National Research University Higher School of Economics
Faculty of Computer Science
Programme ‘Master of Data Science’

MASTER’S THESIS

topic

Optimal parameters of limit orders in high frequency market making strategy

Student: Roman Sharapov

<Full name>

Supervisor: MS, Data science lead of Global Markets department, Andrey Akhmetov

<degree>, <title>, <full name>

Moscow, 2021

Abstract

We consider problem of providing two-side quotes on financial markets (market-making or dealing). Problem is actual mostly for the institutions who have obligations to support quotes for some instruments by placing bid and ask limit orders into the limit order book (also known as Depth of Market). Dealer or market-maker usually has to support both sides simultaneously during main trade session (full or part) with limit on maximum spread. The question is at what prices market-maker should set limit orders to maximize an expected utility (profit) and minimize risks.

From theoretical point of view, it is a stochastic optimal control problem. Optimal strategy heavily depends on market microstructure properties and there are simple solutions in some cases. In this work we investigate market microstructure properties of Russian currency market (USD/RUB) and test one implementation of such market-maker strategy. We test it on historical and generated data and compare with simplest possible strategy, in which dealer sets orders at constant distance from mid-price.

Key words: market-making, dealing, optimal pricing strategy, limit orders pricing, high-frequency trading.

Source code of the project is available on GitHub: https://github.com/rasharp/FinalMDS

Table of contents

Introduction

In financial market dealer or market-maker (in this work we consider both this terms as synonyms) provides liquidity by quoting two-side limit order prices (bid and ask) at which they are willing to buy (bid) or sell (ask) asset. Even in order-driven markets often there are participant who tries to trade on spread. Key issue here that dealer is ready to trade immediately any side without long-term preferences in direction. It differs them from as short-term traders as investors who makes bet on some price direction. Profit arises when market-maker trades a lot of deals without strong directed price movement. So, the key risk of the dealer is having too big position (inventory) in the asset during adverse price movement (e.g., price falling having a lot of stocks). It is so-called inventory risk, and we focus on it in this work. Another kind of risk is asymmetric information risk arising from informed participants and it is not considered here.

Prices are changing extremely quickly. For liquid assets stock exchange processes dozens of limit and market orders a second. So, dealer must react on this condition changes as soon as possible to reduce risk. Such trading called high-frequency trading (HFT). Strategies considered here are such kind. We don't consider technical issues of implementation HFT strategies in real-market conditions (it requires special low-latency hardware and software infrastructure) and confine research to simulations on generated and historical data.

We consider single dealer market who trades single asset. The generalization to many assets is straightforward and basically not difficult. We consider strategies when dealer is always "in market", i.e., they have obligations to quote two-sides almost permanently. This obligation is quite natural: for example, ETFs on Russian market usually has requirements for liquidity providers to quote both sides with limitation on time and maximum spread. The logical consequence of such requirements is simple spread strategy, which is natural benchmark for this reason. In this strategy dealer places limit orders equidistantly from current mid-price (medium price between best bid and best ask prices of current limit order book). Distances are constant in time and equal for bid and ask prices. So, the only task is to fresh bid and ask prices in time. Our goal here is to construct and test a more complicated strategy and compare it with this simplest one.

We suppose several light simplifications:

- Dealer always has sufficient cash to satisfy collateral requirements.
- We don't consider problem of definition "true" or fair value of the asset.
- Only intraday interval of strategy. It is quite realistic because most dealer prefer set inventory to zero before market closing (the only exception may be a FOREX which trades continuously 5 days a week).
- Logical consequence of previous point: we don't consider risk-free rate on cash because we don't trade overnight. Also, we don’t think about inventory returns like dividends.
- We don't consider transaction costs here. Market-makers usually has cheap cost for transaction as key institutional participants of the financial markets. But it is the interesting point for further analysis.

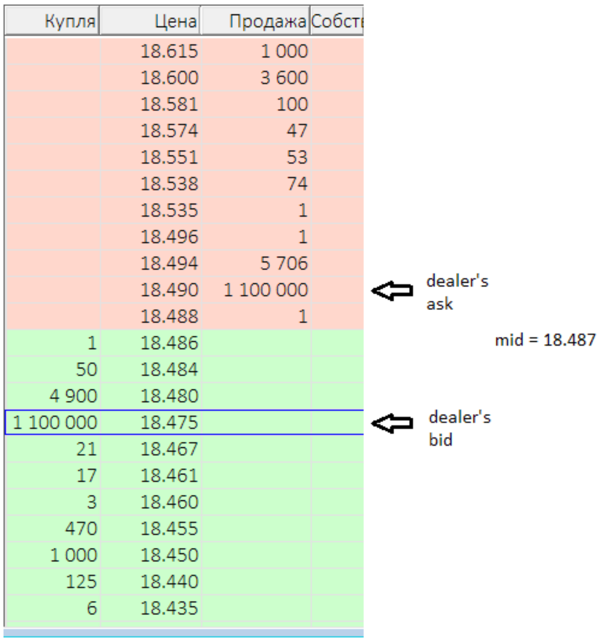
Key and substantial simplification here is that we don’t consider impact of dealer orders on prices and don’t consider partial execution of limit order, which is crucial issue for large size limit orders. So, we focused only on the question where is better to set orders, not which size is better to set.

One of the main problems for any HFT strategy is testing. Order book data consume a lot of space and are not so accessible as price data. Obvious solution here is data generation. One more benefit from this approach is possibility to assess statistical properties of strategy. We use two different approaches of simulation besides usual historical backtest. Another problem is data availability for academic research: limit order book history, market order book history and tick quotes usually are either unavailable at all or available for a fee.

Work is structured following way: in Chapter I we consider a formalism for optimal market-maker strategy. Also, we formulate simplest equidistant dealer strategy mentioned above which is used as a benchmark. In Chapter II we provide the market microstructure research, necessary to implement and test strategies on USD/RUB pair traded on Moscow exchange (most traded pair). In Chapter III we perform several types of simulation and compare strategies results.

Chapter 1. Task definition and theoretical models

We need to introduce some notation to go further. Consider regular limit order book (or depth of market, Fig.1.).



Купля	Цена	Продажа	Собст
	18.615	1 000	
	18.600	3 600	
	18.581	100	
	18.574	47	
	18.551	53	
	18.538	74	
	18.535	1	
	18.496	1	
	18.494	5 706	
	18.490	1 100 000	
	18.488	1	
1	18.486		
50	18.484		
4 900	18.480		
1 100 000	18.475		
21	18.467		
17	18.461		
3	18.460		
470	18.455		
1 000	18.450		
125	18.440		
6	18.435		

Fig.1. Depth of Market. Dealer limit orders.

There are limit orders to buy (bids) and to sell (asks). When exchange gets market order it match it with limit orders and deals happen. On image above dealer's limit orders are easily distinguishable by volume (large sizes). We denote current mid-price (or briefly mid) as $\boldsymbol{s_t}$.

Dealer prices for limit orders we denote as $\boldsymbol{p_b}$ for bid limit order and $\boldsymbol{p_a}$ for ask limit order. Also, it is convenient to introduce distances from mid: $\boldsymbol{\delta_b = mid - p_b}$ and $\boldsymbol{\delta_a = p_a - mid}$.

We consider mid-price as stochastic process that follows Brownian motion: $\boldsymbol{dS = \mu dt + \sigma dZ}$ (see assumptions of model below). We don't consider how market order execution affects mid-price (so-called temporal price impact), for simplicity we consider mid-price evolving as independent process.

Current inventory (or position) and current cash are denoted as $\boldsymbol{q_t}$ and $\boldsymbol{x_t}$ accordingly. Current wealth of the dealer therefore is $\boldsymbol{W_t = x_t + q_t s_t}$. Basically, our goal is to maximize some utility function of final wealth.

Virtual or instant price impact is the maximum price shift during market order execution: $\boldsymbol{\Delta p}$ = ***abs* (*best execution price − worst execution price*)**.

We don't consider partial execution of orders in this work, and it is significant simplification. If price impact Δp is greater than limit order distance δ then we treat it as filled. We assume that market orders will fill our limit orders with some Poisson intensity $\boldsymbol{\lambda}$ which obviously depends on order price distance δ. The further order placed the rarely it will be filled. We denote these intensities (also called *arrival rates*) as $\boldsymbol{\lambda_b(\delta_b)}$ and $\boldsymbol{\lambda_a(\delta_a)}$ for bid and offer orders accordingly. Moreover, we assume these Poisson processes are stable in time (intensities doesn't change for a time horizon).

Task here is to find the optimal way to set limit order in the process of market-making. Here we consider a formalism of this process: exponential utility maximization strategy introduces in [Avellaneda, Stoikov 2008]. A very brief explanation of the model is provided next section.

In this work we research this strategy applicability in more details. The goals of this research are:

- consider the model applicability for the USD/RUB (TOM) currency pair, i.e., check model assumptions;
- estimate market microstructure parameters;
- test model on historical and generated data;
- compare results with benchmark strategy.

Avellaneda-Stoikov model.

Brief model description

Firstly, we introduce expected exponential utility $\boldsymbol{U(W_t)} = \boldsymbol{E[-\exp(-\gamma * W_t)]}$, where $\boldsymbol{\gamma}$ is risk aversion of the dealer (strategy parameter). The goal (optimal performance function) is to maximize final expected utility by control distances δ_b and δ_a:

$$J(x, q, s, t) = \max_{\delta_b, \delta_a} E\left[-e^{-\gamma * W_T}\right]$$

where $\boldsymbol{T}$ is time-horizon of the strategy.

As it is shown in [Ho, Stoll 1981] and [Davis 1977], this task can be reformulated in terms of dynamic programming (Hamilton-Jacobi-Bellman equation):

$$J_t + J_s \mu_s s + \frac{1}{2} J_{ss} \sigma_s^2 s^2 + \max_{\delta_b} \lambda^b(\delta_b)[J(x - s + \delta_b, q + 1, s, t) - J(x, q, s, t)] \\ + \max_{\delta_a} \lambda^a(\delta_a)[J(x + s + \delta_a, q - 1, s, t) - J(x, q, s, t)] = 0$$

With border condition $J(x, q, s, T) = U(W_T)$.

Next key issue is an assumption about decomposition of utility function:

$$U(x, q, s, t) = -\exp(-\gamma x) \exp(-\gamma \theta(q, s, t))$$

for some $\boldsymbol{\theta(q, s, t)}$. It should be notices that it is only relevant in case of simple Brownian motion (not geometric). So, we can simplify partial derivative equation and exclude dependence on cash. It is quite logical that market order parameters depend only on inventory:

$$\frac{\partial \theta}{\partial t} + \mu \frac{\partial \theta}{\partial s} + \frac{1}{2} \sigma^2 \left(\frac{\partial^2 \theta}{\partial s^2} - \gamma \left(\frac{\partial \theta}{\partial s} \right)^2 \right) + \frac{1}{\gamma} \max_{\delta_b} \lambda^b(\delta_b) \left(1 - e^{-\gamma(r_b - p_b)}\right) \\ + \frac{1}{\gamma} \max_{\delta_a} \lambda^b(\delta_b) \left(1 - e^{-\gamma(r_b - p_b)}\right) = 0$$

With border condition $\theta(q, s, t) = qs$.

Here p_a and p_b are ask and bid prices of limit orders and $\boldsymbol{r_b}$ and $\boldsymbol{r_a}$ are so-called reservation prices:

$$r_b = \theta(q+1, s, t) - \theta(q, s, t)$$

$$r_a = \theta(q, s, t) - \theta(q-1, s, t)$$

The meaning is the following: reservation prices are prices of dealer indifference to inventory changing. For example, definition for bid reservation price is $U(x, q, s, t) = U(x-p_b, q+1, s, t)$.

Using first-order optimality condition for *max* expressions in PDE above we can obtain following equations for δ_b and δ_a:

$$s - \delta_b = r_b - \frac{1}{\gamma}\ln\left(1 - \gamma\frac{\lambda^b(\delta_b)}{\frac{\partial\lambda^b}{\partial\delta_b}(\delta_b)}\right)$$

$$s + \delta_a = r_a + \frac{1}{\gamma}\ln\left(1 - \gamma\frac{\lambda^a(\delta_a)}{\frac{\partial\lambda^a}{\partial\delta_a}(\delta_a)}\right)$$

Assumptions of the model

To solve mentioned above PDE we need some simplifications and assumptions about market microstructure to estimate $\lambda_a(\delta_a)$ and $\lambda_b(\delta_b)$:

- Brownian motion of prices:

$$dS = \mu dt + \sigma dZ$$

As mentioned above, we consider simple Brownian motion instead of geometric one. For very short intervals like several hours or a day it is not crucial whether geometric or simple Brownian motion to use, but the last one is simpler for analytical solution research.

- Volume distribution follows power law.

Distribution of market and limit order size are well studied. See, for example [Gopikrishnan et al. 2000], [Gabaix, X. 2006] and [Maslow, Mills 2001]. Consensus here is that market order size obeys power law. Distribution of limit order sizes is not so clear and seems follows log-normal distribution, see also [Maslow, Mills 2001]. Model assumption for density of market order size is

$$f_V(x) \propto x^{-(\alpha+1)}$$

where $\boldsymbol{\alpha}$ is so-called tail exponent or tail index.

- Price impact on average is proportional to log of volume of market order.

There are several studies of price impact like [Potters, Bouchaud, 2003] and [Weber, P. and Rosenow, B. 2003]. They focused mostly on temporal structure of impact. There is no consensus about price impact dependency on volumes, but there are two hypothesis:

$$\Delta p \propto V^{-\beta} \quad \text{and} \quad \Delta p \propto \log V$$

Exhaust research of price impact from theoretical point of view for different types of participants (regular and large investors) is provided in [Gabaix, X. 2006], where theoretical estimation for β are obtained.

In Avellaneda-Stoikov model the logarithmic relationship is supposed.

- Intensity of market orders is constant.

We consider market order appearance as Poisson process with constant intensity. This assumption is not true in strict sense, but for short time horizon it is not so implausible. We denote intensity as $\boldsymbol{\Lambda}$ and not differ intensity for buy and sell orders for simplicity. Using this assumptions, intensities (arrival rates) for buy and sell market orders are:

$$\lambda(\delta) = \Lambda P(\Delta p > \delta) = \frac{\Lambda}{\alpha} e^{-\alpha\delta/K} = Ae^{-k\delta}$$

Here $\boldsymbol{A} = \boldsymbol{\Lambda}/\boldsymbol{\alpha}$ is intensity parameter and $\boldsymbol{k} = \boldsymbol{\alpha}/\boldsymbol{K}$ is sensitivity to spread.

- Quadratic asymptotic expansion for $\theta(q, s, t)$ (Taylor series):

$$\theta(q, s, t) = \theta_0(s, t) + \theta_1(s, t)q + \frac{1}{2}\theta_2(s, t)q^2$$

Using all above assumptions it is possible to simplify PDE and find analytical solution.

Solution

Under assumptions Avellaneda-Stoikov model (or simply AS-model) PDE solution is:

$$\theta_1 = s + \mu(T-t)$$

$$\theta_2 = -\sigma^2\gamma(T-t)$$

Optimal spread is:

$$\delta_b + \delta_a = -\theta_2(s, t) + \frac{2}{\gamma}\log\left(1 + \frac{\gamma}{k}\right) = \sigma^2\gamma(T - t) + \frac{2}{\gamma}\log\left(1 + \frac{\gamma}{k}\right)$$

which is around the mid reservation price:

$$r\text{-}price = \frac{r_b + r_a}{2} = \theta_1(s, t) + \theta_2(s, t)q = s + \mu(T - t) - \sigma^2\gamma(T - t)q$$

Respectively optimal bid and ask prices are:

$$p_b = r\text{-}price - (\delta_b + \delta_a)/2$$

$$p_a = r\text{-}price + (\delta_b + \delta_a)/2.$$

Benchmark strategy

In this strategy we set orders on some constant equal distance from the current mid-price. Mathematically it is rather simple: just use constant spread around the mid-price:

$$\delta_a + \delta_b = const$$

$$r\text{-}price = mid$$

Optimal bid and ask prices are the same:

$$p_b = r\text{-}price - (\delta_b + \delta_a)/2$$

$$p_a = r\text{-}price + (\delta_b + \delta_a)/2.$$

Two make strategies results comparable it makes sense to set spread here equals to average spread of AS-model.

Chapter 2. Market microstructure research

We need to check model assumptions and estimate market microstructure parameters. In this section we provide research for USD/RUB currency pair traded on Moscow Exchange. Assumptions to be checked:

- distribution of volumes obeys power law;
- logarithmic relationship between average price impact and order size.

If assumptions are true, we need to estimate parameters:

- power law tail index $\boldsymbol{\alpha}$;
- logarithmic regression slope coefficient $\boldsymbol{K}$;
- intensity of market orders $\boldsymbol{\Lambda}$.

Preprocessing data

Table of all deals for USD/RUB (TOM) was used as an input. Data contains information about date and time of the deal, its size, price, and the side of initiator of the deal (buyer or seller). Also, limit order book history information was used, which contains full bid and ask quotes history. Preprocessing consists of three steps:

1. grouping deals to market order with price impact calculation (using deals table).
2. add info about best bid and best ask prices at deal moment (using limit order book history files) for historical simulation.
3. add features for further analysis and filtering: day of week, hour.

Market order volume distribution

Trade session for USD/RUB (TOM) instrument on MOEX is quite long: from 07:00 till 23:00. But most active hours are from 10:00 (historical open time) till 17:45 (TOD session closing). Therefore, we filter data for these most active trade hours (from 10:00 till 17:59) before plot the histogram (Fig.2).

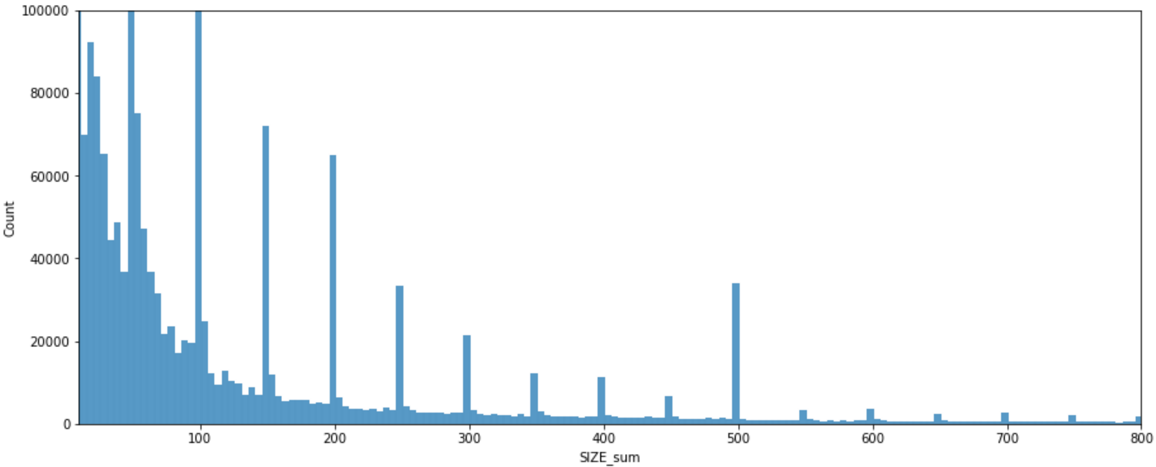


Fig.2. Market order volume distribution

It is easy to notice peaks of volumes of factor of 50 lots (1 lot = 1000$). Also, there are high peaks at 1000, 2000, 2500, 3000 and 5000 lots. We suppose market order size as random variable consists of three components:

- non-rounded lots caused by regular participants $\boldsymbol{V_r}$;
- medium size rounded to 50 (factor of 50 orders or briefly x50) $\boldsymbol{V_{x50}}$;
- large size rounded to 500/1000 (x500 orders) $\boldsymbol{V_{x500}}$.

Total count on histogram for any size $V = V_r + V_{x50} + V_{x500}$. We suppose that it is caused because bank dealers prefer to round their volumes to "nice looked" values. And the larger bank the more open currency position it can get and the rougher rounding it will use.

This structure of distribution make analysis a bit complicated. We need to decompose distribution into parts. For this we used rather simple smoothing technique for deals count:

- ✓ to estimate "true" regular volumes count V_r for factor of 50 orders we used simple average of neighbors: $V_r(50x) = (V_r(50x-1) + V_r(50x+1)) / 2$.
- ✓ extra counts for x50 orders is "true" V_{x50} volume.
- ✓ same procedure is performed for V_{x50} vs V_{x500} volumes.

Decomposition is presented on Fig.3.

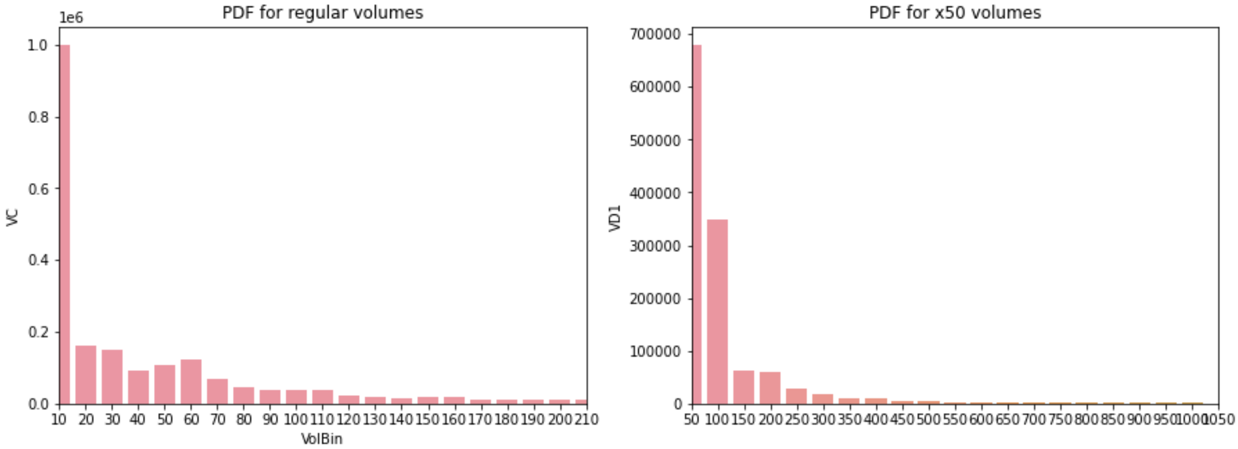


Fig.3. Volume distributions for regular and x50 orders

There are plenty of method for tail exponent estimation. Usually Hill's estimator is used:

$$\alpha(k) = k \sum_{j=1}^{k} \log \frac{x_{N-j+1}}{x_{N-k}}$$

But method is quite sensitive to start index k selection, and as it is shown in [I.I. Komarov, H.L. Chen. 2010] it works well mostly for exponent less than 1.5. Exhaustive review of tail index estimation method is provided in [Munasinghe, R. et al., 2020].

We used method of linear regression for log complement cumulative distribution function:

$$CCDF(V) \propto V^{-\alpha},$$

where *CCDF(x) = P(X > x) = 1 – CDF(x)*. For more details about this method see for example [Maslow, Mills 2001] or [Gabaix X. et al., 2006].

Using linear regression in log-log scale it is easy to estimate **α** as slope coefficient. We use GLS (generalized least squares) to estimate slope coefficient. We plot regression for all data and decomposed data as well. Results are displayed in Table 1. See Fig.4 and 5 for regression plots.

	All orders	Regular orders	x50 orders
R-squared	0.991	0.992	0.993
F-statistics	5.3e+04	6.3e+04	6.7e+04
P-value (F-statistics)	0.00	0.00	0.00
Slope	**-1.9950**	**-1.965**	**-2.305**
t-statistic for slope	-230.56	-249.96	-259.8
P-value (slope)	0.000	0.000	0.000

Table 1. Regression results.

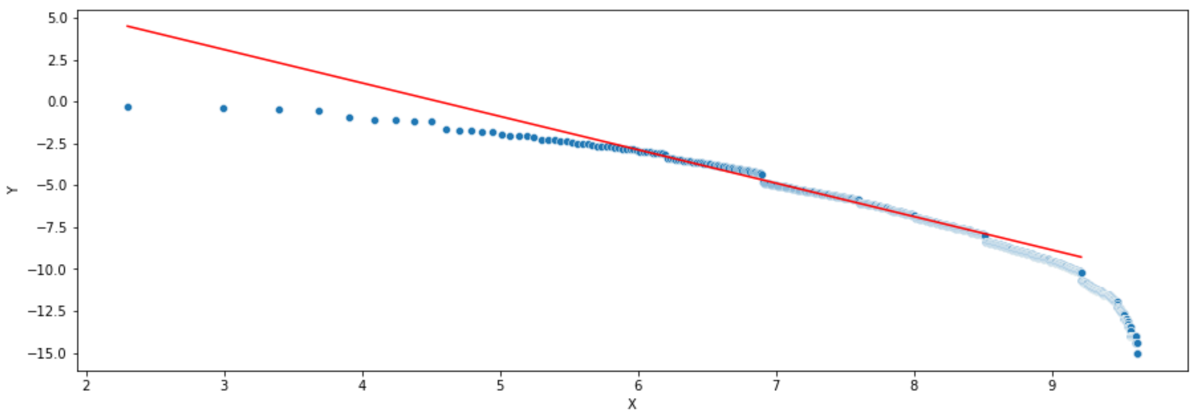


Fig.4. CCDF regression for all market orders

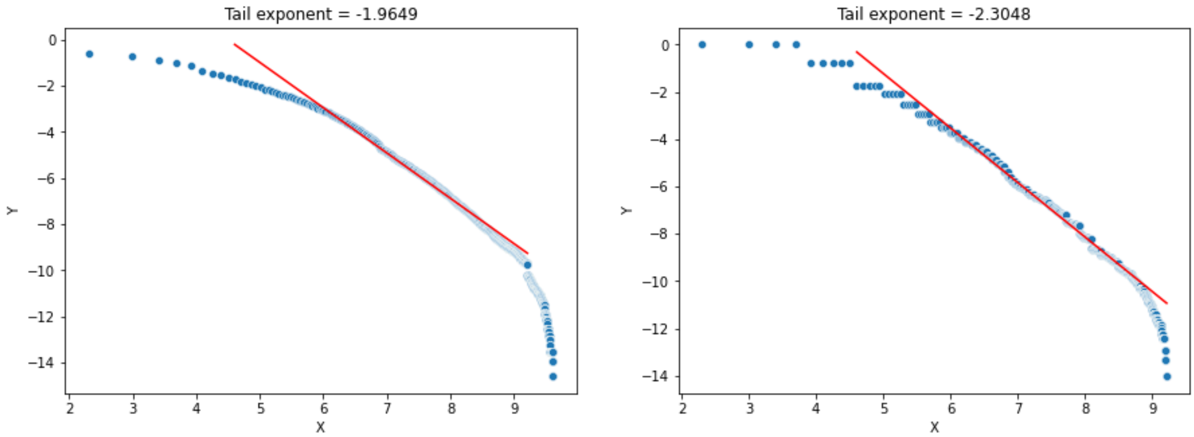


Fig.5. CCDF regression for regular (left) and x50 (right) market orders

For simulations we used **α** equals to **2.0**.

Detailed regression info is given in Appendix. Some other estimations with method descriptions are also given in Appendix.

Price impact

Next issue is to check hypothesis of log dependency of price impact and try to estimate regression parameters. Remind that model uses coefficient of regression

$$\overline{\Delta p} = K \log V$$

We used sliding window to average volumes and additional filtering to exclude impact anomalies in data as well (like very large order sizes and impact outliers). See the plot of average impact on volume in Fig.6.

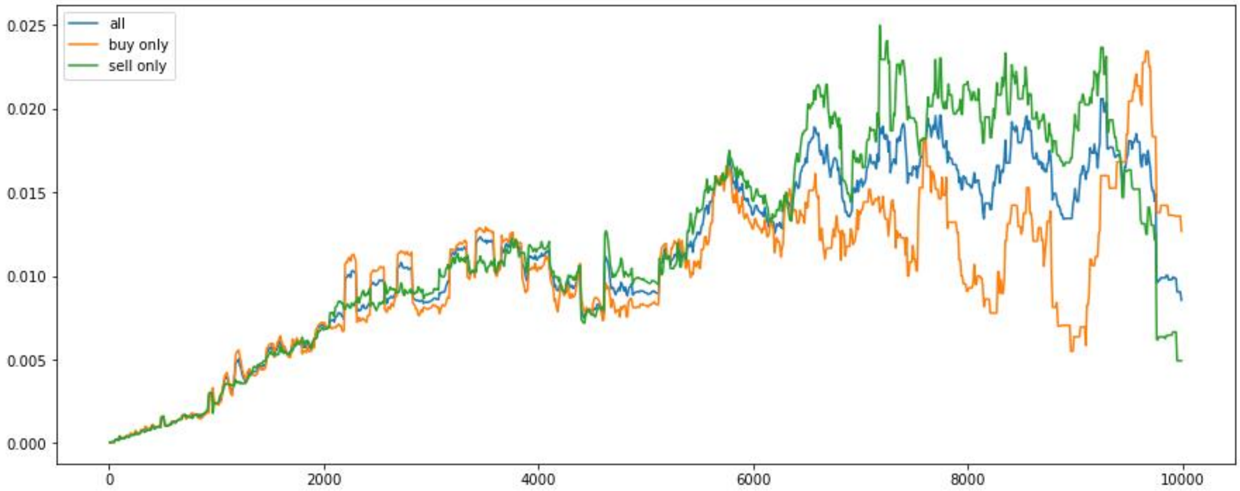


Fig.6. Average impact – volume plot

It looks like linear for small impacts/volumes and likely logarithmic for large ones. Checking it we plotted linear regression (Δp ~ V) for volumes less than 800K lots and log-linear one (Δp ~ log V) for volumes greater than 800K lots, see fig.7.

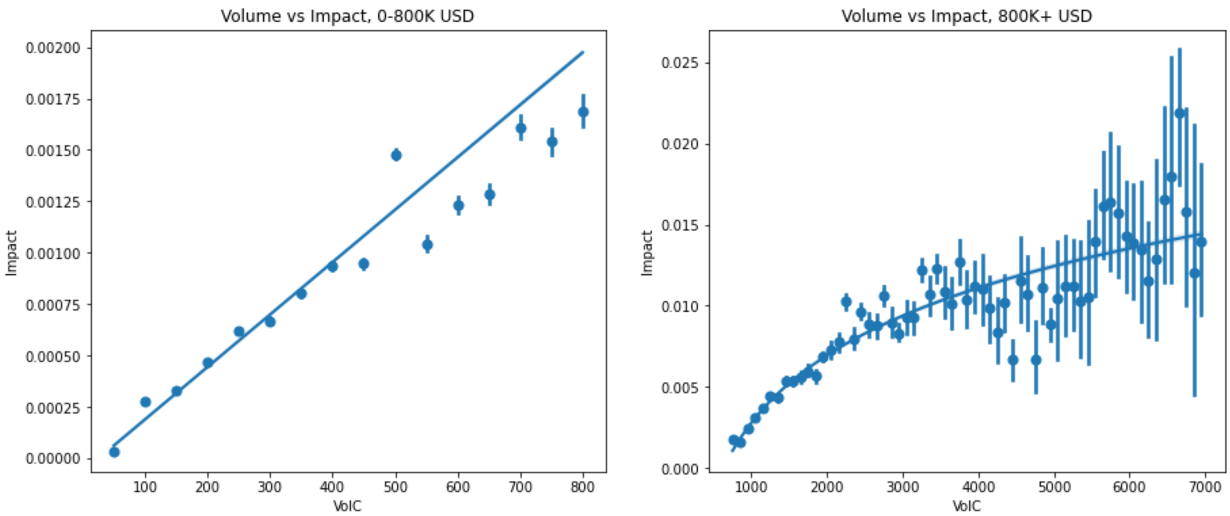


Fig.7. Average impact – volume regression plots: linear for small volumes (left) and log-linear for larger ones (right).

Regression results are displayed in Table 2.

	Volumes 0-800K, linear	Volumes 800K+, log-linear
R-squared	0.947	0.819
F-statistics	1387	3250

P-value (F-statistics)	0.00	0.00
Slope	2.1e-06	**0.0067**
t-statistic for slope	37.3	57.0
P-value (slope)	0.000	0.000

Table 2. Regression results for all orders.

Both regressions are significant with high enough determination coefficient.

We interested mostly large and more frequent market orders, therefore this relationship was tested for x50 orders as well. Results are mostly the same (see Fig.9 and Table 3).

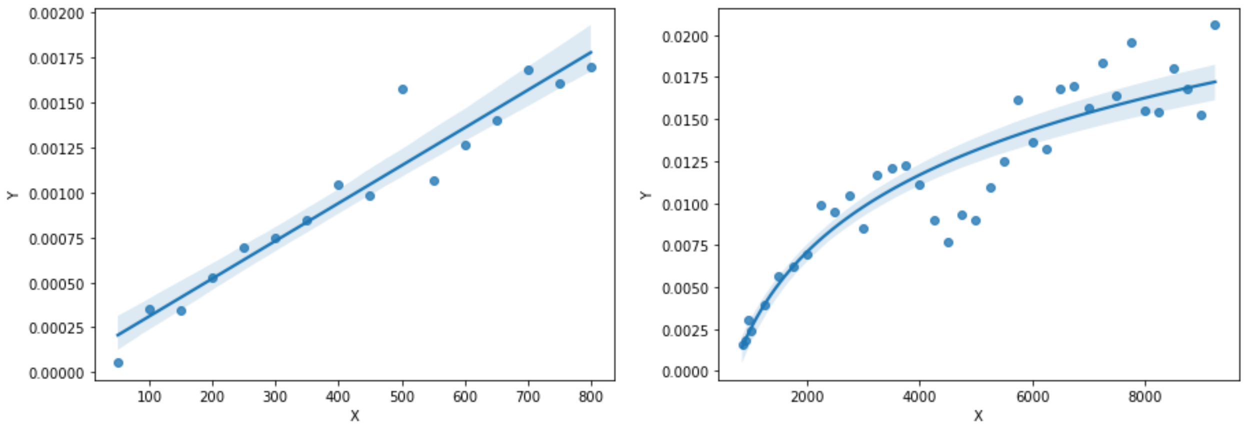


Fig.7. Average impact – volume regression plots for x50 market orders.

	Volumes 0-800K, linear	Volumes 800K+, log-linear
R-squared	0.924	0.860
F-statistics	170	216
P-value (F-statistics)	0.00	0.00
Slope	2.1e+06	**0.0066**
t-statistic for slope	13.0	14.7
P-value (slope)	0.000	0.000

Table 3. Regression results for x50 orders.

Again, regression is significant with high determination coefficient. Slope is equal to 0.0066 that is almost the same as previous one. We used value **0.0067** for simulations.

Market volatility

It is the last internal parameter of the model. We used standard procedure for volatility estimation. Daily quotes for USD/RUB TOM (close prices) for period from Jan'2021 till Nov'20201 are used, then daily basis standard deviation of first differences of prices were calculated with periods equals to 21, 50 and 100 days (see Fig.8):

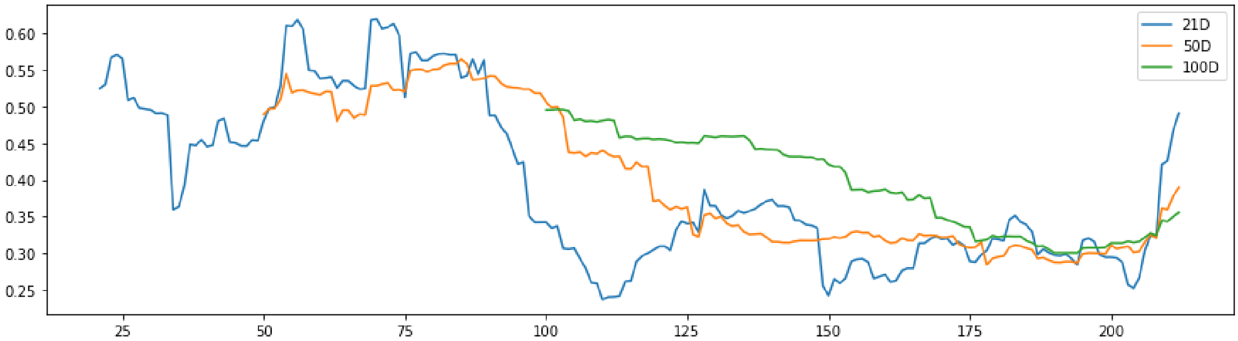


Fig.8. Market volatility.

For simulation value **0.50** (daily basis) or the same **0.17** (hour basis) is used (last value for 21-day standard deviation).

Some other market parameters

Market order frequency and buy orders share

Order frequency is not the part of the AS-model solution itself, but it is useful for market simulations. Also, it is interesting to check how wide is the assumption about frequency constancy from the truth. We used filtering to exclude extra active first 10 minutes of trades of main session.

Orders average intensity depends on time and date and buy share is also not constant. E.g., first and last trade hours are most active which is usual for financial markets. See Fig.9.

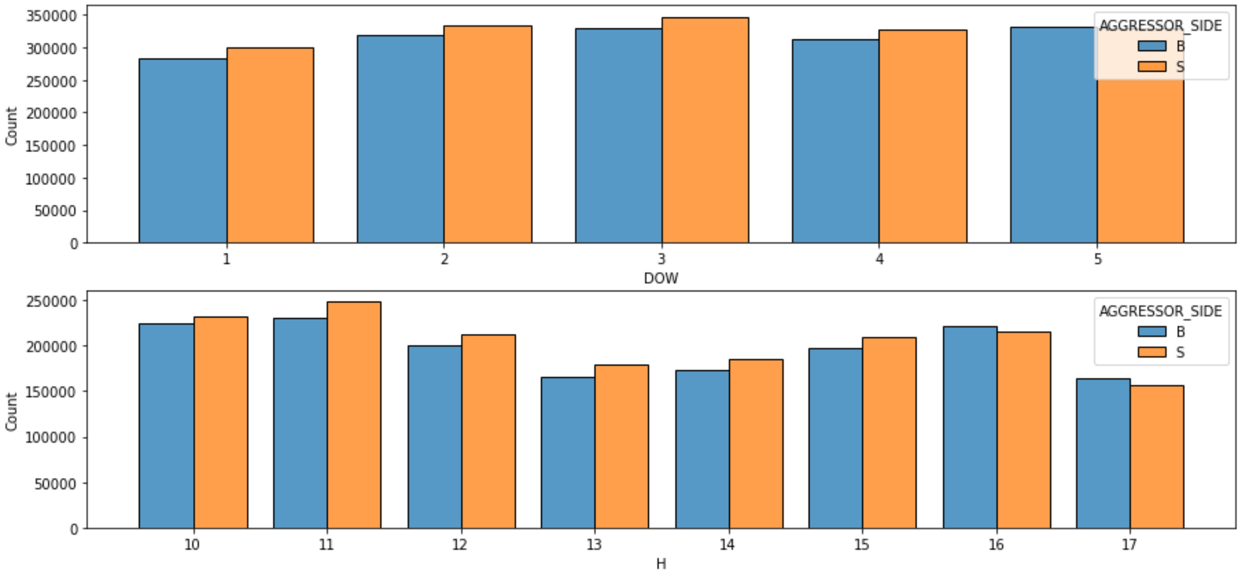


Fig.9. Average count of orders depends on day of week and hour.

Average count of orders per hour and buy orders share changing over a day is illustrated on Fig.10.

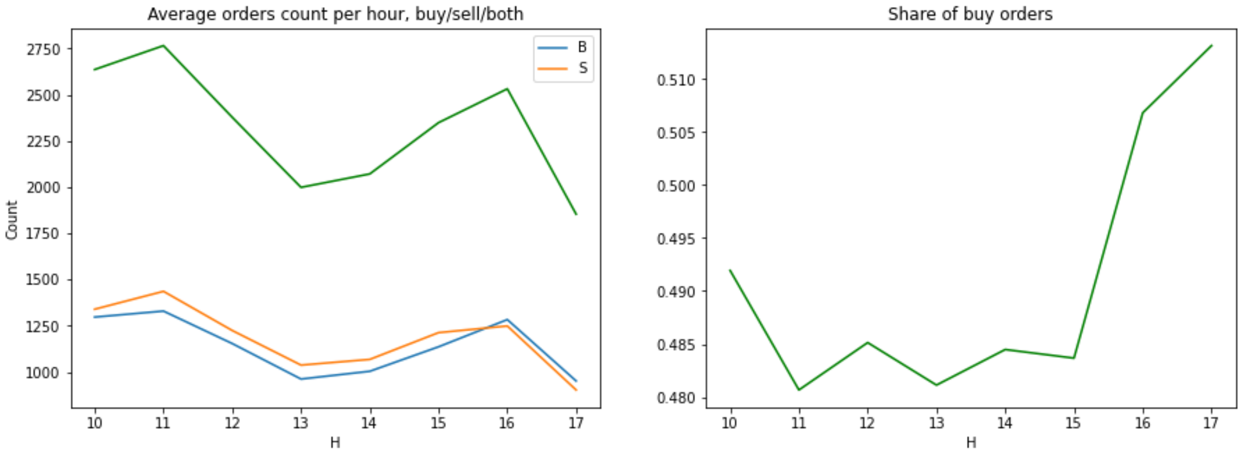


Fig.10. Average count of orders per hour and buy orders share.

We used 2300 orders an hour as orders intensity in simulation, and we don't distinguish frequency of buy and sell orders and put buy/sell ratio to 1 (50/50) during simulations.

Spread distribution

Spread distribution is essential for generation bid and ask prices separately. We used it in advanced simulation. Spreads distribution is displayed on Fig.11.

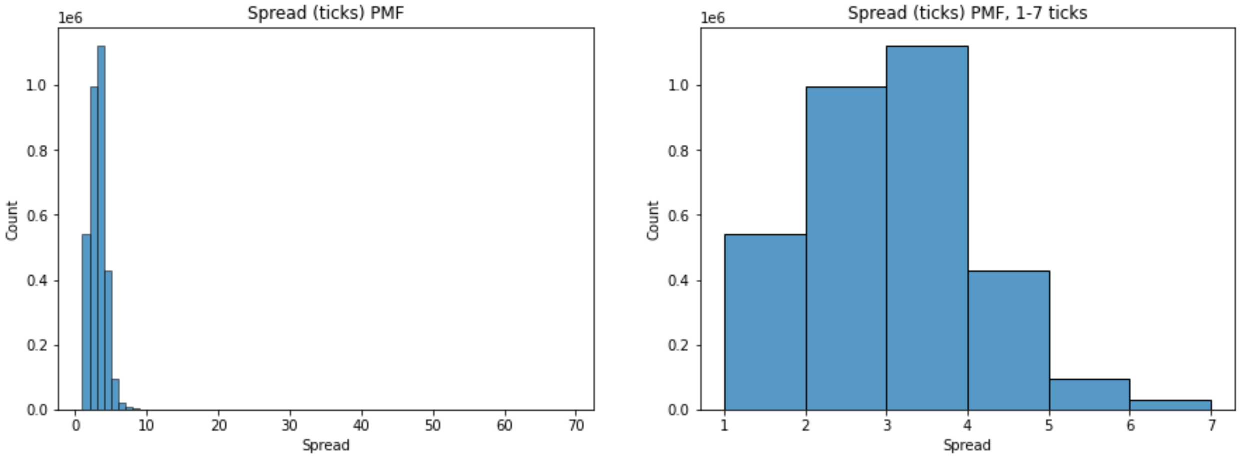


Fig.11. Spread distribution.

More than 99% of time spread is less than 7 ticks.

Summary

We can make following conclusions:

- Power law fits empirical distribution of order size sufficiently well. Tail index **α** = **2.0**.
- Price impact for large orders is proportional to logarithm of size on average. The regression slope coefficient K is equal to **0.0067**.
- Hour basis standard deviation of prices **σ** is about **0.17**.
- Average frequency of market orders **Λ** is about **2300** per hour.
- Share of buy orders roughly is close to **50%**.

Chapter 3. Strategies tests

In this chapter we provide results of strategy tests and comparison with the benchmark strategy.

Simple base model

First of all, we tested model in the same conditions as in original paper with the same simulation algorithm. Algorithm is rather simple:

0. Start with $t = 0$, choose small enough dt.
1. For the next time moment $t_{i+1}=t_i + dt$ generate next mid-price s_{t+1} in accordance with Brownian motion process.
2. With intensity Λdt (Poisson stochastic process) generate market order event. With probability 50% it is a buy order.
3. Compute optimal market order prices p_b, p_a, distances δ_b, δ_a and reservation price $r\text{-}price$ based on current market state.
4. With probability $\lambda(\delta) = Ae^{-k\delta}$ market order fills limit order.
5. If order executed, then update cash and inventory.

Algorithm for equidistant strategy is mostly the same with only difference in step 3:

3. Compute optimal market order prices p_b, p_a using average spread from AS-model (for comparability of results).

Parameters for simulation are the following:

- Simulation parameters: time horizon T = 1, dt = 0.005.
- Market parameters: α = 1.5, Λ = 210, K = 1, σ = 2, μ = 0.
- Start price is 100.

Fig.12 illustrates how the AS-model strategy works providing compare of limit order prices and inventory.

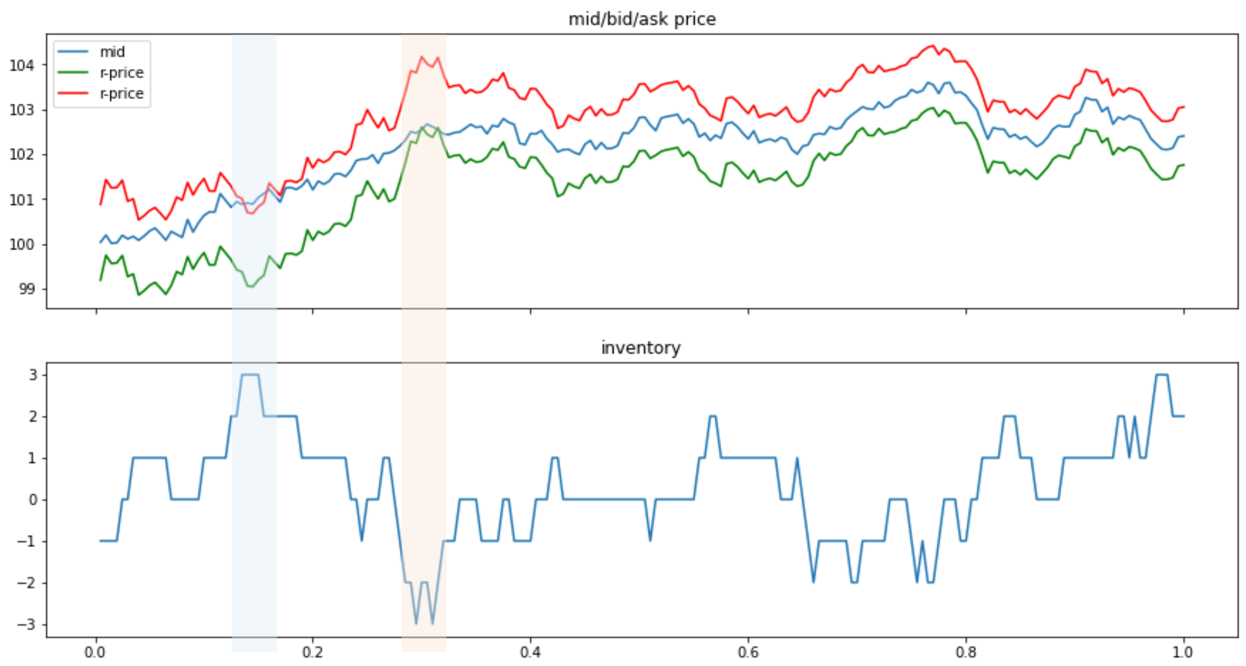


Fig.12. Limit order prices and inventory.

When inventory is high by absolute value then model shifts limit order prices: it shifts them down for positive inventory (blue region in figure) and down for negative ones (red region). It is quite simple and logical: if you have excess of inventory then just quote attractive prices to sell and vice versa.

Profit/loss curve is smooth. Spread is decreasing linearly till the end of time horizon, which is the consequence of less risks at the end of trading period (see Fig. 13).

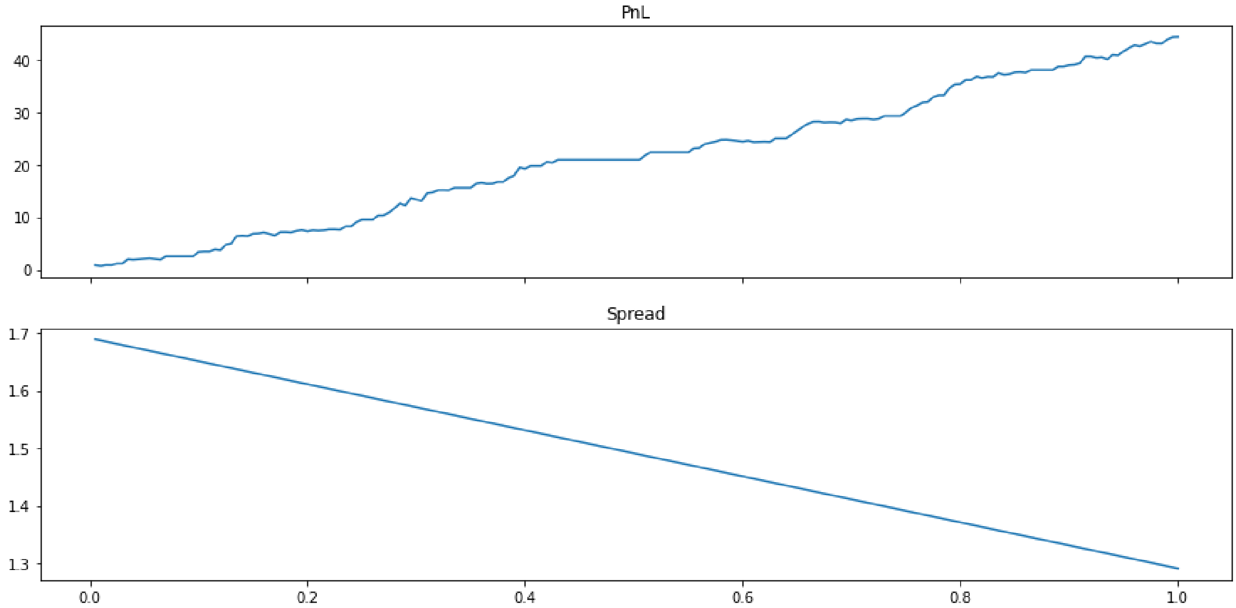


Fig.13. PL(top) and spread(bottom).

Then, we run 1000 simulations to obtain profit-loss distribution and final inventory distribution (see Fug.14 and Table 4).

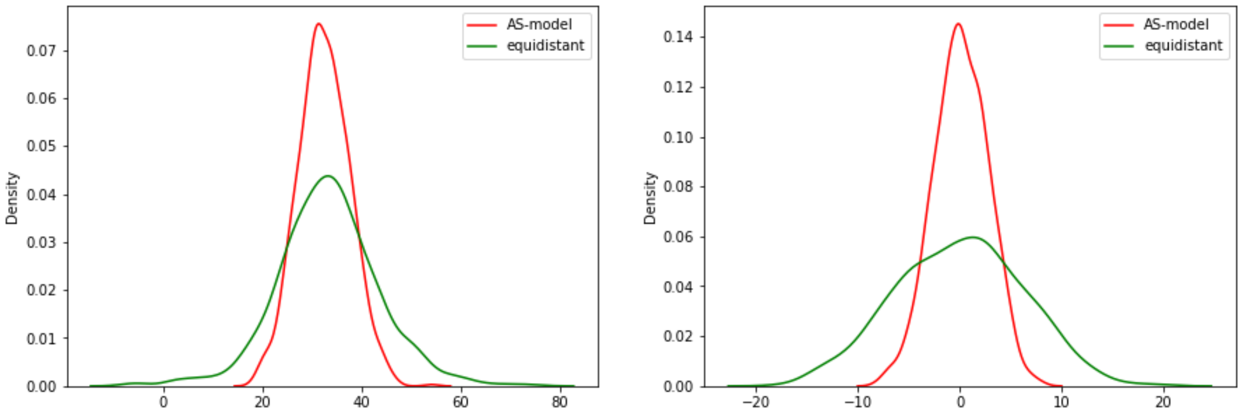


Fig.14. PL(left) and final inventory(right) distribution.

Strategy	Average spread	Mean PL	Std dev PL	Mean final Q	Std dev final Q	Mean PL/deal	Std dev PL/deal
AS-model	1.47	32.4	5.3	-0.12	2.66	0.70	0.07
Benchmark	1.48	34.5	10.2	-0.14	6.59	0.80	0.22

Table 4. Simulation results, $\gamma = 0.1$.

Here "std dev" is standard deviation, "final Q" is final inventory, "PL/deal" is average profit a deal. AS-model strategy provides much more concentrate profit and inventory profile with almost the same mean in comparison with benchmark strategy, so, as expected, model is less risky.

It should be notice that increasing free parameter **γ** makes strategy more conservative (risk averse). Particularly, spread becomes wider (and therefore deals happens rarely), price shift becomes more aggressive. For illustration we run same 1000 simulations with **γ = 0.5**. Results are displayed below (Fig. 15 and Table 5).

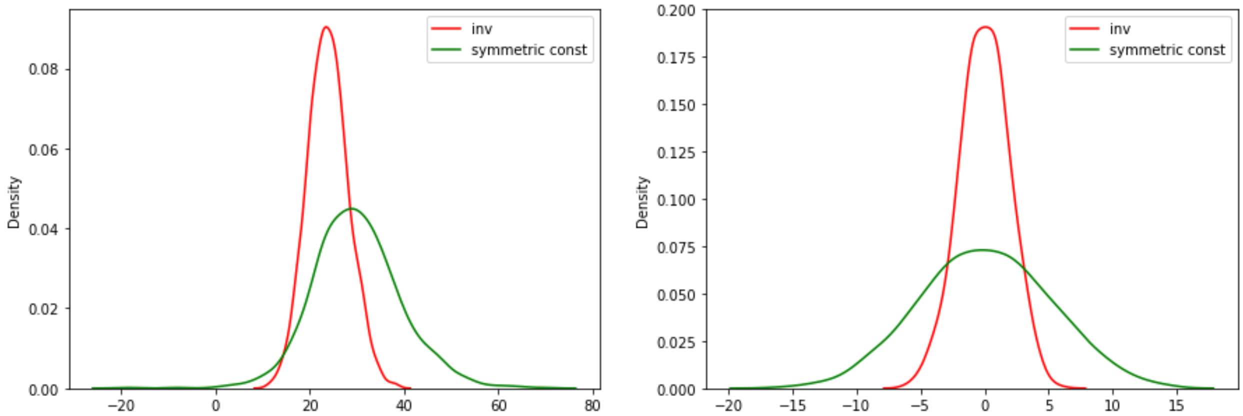


Fig.15. PL(left) and final inventory(right) distribution, **γ** = 0.5.

Strategy	Average spread	Mean PL	Std dev PL	Mean final Q	Std dev final Q	Mean PL/deal	Std dev PL/deal
AS-model	2.13	23.9	4.3	-0.01	1.9	0.6	0.06
Benchmark	2.12	30.0	9.3	0.06	5.1	1.1	0.27

Table 5. Simulation results, **γ** = 0.5.

Profits became less, but AS-model still is more concentrated in PL and inventory and thus less risky.

Base model for USD/RUB

In this simulation we used parameters estimated above and took into account the tick size for USD/RUB quotes (by rounding prices). Tick size is equal to 0.0025 ruble. Algorithm is the same as in the previous one.

Parameters for simulation are the following:

- Simulation parameters: time horizon T = 1 hour, dt = 0.5 second.
- Market parameters: $\alpha = 2.0$, $\Lambda = 2300$, $K = 0.0067$, $\sigma = 0.17$, $\mu = 0$.
- Start price is 70.
- Risk aversion $\gamma = 0.1$.

Again run 1000 simulations to obtain and compare profit-loss and inventory distributions. See results in Fig. and Table 6.

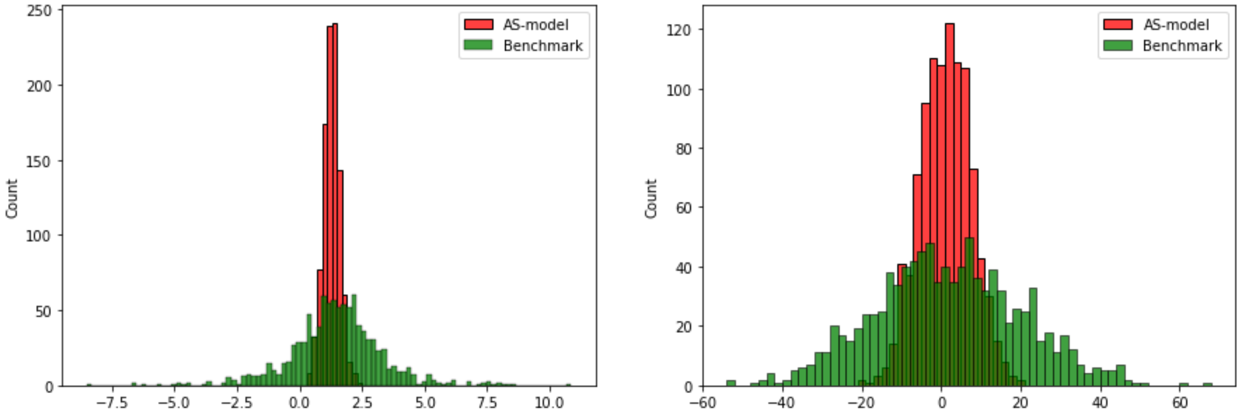


Fig.15. PL(left) and final inventory(right) distribution, $\gamma = 0.1$.

Strategy	Average spread	Mean PL	Std dev PL	Mean N	Mean final Q	Std dev final Q	Mean PL/deal	Std dev PL/deal
AS-model	0.0075	1.27	0.31	407	0.53	6.4	0.0031	0.0008
Benchmark	0.0075	1.48	1.93	370	1.11	18.6	0.0040	0.0052

Table 6. Simulation results, $\gamma = 0.1$.

where “Mean N” is average number of deals.

Overall looks similar, interesting issue that profit per deal is quite small and comparable with tick size. Again, AS-model produces more concentrated results with similar mean values.

For higher risk aversion ($\gamma = 0.5$) results are the following (Fig. 16 and Table 7).

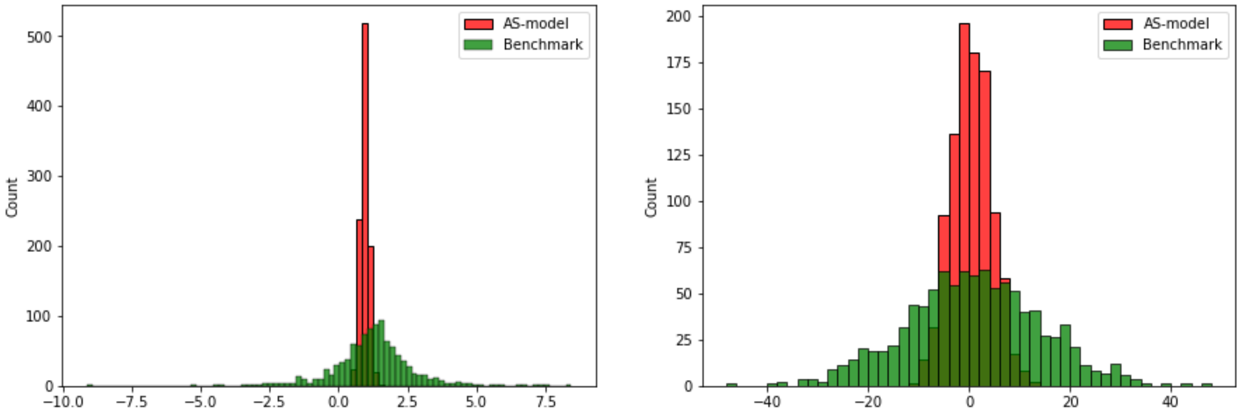


Fig.16. PL(left) and final inventory(right) distribution, $\gamma = 0.5$.

Strategy	Average spread	Mean PL	Std dev PL	Mean N	Mean final Q	Std dev final Q	Mean PL/deal	Std dev PL/deal
AS-model	0.0117	0.95	0.15	291	-0.08	3.96	0.0032	0.0005
Benchmark	0.0125	1.23	1.43	199	0.31	13.4	0.0062	0.0071

Table 7. Simulation results, $\gamma = 0.5$.

Again, AS-model provides more concentrated results and a bit less mean profit. Spread became wider and deals became sparser.

Historical simulation

Base algorithm of simulations is simple and fast, but mostly based on our assumptions about market microstructure. Much more interesting test strategies on real data. Here we provide results of historical simulations.

We ran simulation based on 500 randomly chosen trade hours from history of market order book for the period from Feb'2021 to Oct'2021. Tick size is 0.0025, benchmark strategy spread is equals to 4 ticks. Other parameters for AS-model are the following:

- Market parameters: $\alpha = 2.0$, $K = 0.0067$, $\sigma = 0.17$, $\mu = 0$.
- Risk aversion $\gamma = 0.2$ (to make average spread close to benchmark).

Results displayed in Fig. 17 and Table 8.

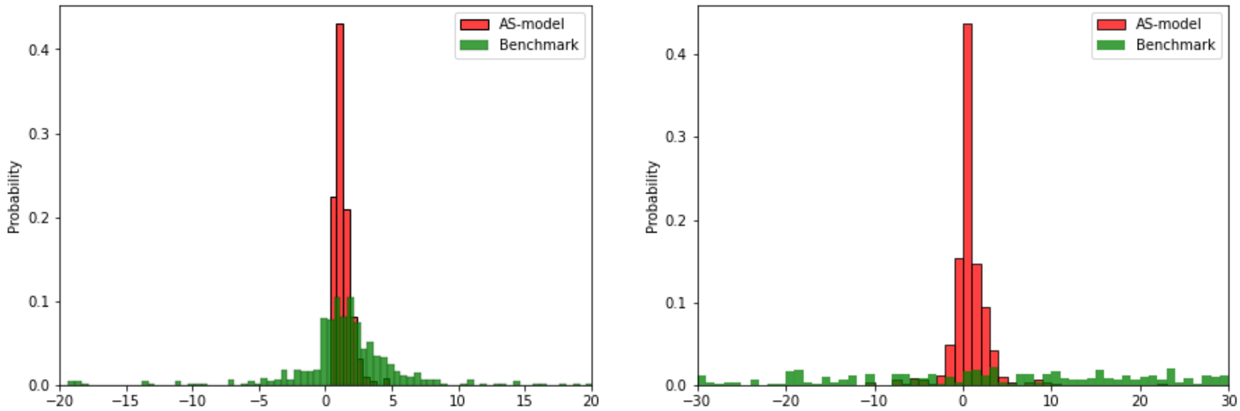


Fig.17. PL (left) and final inventory (right) distribution, $\gamma = 0.2$.

Strategy	Average spread	Mean PL	Std dev PL	Mean N	Mean final Q	Std dev final Q	Mean PL/deal	Std dev PL/deal
AS-model	0.0097	1.28	**0.62**	693	0.35	**3.0**	0.0019	0.0003
Benchmark	0.0114	1.68	**7.16**	524	9.80	**100.1**	0.0041	0.0079

Table 8. Simulation results, $\gamma = 0.5$.

Average spread is close to 0.01 (4 ticks), mean profit a bit better for benchmark. But surprisingly volatility of final inventory is huge for benchmark strategy. Volatility of profits is also much larger on historical data. It means that our simple simulation algorithm is far from reality. Meanwhile volatility of profit and inventory of AS-model is almost the same as in the base simulation.

Fig. 18 illustrates distribution of losses and maximum absolute inventory for strategies. Both metrics are significantly less and concentrated for AS-model.

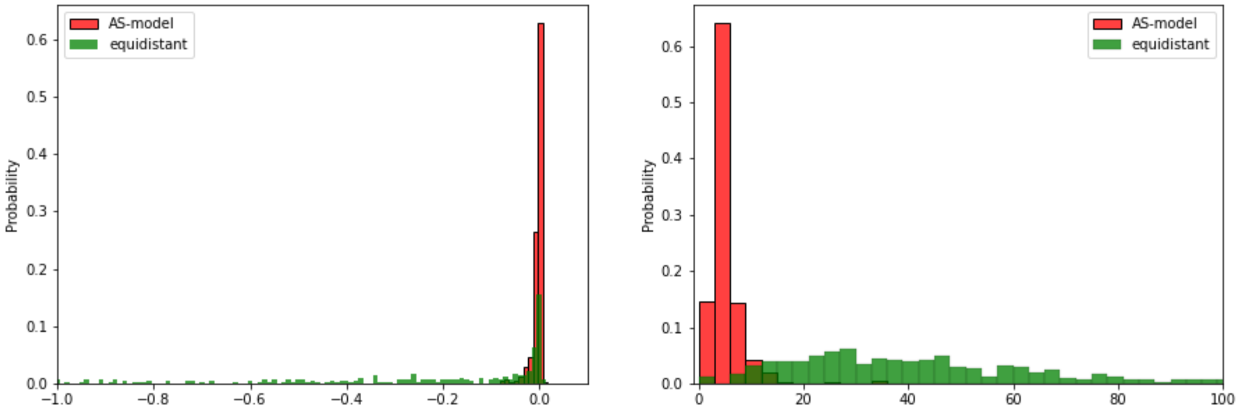


Fig.18. Max loss(left) and maximum inventory(right) distribution.

Let us look closer to an example of price path and corresponding inventory (7’Jul 2021, 11.00-12.00) at Fig.19.

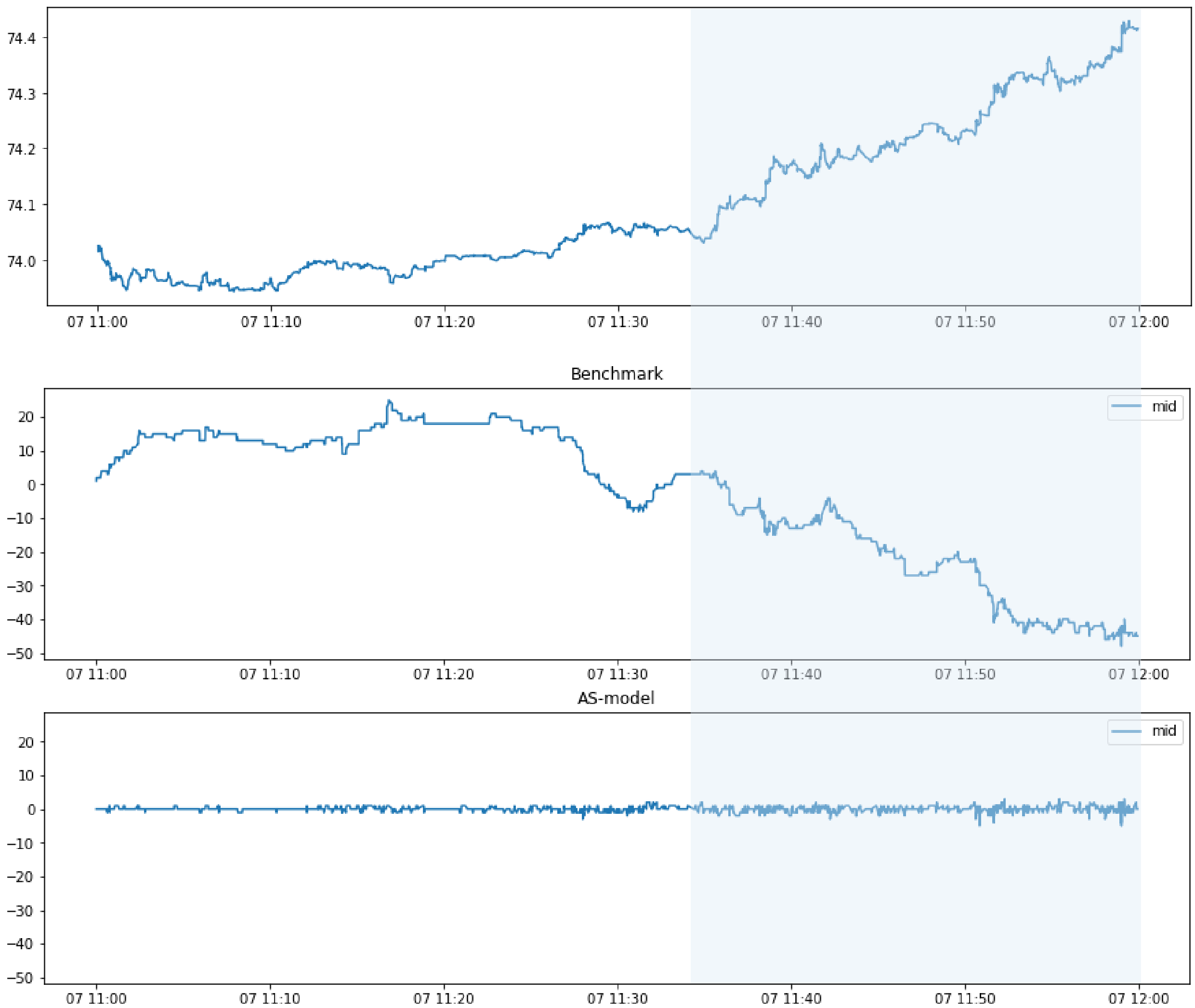


Fig.19. Price (top) and final inventory for benchmark (mid) and AS-model (bottom) strategies.

So, the key issue here is that AS-model efficiently avoids large inventory for the periods of strong trend price movement (blue region on a plots) while equidistant strategy tends to deal in countertrend manner.

Advanced simulation

As it is shown in historical simulation, simple simulation is far from reality. So, we try to improve simulation procedure to obtain comparable results. First idea is to use two-component volume-impact distribution: one for regular orders and another for x50 orders. Second idea is to generate both best bid and best ask prices as combination of Brownian motion and spread random variable.

Algorithm is following:

1. Using Poisson process with constant intensity Λ obtain series of time of market order events. This part is implemented using simple fact, that time between to events in Poisson process obeys exponential distribution with same parameter:
$$\Delta t \sim \text{Exponential}\ (\Lambda)$$
2. With probability 50% (close to historical share) new order is buy order.
3. With probability $\boldsymbol{P_{x50}}$ new order is x50-order, otherwise it is regular. Probability is estimated based on historical data and is equal to **36%**. We estimated this probability simply as share of x50 orders in market order book.
4. Using Brownian motion process obtain bid prices and round them to ticks. It seems to be simplification and probably fractional Brownian motion more precise.
5. Using empirical PMF for spread distribution generate spread for each bid and obtain ask prices.
6. Using empirical PMF for x50 order volumes/impacts generate x50 order parameter (volume, impact) OR using Kernel density estimation (KDE) model for regular order volumes/impacts generate regular order parameters.
It is an implementation of the idea of composite distribution. We consider here 2-dimensinal distribution of random variable (volume, impact) to make process simpler and straightforward. PMF is limited with 12 ticks and $5 mln. KDE used “tophat” kernel with preprocessing to make scale of both dimensions close.

We ran 500 simulations. Parameters are the same as for historical test:

- Market parameters: $\alpha = 2.0$, $K = 0.0067$, $\sigma = 0.17$, $\mu = 0$.
- Spread for benchmark strategy is equal to 4 ticks.
- Risk aversion $\gamma = 0.2$.

Below test results of advanced generator are provided (see Fig. 20 and Table 9). Fig. 21 illustrates distribution of losses and maximum absolute inventory for strategies.

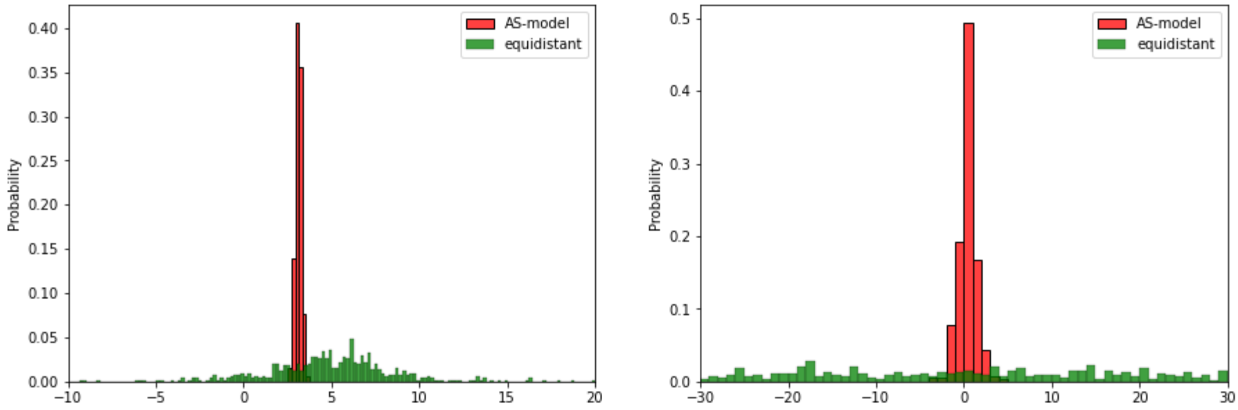


Fig.20. PL(left) and final inventory(right) distribution.

Strategy	Average spread	Mean PL	Std dev PL	Mean N	Mean final Q	Std dev final Q	Mean PL/deal	Std dev PL/deal
AS-model	0.012	3.1	0.17	1014	-0.09	1.1	0.0031	0.0001
Benchmark	0.011	5.0	4.00	986	-0.68	30.0	0.0050	0.0041

Table 9. Simulation results.

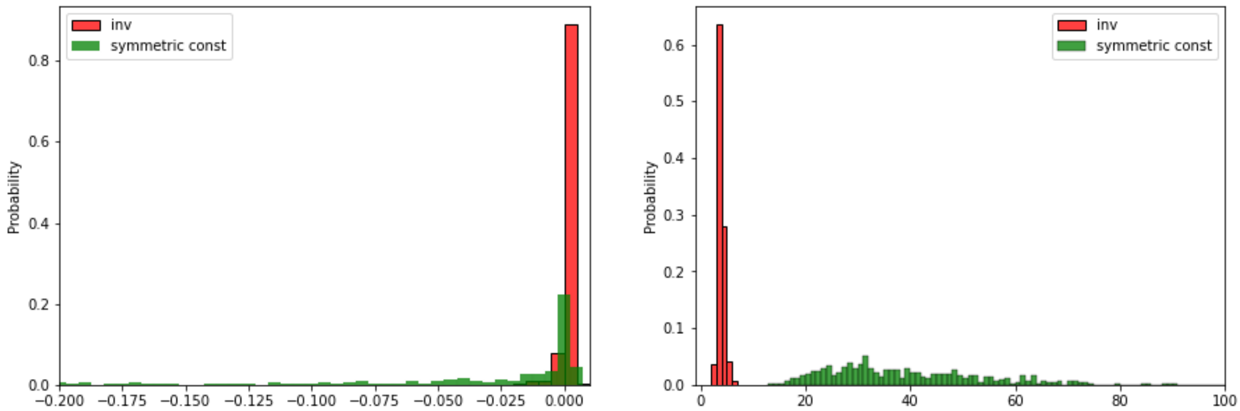


Fig.21. Max loss(left) and maximum inventory(right) distribution.

Results are quite similar to historical simulations, but volatility of profit and inventory for benchmark decreased. Overall, AS-model again provide significantly less risky and concentrated results.

Conclusions

We have researched microstructure properties of Russian currency market by the USB/RUB example (most traded pair on the Moscow Exchange). We find that distribution of volumes of large market orders follows power law with tail index equals to 2.0. Also, we find that distribution highly likely consists of several components, and estimation of these components separately provides similar results.

We investigate relationship between average instant price impact and market order size: it is linear for small-size orders and logarithmic for large-size ones. Analysis shown that mentioned above components have the same relationships. It is crucial for us because we interested in large and frequent orders (like 500 lots).

Finally, we have made simulations to test and compare results of benchmark equidistant strategy and strategy based on quasi-optimal Avellaneda-Stoikov model solution. We can conclude that advanced model has comparable mean profit and mean final inventory, but also significantly less volatility of these metrics. It means that AS-model strategy is much less risky because of less chance of adverse inventory revaluation. One more advantage of this model is solution simplicity.

We must mention some drawbacks of our simulation process:

1. Transaction costs are not included, but it is rather simple to include. It can affect model results dramatically if their size is higher than average profit a deal. It is fair to say that it is not issue of the AS-model only, it is an issue for any HFT market-making strategy.
2. Partial execution of market orders. Touching the price by market order does not guarantee limit order execution. It is a substantial drawback of simulation.

Possible further steps are:

1. Test on real market is logical next step to approve model applicability in real conditions. But it is quite difficult task in terms of required infrastructure.
2. Improve model. It is possible to avoid some assumptions by solving PDE numerically. But the problem here is how to estimate and to model arrival rates $\lambda(\delta)$.
3. Improve simulation with more complicated model of market microstructure. E.g., considering partial order execution which is the weakest point in simulation.
4. Improve simulation with fractional Brownian motion instead of simple one.
5. Include transaction costs.

References

1. Avellaneda, M., Stoikov S. High-frequency trading in a limit order book. Quantitative Finance, 2008.
2. Ho, T. and Stoll, H., Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics, 1981.
3. Davis, M.A.A., Linear estimation and stochastic control, 1977.
4. Gopikrishnan, P., Plerou, V., Gabaix, X. and Stanley, H.E. Statistical properties of share volume traded in financial markets. Physical review E., 2000.
5. Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H.E., Institutional investors and stock market volatility. Quartely Journal of Economics, 2006.
6. Maslow, S. and Mills, M., Price fluctuations from the order book perspective: empirical facts and a simple model. Physica A, 2001.
7. Potters, M. and Bouchaud, J-P., More statistical properties of order books and price impact. Physica A, 2003.
8. Weber, P. and Rosenow, B., Order book approach to price impact. Quantitative Finance, 2005.
9. I.I. Komarov, H.L. Chen. Application of the Hill method for the tail index estimation. Informatica, 2010.
10. Munasinghe, R., Kossinna P., Jayasinghe D., Wijeratne D. Tail Index Estimation for Power Law Distributions in R. Finance & Stochastic Processes, 2020.

Appendix.

1. Tail index regression.

```
                            GLS Regression Results
==============================================================================
Dep. Variable:              np.log(Y)   R-squared:                       0.991
Model:                            GLS   Adj. R-squared:                  0.991
Method:                 Least Squares   F-statistic:                 5.316e+04
Date:                Tue, 21 Dec 2021   Prob (F-statistic):               0.00
Time:                        19:24:07   Log-Likelihood:                 259.17
No. Observations:                 481   AIC:                            -514.3
Df Residuals:                     479   BIC:                            -506.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      9.0892      0.067    136.667      0.000       8.958       9.220
np.log(X)     -1.9950      0.009   -230.558      0.000      -2.012      -1.978
==============================================================================
Omnibus:                      119.447   Durbin-Watson:                   0.063
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              733.159
Skew:                          -0.918   Prob(JB):                    6.26e-160
Kurtosis:                       8.763   Cond. No.                         80.6
==============================================================================
```

Regression summary for all volumes.

```
Regression for regular volumes:
                            GLS Regression Results
==============================================================================
Dep. Variable:              np.log(Y)   R-squared:                       0.992
Model:                            GLS   Adj. R-squared:                  0.992
Method:                 Least Squares   F-statistic:                 6.248e+04
Date:                Tue, 21 Dec 2021   Prob (F-statistic):               0.00
Time:                        19:25:04   Log-Likelihood:                 305.33
No. Observations:                 481   AIC:                            -606.7
Df Residuals:                     479   BIC:                            -598.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.8365      0.060    146.250      0.000       8.718       8.955
np.log(X)     -1.9649      0.008   -249.955      0.000      -1.980      -1.949
==============================================================================
Omnibus:                      230.044   Durbin-Watson:                   0.004
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1355.265
Skew:                          -2.033   Prob(JB):                    5.11e-295
Kurtosis:                      10.147   Cond. No.                         80.6
==============================================================================
```

Regression summary for regular volumes.

2. Tail index estimations.

Here is a brief overview of tail exponent estimation methods.

1. Maximum Likelihood Estimation. Unbiased MLE estimation is:

$$\hat{\alpha} = (N-2)\sum_{i=1}^{N}\frac{x_i}{\hat{x}_{min}}$$

where x_{min} is Pareto distribution minimum data value. Usually, sample minimum is used.

2. Weighted Least Squares. Sample should be sorted in ascending order. Let r_i is number of points greater then x_i (rank). Tail index estimation is:

$$\hat{\alpha} = -\frac{\sum_{i=1}^{N}\log r_i/N}{\sum_{i=1}^{N}\log x_i/\hat{x}_{max}}$$

3. Percentile methods. It is a bunch of quite simple methods which uses percentiles:

Percentile method (PM): $\hat{\alpha} = \frac{\log 3}{\log P_{75} - \log P_{25}}$

Modified percentile method (MPM): $\hat{\alpha} = \frac{\log 2}{\log P_{75} - \log P_{50}}$

Geometric mean percentile method (GMPM): $\hat{\alpha} = \frac{1 - \log 4}{\frac{1}{N}\sum_{i=1}^{N}\log x_i - \log P_{75}}$

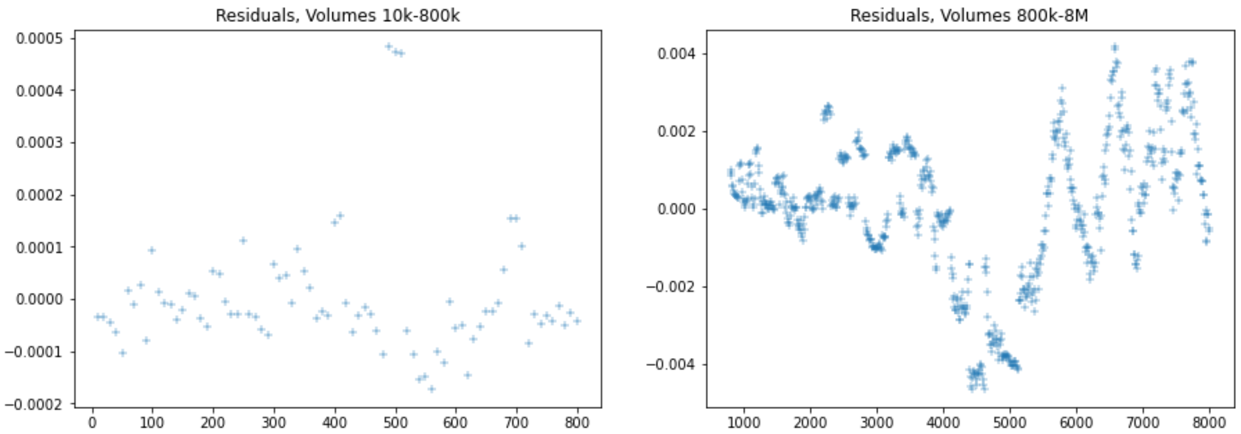
4. Momentum method:

$$\hat{\alpha} = \frac{\sum_{i=1}^{N} x_i}{\frac{1}{N}\sum_{i=1}^{N} x_i - N\hat{x}_{max}}$$

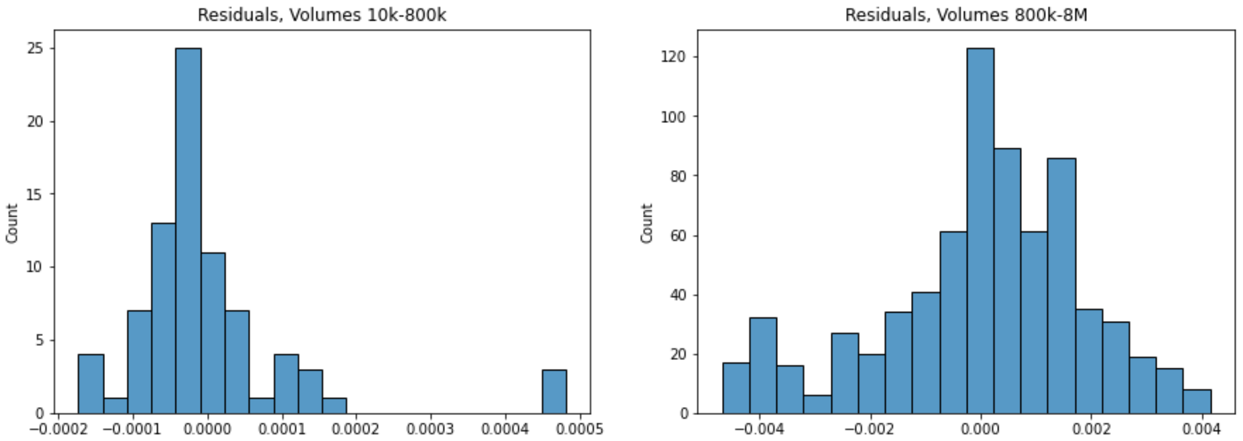
Results are in Table below:

Method	All data	Regular volumes	Factor of 50 volumes
MLE	1.70	1.87	1.68
WLS	1.70	1.87	1.68
PM	1.71	2.18	2.54
MPM	2.17	2.60	5.37
GMPM	1.66	2.40	3.88
MoM	1.97	2.07	1.84

3. Price impact regression results.



Regression residuals plot.



Regression residuals distribution

```
Regression for high volumes (800k+), log linear, summary

                            GLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared:                       0.819
Model:                            GLS   Adj. R-squared:                  0.819
Method:                 Least Squares   F-statistic:                     3250.
Date:                Tue, 21 Dec 2021   Prob (F-statistic):          6.28e-269
Time:                        20:45:37   Log-Likelihood:                 3514.3
No. Observations:                 721   AIC:                            -7025.
Df Residuals:                     719   BIC:                            -7015.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0439      0.001    -45.383      0.000      -0.046      -0.042
np.log(X)      0.0067      0.000     57.006      0.000       0.006       0.007
==============================================================================
Omnibus:                       29.723   Durbin-Watson:                   0.053
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               32.636
Skew:                          -0.521   Prob(JB):                     8.19e-08
Kurtosis:                       3.029   Cond. No.                         118.
==============================================================================
```

Regression summary.