \leq V \leq V_2$.

Proof. Denote $\rho_1 = \rho(V_1)$, $\rho_2 = \rho(V_2)$, $F_1 = F(\rho(V_1)-)$, $F_2 = F(\rho(V_2)-)$. Since $F(D)$ is right-continuous, there exists $\varepsilon > 0$: $F(D) < V_2$ for $D \in [D_1, D_1 + \varepsilon)$. Hence, $\rho_2 \geq \rho_1 + \varepsilon$. Due to $C_I(0) = 0$ it is sufficient to show that $\frac{C_I(V_2) - C_I(V_1)}{V_2 - V_1} > \frac{C_I(V_1)}{V_1}$. Indeed,

$$C_I(V_2) = \int\limits_0^{\rho_2} (P_a + D) dF(D-) + V_2\rho_2 - F_2(P_a + \rho_2) =$$

$$= \int_0^{\rho_1} (P_a + D)dF(D-) + V_1\rho_1 - F_1(P_a + \rho_1) +$$

$$+ \int_{\rho_1}^{\rho_2} (P_a + D)dF(D-) + V_2\rho_2 - V_1\rho_1 + F_1(P_a + \rho_1) - F_2(P_a + \rho_2) =$$

$$= C_I(V_1) + \int_{\rho_1}^{\rho_2} (P_a + D)dF(D-) + V_2\rho_2 - V_1\rho_1 + F_1(P_a + \rho_1) - F_2(P_a + \rho_2).$$

1) Let $F(D) \equiv V_1$ for $D < D_2$. Then $V_2 > F_2 = V_1$ and

$$C_I(V_2) \geq C_I(V_1) + (P_a + \rho_1)(F_2 - F_1) +$$

$$+ V_2\rho_2 - V_1\rho_1 - (F_2 - F_1)P_a - F_2\rho_2 + F_1\rho_1 =$$

$$= C_I(V_1) + V_2\rho_2 - V_1\rho_1 - F_2(\rho_2 - \rho_1) = C_I(V_1) + (V_2 - V_1)\rho_1 + (V_2 - F_2)(\rho_2 - \rho_1).$$

Hence,

$$\frac{C_I(V_2) - C_I(V_1)}{V_2 - V_1} \geq \rho_1 + \frac{(V_2 - F_2)(\rho_2 - \rho_1)}{V_2 - V_1} > \rho_1.$$

2) Assume $F(D)$ is not identically constant for $D < D_2$. Since $F(D)$ is non-decreasing, $\int_{\rho_1}^{\rho_2} (P_a + D)dF(D-) > P_a(F_2 - F_1)$, and we similarly obtain

$$C_I(V_2) > C_I(V_1) + (P_a + \rho_1)(F_2 - F_1) +$$

$$+ V_2\rho_2 - V_1\rho_1 - (F_2 - F_1)P_a - F_2\rho_2 + F_1\rho_1 =$$

$$= C_I(V_1) + (V_2 - V_1)\rho_1 + (V_2 - F_2)(\rho_2 - \rho_1).$$

Therefore

$$\frac{C_I(V_2) - C_I(V_1)}{V_2 - V_1} > \rho_1 + \frac{(V_2 - F_2)(\rho_2 - \rho_1)}{V_2 - V_1} \geq \rho_1.$$

In both cases we have $\frac{C_I(V_2) - C_I(V_1)}{V_2 - V_1} > \rho_1$.

Definition of $C_I(V)$ yields

$$C_I(V) = \int_0^{\rho} (P_a + D)dF(D-) + V\rho - F(\rho-)(P_a + \rho) \leq$$

$$\leq \int_0^{\rho} (P_a + \rho) dF(D-) + V\rho - F(\rho-)(P_a + \rho) = V\rho.$$

Therefore $\frac{C_I(V_2) - C_I(V_1)}{V_2 - V_1} > \rho_1 \geq \frac{C_I(V_1)}{V_1}$.

$\square$

Theorem 3. *For any distribution $F(D)$ the following statements are equivalent:*

1. *$C_I(V)$ is linear for $0 \leq V < V^*$;*

2. *sell side of the limit order book has absolute liquidity level V^*;*

3. *$C_I(V) \equiv 0$ for $0 \leq V < V^*$.*

Proof. $2 \Leftrightarrow 3$ is proven in Theorem 2.

$2 \Rightarrow 1$ since $2 \Rightarrow 3$, and constant is a special case of linear function.

$1 \Rightarrow 2$: continuity of $C_I(V)$ and linearity for $0 \leq V < V^*$ yield linearity for $0 \leq V \leq V^*$. Then Lemma 2 implies that $F(D)$ cannot accept more than one value in $(0, V^*]$. Hence, there are two possible choices for $F(D)$:

1. $F(0) \geq V^*$ which is equivalent to 2).

2. $F(D) = \begin{cases} 0, D < D^*, \\ \geq V^*, D \geq D^*. \end{cases}$ This behavior contradicts definition of $F(D)$ since if sell side is not empty than at least one order is denoted the best, therefore, $F(0) > 0$. This completes the proof.

$\square$

Theorem 3 implies that linear implicit costs model cannot be used if market price coincides with best quoted price. The only exception is zero implicit costs for absolutely liquid book side. Should market and best prices differ, linearity is possible:

Theorem 4. *For any distribution $F(D)$ the following statements are equivalent:*

1. *$C_I(V; P_a^*)$ is linear in V for $0 \leq V < V^*$;*

2. *sell side of the limit order book has absolute liquidity level V^*;*

3. *$C_I(V; P_a^*) = V(P_a - P_a^*)$ for $0 \leq V < V^*$.*

Proof. Proof is obvious from Theorem 3 applied to $C_I(V; P_a)$, and formula (2). □

Corollary 2. *For any distribution $F(D)$, if $C(V; P_a^*)$ is linear in V for $0 \leq V < V^*$ then sell side of the limit order book has absolute liquidity level V^*.*

Proof. Since $C(V; P_a^*) = C_E(V) + C_I(V; P_a^*)$, we use convexity and non-decreasing property of $C_E(V)$, $C_I(V; P_a^*)$ to derive that linearity of $C(V; P_a^*)$ is equivalent to linearity of both $C_E(V)$ and $C_I(V; P_a^*)$ on the same interval. Thus, by virtue of Theorem 4, we derive the main statement. □

Theorem 4 implies that linear approximation of implicit costs function $C_I(V)$ does not match any càdlàg distribution function, hence, no real non-empty limit order book. In this regard, a more appropriate parametric form of $C_I(V)$ is required to asses costs of large deals adequately. It is noteworthy that linear form might be appropriate if $C_I(V)$ can be *considered* linear within allowed tolerance, for example, for small-volume deals. For more substantial deals that match in-depth limit orders, costs can be underestimated due to linearity. This fact has been noticed, for example, in [Андреев et al., 2011] for high-frequency MICEX data: polynomial interpolation $C_I(t, V) = a(t)V^2 + b(t)V^3$ was introduced since it contained a minimal set of parameters while providing acceptable fitting.

Transaction costs function on MICEX stock market

In this section, we introduce a parametric form of the implicit costs function based on empirical properties of MICEX data for the blue chip stocks of Lukoil (LKOH), Rostelecom (RTKM) and Gazprom (GAZP) through the period of 2006-2007. The chosen stocks were among the most actively traded during the period (judging by average number of market events per trading day) and provide particularly good statistics. Empirical studies show that in most cases the distribution is quite well fitted by a scaled gamma-distribution function which coincides with the findings of [Bouchaud et al., 2002] for Paris Bourse. This is an example of how empirical function can be approximated by a well-known absolute continuous distribution. Fig. 1 - 4 demonstrate fitting results for LKOH and RTKM. Figures show good fitting results, with precision decreasing for large depth values. This is systematic error which occurs due to the in-depth "inactive" limit orders — orders which will be matched with very small probability but at great price for the seller. A participant places such orders to make use of a sudden market movement or an arrival of extremely large market order (for example, during restructures of large portfolios). Should considerable volume at better prices be matched, some part of the order might reach the “inactive area”. We do not consider “inactive” area in this work, aiming for average-sized portfolio and assuming that market order’s trades never reach such depth. Besides, placement of “inactive” orders is a rare event on MICEX which renders useless any statistical methods for estimation.

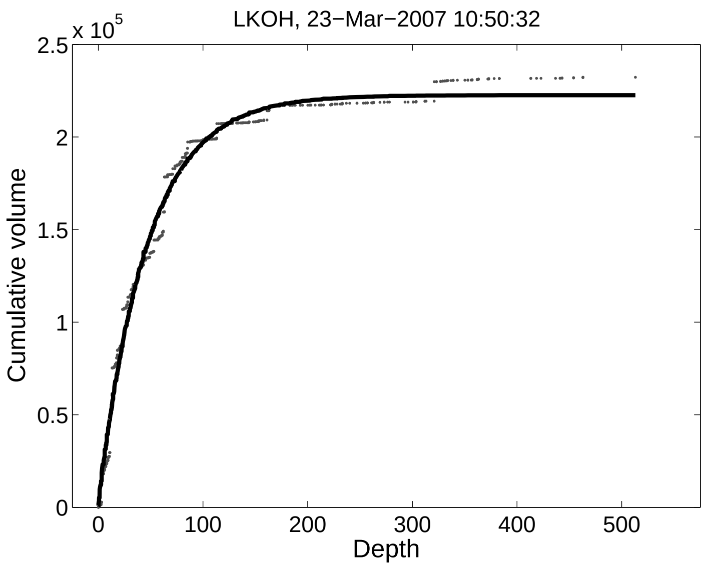
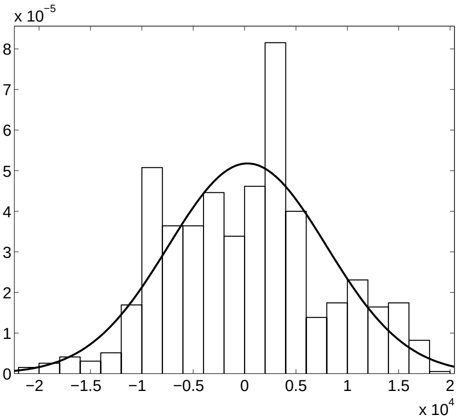
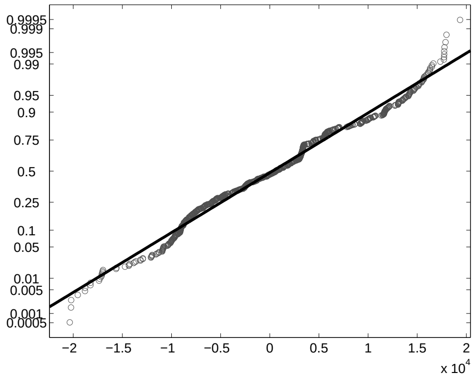


Figure 1: Approximation of trading volume distribution function by a scaled gamma-distribution function for LKOH. Dots denote empirical values, solid line denotes approximation curve.

To derive costs function in this case, denote distribution function of the sell side $F(D) = V_{max}F_\gamma(D)$ where $V_{max} < +\infty$ is available sell volume on the market, and $F_\gamma(D)$ is gamma-



(a) Distribution of residuals and normal approximation (solid line).



(b) Distribution of residuals on normal quantile plot. Dots denote quantiles of empirical distribution.

Figure 2: Statistical properties of residuals during approximation of LKOH trading volume distribution.

distribution function. Denote gamma-density

$$q_\gamma(D) = \frac{1}{\Gamma(k)\theta^k} D^{k-1} e^{-\frac{D}{\theta}}, \quad k > 0,\, \theta > 0.$$

By definition, $q(P_a + D) \overset{\text{a.s.}}{=} q_\gamma(D)$, $q(D) > 0$ where $D > 0$. For absolute continuous distribution we have

$$\int\limits_0^{\rho(V)} q_\gamma(D) dD \equiv V, \quad V \geq 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \rho(V) \equiv \frac{1}{q_\gamma(\rho(V))};$$

$$\rho''(V) q_\gamma(\rho) + {\rho'}^2(V) q'_\gamma(\rho) \equiv 0, \quad V \geq 0. \tag{8}$$

Lemma 1 yields

$$C_I(V) = \int\limits_0^{\rho(V)} D q_\gamma(D)\, dD \quad \Rightarrow$$

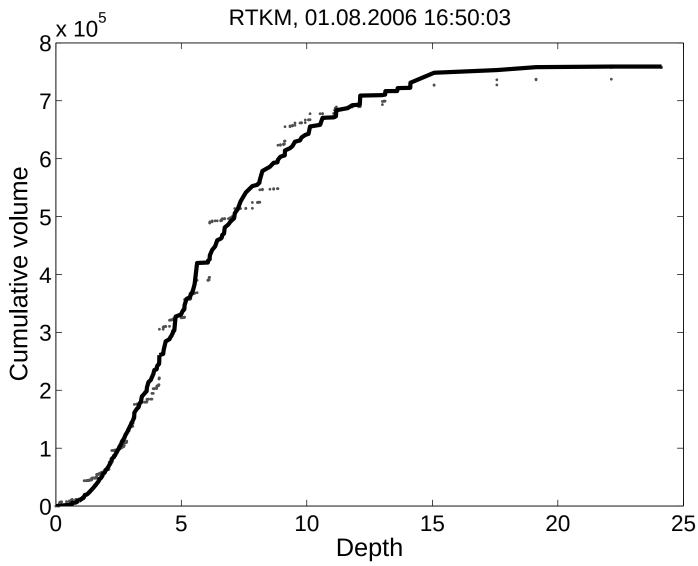


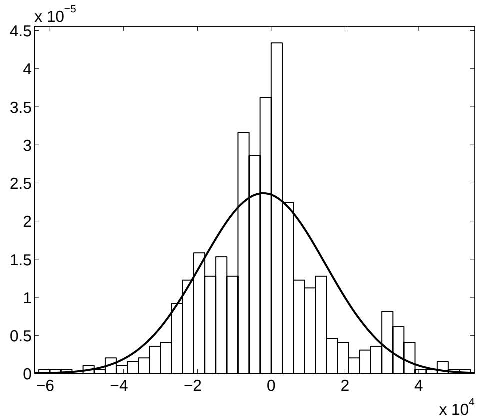
Figure 3: Approximation of trading volume distribution function by a scaled gamma-distribution function for RTKM. Dots denote empirical values, solid line denotes approximation curve.

$$\Rightarrow \quad C_I'(V) = \rho'(V)\rho(V)q_\gamma(\rho(V)) = \rho(V). \tag{9}$$

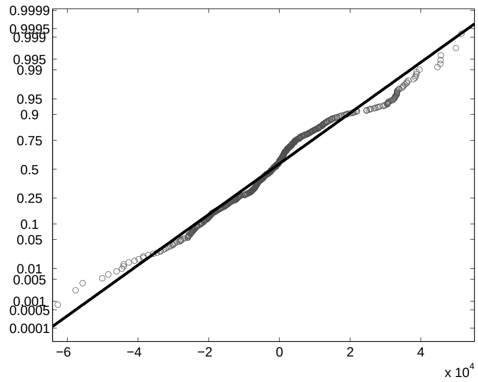
The study of cost function properties can be conducted through the properties of $\rho(V)$. In view of (8), we can see that $\rho''(V) > 0$ is equivalent to $q_\gamma(\rho(V)) < 0$, and $\rho''(V) = 0$ is equivalent to $q_\gamma(\rho(V)) = 0$. Since

$$q'(\rho) = \frac{V_{max}}{\Gamma(k)\theta^k}\left[(k-1)\rho^{k-2} - \frac{1}{\theta}\rho^{k-1}\right]e^{-\frac{\rho}{\theta}},$$

for all $k \leq 1$ we have $q' < 0$ for all $D > 0$, therefore, $\rho(V)$ is strictly monotonically increasing



(a) Distribution of residuals and normal approximation (solid line).



(b) Distribution of residuals on normal quantile plot. Dots denote quantiles of empirical distribution.

Figure 4: Statistical properties of residuals during approximation of RTKM trading volume distribution.

with no inflection points. For $k > 1$

$$q'_\gamma(\rho(V^*)) = 0 \quad \Leftrightarrow \quad \rho(V^*) = (k-1)\theta.$$

$$\theta\rho(V) = F_\gamma^{-1}(\frac{V}{V_{max}}) \quad \Rightarrow \quad \frac{V^*}{V_{max}} = F_\gamma((k-1)\theta).$$

Thus, $\rho(V)$ has exactly one inflection point. Therefore, polynomial approximation of $\rho(V)$ on MICEX is adequate to the data only for polynomials of third or higher degree, while (9) implies forth or higher degree polynomials for $C_I(V)$. By virtue of (9), we obtain feasible parametric form of total cost function on MICEX stock market:

$$C(V; P_a^*) = C_E(V) + V(P_a - P_a^*) + \int_0^V F_\gamma^{-1}(\frac{V}{V_{max}}; \theta, k)\, dV. \tag{10}$$

As shown in previous chapter, cost function can be considered both non-zero and linear only in the market with absolutely liquidity level greater than the market order's size. In practice, one can use linear approximation if underestimation of costs due to incorrect model is negligible. We illustrate this for MICEX market with cost function (10). Denote market price by P_a^*. Assume that investment strategy neglects implicit costs for volume V if they do not exceed $\alpha P_a^* V$, $0 < \alpha < 1$, i. e. predefined fraction of the deal's market value. Denote by V^* a non-zero solution of

$$\int_0^V F_\gamma^{-1}(\frac{V}{V_{max}}; \theta, k)\, dV = \alpha P_a^* V. \tag{11}$$

Since $F_\gamma^{-1}(0; \theta, k) = 0$ and continuous, then there is V' such that

$$\int_0^V F_\gamma^{-1}(\frac{V}{V_{max}}; \theta, k)\, dV \leq \alpha P_a^* V \text{ for } V \leq V'.$$

Should the inequality hold for all $V \leq V_{max}$, non-linear part in (10) is negligible according to strategy's policy, and cost function can be considered linear. If the inequality is violated, there is $V^* > 0$ — minimal non-zero solution of (11). Therefore, costs function can be considered linear for the deal with volume $V \leq V^*$. For example, for $\theta = 4$, $k = 2$, $P_a^* = 1$, $V_{max} = 100,000$, $\alpha = 0.02$, we have $V^* \approx 678$ — non-linear part of total costs value does not overcome 2% of market value for deals with this volume or less. Note that the result depends on the accuracy of estimates for θ and k.

Conclusion

The collapse of LTCM in 1998 and financial crisis of 2008 drew attention to the problem of liquidity risk in portfolio management, defined via transaction costs as a function of deal volume. Many contemporary works in optimal portfolio selection and management theory assume linearity of the cost function due to both the possibility of a closed-form solution in many theoretical frameworks and estimation of parameters by common econometric methods during practical use. While justified for explicit costs, for implicit part, which depends on market properties, the assumption needs verification or theoretical explanation. We show that linearity of costs is equivalent to the introduced form of absolute liquidity thus valid only for small deal volumes. Meanwhile, assumption of linearity for any volume does not match any real limit order book and can yield severe underestimation of liquidity risk. For MICEX stock market, we present a more adequate form of transaction costs function and provide an example of estimating an upper bound of deal's volume for which costs can be considered linear within tolerance.

References

[Андреев et al., 2011] Андреев, Н., Лапшин, В., Науменко, В., and Смирнов, С. (2011). Определение ликвидационной стоимости портфеля акций с учетом особенностей микроструктуры рынка (на примере ММВБ). *Управление риском*, 2(58):35–53.

[Alfonsi et al., 2010] Alfonsi, A., Fruth, A., and Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157.

[Biais et al., 1995] Biais, B., Hillion, P., and Spatt, C. (1995). An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse. *The Journal of Finance*, 50(5):1655–1689.

[Bouchaud et al., 2002] Bouchaud, J., Mézard, M., and Potters, M. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256.

[Coase, 1937] Coase, R. (1937). The nature of the firm. *Economica*, 4(16):386–405.

[Constantinides, 1979] Constantinides, G. (1979). Multiperiod consumption and investment behavior with convex transactions costs. *Management Science*, 25(11):1127–1137.

[Constantinides, 1986] Constantinides, G. (1986). Capital market equilibrium with transaction costs. *The Journal of Political Economy*, 94(4):842–862.

[Davis and Norman, 1990] Davis, M. H. A. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):676–713.

[Framstad et al., 2001] Framstad, N., Øksendal, B., and Sulem, A. (2001). Optimal consumption and portfolio in a jump diffusion market with proportional transaction costs. *Journal of Mathematical Economics*, 35(2):233–257.

[Glosten and Milgrom, 1985] Glosten, L. and Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14:71–100.

[Jorion, 2000] Jorion, P. (2000). Risk management lessons from Long-Term Capital Management. *European financial management*, 6(3):277–300.

[Keim and Madhavan, 1998] Keim, D. B. and Madhavan, A. (1998). The Cost of Institutional Equity Trades. *Financial Analysts Journal*, 54(4):50–69.

[Klaes, 2000] Klaes, M. (2000). The history of the concept of transaction costs: neglected aspects. *Journal of the History of Economic Thought*, 22(2):191–216.

[Kyle, 1985] Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 53(6):1315–1335.

[Ma et al., 2013] Ma, J., Song, Q., Xu, J., and Zhang, J. (2013). Optimal Portfolio Selection Under Concave Price Impact. *Applied Mathematics & Optimization*, 67(3):353–390.

[Magill and Constantinides, 1976] Magill, M. and Constantinides, G. (1976). Portfolio selection and transactions costs. *Journal of Economic Theory*, 13(2):245–263.

[Marschak, 1950] Marschak, J. (1950). The Rationale of the Demand for Money and of “Money Illusion”. *Metroeconomica*, 2(2):71–100.

[Ø ksendal and Sulem, 2010] Ø ksendal, B. and Sulem, A. (2010). Singular stochastic control and optimal stopping with partial information of jump diffusions. *preprint*, pages 1–31.

[Obizhaeva and Wang, 2013] Obizhaeva, A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32.

[Predoiu et al., 2011] Predoiu, S., Shaikhet, G., and Shreve, S. (2011). Optimal execution in a general one-sided limit-order book. *SIAM Journal on Financial Mathematics*, 2(1):183–212.

[Revuz and Yor, 1999] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*. Springer, Berlin, 3 edition.

[Shreve and Soner, 1994] Shreve, S. and Soner, H. (1994). Optimal investment and consumption with transaction costs. *The Annals of Applied Probability*, pages 609–692.

[Vath et al., 2007] Vath, V. L., Mnif, M., and Pham, H. (2007). A model of optimal portfolio selection under liquidity risk and price impact. *Finance and Stochastics*, 11(1):51–90.

Author:

Nikolay Andreev, National Research University Higher School of Economics (Moscow, Russia). Laboratory for Financial Engineering and Risk Management. Researcher;

E-mail: nandreev@hse.ru

Author prefers to publish this text without proofreading.