

Market Microstructure and Order Book Dynamics in Cryptocurrency Exchanges

Mate Puljiz, Stjepan Begušić, and Zvonko Kostanjčar

Laboratory for Financial and Risk Analytics
Faculty of Electrical Engineering and Computing
University of Zagreb
Unska 3, 10000 Zagreb, Croatia

Email: mate.puljiz@fer.hr, stjepan.begusic@fer.hr, zvonko.kostanjcar@fer.hr

Abstract—We study investor behavior in the Bitcoin/USD market based on a continuous-time stochastic model for the order book dynamics. The simplicity of the model allows for straightforward parameter estimation from the data and a comparison with developed exchanges such as stock markets. The analysis is performed on the order book data stream captured over a continuous period of 300 hours. Our results suggest that the model is a good match for the considered cryptocurrency exchange. We report a large number of front runners in the market, which may artificially inflate both incoming and outgoing order intensities. In addition, we find that symmetrizing bid and ask intensities improves the model accuracy. Based on the model we are able to calculate probabilities of an upward mid-price movement, and shed new light on the market microstructure in cryptocurrency exchanges.

I. INTRODUCTION

Since its inception, Bitcoin has brought the blockchain technology into the spotlight of many academic and industrial endeavours, forming the world of cryptocurrencies. By transforming the financial engineering process of *securitization* into the decentralized mechanism of *tokenization*, cryptocurrencies introduced a novel asset class available to investors worldwide [1]. Over the course of this transformation, the exchange mechanisms have evolved – from enthusiasts meeting in person and using over-the-counter markets, to organized double-auction markets, implemented as online cryptocurrency exchanges featuring simple administration and low transaction fees [2]. Due to their pervasiveness and a generally highly permissive nature, cryptocurrency exchanges quickly became a popular destination for many traders and investors, achieving rampant growth in volume over the past years.

However, while exchanges in developed markets feature liquidity mechanisms and safety procedures (e.g. volatility interrupts) based on the known behavior of investors, this sort of governance is still lacking in most cryptocurrency exchanges. A critical point in this process is understanding the investor dynamics in the double-auction system, which in cryptocurrency markets is yet to be analyzed [3]. Understanding market microstructure and investor behaviour is a key step towards safer and improved implementations of cryptocurrency exchanges, especially when considering the risks of extreme price fluctuations [4], [5], and the security of cryptocurrency exchanges themselves [6].

Up until recently the infinite depth order book data in financial markets was generally hard to obtain. With the arrival of cryptocurrency exchanges with open APIs it became

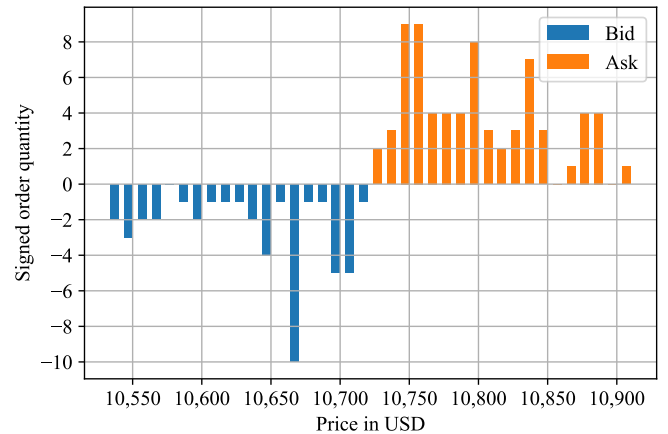


Fig. 1. A snapshot of the limit order book. The limit buy orders (bid) are represented with negative, while the limit sell orders (ask) are represented with positive values.

possible to capture all the incoming orders and reconstruct the complete order book, inciting new insight on the investor behavior and liquidity dynamics in cryptocurrency markets [7]. In addition, this development allowed anyone with a computer and some savings to act as a market maker on these exchanges. As the structure of traders shifted, an open and interesting question is whether the well established models for order book dynamics still provide a satisfactory description of the observed dynamics [8].

In this paper we employ a continuous-time order book model of Cont *et al.* [9] to analyze the market microstructure in a Bitcoin/USD market. The model views each price level in the order book as a queue at which orders arrive and leave at certain rates, estimated as model parameters [10]. Our dataset consists of data captured from Bitstamp’s [11] order book data stream in a continuous period of 300 hours from 2018-01-24 06:00:00 UTC until 2018-02-05 18:00:00 UTC. Due to the simplicity of the model, parameter estimation is quite straightforward, allowing us to perform an empirical analysis of the order book dynamics and estimate probabilities of an upward mid-price move [12].

II. LIMIT ORDER BOOK MODEL

Let us recall the model presented in [9]. At any given time $t \geq 0$ the state of the order book is represented by a vector of

integer entries $(X_1(t), X_2(t), \dots, X_n(t))$ where $\{1, 2, \dots, n\}$ is a fixed price grid and each positive $X_i(t) > 0$ represents the number of limit sell orders at the price level i and for each negative $X_j(t) < 0$, $|X_j(t)|$ represents the number of limit buy orders at price level j .

Not all integer vectors represent a valid state of the order book. The ask price at time t is defined as the lowest price level for which there is a limit sell order

$$p_A(t) = \inf\{1, 2, \dots, n | X_i(t) > 0\}, \quad (1)$$

and similarly the bid price at time t is the highest price level for which there is a buy limit order

$$p_B(t) = \sup\{1, 2, \dots, n | X_i(t) < 0\}. \quad (2)$$

Clearly, the only valid representations are those for which $p_B(t) < p_A(t)$. Figure 1 represents one valid state vector. Note that the price level i in that figure corresponds to the price $i \cdot 10\text{USD}$ as the tick size we used was 10USD.

The set of all admissible vectors forms a state space for a continuous Markov process that provides the dynamics in this model. Given that the process is in the state $x_t = (X_1(t), X_2(t), \dots, X_n(t))$, it is only allowed to transition with non-zero intensity to the states $x_{t'}$ for which $x_{t'} = x_t \pm e_i$, where $e_i = (0, \dots, 1, \dots, 0)$ is i^{th} vector of the canonical basis.

To put it precisely, let $\lambda(i) > 0, 1 \leq i \leq n$ be the intensities of arrivals of limit orders at the price level i ticks away from the opposite best, $\theta(i) > 0, 1 \leq i \leq n$ the intensities of cancellations of limit orders at the price level i ticks away from the opposite best, and $\mu > 0$ the intensity of market orders. These parameters are assumed to be fixed in the observed period and they completely determine the dynamics as follows.

Given that the process at time t is in the state x , denoted by p_B, p_A the best bid and ask respectively, then the transition rates to the state $x' = x + e_i$ are:

$$q_{x \rightarrow x'} = \begin{cases} |X_i| \theta(p_A - i), & i < p_B < p_A \\ \mu + |X_{p_B}| \theta(p_A - p_B), & i = p_B \\ \lambda(i - p_B), & p_B < i \end{cases} \quad (3)$$

to the state $x' = x - e_i$:

$$q_{x \rightarrow x'} = \begin{cases} \lambda(i - p_A), & i < p_A \\ \mu + |X_{p_A}| \theta(p_A - p_B), & i = p_A \\ |X_i| \theta(i - p_B), & p_B < p_A < i \end{cases} \quad (4)$$

and 0 to all the other states.

Note that this model assumes independent and exponentially distributed waiting times between orders.

III. PARAMETER ESTIMATION

As we mentioned before, our dataset comes from capturing a live stream of orders provided by Bitstamp exchange [11]. From this data we could reconstruct a good portion of the limit order book at any point in time. The next step was to aggregate the data using a tick size of 10USD. The bid prices were rounded down, while the ask prices were rounded up.

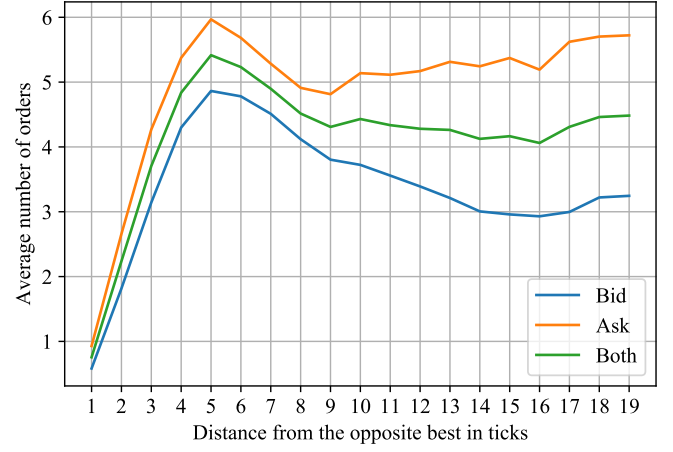


Fig. 2. Approximation of the steady state shape of the book. The model we use is known to be ergodic and it is therefore meaningful to take time averages of snapshots of the order book to obtain the steady state shape of the book.

The intensities $\lambda(i)$ were then estimated to be

$$\lambda(i) = \frac{N_l(i)}{2T}, \quad (5)$$

where $N_l(i)$ was the number of limit buy or sell orders that arrived at the distance i ticks from the opposite best, and T the length of the observed period in seconds.

Similarly $\theta(i)$ was estimated as

$$\theta(i) = \frac{N_c(i)}{2TQ(i)}, \quad (6)$$

where $N_c(i)$ was the number of cancellation buy or sell orders that arrived in the observed period and $Q(i)$ the average number of orders at the distance i from the best bid/ask. In Figure 2 this value is plotted in green.

Finally μ was estimated as

$$\mu = \frac{N_m(i)}{2T} \frac{S_m}{S_l}, \quad (7)$$

where $N_m(i)$ was the number of market buy or sell orders that arrived in the observed period, S_m was the average size of market orders, and S_l the average size of orders at the best price levels. This correction is needed as market orders are on average smaller in size than limit orders.

In fact we were able to obtain a slightly better results when, instead of using one unique parameter μ , we used different normalisations for buy and sell market orders. To be precise, we used

$$\mu^B = \frac{N_m(i)}{2T} \frac{S_m^B}{S^A(1)}, \quad (8)$$

as market buy intensity where S_m^B was the average size of market buy orders and $S^A(1)$ the average size of sell orders at the best ask. Similarly

$$\mu^A = \frac{N_m(i)}{2T} \frac{S_m^A}{S^B(1)}, \quad (9)$$

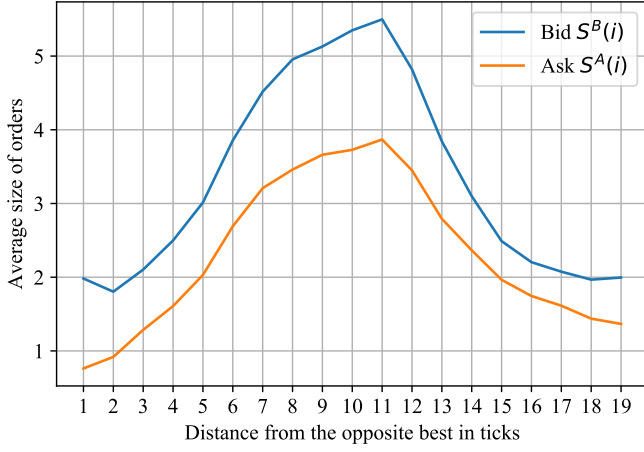


Fig. 3. The average size of buy (Bid $S^B(i)$) and sell (Ask $S^A(i)$) orders at the given distance from the opposite best order.

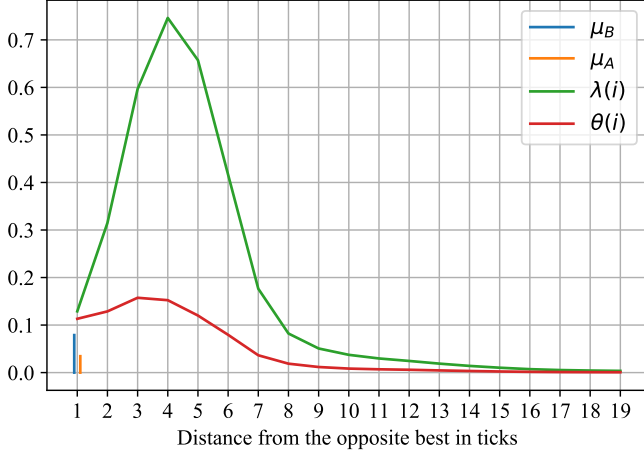


Fig. 4. Estimated model parameters: market buy (sell) order μ_B (μ_A), symmetrized limit $\lambda(i)$ and cancellation $\theta(i)$ intensities.

was estimated market sell intensity where S_m^B was the average size of market buy orders and $S^B(1)$ the average size of buy orders at the best bid. In Figure 3 we see graphs of both S^B and S^A . All the estimated parameters are shown in Figure 4.

IV. PROBABILITY OF AN UPWARD PRICE MOVE

Once the parameters have been successfully estimated, the model can be used to produce predictions which can then be put to test.

One of the predictions that the authors in [9] are able to compute is the expression for the probability of the next upward price move. A particularly simple case is that of spread being smaller than tick size.

Recall that the spread at time t in this model is $p_A(t) - p_B(t)$ ticks and the mid-price level is given as

$$p_M(t) = \frac{p_A(t) + p_B(t)}{2}. \quad (10)$$

If $p_A - p_B = 1$ then the only way that the mid-price can change is either if the number of orders at the best bid $|X_{p_B}|$

TABLE I. PREDICTED FREQUENCIES OF THE UPWARD MID-PRICE MOVE

		a				
b		1	2	3	4	5
1		0.533	0.385	0.317	0.284	0.261
2		0.694	0.547	0.483	0.441	0.409
3		0.773	0.634	0.562	0.530	0.501
4		0.819	0.698	0.621	0.577	0.558
5		0.851	0.746	0.672	0.621	0.591

TABLE II. OBSERVED FREQUENCIES OF THE UPWARD MID-PRICE MOVE

		a				
b		1	2	3	4	5
1		0.521	0.310	0.249	0.207	0.184
2		0.741	0.534	0.450	0.428	0.403
3		0.816	0.631	0.522	0.477	0.495
4		0.848	0.671	0.592	0.564	0.606
5		0.867	0.716	0.639	0.615	0.609

go to 0, or if the number of orders at the best ask $|X_{p_A}|$ go to 0. Indeed, in former case the mid-price will move downwards and in the latter upwards. Thus, the probability of the upward move is equal to the probability that the queue of orders at the price level p_A reaches zero before the queue of orders at the level price p_B .

It is then known from the queuing theory that the required probability can be computed using incoming and outgoing order intensities of those two queues. The precise expression for this probability is given in [9, Proposition 3].

V. RESULTS

We used the estimated parameters to compute the probabilities that the mid-price moves upwards given that the spread is of size less than one tick and that the number of orders at the best bid and ask are $1 \leq a \leq 5$ and $1 \leq b \leq 1$ respectively. The results are given in Table I.

But from the data we can in each instance, for which the spread was under the tick size, top ask consisted of a orders, and top bid of b orders, calculate whether the mid-price actually went up or down. Taking then the relative frequency of the positive outcomes should, if the model is right, match the predicted probabilities.

The observed frequencies are given in Table II below and we can see that they match those in Table I quite nicely. In fact the matrix distance between the two is around 0.199. This accuracy is comparable to 0.48 that was obtained in [9] for conventional exchanges.

VI. CONCLUSION

We tested alignment of the limit order book model developed in [9] with the data coming from one of the biggest Bitcoin exchanges and found a good match with the model when it comes to computing probabilities of the upward mid-price movement.

It was proven in [9] that this model is an ergodic Markov process and therefore it is meaningful to take time averages and compute steady state distributions. This also means that as

we increase the amount of observed data, one expects to see more accurate matching with the model and this effect could explain the better accuracy we see when compared with that in the original paper.

There is plenty of space for improvements however. For example, we tried separating buy and sell orders and then estimating arrival and cancellation rates separately for each, but the results were not as good as when they were taken together. This could be due to the effect that symmetrizing the intensities by taking them together pushes all the probabilities towards $\frac{1}{2}$ which leads to a better model or it could be that by halving $\frac{1}{2}$ the amount of data we use to estimate each parameter we inevitably lose on accuracy, or both.

Another problem when analysing these markets is a large number of front runners, traders who most of the time submit just a tiny bit better offer in front of the best order to make sure their order is executed first. This artificially inflates both incoming and outgoing intensities and results that they almost match. A good way to exclude this effect could potentially lead to a more accurate parameter estimation and a better model.

ACKNOWLEDGMENT

This research was supported by the Croatian science foundation project ASYRMEA (5349).

REFERENCES

- [1] R. Böhme, N. Christin, B. Edelman, and T. Moore, "Bitcoin: Economics, Technology, and Governance," *Journal of Economic Perspectives*, vol. 29, no. 2, pp. 213–238, 2015. [Online]. Available: <http://pubs.aeaweb.org/doi/10.1257/jep.29.2.213>
- [2] T. Kim, "On the transaction cost of Bitcoin," *Finance Research Letters*, vol. 23, pp. 300–305, 2017.
- [3] M. Brandvold, P. Molnár, K. Vagstad, and O. C. Andreas Valstad, "Price discovery on Bitcoin exchanges," *Journal of International Financial Markets, Institutions and Money*, vol. 36, pp. 18–35, may 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S104244311500027X>
- [4] J. Osterrieder and J. Lorenz, "A Statistical Risk Assessment of Bitcoin and its Extreme Tail Behavior," *Annals of Financial Economics*, vol. 12, no. 01, mar 2017. [Online]. Available: <http://www.worldscientific.com/doi/abs/10.1142/S2010495217500038>
- [5] S. Begušić, Z. Kostanjčar, H. E. Stanley, and B. Podobnik, "Scaling properties of extreme price fluctuations in Bitcoin markets," *Preprint submitted to Physica A: Statistical Mechanics and its Applications*, 2018.
- [6] T. Moore and N. Christin, "Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk," in *International Conference on Financial Cryptography and Data Security*. Springer, Berlin, Heidelberg, 2013, pp. 25–33. [Online]. Available: http://link.springer.com/10.1007/978-3-642-39884-1_3
- [7] J. Donier and J. P. Bouchaud, "Why do markets crash? Bitcoin data offers unprecedented insights," *PLoS ONE*, vol. 10, no. 10, oct 2015. [Online]. Available: <http://dx.plos.org/10.1371/journal.pone.0139356>
- [8] R. Cont, "Statistical modeling of high-frequency financial data," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 16–25, 2011.
- [9] R. Cont, S. Stoikov, and R. Talreja, "A stochastic model for order book dynamics," *Operations research*, vol. 58, no. 3, pp. 549–563, 2010.
- [10] C. d. J. L. de Abreu, "Stochastic modeling of order book dynamics," Master's thesis, Delft University of Technology, Delft, 2010.
- [11] Bitstamp. [Online]. Available: <https://www.bitstamp.net>
- [12] R. Cont and A. De Larrard, "Price dynamics in a markovian limit order market," *SIAM Journal on Financial Mathematics*, vol. 4, no. 1, pp. 1–25, 2013.