Limit Order Book reconstruction, visualization and statistical analysis of the order flow

May 31, 2014

Julien Schroeter
D-MATH

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Supervisors
Prof. Dr. Vladimir Filimonov
Prof. Dr. Didier Sornette
Prof. Dr. Matthias Troyer

Contents

1 Introduction

In recent years, most exchanges have switched from floor-based markets to automated electronic markets. This structural change had a great impact on numerous aspects of trading:

On the one hand, the size of orders has substantially been reduced; this is linked to the fact that many more orders are posted nowadays. Furthermore, the time has now to be considered on a millisecond scale; everything happens much more quickly. For instance, a substantial majority of orders are canceled within a second. Hence, the dynamics are far more complex than they were before, and the changes are more sudden; all those factors make it more complicated to analyze the behavior of the markets.

On the other hand, each order submission, each cancellation or each trade execution can be, unlike before, recorded; many databases can even be accessed on the internet. This is the greatest advantage of electronic systems: nothing is lost and everything can become matter of investigation.

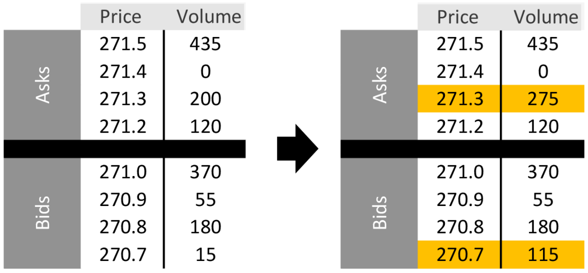
The study of the market dynamic is a central concern for the financial world. For instance, academics are willing to understand the mechanisms behind the market as well as to model its behaviors. As another example, the precise investigation of the limit order book and of the dynamics of the order flow is essential in order to build accurate algorithmic trading strategies. Even regulation authorities could benefit from a deeper understanding of the dynamics and the behavior behind the trades.

This project will contain three principal parts. First, an order book reconstructor will be implemented in order to get access to the essential structure which is the limit order book. In a second time, visualization method for this complex system will be developed and presented, and finally a small statistical analysis will be conducted.

2 The limit order book

First, it is necessary to discuss how orders are posted to the centralized trading systems and how trades are then executed. Indeed, there exist two different ways market participants can choose from in order to place a buy or a sell order.

On the one hand, they can decide to make a proposal to trade securities at a given price; such a request is named *limit order*. In this case, submitted orders stay in the system as long as they are not executed nor canceled; those pending orders can all be saved into a common structure: the *limit order book*. As shown in Fig.1 below, this limit order book is usually represented by the total amount of volume sitting on each price level at a certain time. On the left side of the figure, a small fictional limit order book is displayed; it consists only of a few different price levels ranging from 270.7 to 271.5. It can be observed in this example that, at this particular moment, the highest price market participants are willing to offer for this specific asset is 270. On the ask side of the book, market participants seem not ready to give up their shares for less than 271.2. Thus, there is here an obvious gap between the ask sideand the bid side and, therefore, no match for a trade can be found between the two fronts. Note that, in the same figure, the changes implied by a buy order of 100 shares at a price of 270.7 as well as by a sell order of 75 at a price of 271.3 can also be observed.

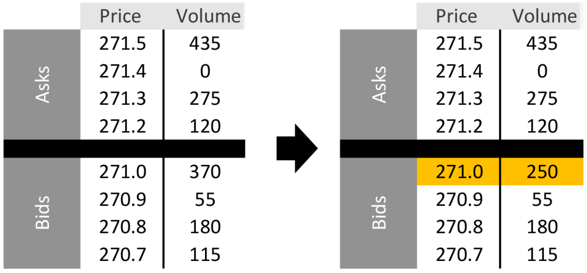


	Price	Volume
Asks	271.5	435
	271.4	0
	271.3	200
	271.2	120
Bids	271.0	370
	270.9	55
	270.8	180
	270.7	15

	Price	Volume
Asks	271.5	435
	271.4	0
	271.3	275
	271.2	120
Bids	271.0	370
	270.9	55
	270.8	180
	270.7	115

Figure 1: *Limit orders. Buy 100 shares at 270.7 and sell 75 shares at 271.3.*

On the other hand, market agents can prefer to post a *market order*; such an order is directly executed at the best price available in the book. Fig.2 shows a market sell order of 120 shares, which is immediately executed at the best price available, namely 271. Of course, some complications can arise if the size of the market order exceeds the volume available at the best price. In this situation, the entire quantity sitting at the best price is traded and then the biggest possible part of the remaining volume is exchanged at the next best price level available, and so on, until the whole initial volume is executed. A direct consequence of this splitting process is that not every share of a certain market order is traded at the same price. For example, if in Fig.2 the volume of the market sell order had not been of 120 but rather of 400 shares, then the first 370 shares would have been sold at the price of 271 and the remaining 30 shares at 270.9.

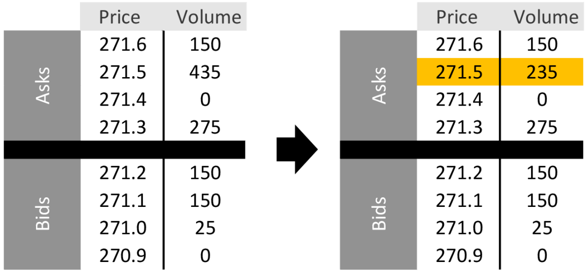


	Price	Volume
Asks	271.5	435
	271.4	0
	271.3	275
	271.2	120
Bids	271.0	370
	270.9	55
	270.8	180
	270.7	115

	Price	Volume
Asks	271.5	435
	271.4	0
	271.3	275
	271.2	120
Bids	271.0	250
	270.9	55
	270.8	180
	270.7	115

Figure 2: *Market order. Sell 120 shares at best price available.*

Since market orders are immediately executed, no cancellation can be made once they have been posted. This is obviously not the case for limit order: at any time, market participants are allowed to cancel partially or completely the volume of a limit order as long as it remains in the book. Fig.3 shows an example of some limit order cancellation.

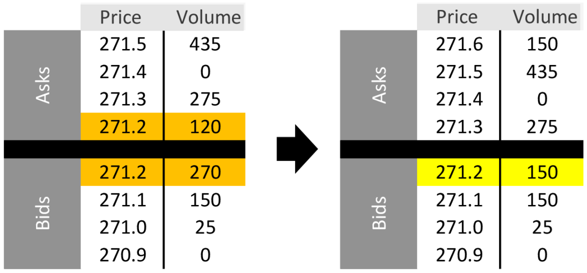


	Price	Volume
Asks	271.6	150
	271.5	435
	271.4	0
	271.3	275
Bids	271.2	150
	271.1	150
	271.0	25
	270.9	0

	Price	Volume
Asks	271.6	150
	271.5	235
	271.4	0
	271.3	275
Bids	271.2	150
	271.1	150
	271.0	25
	270.9	0

Figure 3: *Limit order cancellation. Remove 200 shares of a sell order posted at 271.5.*

As previously mentioned, limit orders - in contrast to market orders - are not directly executed. However, as soon as a match between the ask price and the bid price is found, a trade occurs. In Fig.4, for example, both sides coincide at the price of 271.2 and, therefore, the maximum possible amount of volume - namely 120 - is exchanged. Nevertheless, it is important to note that a perfect match is not necessarily required in order to trigger a trade: clearly, trades can happen as well when the price of a buy order exceeds the price of any sell orders. Yet, after the diverse executions, a gap will once again separate the bid from the ask side.



	Price	Volume
Asks	271.5	435
	271.4	0
	271.3	275
	271.2	120
Bids	271.2	270
	271.1	150
	271.0	25
	270.9	0

	Price	Volume
Asks	271.6	150
	271.5	435
	271.4	0
	271.3	275
Bids	271.2	150
	271.1	150
	271.0	25
	270.9	0

Figure 4: *Trade execution. Exchange 120 shares at 271.2.*

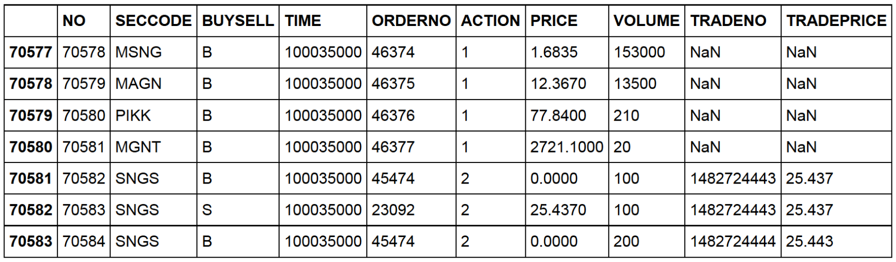
The limit order book is constantly evolving; each cancellation, each order placement and each execution has a direct impact on the shape of the book. The expansion of electronic trading has drastically increased the number of orders as well as the number of cancellations, and has also accelerated the speed of the various transactions; nowadays, the majority of the limit orders are even canceled within a second. Hence, this type of procedure transforms the book permanently, even several times within a single millisecond!

As stated above, the analysis of the limit order book is key step towards a better understanding of the dynamics underlying the markets. Therefore, this structure will be the corner stone of this project. The first part will consist of finding ways to reconstruct the book from some data flow which will be presented in detail below. The second task will deal with different means of visualizing such a complex system in order to get a first sense of the diverse dynamics. And, finally, a few specific aspects of the book will be investigated.

3 Data

As hinted earlier, it is now important to talk about the data at disposal. For this project, the available data consist of all orders and executions made during the year 2012 by all companies of the MICEX — Moscow Interbank Currency Exchange. Given the considerable size of the data set, the analysis will focus on the first two months of the year and only a few companies will be investigated.

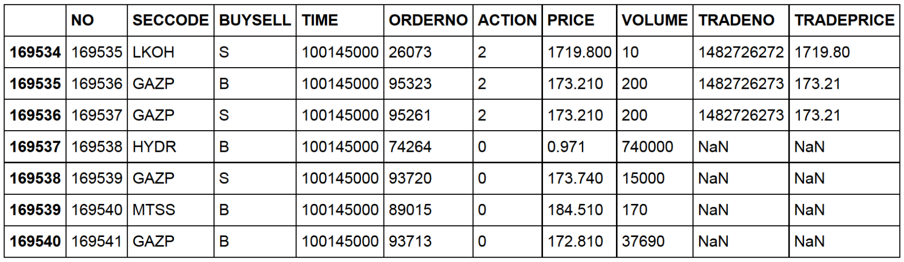
3.1 Presentation



	NO	SECCODE	BUYSELL	TIME	ORDERNO	ACTION	PRICE	VOLUME	TRADENO	TRADEPRICE
70577	70578	MSNG	B	100035000	46374	1	1.6835	153000	NaN	NaN
70578	70579	MAGN	B	100035000	46375	1	12.3670	13500	NaN	NaN
70579	70580	PIKK	B	100035000	46376	1	77.8400	210	NaN	NaN
70580	70581	MGNT	B	100035000	46377	1	2721.1000	20	NaN	NaN
70581	70582	SNGS	B	100035000	45474	2	0.0000	100	1482724443	25.437
70582	70583	SNGS	S	100035000	23092	2	25.4370	100	1482724443	25.437
70583	70584	SNGS	B	100035000	45474	2	0.0000	200	1482724444	25.443

Figure 5: *Diverse order posts and trade executions on January 3rd.*

Fig.5 shows a sample of the available data; each line of the table describes an order placement, a cancellation or an execution. Before going into more detail regarding the meaning of each column, it can be noticed that some of them are straightforward and easily understood: *TIME*, *VOLUME*, and *NO*. Nevertheless, the other columns may require some deeper explanation. To begin with, the column *ACTION* specifies the nature of the transaction: the number 0 points out that the line contains an order cancellation, 1 stands for an order placement and 2 shows that the line describes a trade execution. Secondly, the column *BUYSELL* indicates whether the line concerns a buy transaction or a sell transaction; obviously, B stands for buy and S for sell. Thirdly, the order number of a given transaction can be found in the *ODERNO* column. Moreover, it is essential to understand that the order number of a trade or a cancellation is identical to the order number assigned to its corresponding placement. This helps greatly to keep track of the different volume changes for each order, and also facilitates the computation of the time between an order placement and either its cancellation or its trade execution. Next, the *PRICE* column indicates at which price level the order has been posted into the book. Note that if the price is set to zero, it designates not a limit order but rather a market order, which means that the volume must be executed immediately at the best price available. Finally, the last two columns - *TRADEPRICE* and *TRADEVOLUME* - uniquely concern the trade executions. It is important to note that sometimes the trade price and the placement price can differ; this might happen for instance if an over advantageous order is posted. Indeed, in such a case, this order may initiate a trade at a price that differs from the initial one. Furthermore, the trade volume can also differ from the initial volume, if, on the one hand, a partial cancellation occurred in the interval or, on the other hand, if a part of the volume placed was not visible.



	NO	SECCODE	BUYSELL	TIME	ORDERNO	ACTION	PRICE	VOLUME	TRADENO	TRADEPRICE
169534	169535	LKOH	S	100145000	26073	2	1719.800	10	1482726272	1719.80
169535	169536	GAZP	B	100145000	95323	2	173.210	200	1482726273	173.21
169536	169537	GAZP	S	100145000	95261	2	173.210	200	1482726273	173.21
169537	169538	HYDR	B	100145000	74264	0	0.971	740000	NaN	NaN
169538	169539	GAZP	S	100145000	93720	0	173.740	15000	NaN	NaN
169539	169540	MTSS	B	100145000	89015	0	184.510	170	NaN	NaN
169540	169541	GAZP	B	100145000	93713	0	172.810	37690	NaN	NaN

Figure 6: *Sample of trade executions and order cancellations on January 3rd.*

Although this section might seem futile, it was important to present all those columns, since this information will be used at some point in the analysis.

3.2 Comments

At this point an important issue concerning the data has to be raised. Indeed, the specific moment actions occur is reported in hours, minutes and seconds, yet the milliseconds are omitted. Moreover, the actions taking place within a given second seem to be ordered not by time - unfortunately - but rather by their nature: order placements are listed first, and then trade executions and finally cancellations. Thus, the exact development of the book within a particular second cannot be retrieved. This could explain why it can be witnessed that the ask price and bid price cross each other frequently within a given second, although this is not consistent. This problem raises some concerns about the general accuracy and coherence of the data. Nevertheless, if the time resolution of the analysis is based on seconds and not milliseconds, this predicament should then not arise.

4 Reconstruction of the limit order book and technical aspects

Needless to say, a thorough investigation of the dynamics underlying the limit order book cannot be conducted without access to this specific book. In that sense, the first goal of this project will be to code a limit order book constructor which takes as input the data presented in the previous section, and process them in order to output the entire state of the book at any given time.

4.1 Structure

A limit order book — as mentioned in Section 2 — is represented by the amount of volume sitting on each price level. More specifically, each price level contains all limit orders that have been made during the day at this respective price and that have not been executed nor totally canceled yet. A key feature of usual order books is the fact that a substantial majority of price levels are empty and do not contain any volume. This main characteristic leads to the necessity of using nodes to represent each price level, and of finding a way to connect them — without ever having to consider the unused price levels. In addition, this last remark explicitly rules out the idea of using a simple list to represent the book.

In order to build these nodes, it is important to understand which information they need to carry. First, each price node contains a dictionary of all orders still sitting at this respective price. The order number — held in the *ORDERNO* column — operates as the key in this dictionary and is only associated with the remaining volume of the order. Of course, the time of the order post could have been included as well; however, although such an expansion could be useful to investigate the cancellation and the execution time, it has been decided here to proceed differently. Indeed, it would have otherwise considerably slowed down the program execution. Second, besides the dictionary, another important attribute has to be saved in the node: the aggregated volume still sitting at this price level. Though this value could be directly retrieved from the dictionary by summing up all remaining volume, it seems better to keep track of it separately in order to boost the output process. Finally, each node has to contain links connecting it to the other price levels. These links are dependent on the chosen structure used to arrange the nodes and, therefore, will be discussed in further detail below.

Consequently, the second and central concern of the book reconstruction lays in the choice of a suitable data structure which can hold and connect together the different price nodes. The first idea consists in using two different structures, one for the bid and one for the ask side; indeed, it may be far too complicated to find a convenient common structure. A tree would theoretically yield good results; however, since most orders are posted at a price close to the best price, it would be wise to get the best price and its closest nodes near the root. Unfortunately, this would produce a highly unbalanced tree and the advantages specific to trees would be diminished by this constraint. Thus, there is another structure that needs to be considered: that of tree combined with an array. In this set-up — represented in Fig.7, the first n price levels could be stored into an array while the remaining nodes would be saved into a tree. This idea will not be explored more deeply, though it would certainly have been interesting to test such a structure.

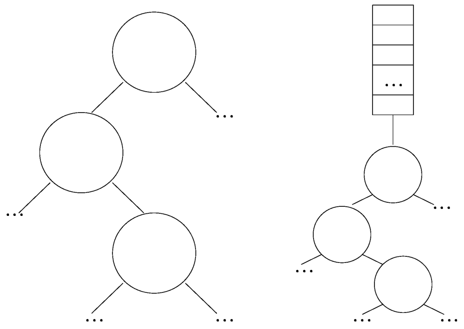


Figure 7: *On the left side, representation of a standard tree. On the right side, display of a list, which could contain the first price levels, combined with a tree for the remaining levels.*

For this project, two simple linked lists are used to represent the bid and the ask side of the limit order book separately. In each of these two lists, nodes are ordered by price, in descending and ascending order respectively. Such a structure allows fast insertion and deletion of nodes, handy output, and even immediate access to the ask price and the bid price, since they are respectively the first element of each list. The computational intensity is much lower than it would be expected for such a structure. Indeed, as mentioned above, many orders are made close to the best price, and thus less time is required to access the corresponding price node. Another important advantage of such a structure is its simplicity; no re-balancing has to be made — in contrast to a tree — and each price level is relevant — in opposition to a list.

As an expansion of this simple structure, it may be wise to temporarily save each limit order into a queue before entering it, about a second later, into the book. In fact, this process might even be much more efficient than using simple linked lists, since more than half of the posted limit orders are canceled within a second. Although it would have been surely interesting to investigate this idea further, a simple structure made of two linked lists seems efficient enough for the purpose of this project, and therefore will officiate as the main structure throughout this paper.

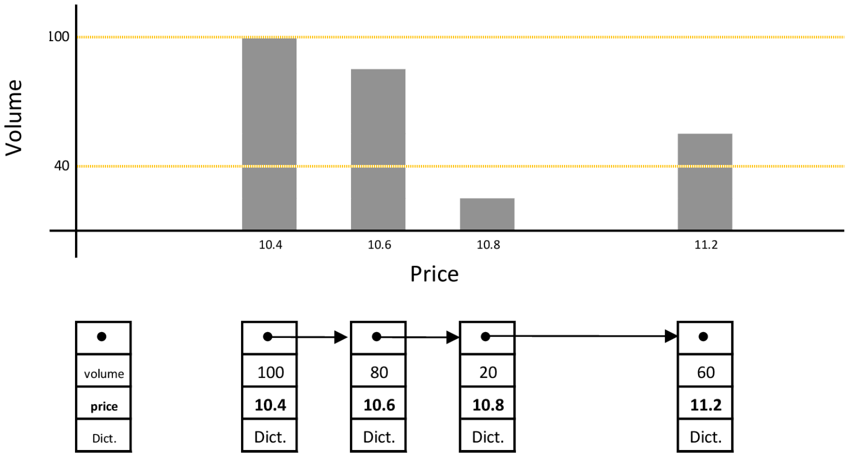


Figure 8: *Example of the linked list structure. On the upper side, a few price levels of a limit order book which are the saved in the nodes depicted on the lower side.*

4.2 Reconstruction process

Now that a suitable structure has been chosen, it is time to discuss the reconstruction process of the limit order book. In summary the constructor examines each line of the available data one after another, in order to rebuild the book up until a specific point in time. It processes those lines differently depending on their nature — order placement, order cancellation or trade execution.

In the first case, supposing that it comes upon an order placement, the constructor scans either the linked list of the bid side or the one of the ask side and checks if the node of the corresponding price already exists. Provided that the node is already there and contains some volume, a new entry is directly added to the dictionary of orders and the aggregated volume of the price node is updated. If, however, there is no price level yet corresponding to the price of the placement, a new node is inserted into the linked list and a new entry — containing the order number and the volume — is added to the newly created dictionary. Note that the scanning process takes more time when considering either longer or denser linked list or nodes that are far away from the best price.

In the second case, namely order cancellation, the procedure is similar to the one of an order placement. First, the constructor scans the corresponding list up to the desired price node and then searches in the dictionary the entry corresponding to the given order. At this point, two different cases have to be considered: on the one hand, if the remaining volume of the order is equal to the volume of the cancellation, then the entry in the dictionary is deleted and the aggregated volume is updated. On the other hand, if the remaining volume exceeds the volume of the cancellation — if this is only a partial cancellation — then the remaining volume in the dictionary as well as the aggregated volume are simply updated.

Finally, conceding that the constructor encounters a trade, the process goes exactly as in the case of a cancellation except that the price to look for in the linked list is not the traded price but rather the price of the order placement. As those two prices can be different, it is important to update the node containing the corresponding order post.

Besides, whenever the volume of a price node reaches zero, the node is deleted and the links to its neighbors are updated. This process allows the linked lists to be kept to a reasonable size even after a thousand iterations. Furthermore, it ensures that the multiple scanning of the list is not too intensive computationally.

4.3 Output

The use of linked lists allows a direct access to the first bid and ask price level. The remaining levels of the limit order book can be reconstructed by scanning through all available price levels in both linked lists. However, a few applications do not require the entire book and the scanning procedure can, in that case, be stopped at any moment in order to output only the first few levels. Besides, since the linked lists are already ordered by the price, the output procedure does not need too much computation power. This is one of the great advantages of this simple structure over the different variants of trees that have been mentioned earlier. Moreover, they can save a great amount of time considering the fact that an output can be required for each millisecond of the trading session.

5 Visualization

One of the first important steps towards a better understanding of the dynamics underlying the limit order book is the visualization of such a complex structure.

Throughout this section, the limit order book of three main companies — which illustrate at best the possible disparities between different securities — will be taken as an example and compared with each other. In order to simplify the analysis only one day of data for each firm will be considered.

First, the book evolution of the PIK Group — PIKK — on the 3rd January 2012 will be investigated; during that specific day, this ticker got a total amount of 815879 orders, among which only 1860 concerned trade executions. Therefore, with approximately one trade execution each five minutes, this security can be considered as highly inactive or — using financial terminology — highly illiquid. Secondly, the limit order book of the oil company Rosneft — ROSN — on the 9th of January 2012 will be examined: with a number of 26580 trades for a total of 1164967 orders, this company can be considered as average in comparison to the other MICEX companies. Finally, Sberbank — SBER — the largest bank in Russia, with its considerable number of trade executions — namely 146548 — during the 22nd of February 2012, is the perfect illustration of what is called a liquid asset; it will therefore be taken as third example.

Thus, the similarities and more importantly the differences between those three differing securities will be investigated in order to analyze the links between liquidity and the various characteristics of the limit order book.

5.1 Trade execution heat map

A first visual approach, which might help to get a first impression of the data, can consist in creating for each firm heat maps of the trade executions, the order placements and the cancellations.

In Fig.9, for example, a heat map of the number of trade executions is plotted; each point on the figure represents the number of trades for a limited time range and for a few specific volume levels. In this same figure, a substantial difference in the number of trades between the three above-cited examples can be observed. Indeed, as noticed before, the trade activity of SBER is significantly higher compared to other companies. Those trades are also spread out over the entire heat map which indicates, on the one hand, that a wide range of volumes are concerned and, on the other hand, that no important decline in activity can be observed over that day. In contrast, the plot of PIKK trades — with its 1860 trade executions — seems nearly empty.

In addition, peaks of activity can directly be examined through this kind of visualization. For example, a big trading intensity can be noticed in SBER example at around 15pm; it will be show below that this time range coincides with different events.

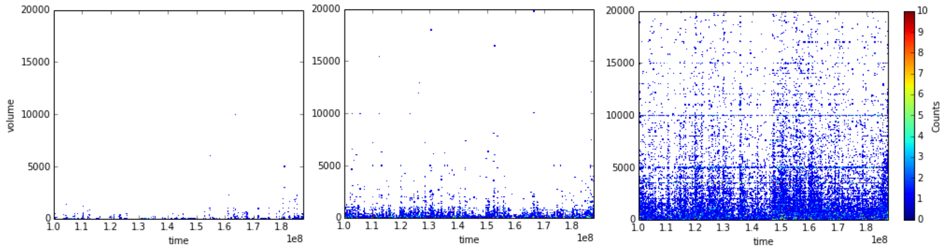


Figure 9: *Trades heat map. From left to right, the trade activities of PIKK, ROSN and SBER are plotted.*

5.2 Price volume heat map

First of all, there is an important point to take into account: visualizing the order book over time requires at least three dimensions — time, price level and volume sitting at each price level. A first approach consists in producing a two dimensional graph and including the third dimension using a color scale.

As a first example, Fig.10 represents the evolution of the PIKK order book during January 3rd 2012 through the heat map defined above. Some general comment can be made about this picture. First, the non-empty price levels are sparsely distributed; indeed, many price levels do not contain any volume. Secondly, volumes are relatively small in comparison with other companies of the MICEX. Thirdly, the *spread* — the difference between the ask side and bid side — is significant. Finally, the stepwise shape of the ask and bid side must be pointed out. It may be due to the fact that the volume sitting on each price level is small and that many price levels are empty; accordingly, any important market buy or sell order can have a huge impact on the best price. As mentioned above, this specific ticker is particularly inactive during that day. This could explain the large spread; apparently, sellers and buyers of the share did not often reach a consensus during that day.

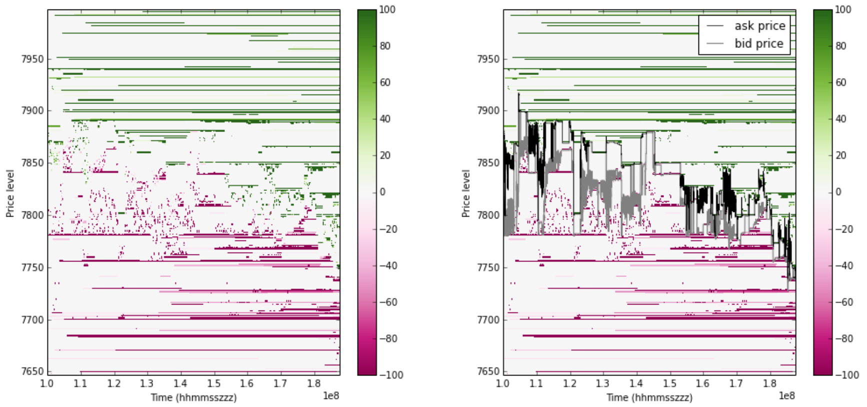


Figure 10: *Representation of the limit order book of PIKK over the entire 3rd of January 2012.*

The second example worth mentioning is the order book of ROSN shown in Fig.11. The spread is less important than in the previous example, and the price changes are also less abrupt. As a matter of fact, this company activity was already considered as average.

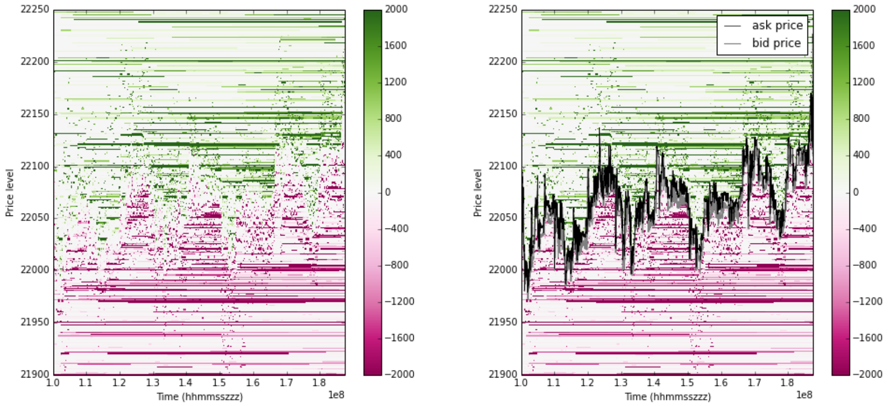


Figure 11: *Limit order book of ROSN on the 9th of January 2012.*

Finally, Fig.12 depicts the order book of SBER, the third investigated company. The scale on the right hand side of the figure must be pointed out: indeed, the volumes sitting in the book are enormous in comparison with the other two examples. This is not surprising, since this firm has already been said to be highly active.

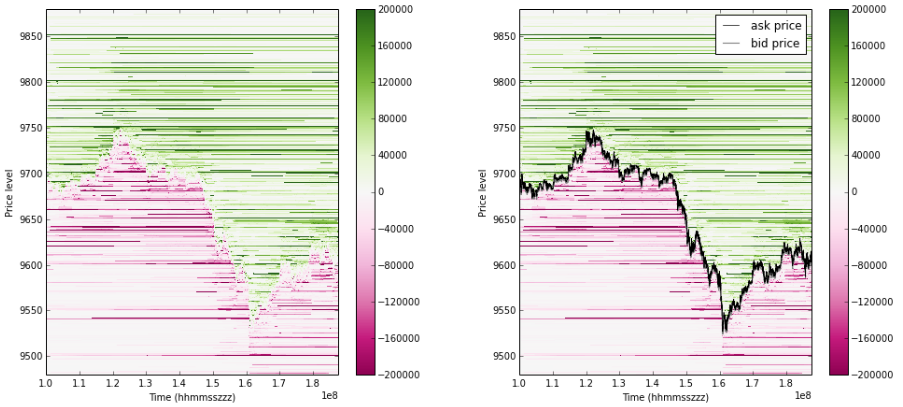


Figure 12: *Visualization of the book of SBER on the 22nd of February 2012.*

The progression between those three examples is manifest. First, the spread is highly significant in the first example and yet almost non-existent in the last one. Second, variations in the curves are completely different throughout the three examples: it ranges from stepwise shape — for PIKK — to an almost smooth line — for SBER. Finally, the small proportion of price levels containing volume in the first example contrasts sharply with the high density observed in the last figure. For those heat maps, an undeniable link between the number of trades throughout the day and various properties of the order book — the spread and the smoothness of the ask price and bid price changes — can be observed. Deeper analysis could be useful to quantify the intensity of the correlation between those elements, but will not be conducted in this project.

5.3 Average and mean book shape

A third useful visualization consists in showing the mean or the median — over a day —- of the limit order book shape. Thus, in the following pictures each column represents the mean or the median volume aggregated over twenty price levels, except for the bid and the ask price which are both displayed separately. In Fig.13, for instance, the two first bars — in the upper as well as in the lower graph — stand for the mean and median volume sitting respectively at the ask and at the bid price. Then, the two bars on their right depict the aggregated volume sitting on the first twenty price levels — that is starting from the best price — and after that the two next bars display the next twenty price levels, and so on. This approach has made it possible to give a visual account of the 160 first ask levels as well as of the 160 first bid levels; that corresponds to the central part of the limit order book.

Hence, the mean and median book shape of PIKK are represented in Fig.13 in gray and in orange respectively; it is important to recall that the ask side is displayed on the upper side of the figure and the bid side on the lower side. The significant difference between the volume levels of the mean and the median books hints at the fact that the distribution of these books is potentially skewed. Besides, disparities in the column size can be perceived: there is more mass standing on the hundred first levels of the ask side than on those of the bid side. This could thus explain the overall decrease in price of the asset over the day, observed in Fig.10.

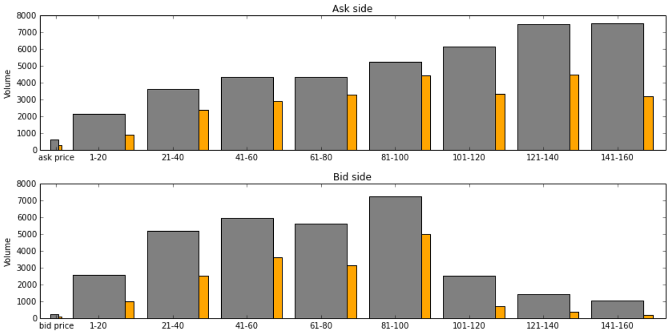


Figure 13: *PIKK. Average and mean limit order book shape respectively displayed in gray and orange.*

In Fig.14, there is, between the ask side and on the bid side, a significant difference in the amount of volume sitting on the first levels. Again, this could be one of the reasons why a global increase in price over the day could be observed in Fig.11.

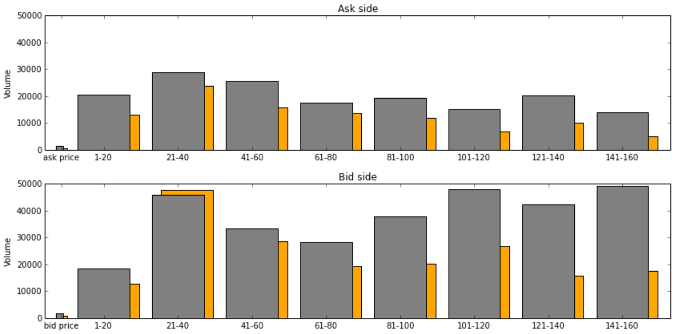


Figure 14: *ROSN. Average and mean book shape.*

Once more, a correspondence between the volume difference of the ask side and bid side and the price evolution over the day can be noticed in Fig.15. Moreover, the median volume is here equivalent to the mean volume, in opposition to the other two examples.

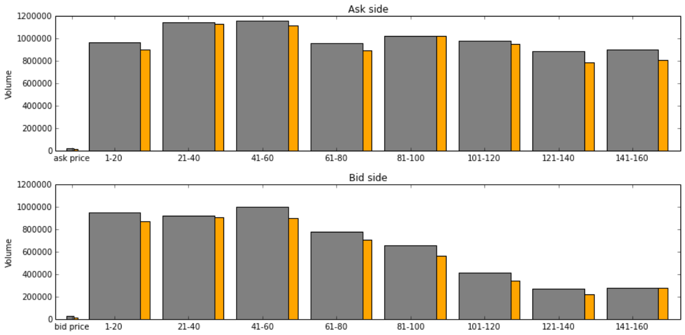


Figure 15: *SBER. Average and mean book shape.*

Accordingly, deeper analysis could be conducted so as to estimate the relation between the median or mean book shape and the price direction throughout a day.

6 Liquidity measures

A security is usually defined as liquid if buying or selling does not affect too much its price. However, there is no official benchmark when it comes to qualify a market as liquid or illiquid. Provided that very different cases are brought into comparison, it is nevertheless possible to get a sense of what liquidity of a ticker means. In this end, it seems interesting to recall, on the one hand, Fig.12 which shows a perfect example of a liquid share and, on the other hand, Fig.10 which, by its wiggliness and the size of spread — difference between the ask and the bid price — is a good representation of what an illiquid ticker looks like. In this section, two different liquidity measures will be presented in order to quantify the slightly subjective concept of liquidity.

6.1 Spread

A first way to measure liquidity consists in examining the spread. For example, on the left side of Fig.16, the spread of PIKK is plotted over time; it is interesting to notice the evolution in the second part of the day.

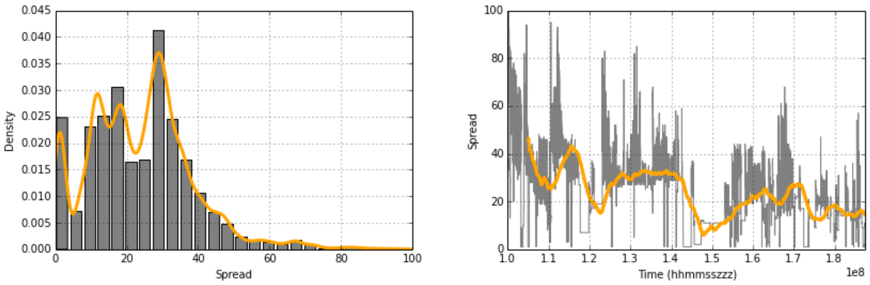


Figure 16: *PIKK. Visualization of the spread over the entire trading session.*

In contrast to the previous example the spread in Fig.17 is much smaller throughout the day and its variability is also significantly reduced.

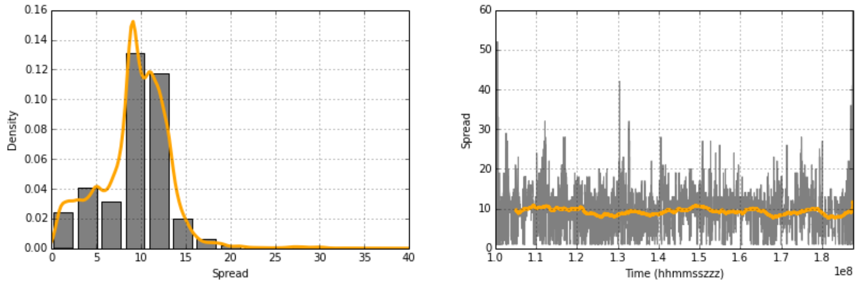


Figure 17: *ROSN. Histogram and plot of the spread.*

Fig.18 confirms the observation made about Fig.12 which stated that the spread in the SBER example was almost nonexistent. Besides, it can be noticed that there is almost no measurable difference as to the spread size over the day. Again, the striking dissimilarity between PIKK to SBER must be pointed out.

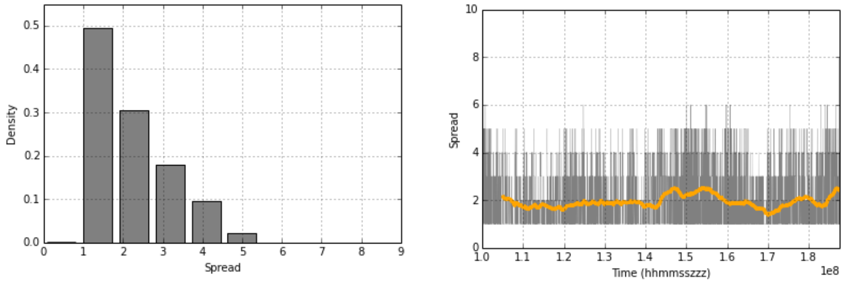


Figure 18: *SBER. Display of the spread*

In conclusion, those observations point all in the same direction as the ones made in the previous section. This second approach however leads to a precise quantification — a measure — of the subjective observations made previously.

6.2 Round trip cost

The *round trip cost* is a second liquidity measure. It is defined as the cost of buying a certain number of shares and of selling the exact same amount at a given time. In other words, this measure consists in calculating the difference in price between a market sell order and a market buy order of massive size. The quantity that needs to be bought and sold is set arbitrarily, yet it has to depend on the average volume sitting in the book. Indeed, this quantity cannot be too small since a certain minimum number of price levels — and not just the best price — should be impacted by this process; this would otherwise yield similar results as with the simple spread measure. Accordingly, the benchmark quantities for the standard examples were set to the mean volume sitting on the first twenty price levels, which could directly be read from Fig.13 to Fig.15.

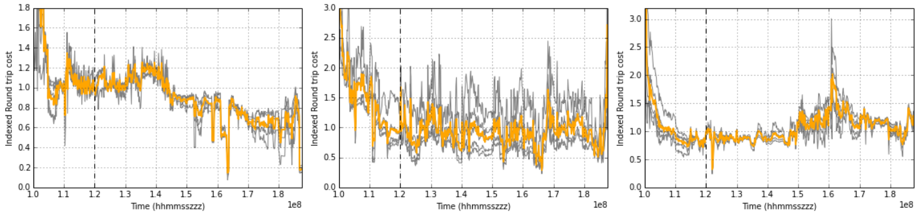


Figure 19: *Indexed round trip costs. From left to right, PIKK, ROSN and SBER.*

For each company the round trip cost was computed five different times using respectively the first five multiples of the benchmark quantity as the amount of shares to sell and buy. The — at 12pm indexed — round trip costs for the three examples are plotted in Fig.19. It is noteworthy that the indexed round trip costs for the various quantities are nearly similar. The indexation does unfortunately not allow comparing the three companies with each other, but interesting likeness in the dynamics of the spreads presented above can be noticed. For example, the drop in the spread size of PIKK at the middle of the day — observed on Fig.16 — coincides closely with the decrease in the round trip cost. However, new elements that did not appear in the spread can be perceived as, for example, the sudden increase in the round trip cost of SBER at around 16pm. This peek interestingly occurs at the same time than an abrupt and swift change in the price direction shown in Fig.12.

Of course, the two introduced measures are more or less correlated, since, on the one side, the spread is the difference in the ask and bid price and, on the other side, the round trip depends greatly on the size of the price gap between the ask and bid side.

7 Statistical analysis of cancellations

Cancellation is an essential aspect of modern trading. As mentioned earlier, a large proportion of orders are canceled at the very second they are posted. This section will therefore focus on limited order cancellation; more specifically a few statistical analysis will be conducted in order to get a better understanding of this process.

7.1 Cancellation heat map

First, as seen for trades, a heat map of cancellations can be drawn. Fig.20. shows such a heat map for the three recurring companies — PIKK, ROSN and SBER. As already mentioned about trade executions, the order cancellations of SBER are numerous and spread throughout a wide range of volumes. However, in contrast to the observations made earlier, there are no significant differences in the intensity of the activity between ROSN and PIKK. On the left side of the figure, the order cancellations of PIKK can be examined; it is interesting to point out the sudden decline in the pace of cancellation activity taking place in the middle of the day. Surprisingly, this specific decrease in the number of cancellations coincides with both the drop in the spread and in the round trip cost observed before. However, similarities of behavior between the cancellation activity and the different liquidity measures cannot be as easily perceived from the cancellation heat map of the two other companies; indeed, no significant changes can be seen in those cases. It would surely be interesting to investigate further the link between those different concepts; nevertheless this will not be conducted in this paper. In conclusion, this approach mainly helps to get a first look at the distribution of the cancellations throughout the day.

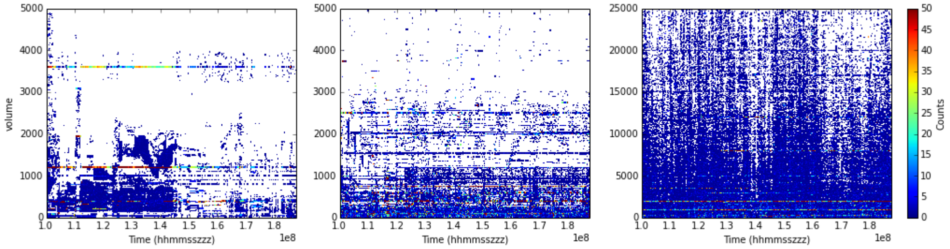


Figure 20: *Order cancellations heat map. From left to right, the cancellation activity of PIKK, ROSN and SBER.*

7.2 Cancellation versus distance from the mid price

The previous section presented a heat map of the cancellation count which depended on time as well as on cancellation volume. A similar approach is used in Fig.21; however, in this representation, the number of cancellation is plotted based on time and on the concerned price level. Two main comments have to be made about it: on the one side, the decline in cancellation activity for PIKK assets — already observed in Fig.20 — can be noticed once again at the middle of the day. On the other side, the two other companies show an interesting cancellation pattern far below the mid price. More specifically, sell orders are posted over a wide range of price levels — all standing far away from the mid price — and are then canceled after a short period of time; this procedure is repeated many times throughout the entire day. This could be part of some sort of trading or signaling strategy attempting to influence the price of the security. Nonetheless, the fact that this could also be due to a mistake of an automated trading strategy cannot be ruled out. Once again, as seen in Fig.20, interesting patterns can be observed through a simple heat map.

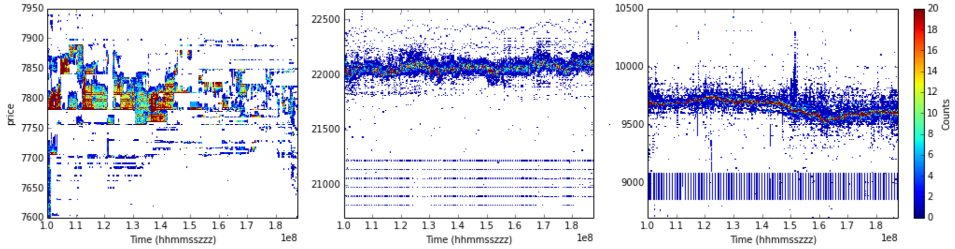


Figure 21: *Another example of order cancellations heat map. From left to right, the cancellation activity of PIKK, ROSN and SBER*

Given that this previous approach fails to give a visual account of the volume and the number of cancellations according to the mid price a new method is required. Below, the number of cancellations as well as the corresponding volume has been plotted over the price centered at the mid price. First, cancellations of PIKK assets are significantly more frequent on the bid side than on the ask one; this leaning was not obvious on the previous figure though. Besides, cancellations are made moderately close from the mid price provided that the spread is considered.

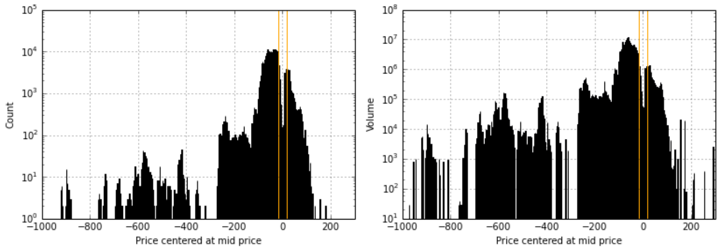


Figure 22: *PIKK. Histogram of the cancellation volume and count centered at the mid price. The mean spread of the day is also depicted in orange.*

Secondly, the surprising cancellation patterns occurring far below the mid price, as pointed out above, seem to have indeed a direct impact on the distribution of cancellation number and volume, as it can be noticed in Fig.23. The same remark can be made about SBER in Fig.24. However, in contrast to the previous figure, the cancellations — except for the weird patterns — seem to be symmetrically spread around the mid price.

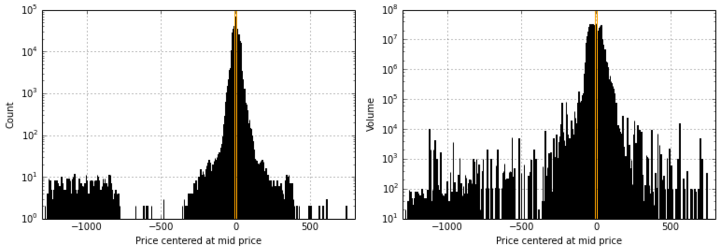


Figure 23: *ROSN. Histogram of cancellations centered at mid price.*

Thirdly, a substantial majority of the cancellations of SBER securities are made close to the mid price if the absolute distance is considered. If, however, this distance is compared to the mean spread, opposite conclusions can be drawn. Nevertheless, this is simply due to the fact that there are little or no differences between the ask side and the bid side.

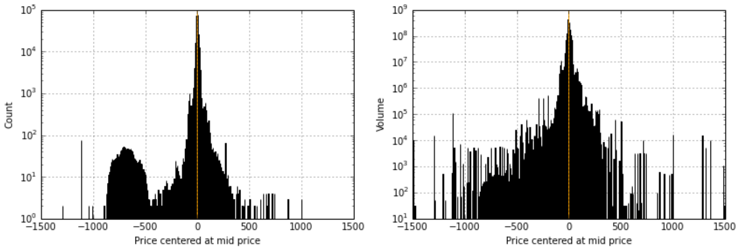


Figure 24: *SBER. Histogram of cancellations centered at mid price.*

A possible expansion could make use of this kind of distributions in order to model cancellation dynamics in the market.

7.3 Conditional cancellation count and volume

In this section, the volume and the number of cancellations, depending on the direction of the mid price, will be investigated. The data is produced as following: each second the change in the mid price is recorded as well as the number of cancellation which occurred during that same second. Afterward, the number of cancellations in the previous time interval as well as in the following time interval are added to each line. This data can easily be used to get the conditional mean number of cancellations for the previous time interval — $t-1$ —

the given time interval — t — as well as for the following time interval — $t+1$ — given that the price went up, down or stayed stable during the interval t.

In Fig.25, the conditional mean and median of number of cancellation — on the left – as well as the conditional mean and median of the cancellation volume — on the right — are represented in dark gray and orange respectively. Moreover, the weighted — according to the intensity of the price change — average is plotted in light gray. The difference in the number of cancellation between the ask price and the bid price is striking; this was already mentioned while observing Fig.22.

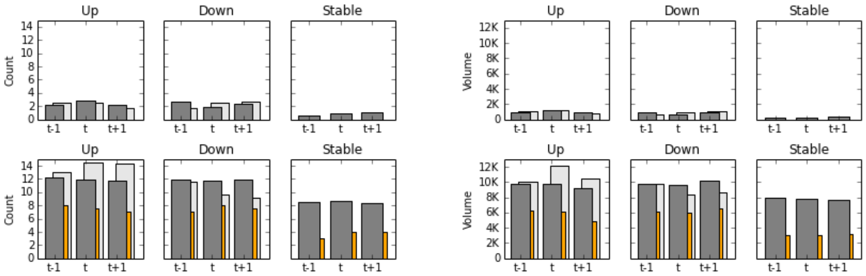


Figure 25: *PIKK. Conditional mean, median and weighted average of the cancellations count and volume, respectively in dark gray, orange and light gray. At the top, the ask side is presented and at the bottom the bide side. The labels "up", "down" and "stable" indicate the direction of the price change in the interval t.*

The disparities between these different approaches — the mean, the median and the weighted mean — could be discussed further. Nevertheless, a more interesting aspect of this figure has to be investigated: whether the conditional mean between periods is relevant. For this purpose, a Welch's t-test can be used to test if the means of two samples are significantly different. On the 5% level, the difference in condition mean between the cancellation numbers or volumes at different time is — unfortunately — statistically not significant, according to the Welch's test. However, it can be observed whether the weighted mean — the light gray bars — is considered a significant peek in volume canceled on the bid price during the periods of price increase. Since there is no difference in the unweighted mean, this indicates that the volume canceled on the bid side is more important if the increase in price is stronger.

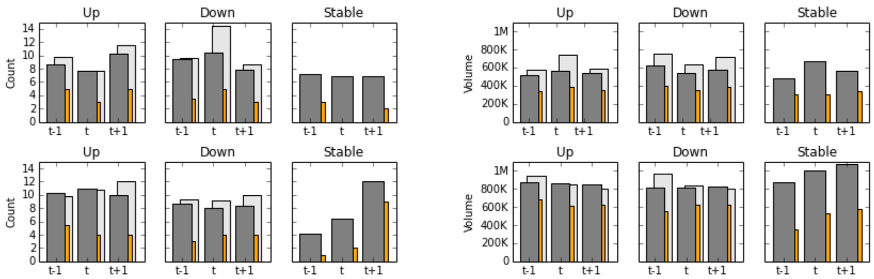


Figure 26: *ROSN. Conditional mean, median and weighted average of the cancellations count and volume.*

Welch's tests have been run for all cases and unfortunately none of them show significant differences in the means.

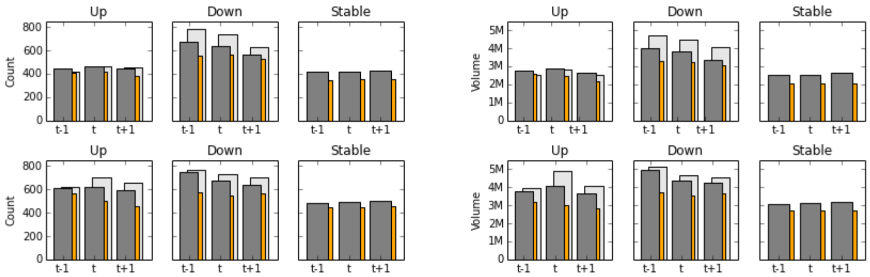


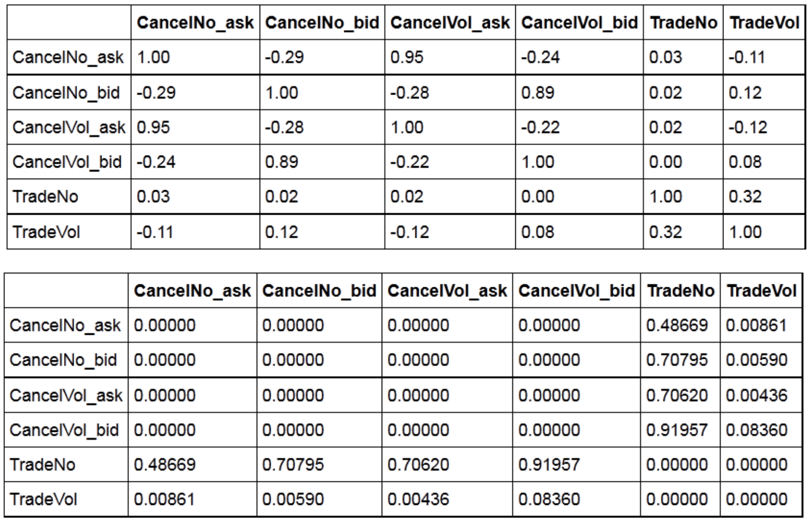
Figure 27: *SBER. Conditional mean, median and weighted average of the cancellations count and volume.*

This approach could be expanded to multiple days, and numerous aspects could be tested. However, the aim of this section was simply to, first, present the three methods and then to show few interpretations that could be read out of those plots.

7.4 Correlation analysis

Pearson test can be used to investigate the correlation between the cancellations volume or number and the trades volume or number. The figures in this section use values — for the diverse number and volumes — aggregated over a range of one minute. More specifically, the sample containing the information about the cancellation and the trades holds data describing these factors for interval of a minute of the trading session.

At the top of Fig.28, a correlation matrix with the Pearson correlation coefficients for PIKK is displayed, and at the bottom of the figure the corresponding double tailed p-values are depicted. Firstly, without surprise, the number of cancellation is highly correlated with the volume canceled on the bid side as well as on the ask side. Secondly it can be noticed, interestingly enough, that the number and the volume of cancellation are slightly negatively correlated with the trade volume. Finally, a negative correlation between the ask and bid volumes and numbers can also be noted.

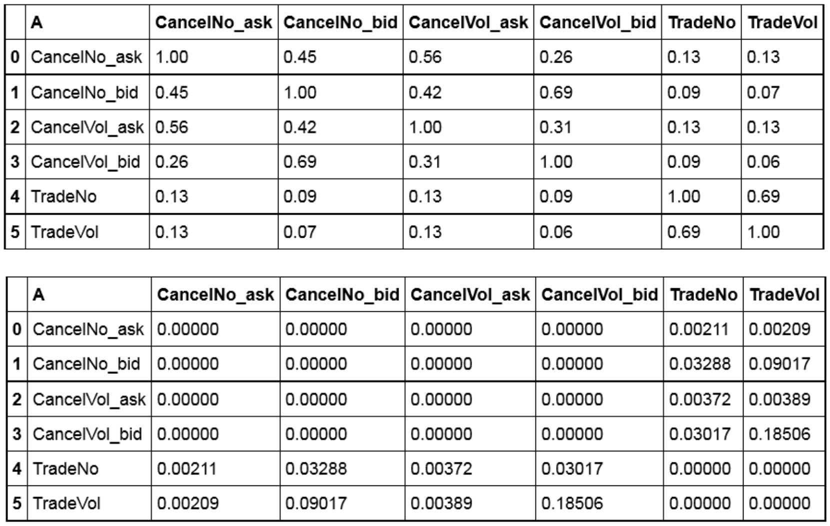


	CancelNo_ask	CancelNo_bid	CancelVol_ask	CancelVol_bid	TradeNo	TradeVol
CancelNo_ask	1.00	-0.29	0.95	-0.24	0.03	-0.11
CancelNo_bid	-0.29	1.00	-0.28	0.89	0.02	0.12
CancelVol_ask	0.95	-0.28	1.00	-0.22	0.02	-0.12
CancelVol_bid	-0.24	0.89	-0.22	1.00	0.00	0.08
TradeNo	0.03	0.02	0.02	0.00	1.00	0.32
TradeVol	-0.11	0.12	-0.12	0.08	0.32	1.00

	CancelNo_ask	CancelNo_bid	CancelVol_ask	CancelVol_bid	TradeNo	TradeVol
CancelNo_ask	0.00000	0.00000	0.00000	0.00000	0.48669	0.00861
CancelNo_bid	0.00000	0.00000	0.00000	0.00000	0.70795	0.00590
CancelVol_ask	0.00000	0.00000	0.00000	0.00000	0.70620	0.00436
CancelVol_bid	0.00000	0.00000	0.00000	0.00000	0.91957	0.08360
TradeNo	0.48669	0.70795	0.70620	0.91957	0.00000	0.00000
TradeVol	0.00861	0.00590	0.00436	0.08360	0.00000	0.00000

Figure 28: *PIKK. Correlation matrix of different aspects of the book. At the top, the Pearson correlation coefficients and at the bottom, the corresponding double tailed p-values.*

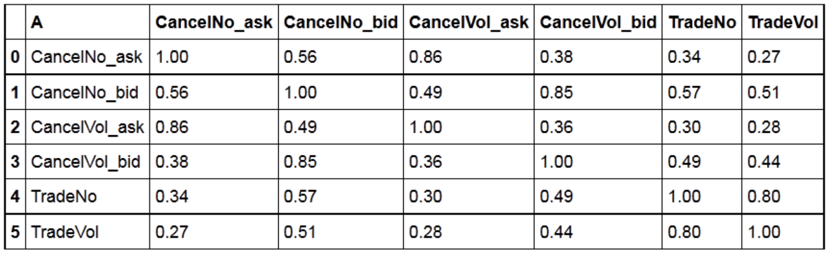
The Pearson coefficient table for ROSN — Fig.29 — reveals a much stronger correlation between the ask side and bid side than it has been observed above for PIKK. Another difference with the previous example is that no negative correlation can be perceived; in this case, the correlation between the number cancellation and the volume canceled is much smaller. The final correlation table — Fig.30 — concerns SBER. The most interesting difference compared with the other examples is the fact that, in this sample, the Pearson coefficients between the trade count or volume and the cancellation number or volume are far more significant. This is even more striking on the bid side; indeed cancellation orders occurring on this side seem closely correlated to the overall trade executions. The last two examples share a few common features: on the one hand, the number of trade and the volumes traded are highly correlated which was not the case in the first example. One the other hand, the number and the volumes of cancellation for both sides are positively correlated.



	A	CancelNo_ask	CancelNo_bid	CancelVol_ask	CancelVol_bid	TradeNo	TradeVol
0	CancelNo_ask	1.00	0.45	0.56	0.26	0.13	0.13
1	CancelNo_bid	0.45	1.00	0.42	0.69	0.09	0.07
2	CancelVol_ask	0.56	0.42	1.00	0.31	0.13	0.13
3	CancelVol_bid	0.26	0.69	0.31	1.00	0.09	0.06
4	TradeNo	0.13	0.09	0.13	0.09	1.00	0.69
5	TradeVol	0.13	0.07	0.13	0.06	0.69	1.00

	A	CancelNo_ask	CancelNo_bid	CancelVol_ask	CancelVol_bid	TradeNo	TradeVol
0	CancelNo_ask	0.00000	0.00000	0.00000	0.00000	0.00211	0.00209
1	CancelNo_bid	0.00000	0.00000	0.00000	0.00000	0.03288	0.09017
2	CancelVol_ask	0.00000	0.00000	0.00000	0.00000	0.00372	0.00389
3	CancelVol_bid	0.00000	0.00000	0.00000	0.00000	0.03017	0.18506
4	TradeNo	0.00211	0.03288	0.00372	0.03017	0.00000	0.00000
5	TradeVol	0.00209	0.09017	0.00389	0.18506	0.00000	0.00000

Figure 29: *ROSN. Correlation matrix.*



	A	CancelNo_ask	CancelNo_bid	CancelVol_ask	CancelVol_bid	TradeNo	TradeVol
0	CancelNo_ask	1.00	0.56	0.86	0.38	0.34	0.27
1	CancelNo_bid	0.56	1.00	0.49	0.85	0.57	0.51
2	CancelVol_ask	0.86	0.49	1.00	0.36	0.30	0.28
3	CancelVol_bid	0.38	0.85	0.36	1.00	0.49	0.44
4	TradeNo	0.34	0.57	0.30	0.49	1.00	0.80
5	TradeVol	0.27	0.51	0.28	0.44	0.80	1.00

Figure 30: *SBER. Correlation matrix.*

In conclusion, those matrices reveal sometimes important correlations between elements that did not appear to be connected at first. Thus, considering such simple analysis before attempting to model the volumes traded could be useful.

8 Conclusion

The reconstruction, the visualization and the analysis of the limit order book is key in order to get a better understanding of the dynamics underlying the price changes of an asset.

Therefore, the first part of this project consisted of implementing a program which could reconstruct the order book from the limit orders flow. First, the price levels were represented by nodes that were then saved into a common structure. Second, it was necessary to choose a suitable structure in order to achieve a precise reconstruction in moderate time. Thus, at the end of the discussion, two simple linked lists, representing the ask side and bid side separately, were selected. This system allows the reconstruction of the limit order book at any given time; this flexible framework opens a wide range of practical applications.

The second part focused on finding ways to visualize the limit order book so as to get a first sense of the available data as well as to get a first impression of the dynamics behind the limit order book. Diverse heat maps were plotted and analyzed, and the full evolution of the book over the entire days was presented as well. From those pictures, interesting connections could be made between the trade activity and the diverse aspects of the limit order book: for example, the spread or the wiggliness of the ask and bid curves. Through investigation of the aggregated mean and median book shape, some conjectures could be made about a hypothetical link between the book shape and the price direction.

The third part was devoted to a small statistical analysis of the order cancellations. Cancellation heat maps helped to provide a global overview of the dynamics. Based on further plots, the mean canceled volume depending on the direction of the mid price level was investigated. For this purpose, the Welch's t-test was used to test if a sudden change in cancellation activities occurs before or at the time of a price increase or decrease; unfortunately, this did not yield any significant results for the examples used in this paper. However, significant correlation between trade volumes and cancellation volumes through Pearson correlation coefficients could be proven in some cases.

In conclusion, this project offers different tools providing a first sense of the data and the dynamics of the limit order book, which could help to conduct deeper analysis. Numerous other expansions could be added to the computational tools to get a wider range of investigation perspective. The next step would be to use those tools and visualizations to model diverse aspects of the financial system so as to get an even more precise understanding of the dynamics underlying the markets.

References

[1] Ruihong Huang, Tomas Polak, *LOBSTER: Limit Order Book Reconstruction* System, 2011

[2] Cont, Rama, *Statistical Modeling of High Frequency Financial Data: Facts, Models and Challenges*, 2011

[3] Tse Jonathan, Lin Xiang, Vincent Drew, *High Frequency Trading - Measurement, Detection and Response*, Working Paper, 2012

[4] Tse Jonathan, Lin Xiang, Vincent Drew, *High Frequency Trading - The Good, The Bad, and The Regulation*, Working Paper, 2012

[5] Marco Avellaneda & Sasha Stoikov, *High-frequency trading in a limit order book*, Quantitative Finance, Taylor & Francis Journals, vol. 8(3), pages 217-224, 2008