MODELING THE BEHAVIOR OF PARTICIPANTS ON WEBMONEY EXCHANGER MARKET

by

Igor Lytvynenko

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis

Kyiv School of Economics

2013

Thesis Supervisor ______________________ Rostyslav Hryniv

Approved by ______________________

Head of the KSE Defense Committee, Professor Wolfram Schrettl

____

____

____

Date ______________________

Kyiv School of Economics

Abstract

MODELING THE BEHAVIOR OF PARTICIPANTS ON WEBMONEY EXCHANGER MARKET

by Igor Lytvynenko

Thesis Supervisor Rostyslav Hryniv

This study proposes an approach to evaluating and comparing trader strategies on exchange market that uses Central Limit Order Book mechanism of clearing orders. The methodology is based on Monte Carlo simulations of the market with the injected trader, who is implementing some strategy of exchange. Simulation results summarize historical information available from the market and extend trader's information set.

All the numerical analysis is based on the data collected for WebMoney Exchanger exchange market. It was shown that nonhomogeneous Poisson process adequately models order flows. In addition to that, significant dependence between order sizes, rates and current market state have been observed in the data. As a result, using the revealed facts, simulation model of the WebMoney Exchanger market was built for one exchange pair (WMR$\leftrightarrow$WMZ). This simulation as an example was traced throughout the work. Comparing outcome distributions among different strategies, analyst can prefer one strategy over another. Such a possibility can be useful for developing and testing strategies for algorithmic trading, which is becoming more and more popular way to work with the markets nowadays.

Table of contents

List of figures

List of tables

Glossary

Central Limit Order Book (CLOB) is a trading method that matches customer orders (buy and sell ones) on a "price time priority" basis. Customer is free to put a limit order into the book and wait or put a market order and match the best limit order already available in the book.

Limit order is an order to buy or sell at a specific price or better. A buy limit order can only be executed at the limit price or lower, and a sell limit order can only be executed at the limit price or higher. A limit order is not guaranteed to be executed.

Market order is an order to buy or sell immediately at the best available price. The price at which a market order will be executed is not guaranteed.

WMU is an equivalent for UAH issued by WebMoney payment system.

WMR is an equivalent for RUR issued by WebMoney payment system.

WMZ is an equivalent for USD issued by WebMoney payment system.

Chapter 1

Introduction

This research proposes an approach to evaluating and comparing trader strategies on exchange market that uses central limit order book mechanism of clearing orders. The methodology is based on Monte Carlo simulations of the market with the injected trader, who is implementing some strategy of exchange. These simulations give us distribution of the amount of destination currency that can be received after the exchange. Such result summarizes historical information available from the market and extends trader's information set. Comparing distributions among different strategies, analyst can prefer one strategy over another. Such a possibility can be useful for developing and testing strategies for algorithmic trading, which is becoming more and more popular way to work with the markets nowadays.

All the numerical analysis is based on the data collected for WebMoney Exchanger exchange market. This is relatively small but alive[1] exchange market. One can use this market to exchange currencies issued by WebMoney[2] payment system. The results of this work can help to make transactions in a reasonably efficient way without continuous manual monitoring of the market.

The way this market works is very similar to a stock exchange markets. Central

[1]For eye-ball evaluation of its size consider the following statistics: 2.4 million WMZ were exchanged in pair WMZ-WMR, and 0.24 million WMZ in pair WMZ-WMU for one working day (05.12.2012, source: `https://wm.exchanger.ru`).

[2]WebMoney is an internet payment system widely known among internet users of post Soviet Union countries. For almost 15 years already this system has been issuing its virtual currencies, which are almost freely exchangeable with their originals. For example, it issues WMZ as a surrogate of US dollar, WMU as a surrogate of Ukrainian hryvnia. Users of this system can make payments in these currencies through the internet. They can buy goods in internet shops or order some specific services (hosting, domain registration, advertisement and other). This system is free to use, but charges transaction costs. WebMoney official statistics reports that during 2011 the system has processed 130 million transactions with total amount up to 13 billion USD (source: `http://www.webmoney.ru/eng/about/statistics/stat_years.shtml`).

Limit Order Book mechanism with different modifications is used at the NYSE, Nasdaq, Tokyo Stock Exchange, etc.

This mechanism works as follows. Suppose one wants to exchange WMU[3] to WMZ[4]. In order to do this he can:

- Buy one or several existing orders on the market and make this exchange immediately (put market order).
- Put his own order with desirable exchange rate and wait until somebody buys it (put limit order).
- Change exchange rate of the existing limit order.
- Buy one or several limit orders from the opposite side of the market for own limit order (i.e. convert limit order to market order).

WebMoney Exchanger does not charge any commission, but WebMoney takes 0.8% commission for every transaction itself. And it is paid only once for exchange, regardless the strategy used. So, in this case transaction cost have no influence on trade's decision. That is why, in this research we can assume that there are no transaction costs that should be taken into account while choosing strategy for the exchange.

This procedure may look strange for an ordinary user who is used to using usual exchange desk with fixed buy and sell prices, but it should be familiar for users that have experience working with stock markets.

For those who consider playing on exchange market too complicated, there are supplementary services that offer a possibility to perform exchange with predefined rates. But on WebMoney Exchanger one can usually achieve better rates.

[3]Equivalent of Ukrainian hryvnia.
[4]Equivalent of US dollar.

The rest of the work is organized as follows. In chapter 2 a reader can find the review of the related literature. Chapter 3 describes simulation setup and its methodology. Dataset and its features are shown in chapter 4. Order flows and trader choice calibration details can be found in chapter 5 (outlined derivation of maximum likelihood estimator of the rate of nonhomogeneous Poisson process can be found in appendix A). Particular case simulation results are reported in chapter 6.

Chapter 2

Literature review

The first notable empirical paper devoted to limit order book is written by Biais et al. (1995). The authors analyze 6 trading days in June/July 1991 and 19 trading days in October/November 1991 of Paris Bourse stock exchange. They describe different insights on supply and demand formation in such markets and make first observations of the specific features of limit order book.

Subsequent literature on limit order book system can be divided into two main groups: empirical studies and models building of limit order book. First we will consider empirical literature.

Relative price. Zovko and Farmer (2002) consider near two million orders from London Stock Exchange. They merge data from 50 stocks and analyze relative limit price (difference between log limit price and log of best price available). The observation is that the distribution of relative limit price (in both buy and sell orders) has power law tails with exponent close to 1.5. Similar results have been obtained by Bouchaud et al. (2002) on Paris Bourse stock exchange market.

Order size. Maslov and Mills (2001) report that size distribution of marketable orders (limit orders that have been sold) has power low tails (with exponent 2.4 ± 0.1). This result was obtained using high-frequency NASDAQ data.

Limit order cancelations. On many markets agents cannot just change the price of theirs limit orders. In those cases it is a common action to cancel the old limit order and create a new one. Hasbrouck and Saar (2001) report that roughly 25% of limit orders were cancelled during two seconds and near 40% during 10 seconds after submitting (their findings are based on 300 largest equities on NASDAQ, 1 October 1999 – 31 December 1999). Thus cancelations are important issues to model.

Conditional events. Order size, relative price, spread, volatility and other characteristics of the limit order and of the market can be related to each other. But, surprisingly, some studies report evidence of independence of some of these events. For example, Gu et al. (2008) state that there is evidence that the relative price of new limit order is independent of spread and volatility of the market (data for Shenzhen Stock Exchange, 2003). On the contrary Lo and Sapp (2010) provide evidence that in more volatile environment order sizes are lower (data is taken from the different exchange markets for 1997 and 2005 years). In addition Gould et al. (2011) state that there is evidence that arrival rate of order heavily depends on events in the market in the closest five minute window. Biais et al. (1995) report that cancelations rate increases after a matching on either side of the market. This whole research puts warnings and pays attention to independence assumptions in the models.

Heavy-tailed returns. Gould et al. (2011) state that mid-price returns have distribution with tails that are heavier than normal. Despite that these distributions are different in different timescales, there is evidence that in the shortest timescale returns distribution tails can be approximated by the power law. Heavy tails of such returns are important feature of the market, which leads to underestimating risks if one uses normal distribution instead.

Autocorrelation of returns. This autocorrelation is present only in short period of time. In longer period it does not persist. Stanley et al. (2008) report that autocorrelation function of the price decreases exponentially and after 20 minutes turns to the noise. Gould et al. (2011) also support this claim and state that this property is observed in different markets.

All the above facts are well studied in the literature and verified for different markets (with different level of success). They do not pretend to be real truth in all the cases, but they are used for reality check of different market models.

Now we turn to the discussion of theoretical models. All developed limit market models can be divided into three groups (Gould et al. 2011): models with perfect rational agents, zero-intelligence models and agent-based models.

Consider several models from different groups.

Roşu (2009) presents the first (according to the author) dynamic model of order-driven market that allows agents freely modify and cancel limit orders. This model has rational but heterogeneous agents. The time is continuous and the horizon is infinite. There is only one asset in the economy and it does not yield dividends. Sellers and buyers arrive randomly, put their limit or market order on one unit of asset and, after matching it, leave the market. Also there is an assumption that all agents have exogenous incentives to trade. Once limit order is placed, it can be cancelled or changed. Traders have different waiting costs, thus they can be patient and impatient.

Rosu model generates the following predictions: first, higher trading activity leads to smaller spread and lower price impact; and second, orders can cluster away from the spread and build hump-shaped order book.

Cont et al. (2010) provide zero-intelligence model, that according to the authors can be easily estimated from the data and replicates various empirical market features. This model is based on continuous-time Markov process, whose state describes the size of queue of orders to be executed for each price. Equally sized orders arrive according to Poisson processes with different intensities for different prices. This paper uses data from Tokyo Stock Exchange Market for numerical experiments. The main advantage of this model is the fact that it is relatively analytically tractable, thus its simulations are relatively simple and fast.

Lukov et al. (2012) describe one more dynamic stochastic model of the order book market. This model is designed to be able to simulate famous market phenomena appearing around the moment of expected issue of new information. The core of this model is also the Markov chain that describes state of the book and additionally two parameter influencing order flows. This allows to model changing flows intensities.

Cont continues to work in this field. In article (Cont and De Larrard 2013) the authors try to include order cancelations into the model.

Huang (2012) extends Cont et al. (2010) and allows to drop annoying assumption of equal order sizes.

As it can be seen, recent achievements in modeling limit order book (Cont et al. (2010); Lukov et al. (2012); Cont and De Larrard (2013); Huang (2012)) are based on standard tools of queuing theory and its extensions. The cornerstone of those models are continuous time Markov processes with special state spaces. Complex state space and Markovian property force the authors to impose very strong assumptions and restrictions on their models. Some of them are as follows:

- Inter-arrival times are independent and identically distributed (Lukov et al. (2012) consider model, where the inter-arrival times are independent, but not identically distributed).

- Order sizes are equal (Huang (2012) drops this assumption).
- No cancelations (Cont and De Larrard (2013) tries to drop it).
- No changing rates of limit orders.

Moreover, Markov processes have limited ability to model intelligent behavior of a trader.

All these restrictions (even if some of them are dropped) can make predictions and implications almost useless.

This work extends considered empirical literature with evidence from WebMoney Exchanger exchange market and extends considered theoretical literature with experience of applying Monte-Carlo simulations for analysing Central Limit Order Book Market.

Chapter 3

Methodology

Dynamics of a limit order book can be studied by Monte Carlo simulation. On the one hand, it is simpler[1] to introduce as many features as needed, but on the other hand, models constructed this way cannot be studied analytically and thus require careful calibration in order to produce meaningful results.

To avoid confusion, we abstract away from a particular exchange pair and refer to the one side of the market as “side A” and the other side of the market as “side B”. If, for example, on side A we have limit orders that requires WMZ and offers WMR, then on side B limit order will require WMR and offer WMZ. For a stock market, side A can contain sell orders while side B consists of buy orders or vice-versa.

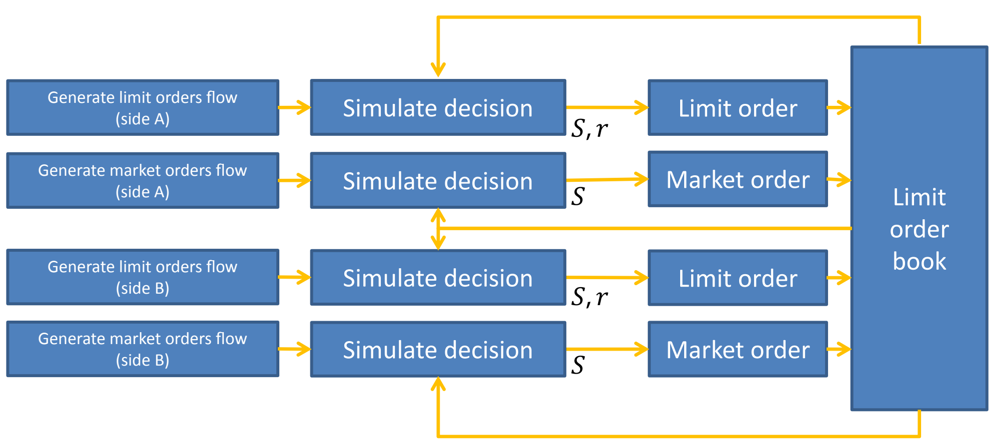


Figure 3.1. Simulation scheme. At the beginning, four order flows are generated. Then in the order of appearance every order chooses its parameters (amount, rate) taking into account current book state and goes into the book for processing.

On the figure 3.1 the schematic representation of the simulation model is shown.

[1] If we compare to modeling using Markov chains.

At the beginning there is an initial order book state with some orders. After that, all four flows start to generate orders according to their calibrated random processes. In the order of appearance each of those orders chooses its parameters (amount, rate) taking into account current book state. Then orders come to the book for processing. After some time of such simulation, one will create all the dynamics of the book and its final state. Repeating these steps as many times as necessary, one will obtain a sample of book states with required characteristics, and theirs empirical distributions.

So, if one wants to check some strategy of exchanging, he can put his limit order at the beginning of the simulation, simulate his strategy together with the market, and, finally, obtain the results of strategy application (for example, the amount of target currency after exchange). Repeating these simulations, one will find distribution of the amount of target currency, conditional to applying particular strategy.

In this model there are two major objects to parameterize:

- Order flows.
- Trader choice.

It is natural to describe every order flow as a sequence of random events. Usually such events sequences are modeled by Poisson process (at least this is the first to by tried). But, unfortunately, in our case it will not work. Traders' activity changes over time and thus average events count per unit of time cannot be assumed constant. That's why we use extension of this process in this work. Input order flows are modeled as Nonhomogeneous Poisson processes with piecewise linear rate function. Such approach allows to capture changes in traders' activity over time. More details about calibration of these flows are described in chapter 5 and appendix A.

When we have rate functions of input flows calibrated, it is easy to generate as many realizations of these flows as required. But every of this event should be assigned with order parameters. Limit order requires to have its size and rate. Market order requires only size. It is logical to assume that these parameters depend on current order book state and on unobservable factors. Exact specifications of these relationships are described in chapter 5. Eventually, orders' sizes and rates are assigned taking into account the current state of the book and drawing disturbance from empirical distribution of the residuals of corresponding models.

Chapter 4

Data

4.1. Data description

Available dataset. WebMoney Exchanger shows currently available bids on the market. There is a possibility to take snapshots of these lists, for example, every minute. Comparing adjacent lists we can reconstruct a lot of events that happened on the market. As a result we can build the dataset close to ideal but with some missing events and some incorrectly interpreted ones (this can happen when the bid disappeared, for example; in this case it is not always clear whether it was bought or cancelled by the owner).

Strictly speaking, we collect the following data:

- Snapshots of lists with active limit orders (only top-50 are available).
- Daily statistics[1] with trade amount and average rates.

Analyzing changes in consecutive snapshots, we can capture the following events:

1. New limit order has appeared.
2. Limit order changed its rate.
3. Limit order changed its amount.
4. Limit order has disappeared

[1] This information is updated every hour. Thus we can calculate hourly statistics observing changes in these "daily" numbers.

Unfortunately, any of these events can be interpreted unambiguously (except changing rate). New limit order can appear in the following circumstances:

1.1. New limit order is created.

1.2. Some limit order disappeared in the past due to exiting from the top-50 and has just re-entered it again.

Changing of limit order amount also can be treated in several ways:

3.1. The limit order has been partially matched.

3.2. Several limit orders have been merged.

If some limit order disappeared, we can have the following reasons for that:

4.1. This limit order has been cancelled.

4.2. This limit order has been matched.

4.3. This limit order has been converted to market order.

4.4. This limit order has been joined with some other one;

4.5. This limit order has left top-50 list.

Unfortunately, it is impossible to distinguish between all these events in deterministic way. But some conclusions can be made.

Cases 1.2 and 4.5 can be distinguished considering all the history[2] of a particular order. If the order disappeared and later we can see it in the future, we can say that the considered order was pushed out from top-50 list. So, having case 1.2 singled out, we can conclude that the rest new orders were just created (case 1.1).

[2]We have id for every order, so we can construct observed history of each order.

Table 4.1. Snapshots series that were successfully collected with gaps no longer than 30 minutes. Series that are shorter than 10 days are omitted.

Begin time	End time	# of snapshots	# of days
2011-01-12 02:23	2011-02-15 17:19	24547	34.6
2011-02-15 17:51	2011-02-28 08:41	8951	12.6
2011-02-28 09:15	2011-03-13 06:55	9289	12.9
2011-03-13 10:46	2011-03-27 02:59	9874	13.6
2011-05-17 16:29	2011-05-28 10:53	7537	10.7
2011-06-28 17:45	2011-07-08 18:29	7044	10.0
2011-10-21 00:28	2011-11-07 18:25	12351	17.7
2011-11-07 19:04	2011-11-20 11:13	8934	12.6
2011-11-20 15:45	2011-12-14 10:09	16693	23.7
2011-12-14 11:43	2011-12-26 14:07	8524	12.1
2011-12-30 21:17	2012-01-11 11:37	8075	11.5
2012-04-16 22:45	2012-05-12 16:29	18089	25.7
2012-06-02 08:33	2012-06-22 03:51	14169	19.8
2012-06-23 22:14	2012-07-04 11:27	7371	10.5
2012-08-07 13:45	2012-08-20 16:09	8597	13.0
2012-08-20 16:47	2012-09-13 11:57	15941	23.7
2012-09-21 15:43	2012-10-16 00:21	16784	24.3
2012-10-28 21:57	2013-01-05 23:15	47731	69.0
2013-01-14 22:33	2013-02-01 18:15	12257	17.8

Cases 3.1 and 3.2 could be distinguished by paying attention to the change in the order amount. If orders have been merged, the amount of the order should increase, but in case 3.1 this amount should decrease.

The most complicated task is to distinguish between cases 4.1, 4.2 and 4.4.

Consider case 4.4 first. For a fixed period of time the number of events of type 3.2 and 4.4 should be equal. So, for each case 3.2 we can try to find the corresponding disappeared order of type 4.4. In ideal situation the amount of those orders should be equal, but if we have overlapping event 3.1, their numbers will be different.

Cases like 4.3 can be captured considering both sides of the market. If we have sold or partially sold one or several orders of total amount equal to that we

Table 4.2. Distribution of identified events. Calculated using all events identified between all available subsequent pairs of snapshots with gaps no longer than 4 minutes.

	Description	Occurences
1.1	New limit order has created	8.80%
1.2	The order has pushed out of top-50	8.68%
2	The rate has changed	44.69%
3.1	The order has been partially matched	21.24%
3.2	The order has been merged	0.23%
4.1	The order has been cancelled	4.41%
4.2	The order has been matched (by the wholesale market order)	2.94%
4.2	The order has been probably matched	2.11%
4.3	The order has been converted to market order	0.17%
4.4	The order has been merged with some other one	0.14%
4.5	The order has re-entered the top-50 list	6.5%

have in disappeared order, we can assume that converting to market order took place here.

The remaining ambiguity lies between cases 4.2 and 4.1. It is really hard to say something for sure about them. Some conclusion can be made taking into account hourly data about amount of trade. Sometimes there are a lot of disappeared orders exactly at the top of the book. They could be considered as sold. The remaining disappeared orders are separated using some heuristic threshold. We have approximately 6% of such events, so such decision should not bring too much distortions in the data.

4.2. Summary statistics

First of all it needs to be mentioned that data collection procedure is time consuming. And it is subject to numerous software and hardware failures. So not always it is possible to take snapshots every minute or two. Snapshots series that were successfully collected with gaps no longer than 30 minutes are described in table 4.1.

Table 4.3. Summary statistics for limit orders data. Built using limit orders data for WMZ→WMR side of the market. `aIn` — size of the order in WMZ. `aOut` — size of the order in WMR. `topAsk` — book top ask rate at the moment of order creating. `topBid` — book top bid rate at the moment of order creating. `avgTopAsk` — book average top ask rate for 20 minutes period before order creating. `avgTopBid` — book average top bid rate for 20 minutes period before order creating. `poissonRate` — value of the rate function of estimated Nonhomogeneous Poisson process for flow of these orders. `rate` — rate of the order WMR/WMZ. `laIn` — log of `aIn`. `laOut` — log of `aOut`. `spread` — `topAsk - topBid`.

Variable	Mean	Std. Dev.	N
aIn	5881.04	282.8706	11558
aOut	188097.75	9049.5897	11558
topAsk	31.923	0.562	11558
topBid	31.824	0.575	11558
avgTopAsk	31.923	0.561	11558
avgTopBid	31.825	0.572	11558
poissonRate	10.191	4.658	11558
rate	32.029	0.577	11558
laIn	9.51	2.581	11558
laOut	12.977	2.58	11558
spread	0.099	0.094	11558

We can see that the longest more or less continuous series is available from the end of October 2012 till the beginning of January 2013. Obviously 30-minute gaps are also too large. But they do not bring too many problems. 95.6% of gaps are not longer than 2 minutes. And only 0.08% of gaps are longer than 4 minutes. So we can state that within the periods listed in table 4.1 our series are good enough.

In the table 4.2 we can see the distribution[3] of identified events in the most dynamic direction of exchange (WMZ→WMR). The most frequent events are changing order rates and partially matching[4] orders.

[3] Events type distribution does not persist over time. And it can be tested formally. But the numbers do not change dramatically. Thus they can help in general understanding of the market.

[4] Under a partially matched order we understand a limit order that has been matched with a corresponding market order but this market order had lower amount than initial limit order. As a result limit order decreases its amount by the amount of matched market order.

On this market we do not have too many cancelations and it is natural due to the fact that user can freely change the order rate, merge orders and convert limit order to market one. Moreover canceling the limit order will lead to 0.8% lose due to WebMoney transaction cost, so it is costly.

The rest of this work mostly uses data for period from 2012-10-28 21:57 till 2013-01-05 23:15. This is the biggest continuous piece of available data. This sample has 407 pieces of 4-hour long time intervals, 11558 limit orders and 35684 market orders at one side of the market and 9482 limit orders and 45415 market orders at the another side of the market.

Table 4.3 reports summary statistics of selected sample for limit orders for one side of the market. These variables are mostly used in chapter 5 while building model for trader choice simulation.

Chapter 5

Calibration

In order to make required simulation, the following processes and their characteristics should be calibrated:

- Order flows of all four order types (side A limit orders, side A market orders, side B limit orders, side B market orders).
- Order amount and rate distributions conditional to current state of the market.

5.1. Order flows representation

In some papers (Cont et al. (2010); Cont and De Larrard (2013); Huang (2012)) order flows are modeled as Poisson process. Lukov et al. (2012) try to use non-homogeneous Poisson process and endogenize its rate. Huang (2012) mentions Hawkes' point process that is also a non-homogeneous Poisson process.

In this work input flows are modeled as non-homogeneous Poisson processes with piecewise linear rate function. These flows are calibrated using maximum likelihood estimator of the rate function using methodology described in appendix A. This approach requires to fix abscissas of the nodes of the rate function. We fix these nodes in a grid with interval between points that equal 4 hours[1]. So, we have 6 nodes for one day of observations. Such regular grid allows us to consider values at nodes as time series (it will be useful for building prediction models).

[1] If we get to large interval between grid points, estimated rate function will not capture daily changes in flows intensities. If this interval is too small, we will have too much grid points (i.e. parameters) and the rate function will be overfitted. Four hours is a trade-off value of this interval.

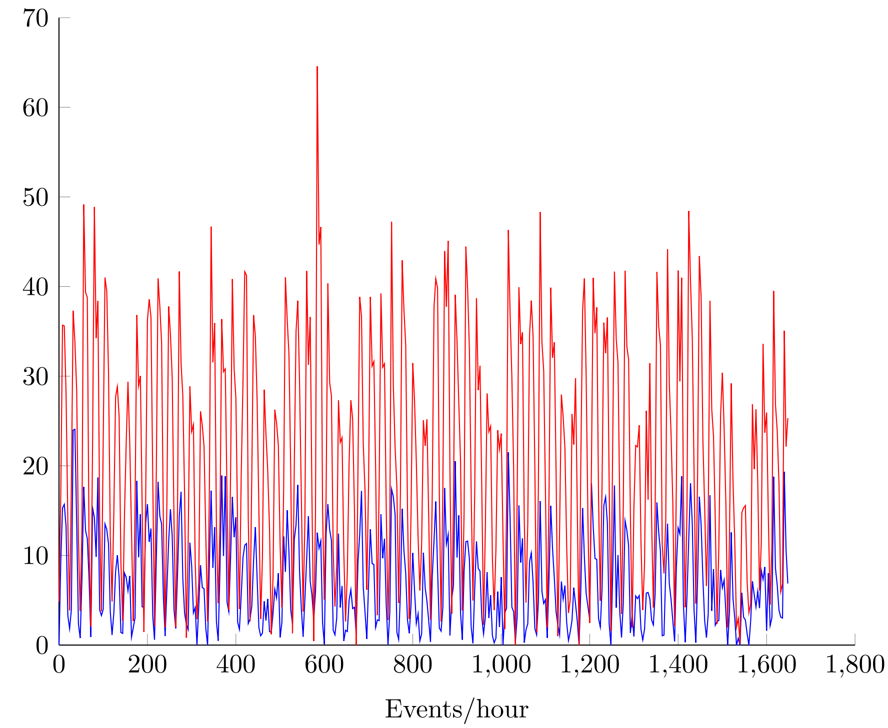


Figure 5.1. Piecewise linear estimation of the Poisson flow rate function for the period from 29 Oct 2012 till 5 Jan 2012 with nodes every 4 hours (for WMZ→WMR side of the market). Blue (lower graph) represents limit orders flow. Red (higher graph) represents market orders flow. Y axis reflects instantaneous intensity of the flow in # of events per hour.

Figure 5.1 shows one estimated input flow rate function during a sample period of time. This rate function has obvious periodic oscillations.

Consider the values of the estimated rate function at its nodes as a time series. Table 5.1 shows the results of the regression of this time series on time related dummies and several autoregressive terms. Insignificant ones state that in those periods the rate function should not significantly differ from its value in the base period (so it looks that Tuesday, Wednesday and Thursday do not significantly differ from Monday). The hypothesis that residuals after this regression have no serial correlation can not be rejected. So, after dropping insignificant dummies, we obtain the model that can be the first approximation of such time series.

Table 5.1. Values of the rate function at nodes are regressed on time dependant dummies and autoregressive terms. `hod` — number of 4-hour period in a day (from 0 to 5). `dow` — number of the day in a week (from 0 to 6). Built using limit orders data for WMZ→WMR side of the market.

Variable	**Coefficient**	**(Std. Err.)**
dow=1	-0.788	(0.616)
dow=2	-1.062†	(0.622)
dow=3	-1.156†	(0.623)
dow=4	-1.285*	(0.616)
dow=5	-3.671**	(0.623)
dow=6	-3.207**	(0.603)
hod=1	-0.931	(0.639)
hod=2	1.677†	(0.855)
hod=3	7.047**	(0.819)
hod=4	4.015**	(0.706)
hod=5	4.345**	(0.596)
1st lag	0.145**	(0.048)
5th lag	0.139**	(0.049)
6th lag	0.204**	(0.050)
Intercept	2.462**	(0.696)

N	407
R^2	0.686
$F_{(14,392)}$	61.146

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Unfortunately, all four rate functions (for different order flows) have different coefficients and models. Some of them require including different lags in order to rule out autocorrelation. But, anyway, in the end, we obtain relatively simple models that are capable to make predictions about future rate function values for several periods of time ahead.

As a result, Nonhomogeneous Poisson process fitting to the data shows robust cyclical structure of the traders' intensity, which is surprisingly stable over substantial period of time. So, using such model allows to capture regular daily and weekly cycles, which is more then enough for short term simulations. There is an example of 24 hour simulation in chapter 6. And strictly speaking,

Table 5.2. Order size regressed on order book characteristics. Dependant variable: log of order size. Regressors are order flow rate, current spread, top ask rate and top bid rate. Several specifications are depicted in columns. Built using limit orders data for WMZ→WMR side of the market.

	(1)	(2)	(3)	(4)
poissonRate	0.0435***	0.0444***	0.0444***	0.0444***
spread	-0.526**	-0.540**	-0.590**	
topAsk		-0.0502		-0.590**
topBid			-0.0502	0.540**
Constant	12.59***	14.18***	14.18***	14.18***
Observations	11,558	11,558	11,558	11,558
R^2	0.007	0.007	0.007	0.007

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

usually there is no reason to consider longer period of time in the same model setup.

5.2. Trader choice estimation

If we consider one side of the market, there are two different types of orders: limit and market ones. Flows for each of them are separated. For the limit one we need to choose size of the order and its rate. For the market order only its size should be chosen.

Consider the amount of the order. Table 5.2 shows the results from several regressions. It is easy to see that current rates are insignificant if spread is taken into account. And if the spread is omitted, coefficients at rates immediately change in order to express that rate (the hypothesis that those coefficients are equal ignoring the sign cannot be rejected with `p-value`=0.245). This evidence allows us to drop current rates from the regression having the spread included.

As a result we can simulate order size as

$$size = exp(C + C_{poissonRate} \cdot poissonRate + C_{spread} \cdot spread + \varepsilon) \qquad (5.1)$$

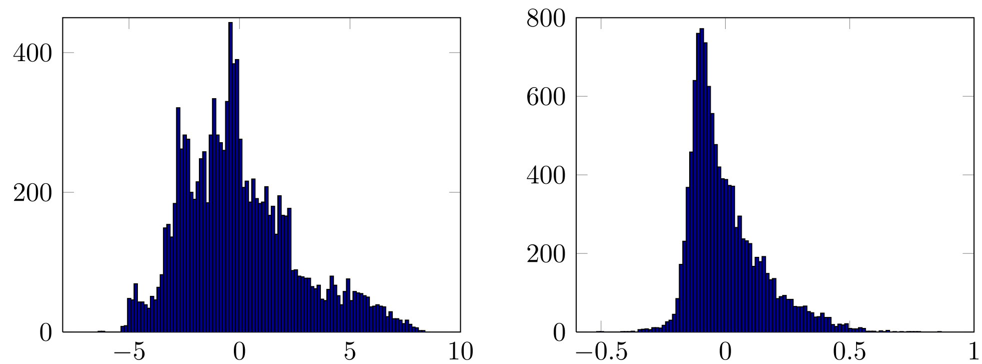
where C, $C_{poissonRate}$ and C_{spread} — coefficients of the corresponding regression, $poissonRate$ and $spread$ — current order flow rate and current book spread, ε — random disturbance distributed according to empirical distribution fitted from the data (histogram for sample data is shown on figure 5.2a).

In formula 5.1 it is very important to pay attention to distribution of disturbances. Due to quite small R^2 of proposed regression these disturbances bring huge amount of variance in the data. And strictly speaking, this is exactly what we expected to have: the order flow rates and current state of the book are not the major determinants of the order size. Major determinants are unobservable and modeled as random.

If there is a limit order, its rate also should be simulated. In literature it is a common approach to tie this rate to ask/bid/mid price (Cont et al. (2010), for example). But such approach leads to undesirable artifacts during simulations[2].

That's why we also use the average ask rate for some short time. Table 5.3 shows that spread seems to be insignificant for the rate of new limit order. Also we can see that an average top rate and current top rate split the influence on new orders: approximately 2/3 of the weight has the average rate and 1/3 the current rate. In addition, it is interesting to observe that the sum of coefficients at average rate and current rate is more than one (the hypothesis that thier sum is equal to the one can be rejected). It means that on average new orders

[2]Consider the book just after successful market order. If new order rate is tied to current rate, the decision becomes based on the rate that is lower than the usual one in the market. If this happens too often, spread starts to widen. And such behavior is not natural and should be ruled out.



(a) Limit order size residuals **(b)** Limit order rate residuals

Figure 5.2. Residuals' histograms for limit order regressions. Built by applying regression using specification from equation 5.2 (a) and equation 5.1 (b) to limit order data for WMZ→WMR side of the market.

are placed on top of the book. Huge R^2 can be explained taking into account the fact that almost all significant variance in rates can be explained by market movements (they are captured by average rate and current rate). But anyway, new orders can choose rate that differs from current by up to approximately 1% (at 90% percentile).

So, the order rate can be simulated as

$$rate = C_{poissonRate} \cdot poissonRate + C_{avgTopAsk} \cdot avgTopAsk + C_{spread} \cdot spread + C_{topAsk} \cdot topAsk + C_{log(size)} \cdot log(size) + \varepsilon \quad (5.2)$$

where C, $C_{poissonRate}$, $C_{avgTopAsk}$, C_{topAsk} and $C_{log(size)}$ are the corresponding regression coefficients, $poissonRate$, $avgTopAsk$, $topAsk$ and $size$ — average rate, current rate and order size, respectively, ε — random disturbance distributed according to empirical distribution fitted from the data (histogram for sample data is shown on figure 5.2b).

As a result, the way of simulating limit and market order parameters have been

Table 5.3. Order rate regressed on order book characteristics without constant. Dependant variable: order rate. Regressors are current rate of order flow, average top ask rate for recent 20 minutes, current top ask rate, current spread and log of the order size.

	rate
poissonRate	0.000955***
avgTopAsk	0.614***
topAsk	0.396***
spread	-0.00804
laOut	-0.0178***
Observations	11,558
R^2	1.000

*** p<0.01, ** p<0.05, * p<0.1

described. This approach generates random decisions, where order sizes and rates are related and partially depend on current order book state.

Chapter 6

Simulation results

In this chapter one specific exchange strategy will be considered. Assume one wants to exchange the amount S at rate r within time T. He puts limit order of size S and rate r and waits until time expires. Then he just converts the remaining limit order to market one and finishes the exchange. The simulation results on sample data are shown on firure 6.1 and firure 6.2.

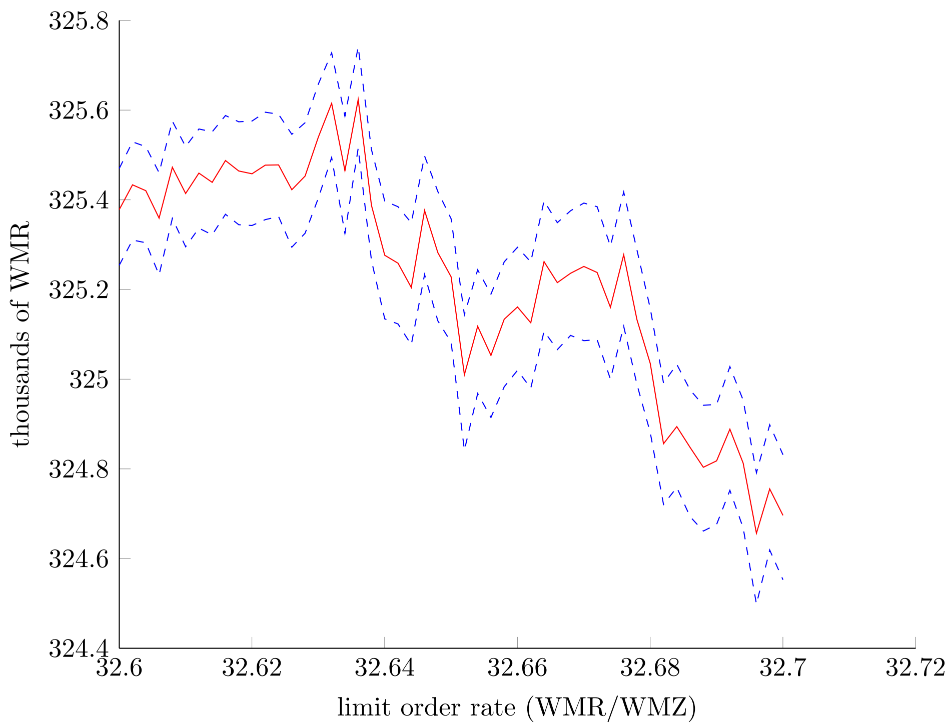


Figure 6.1. Mean of the amount of destination currency received from limit order with different rates. Separate simulations (with 500 iterations each) were done for different values of the rate. It was assumed that we are intended to exchange 10000 WMZ to WMR in 24 hours. Dotted lines show 95% confidence interval for the mean.

Figure 6.1 shows the mean of destination currency for different limit order rates. All the rates are slightly below the top rate of the corresponding side

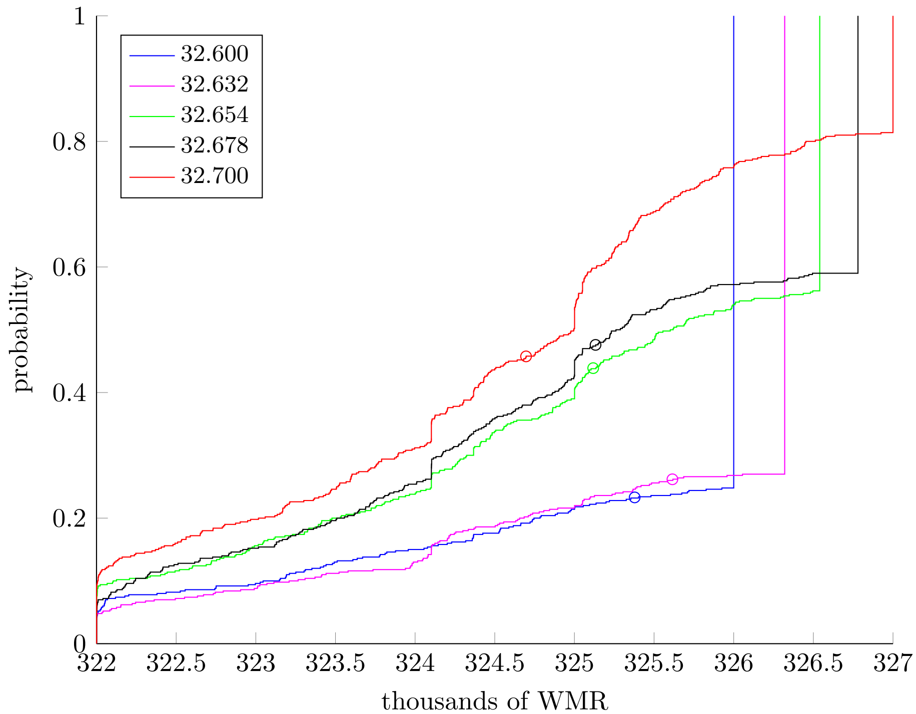


Figure 6.2. CDFs of destination currency for different limit order rates. Dedicated simulation (with 500 iterations each) were done for every value of the rate. It was assumed that we are intended to exchange 10000 WMZ to WMR in 24 hours. The means are indicated by with circles.

of the book. In this graph it is easy to see that in this particular case a good decision would be to choose limit order rate close to 32.632 (magenta CDF on figure 6.2).

More details can be seen from the shape of the cumulative distribution functions shown in figure 6.2. The discrete jump at the right hand side of these distributions shows the probability of matching the whole limit order. It is easy to see that in this case increasing limit order rate from 32.600 to 32.632 or from 32.654 to 32.678 almost has no influence on matching probability, but if we increase order rate from 32.678 to 32.7, matching probability decreases twice as much. These particular results can be explained by structure of starting state of limit order book and by market intensity.

So, as a result, figures 6.1 and 6.2 summarize historical information about the market and try to present this information with the particular exchange case in mind.

In the end, the final choice of limit order rate should be left for the trader, who will take into account his preferences.

Chapter 7

Conclusions

In this work the simulation model of WebMoney Exchanger exchange market (which is an example of Central Limit Order Book market) was built.

Flows of different order types were simulated separately. Each of them is assumed to be a Nonhomogeneous Poisson process. The rate functions of these processes have been calibrated with piecewise linear function using maximum likelihood estimator. As a result, we obtain flows with robust cyclical structure, which is, surprisingly stable over time.

Trader choice was simulated taking into account the revealed dependencies between order parameters (rate and size) and market parameters (current activity level on the market, spread, average rates, top rates). While estimates are statistically significant, the unobservable parameters have responsibility for large part of variation of order sizes. The opposite situation have been observed with rates: they vary around current top rate with skewed small disturbances.

This model differs a lot from its analytical competitors. It does not require assumptions such as equal order sizes or completely random trader behavior. Implementing at least some level of trader intelligence brings higher level of tractability of the model.

Finally, the example of simulation was traced throughout the work. And its result can be used to make decision about strategy of exchange for this particular case. Certainly, all these computations can be made for different point in time, different exchange amount and different strategy using the same approach.

This result summarizes historical information available from the market and extends trader's information set. Comparing outcome distributions among different strategies, analyst can prefer one strategy over another. Such a possibility can be useful for developing and testing strategies for algorithmic trading, which is becoming more and more popular way to work with the markets nowadays.

Future research. Considered model can be extended in several ways taking into account more and more details of traders' behavior.

Due to the fact that we can trace each limit order over time, it is possible to cluster traders by their activity and then use this differentiation during simulations.

From different side, one can consider to review the choice of Nonhomogeneous Poisson process for modeling order flows. For example, there is an evidence (Huang 2012) that after new limit order is putted into the book the expected waiting time for a market order decreases. Such relationships cannot be simulated using current approach. Moreover, one can argue that exponential waiting times (even in case of nonhomogeneous rate function) are not appropriate assumption for order flows. Additional point of pain is a piecewise linearity of the rate function (i.e. in reality it should be more smooth). All this can be probably implemented better replicating more features of the data.

Having additional exogenous data influencing the market, some of simulation model parameters can be linked to them. For example, we can account for holidays while building the forecast of order flows.

While implementing marginal improvements like described above it is important to understand that taking into account too many issues can substantially increase computational resources required to perform simulation. As the same time resulting graphs probably will not differ too much.

Even in current setup calculations takes too much time (comparing to time usually available for decision making). Hopefully there is a lot of space for optimization. First of all Matlab (which is the primary tool in this work) is not efficient in non-matrix manipulations (e.g. queue operations are too slow). At least several times improvement can be achieved substituting Matlab with C#, Java, or C++[1]. From other prospective, simulations could be run on a cluster, shortening waiting period again (hopefully, Monte Carlo simulations can be easily scaled for running on many machines).

[1] C++ is the best choice in terms of performance.

Works cited

Biais, B., P. Hillion, and C. Spatt (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, J., M. Mézard, and M. Potters (2002). Statistical properties of stock order books: empirical results and models. *Quantitative Finance 2*(4), 251–256.

Cont, R. and A. De Larrard (2013). Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics 4*(1), 1–25.

Cont, R., S. Stoikov, and R. Talreja (2010). A stochastic model for order book dynamics. *Operations research 58*(3), 549–563.

Gould, M., M. Porter, S. Williams, M. McDonald, D. Fenn, and S. Howison (2011). Limit order books. *Available at SSRN 1970185*.

Gu, G., W. Chen, and W. Zhou (2008). Empirical regularities of order placement in the Chinese stock market. *Physica A: Statistical Mechanics and its Applications 387*(13), 3173–3182.

Hasbrouck, J. and G. Saar (2001). Limit orders and volatility in a hybrid market: The Island ECN.

Huang, H. (2012). *Modeling order book dynamics using queues and point processes*. Ph. D. thesis, The Florida State University.

Lo, I. and S. Sapp (2010). Order aggressiveness and quantity: how are they determined in a limit order market? *Journal of International Financial Markets, Institutions and Money 20*(3), 213–237.

Lukov, A., S. Myzuchka, and K. Vaninsky (2012). A multi-agent nonlinear markov model of the order book. *arXiv preprint arXiv:1208.3083*.

Maslov, S. and M. Mills (2001). Price fluctuations from the order book perspective—empirical facts and a simple model. *Physica A: Statistical Mechanics and its Applications 299*(1), 234–246.

Massey, W., G. Parker, and W. Whitt (1996). Estimating the parameters of a nonhomogeneous Poisson process with linear rate. *Telecommunication*

Systems 5(2), 361–388.

Pasupathy, R. (2011). Generating Nonomogeneous Poisson Processes. *Wiley Encyclopedia of Operations Research and Management Science*.

Roşu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies 22*(11), 4601–4641.

Stanley, H., V. Plerou, and X. Gabaix (2008). A statistical physics view of financial fluctuations: Evidence for scaling and universality. *Physica A: Statistical Mechanics and its Applications 387*(15), 3967–3981.

Zovko, I. and J. Farmer (2002). The power of patience: a behavioural regularity in limit-order placement. *Quantitative Finance 2*(5), 387–392.

Appendix A

Nonhomogeneous Poisson process estimation

The following estimation procedure is the extension of the maximum likelihood estimator described in (Massey et al. 1996).

Before considering this estimator, we state several facts about nonhomogeneous Poisson processes.

For nonhomogeneous Poisson flow with rate function $\lambda(t)$ the number of events happened throughout the time interval $[a, b]$ can be described as

$$\mathrm{P}\{N(b) - N(a) = k\} = \frac{e^{-\lambda_{a,b}} \lambda_{a,b}^{k}}{k!}, \tag{A.1}$$

where $\lambda_{a,b} = \int_a^b \lambda(t)dt$.

Theorem A.1. *For nonhomogeneous Poisson process with rate function $\lambda(t)$ the distribution of waiting times for the k-th event after time t_0 has the following CDF function*

$$F(x) = 1 - e^{-\int_{t_0}^{t_0+x} \lambda(t)dt} \sum_{i=0}^{k-1} \frac{\left(e^{-\int_{t_0}^{t_0+x} \lambda(t)dt}\right)^i}{i!}$$

Corollary A.2 (Pasupathy (2011))**.** *For nonhomogeneous Poisson process with rate function $\lambda(t)$ the distribution of waiting times for the next event after time t_0 will have the following CDF and PDF functions*

$$F(x) = 1 - e^{-\int_{t_0}^{t_0+x} \lambda(t)dt}$$

$$f(x) = e^{-\int_{t_0}^{t_0+x} \lambda(t)dt} \lambda(t_0 + x)$$

Theorem A.3. *Consider nonhomogeneous Poisson process with rate function $\lambda(t)$. The distribution of time of the last event before time T will have the following CDF and PDF functions*

$$F(x) = \frac{e^{-\int_x^T \lambda(t)dt} - e^{-\int_0^T \lambda(t)dt}}{1 - e^{-\int_0^T \lambda(t)dt}}, \quad x \in [0, T]$$

$$f(x) = \frac{e^{-\int_x^T \lambda(t)\,dt}\lambda(x)}{1 - e^{-\int_0^T \lambda(t)\,dt}}, \quad x \in [0, T]$$

Now, suppose we have a realization of some Poisson flow of events on $[0, T]$, which have happened at times $t_0, \ldots, t_{N-1}$.

Our goal is to estimate the rate function $\lambda(t)$ of such flow.

Note that lengths $t_k - t_{k-1}$ of time intervals (t_{k-1}, t_k) are independent random variables distributed according to the theorem A.2.

Thus, $\{t_1 - t_0, \ldots, t_{N-1} - t_{N-2}\}$ can be considered as the realization of the vector, every component of which is independently distributed according to the theorem A.2.

As a result, we can construct maximum likelihood function as

$$\prod_{i=1}^{N-1} e^{-\int_{t_{i-1}}^{t_i} \lambda(t)dt}\lambda(t_i) \tag{A.2}$$

Then we can consider the maximization problem

$$\operatorname*{argmax}_{\lambda} \prod_{i=1}^{N-1} e^{-\int_{t_{i-1}}^{t_i} \lambda(t)dt}\lambda(t_i) = \tag{A.3}$$

$$= \operatorname*{argmax}_{\lambda} \sum_{i=1}^{N-1} \left(-\int_{t_{i-1}}^{t_i} \lambda(t)dt + \ln \lambda(t_i) \right) = \tag{A.4}$$

$$= \underset{\lambda}{\operatorname{argmax}} \left(\sum_{i=1}^{N-1} \ln \lambda(t_i) - \int_{t_0}^{t_{N-1}} \lambda(t)dt \right) \tag{A.5}$$

Assume that we have the realization of nonhomogeneous Poisson process with piecewise linear rate function

$$\lambda(x) = \begin{cases} \lambda_k(x) = a_i * (x - x_i) + b_i, \quad x \in [x_i, x_{i+1}], i = \overline{0, M-2} \end{cases} \tag{A.6}$$

where $\lambda(x_i) = b_i$ for $i = \overline{0, M-2}$.

Denote $b_{M-1} = \lambda(x_{M-1})$. Then, in order to have (A.6) continuous it is enough to impose

$$a_i * (x_{i+1} - x_i) + b_i = a_{i+1} * (x_{i+1} - x_{i+1}) + b_{i+1}, i = \overline{0, M-2} \tag{A.7}$$

Taking into account that $a_{M-1} * (x_{M-1} - x_{M-2}) + b_{M-1} = \lambda(x_{M-1}) = b_{M-1}$, we have

$$a_i = \frac{b_{i+1} - b_i}{x_{i+1} - x_i}, i = \overline{0, M-1} \tag{A.8}$$

So, in (A.6) we have the following free variables: $b_0, \ldots, b_{M-1}$.

Using (A.6) and setting[1] $x_0 = t_0$ and $x_{N-1} = t_{M-1}$, we can rewrite (A.5) as

$$\underset{\lambda}{\operatorname{argmax}} \left(\sum_{i=1}^{N-1} \ln \lambda(t_i) - \int_{t_0}^{t_{N-1}} \lambda(t)dt \right) = \underset{\{b_0,\ldots,b_{M-1}\}}{\operatorname{argmax}} \left(\sum_{i=1}^{N-1} \ln \lambda(t_i) - \int_{x_0}^{x_{M-1}} \lambda(t)dt \right) =$$
$$= \underset{\{b_0,\ldots,b_{M-1}\}}{\operatorname{argmax}} \left(\sum_{i=1}^{N-1} \ln \lambda(t_i) - \sum_{i=0}^{M-2} \int_{x_i}^{x_{i+1}} (a_i(x - x_i) + b_i)dx \right) =$$
$$= \underset{\{b_0,\ldots,b_{M-1}\}}{\operatorname{argmax}} \left(\sum_{i=1}^{N-1} \ln \lambda(t_i) - \sum_{i=0}^{M-2} \int_{0}^{x_{i+1}-x_i} (a_i x + b_i)dx \right) =$$

$$= \underset{\{b_0,\ldots,b_{M-1}\}}{\operatorname{argmin}} \left(-\sum_{i=1}^{N-1} \ln \lambda(t_i) + \sum_{i=0}^{M-2} \left(\frac{a_i}{2}(x_{i+1}-x_i)^2 + b_i(x_{i+1}-x_i) \right) \right) =$$

$$= \underset{\{b_0,\ldots,b_{M-1}\}}{\operatorname{argmin}} \Big(L(a_0(b_0,b_1),\ldots,a_k(b_k,b_{k+1}),\ldots,a_{M-2}(b_{M-2},b_{M-1}),b_0,\ldots,b_{M-2}) \Big) =$$

$$= \underset{\{b_0,\ldots,b_{M-1}\}}{\operatorname{argmin}} \overline{L}(b_0,\ldots,b_{M-1})$$

Theorem A.4. *The function* $\overline{L}_{x_0,\ldots,x_{M-1}}(b_0,\ldots,b_{M-1})$ *is concave, if*

- $\forall k \exists i \colon t_i \in (x_k, x_{k+1})$*;*
- $\exists k \exists i \neq j \colon t_i \in [x_k, x_{k+1}], t_j \in [x_k, x_{k+1}]$*.*

Proof. Set

$$I_k = \{i \colon t_i \in [x_k, x_{k+1}]\} \tag{A.9}$$

$$\tilde{a}_k = \begin{cases} \sum_{i\in I_k} \frac{1}{\lambda_k^2(t_i)} \cdot \frac{(x_{k+1}-t_i)^2}{(x_{k+1}-x_k)^2}, & k < M-1 \\ 0, & k = M-1 \end{cases} \tag{A.10}$$

$$\tilde{b}_k = \begin{cases} 0, & k = 0 \\ \sum_{i\in I_{k-1}} \frac{1}{\lambda_{k-1}^2(t_i)} \cdot \frac{(t_i-x_{k-1})^2}{(x_k-x_{k-1})^2}, & k > 0 \end{cases} \tag{A.11}$$

$$\tilde{c}_k = \sum_{i\in I_k} \frac{1}{\lambda_k^2(t_i)} \cdot \frac{(t_i-x_k)(x_{k+1}-t_i)}{(x_{k+1}-x_k)^2} \tag{A.12}$$

[1] It is easy to reduce the general case with $x_0 \leq t_0$ and $t_{M-1} \leq x_{N-1}$ to the case with $x_0 = t_0$ and $x_{N-1} = t_{M-1}$ using theorems A.2 and A.3.

Then Hessian matrix of $\overline{L}$ by $b_1, \ldots, b_{M-1}$ will look like

$$\begin{pmatrix} \tilde{a}_0 + \tilde{b}_0 & \tilde{c}_0 & 0 & 0 & \ldots & 0 & 0 \\ \tilde{c}_0 & \tilde{a}_1 + \tilde{b}_1 & \tilde{c}_1 & 0 & \ldots & 0 & 0 \\ 0 & \tilde{c}_1 & \tilde{a}_2 + \tilde{b}_2 & \tilde{c}_2 & \ldots & 0 & 0 \\ \multicolumn{7}{c}{\dotfill} \\ 0 & 0 & 0 & 0 & \ldots & \tilde{a}_{M-2} + \tilde{b}_{M-2} & \tilde{c}_{M-2} \\ 0 & 0 & 0 & 0 & \ldots & \tilde{c}_{M-2} & \tilde{a}_{M-1} + \tilde{b}_{M-1} \end{pmatrix} \quad \text{(A.13)}$$

By definition $\tilde{a}_k > 0$ for $k = \overline{1, M-2}$ and $\tilde{b}_k > 0$ $k = \overline{2, M-1}$ (by theorem assumptions $\forall k \exists i \colon t_i \in (x_k, x_{k+1})$), and $\tilde{c}_k \geq 0$ (here $(t_i - x_k)(x_{k+1} - t_i) > 0$ because of $t_i \in [x_k, x_{k+1}]$).

It can be shown that $\tilde{a}_k \tilde{b}_{k+1} - \tilde{c}_k^2 \geq 0$ and $\exists k$ $\tilde{a}_k \tilde{b}_{k+1} - \tilde{c}_k^2 > 0$

Under these constraints Hessian of $\overline{L}$ is positive definite an thus $\overline{L}$ is concave. ■

While maximizing (A.5) we can consider only positive functions $\lambda(t)$ (because it is the rate function of the Poisson process). It will be enough to impose that $\forall k$ $b_k \geq 0$. In this case $\overline{L}$ has lower bound and the theorem A.4 guarantees the convergence of numerical minimization.

So, having some weak enough conditions on sample size stated in theorem A.4, gradient descent, for example, will converge to the minimum of $\overline{L}$ and will find the maximum likelihood estimator of $\lambda(t)$ in a class of piecewise linear functions with fixed abscissas of the nodes.