Draft: Dec. 2015

Limit order book models and market phenomenology

Jun Hu

Department of Industrial Management,
Tampere University of Technology,
P.O.Box 541, FI-33101 Tampere, Finland

jun.hu@tut.fi

Abstract: In this paper, we propose a dynamical model of the limit order book. After postulating the behavior of order placement, execution and cancellation, Monte-Carlo simulation reconstruct the evolution of the limit order book. Many important aspects of market phenomenology, such as diffusion, jumps, spread and return, emerge from the simple model with less than ten parameters.

Keywords: Limit order book, Monte-Carlo simulation, Diffusion, Jumps, Spread

Contents

1 Introduction

There are several ways of modeling the limit order book.

The research of market micro-structure begins by Garman[1]. Later comes the empirical study of the properties of the order book. O'Hara[2], Keim and Madhavan[3], Coughenour and Shastri[4], Biais, Hillion and Spatt[5] present their empirical analysis of limit order book and order flow in different financial markets. They consider a variety of properties of the market, such as trading cost, public information, average spread, cumulative order distribution and event frequency. Though no real dynamical model is proposed by now. Empirical research provides insights to the later model building.

1.1 Equilibrium Models

The first order-based method is used mostly by financial econometricians. They first propose a simplified structure for the order book and utility functions for different market participants. The problem then reduces to finding the optimal trading strategy which optimizes the utility functions. Such solution is called equilibrium state of the model. With the explicit solution, one can derive the static properties of the market, such as spread, order depth and price distribution etc.

Example of such models are

- Glosten[6] analyses the general condition that utility functions have to satisfy. Since the market structure is not definitely specified, the model can only be partially solved, thus lead to some qualitative facts of the market. When existing a large number of liquidity suppliers, the market exhibits a small spread, and the liquidity suppliers make profit out of the trades.

- Parlour[7] investigates a simple one-tick model, where traders submit orders according to the market conditions and the utility function is clearly defined. Game theoretic equilibrium can be solved in this setting. The most important feature of this model is that, when making decisions, traders know their decision can make an influence on the later traders, and they take this effect into account.

- Rosu[8] assumes that trading is taking place in a bounded region $[A, B]$. Traders in the market are categorized as buyer or seller, patient or impatient. They arrive in the market at different rates, while deciding whether to place a limit order of market order by optimizing their own utility function. Equilibrium can be found with the above assumptions. The model reproduce some of the stylized facts about the market, such as, higher trading volume reduces the spread and dampens the impact from large market orders.

Though specific settings are different, the above models face the same problems. First, utility function plays the central part in these models, however it is different for different trader, and it is hard to express in written forms. The authors often propose it for the sake of solvability, which undermines the applicability of these models. Furthermore, the model structures are also quite simple, which contradict the real world.

1.2 Order-flow Models

The second class of models, many of which made by physicists, focus on the dynamics of the order book. They, instead of guessing what traders think by postulating unreal utility functions, model the net effect of order flow directly. Since the status of order book can be completely derived from the order flow, everything is well-defined after we specify how orders arrive at and leave the order book.

Most of such models share the common structure.

- Buyers and sellers arrive at the market in two independent static Poisson processes.

- With certain probability, the trader would place a market order, otherwise he would place a limit order according to certain distribution.

- Every order is of the same size and can be placed on a price grid.

- Remaining orders on the book is canceled with exponential process.

Though similar, these models do have slight difference.

- Maslov[9][10] proposes that limit order is priced with a uniform distribution in a bounded region away from ask/bid price, and the limit orders can not be canceled.

- Smith et al.[11] use a logarithmic price grid. Limit orders are uniformly distributed in an unbounded price region.

- Luckock[12] proposes a very general limit order distribution without order cancellation.

- Cont et al.[13] assume a linear price grid and the limit order is distributed according to power law $d^{-\alpha}$, where d is the distance to ask/bid price.

Similar models include Bouchaud et al.[14], Bovier et al.[15], Challet and Stinchcombe[16], Daniels et al.[17][18] and Slanina[19].

This kind of mechanical models have several advantages. First, all the parameters, such as arrival rate, cancellation rate and limit order distribution are observable on the market. Therefore, we can use real market data to calibrate the models. Second, though not easy to solve exactly, simulation can be done to make testable predictions with calibrated parameters. Last but not least, we have the freedom to relate the observable parameters to the fundamentals of the market, as opposed to in the equilibrium models, every parameter has a economical interpretation already.

Up to now, only some aspects of such models have been solved theoretically or by simulation, without empirical confirmation. To test whether the assumptions of the model are reasonable, out-of-sample analysis needs to be done with real market data.

1.3 State-based models

Though modeling the state of the order book (usually characterized by cumulative distribution of the order book) is the most intuitive way, there are very few models of this kind.

Malo and Pennanen[20] pioneer the research by characterizing the order book by three parameters, mid-price s, slope of cumulative distribution of sell orders β_s and slope of cumulative distribution of buy orders β_b. The model assumes s to be a geometric Brownian motion and βs follow a 2-dimensional OU process.

The lack of state-based models may due to the fact that they neither investigate the traders' own benefit by their utility functions, nor consider the net effect of what is going on in the real market by order flow. The assumptions seem quite artificial. Therefore, it is hard to convince people that this model is economically more reasonable than others. Moreover, without full specification of the market, such model often omits some properties of the market, such as in Malo and Pennanen[20], information regarding spread and order execution cannot be derived from the model.

Among the three types of models reviewed above, the first one is the most extensively studied, but not suitable for quantitative study. The second type is gaining popularity, and has the potential of being the measure of market activities. While the last one still needs more conceptual justification. The model we propose in this paper falls into the second type.

2 The model setting

In this paper, we propose a dynamic model of limit order book, whose evolution is dependent on the current state of the order book. The model assumptions are:

1. Limit orders can be placed on the whole log-price domain $\mathbb{R}$;
2. Each order (market or limit) is of the same size 1;
3. Limit buy (sell) orders are placed with a constant distribution $p_b(\text{ask}, x)$ ($p_s(\text{bid}, x)$) with respect to the current ask (bid) price with constant exponential arrival rate λ_b (λ_s);
4. Market buy (sell) orders are placed with constant arrival rate μ_b (μ_s);
5. Limit buy (sell) orders remaining on the order book are canceled with universal cancellation rate χ_b (χ_s).

The model is very simple, because only several parameters are needed to model the order book and many of them can be directly observed (calibrated) with the market data, such as limit and market order arrival rates λ and μ and cancellation rate χ. However, the distribution of the limit buy (sell) orders p are not so unambiguously define as the others, because a distribution has to be described by infinitely many parameters.

With different goals in mind, we can define the distribution accordingly. In the following sections, we are going to model several important market properties by this model setting (with different specifications for p).

3 Simulation

Unlike a Brownian motion, whose state can be characterized solely by its position (log-price), the state of the current model is characterized by its whole order book, which is of (potentially) infinite dimensions. Although such a model is not easy to solve analytically, Monte-Carlo simulation is applicable, since the evolution of the order book is well-defined. Simulation for the model can be done as follows.

1. At time t, calculate the sum of $r = \lambda_b + \lambda_s + \mu_b + \mu_s + \chi_b s_b + \chi_s s_s$, where s_b (s_s) is the number of limit buy (sell) orders remaining on the order book.
2. Simulate $\Delta t = -\log(d_1)/r$, where d_1 is a random number uniformly distributed on $[0, 1]$. $t + \Delta t$ is the time of next event.
3. Choose the type of event by rd_2, where d_2 is another $U(0, 1)$ random number.
   (a) If $0 < rd_2 < \lambda_b$, a limit buy order is placed at $a_1 - P_b^{-1}(d_3)$, where a_1 is the current ask price, P_b is the CDF (Cumulative distribution function) for p_b and d_3 is a $U(0, 1)$ random number.
   (b) If $\lambda_b < rd_2 < \lambda_b + \lambda_s$, a limit sell order is placed at $b_1 + P_s^{-1}(d_3)$, where b_1 is the current bid price, P_s is the CDF for p_s and d_3 is a $U(0, 1)$ random number.

(c) If $\lambda_b + \lambda_s < rd_2 < \lambda_b + \lambda_s + \mu_b$, a market buy order is placed, executed at the current ask a_1.

(d) If $\lambda_b + \lambda_s + \mu_b < rd_2 < \lambda_b + \lambda_s + \mu_b + \mu_s$, a market sell order is placed, executed at the current bid b_1.

(e) If $\lambda_b + \lambda_s + \mu_b + \mu_s < rd_2 < \lambda_b + \lambda_s + \mu_b + \mu_s + \chi_b s_b$, a randomly selected limit buy order is canceled.

(f) If $\lambda_b + \lambda_s + \mu_b + \mu_s + \chi_b s_b < rd_2 < \lambda_b + \lambda_s + \mu_b + \mu_s + \chi_b s_b + \chi_s s_s$, a randomly selected limit sell order is canceled.

4. Repeat the above steps to simulate time and event for further times.

4 Diffusion process

In Mathematical Finance, diffusion process is the main tool to introduce uncertainty. The Black-Scholes model assumes the simplest diffusion process, Brownian motion, and manages to explain, with two parameters, many phenomena observed in the real market. Despite its massive success, Brownian motion fails to capture some fundamental properties.

1. Brownian motion is self-similar, while the real market, observed in ultra-microscopic (tick) level, is not diffusive at all. It is comprised of a large number of small jumps between the current bid and ask. Lévy-type models can generate long-term diffusion process by accumulating small jumps at tick scale. The problem is, Lévy models assume i.i.d. jumps at any time, however, at tick scale, the jumps are hardly i.i.d. What is observed in the market is that the price jumps back and forth between bid and ask most of the time.

2. Market depth is another important feature of the real market. For illiquid equities, a single price means very little other than the previous transaction. The state space of most financial models based on stochastic calculus is one-dimensional (price), therefore market depth can only be added artificially.

Limit order book models, on the other hand, provide simple and economically sensible solutions to problems addressed above. For simplicity, we propose that the distribution for limit order placement is half of a normal distribution, whose center is situated at the current bid b_1 and ask a_1 respectively,

$$p_b = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a_1)^2}{2\sigma^2}\right), \quad x < a_1; \tag{4.1a}$$

$$p_s = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-b_1)^2}{2\sigma^2}\right), \quad x > b_1. \tag{4.1b}$$

With the above specification (Figure 1 and Equation 4.1), the model is fully determined and can therefore be simulated.

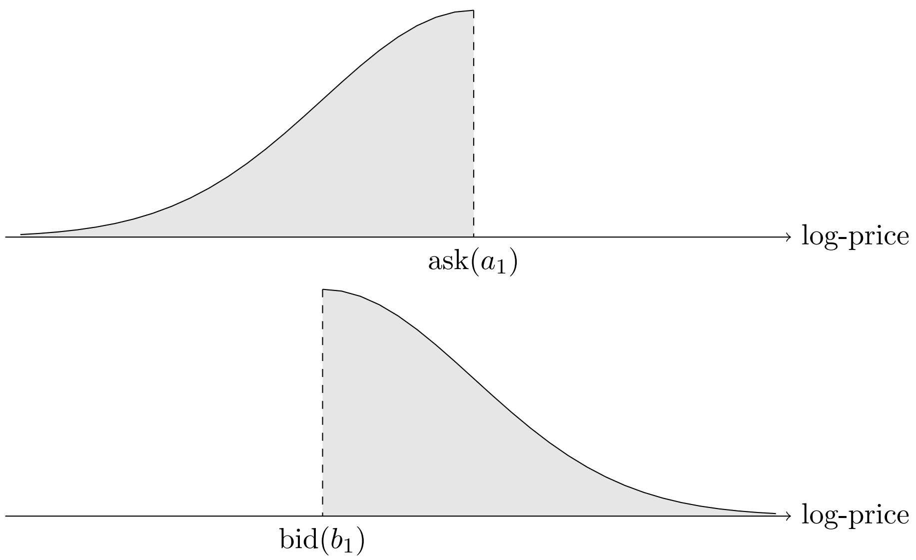

Figure 1. Limit order placement distribution. Buy (sell) limit orders are distributed normally below ask (above bid).

4.1 Price process

Figure 2 shows the price movement for tick and hour scale. At tick scale, the price jumps randomly as market orders arrive. Unlike Lévy jumps, which are often assume to have unbounded range, the order book jumps here happen between the bid and ask at the time, making them less i.i.d. then Lévy jumps.

At hour scale, the price process *appears* very much the same as that of a diffusion. This phenomenon can be explained by Central Limit Theorem. Because the distribution of limit order placement is stationary and moves along with the reference price (bid or ask), the number and distance of the remaining orders with respect to reference price is also stable. For example, for the following parameters

$$\lambda_b = \lambda_s = 1, \quad \mu_b = \mu_s = 0.1, \quad \chi_b = \chi_s = 0.01, \quad \sigma_b = \sigma_s = 1, \tag{4.2}$$

the time average of remaining buy (sell) orders in the book is 90. When the number is below 90, the orders arrive more often than being canceled, vice versa. The mean-reverting process becomes apparent when the market is looked at a long enough scale. Figure 3 shows the correlation of 1-tick return $\mathrm{corr}(p_i - p_{i-1}, p_{i-1} - p_{i-2})$ and 100-tick return $\mathrm{corr}(p_i - p_{i-100}, p_{i-100} - p_{i-200})$, where p_i denotes the ith transaction in a time series. There is clear negative correlation between two consecutive ticks, which contradicts the assumption of the Brownian motion. However, the 100-tick (minutes) return behaves much more random, agreeing with our conjecture of long-term i.i.d.ness. We can further test the conjecture by setting the order placement distribution p to a different distribution, as long as the distribution is stable over time. The result shows that no matter what distribution we choose, the long-time return converges to normal distribution (Figure 4).

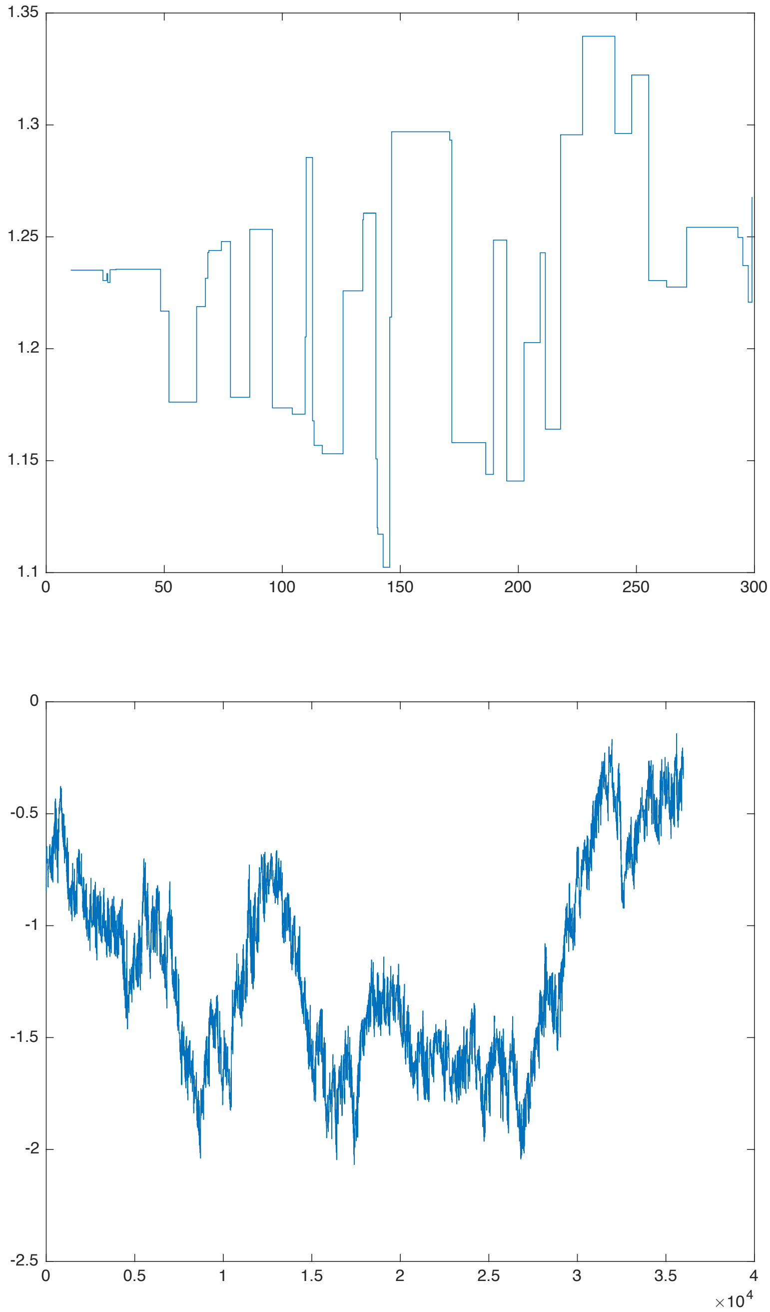

Figure 2. Price processes for tick scale (above) and hour scale (below).

Summing up, the order book model we propose generates a process with the desirable properties listed in Table 1.

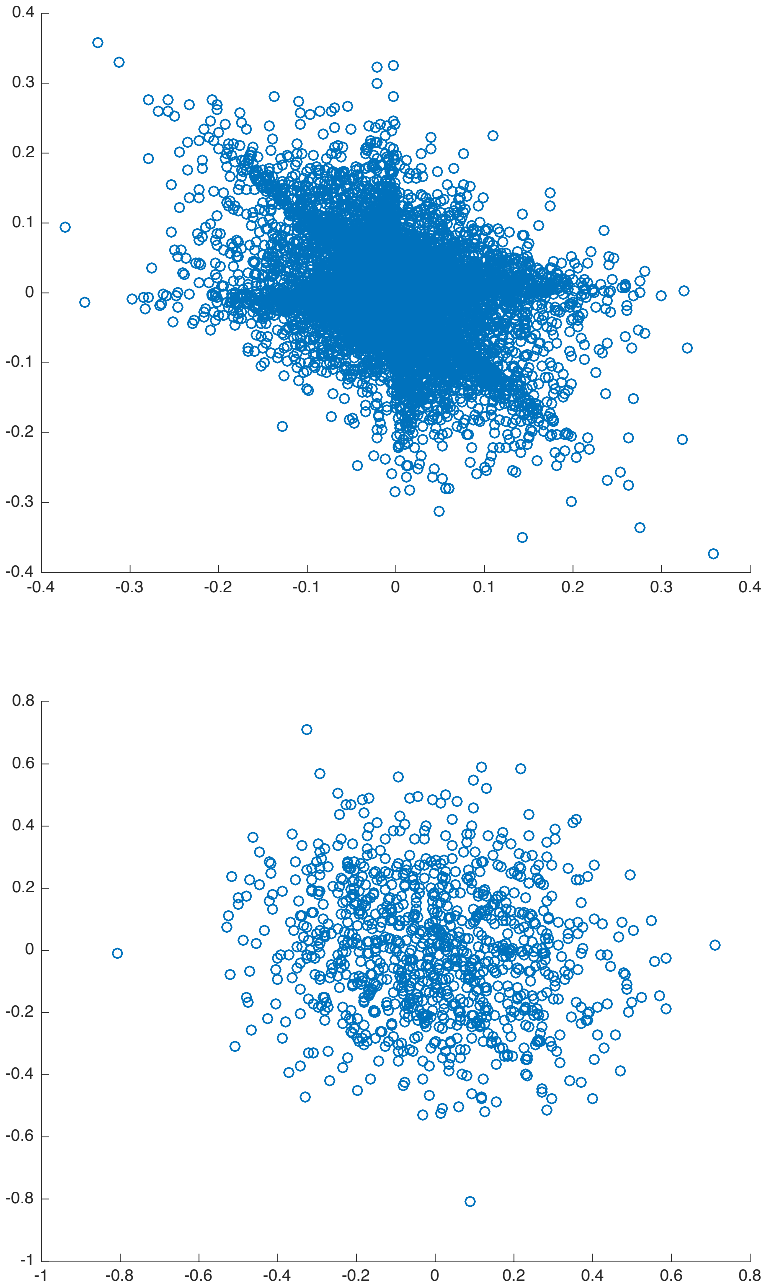

Figure 3. Correlation of 1-tick (above) and 100-tick return (below).

5 Spread and volatility

As mentioned before, the distribution for order placement is not unique. In this section, in order to model spread and volatility, which are also important market statistics, we propose

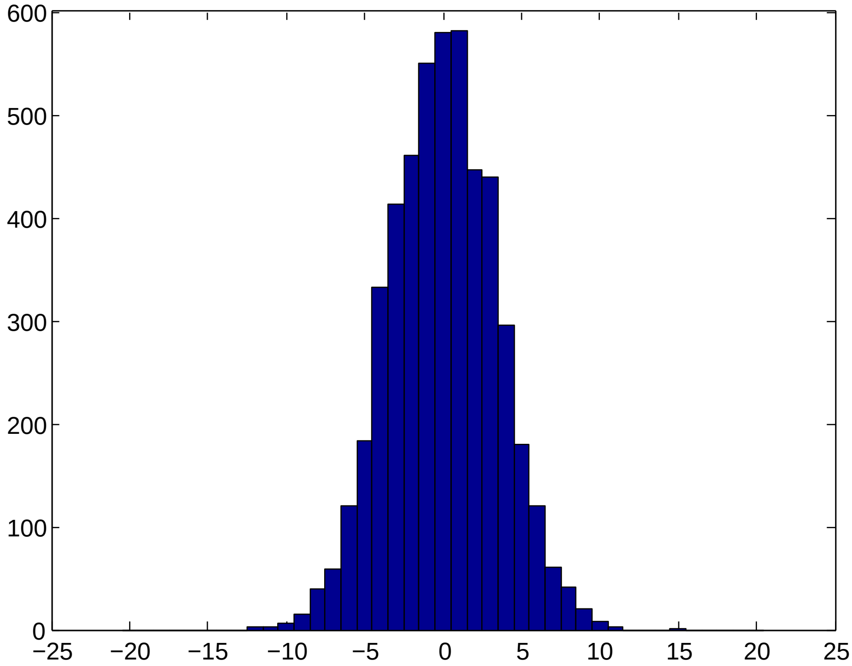

Figure 4. Long-term return distribution.

Property	Short term (tick)	Long term (hour)	Brownian motion
Return distribution	Correlated	I.i.d	I.i.d
Continuity	Non-continuous	Appear continuous	Continuous

Table 1. Comparison of the price process with Brownian motion.

the fixed point distribution,

$$p_b = \delta(a_1 - \sigma_b), \tag{5.1}$$

$$p_s = \delta(b_1 + \sigma_s). \tag{5.2}$$

In this setting, a buy (sell) limit order is always placed σ_b below the ask (σ_s above the bid).

5.1 Calibration

Now that the model has been determined, we can calibrate the model parameters with market data. In this paper, we use the feed data for shares of Google (GOOG) from 6 July to 24 July, 2015. The data includes information for every order placement, order (partial) cancellation and order (partial) execution. The data is downloaded from www.tradingphysics.com.

Order size Although order size does not appear in the model setting, it is used to determine other model parameters. In a period $[t_1, t_2]$, the order size is defined as the average size of trades executed

$$\delta = \frac{1}{n}\sum_{i=1}^{n}\Delta_i, \tag{5.3}$$

where n is the number of executions and Δ_i is the order size of the i-th execution.

Market order arrival rate The market buy and sell order arrival rate are defined as

$$\mu_b = \frac{1}{\delta(t_2 - t_1)}\sum_{i=1}^{n}\Delta_{bi}, \quad \mu_s = \frac{1}{\delta(t_2 - t_1)}\sum_{i=1}^{n}\Delta_{si}, \tag{5.4}$$

where δ is the average order size, Δ_{bi} (Δ_{si}) is the size of the i-th buy (sell) market order in time $[t_1, t_2]$.

Limit order arrival rate The limit buy and sell order arrival rate are defined as

$$\mu_b = \frac{1}{\delta(t_2 - t_1)}\sum_{i=1}^{n}\Delta_{bi}, \quad \mu_s = \frac{1}{\delta(t_2 - t_1)}\sum_{i=1}^{n}\Delta_{si}, \tag{5.5}$$

where δ is the average order size, Δ_{bi} (Δ_{si}) is the size of the i-th buy (sell) limit order in time $[t_1, t_2]$.

Limit order distance The limit order distance is the average distance between the ask price and limit price when the limit order is placed

$$\sigma_b = \frac{\sum_{i=1}^{n}(a_i - p_i)\Delta_{bi}}{\sum_{i=1}^{n}\Delta_{bi}}, \quad \sigma_s = \frac{\sum_{i=1}^{n}(p_i - b_i)\Delta_{si}}{\sum_{i=1}^{n}\Delta_{si}}, \tag{5.6}$$

where a_i (b_i) is the ask (bid) price when the limit order is placed, p_i is the limit order price, Δ_{bi} (Δ_{si}) is the size of the buy (sell) limit order, in $[t_1, t_2]$.

Cancellation rate The cancellation rate is the average number of orders being canceled for every limit order in the book in a unit of time.

$$\chi_b = \frac{\sum_{i=1}^{n}\bar{\Delta}_{bi}}{\sum_{i=1}^{n}(t_i - t_{i-1})\Delta_b i}, \quad \chi_s = \frac{\sum_{i=1}^{n}\bar{\Delta}_{si}}{\sum_{i=1}^{n}(t_i - t_{i-1})\Delta_{si}}, \tag{5.7}$$

where $\bar{\Delta}_{bi}$ ($\bar{\Delta}_{si}$) is the size of a canceled buy (sell) limit order, and Δ_{bi} (Δ_{si}) is the total size of buy (sell) limit orders in the book in period $[t_{i-1}, t_i]$.

Calibration results Since most trading activities of the whole trading day (4:00–20:00) happen in the regular trading hours (9:30–16:00, RTH), we use data in RTH for parameter estimates. The following are the parameter estimates for RTH of 24 July, 2015

$$\begin{array}{llll} \lambda_b = 3.769957 & \lambda_s = 2.804203 & \sigma_b = 6.650542 \times 10^{-4} & \sigma_s = 4.983924 \times 10^{-4} \\ \mu_b = 0.1471042 & \mu_s = 0.238024 & \chi_b = 6.681366 \times 10^{-3} & \chi_s = 2.806163 \times 10^{-3} \end{array}$$

With the model and parameters clearly defined, Monte-Carlo simulation of the order book can be performed. We can study some important properties of the market which emerge directly from the order book model, without further structural assumptions.

5.2 Simulation results

For this model, we try to reproduce three important market statistics of the order book.

- Spread, defined as the difference between the ask and bid.
- Short-term variance, defined as $v = \frac{1}{n}\sum_{i=1}^{n}(p_i - p_{i-1})^2$, where p_i and p_{i-1} are last execution prices separated by 10 seconds.
- Daily return, defined as the difference between the market open and close.

The results are shown in Table 2. The first two columns show that spread can be modeled quite accurately. In other words, limit order distance σ is a critical scale of the market micros-structure. The limit orders can be regarded as placed at their average distance to reference price (bid and ask), in terms of spread modeling.

It is not surprising that the model manages to produce the short-term variance to the right order of magnitude, since short-term variance is closely related to the scale of market micro-structure (spread). While for long-term variance, micro-structure is much less important than the trend in real market, which the order book model does not account for.

The last two columns of Table 2 and Figure 5 show the ability of the model to explain daily return. Especially, clear correlation of market and simulated returns can be observed from the figure. One possible explanation could be that movement of the return is largely determined by the average arrival rates for both market and limit orders. For instance, if the arrival rate for buy market order increases, more sell limit orders are executed and eliminated, leaving less resistance for upward motion. On the order hand, if the arrival rate for sell limit order increases, the resistance increases as well, making the price more likely to move downward.

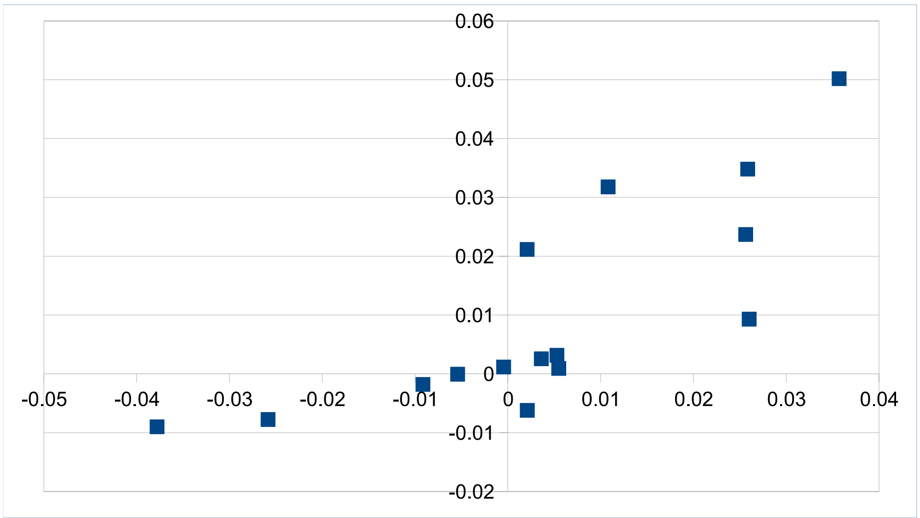

Figure 5. Market-observed and simulated daily returns. Horizontal axis indicates the real market return from data, and vertical axis indicates the model simulated return for the same day.

Date	Spread	Simulated	Variance	Simulated	Return	Simulated
0706	0.0005325983	0.0004502444	3.21e-8	7.29e-8	0.0054835631	0.0009162921
0707	0.0006414786	0.0006173237	5.20e-8	1.60e-7	0.002096517	-0.006206034
0708	0.0007351041	0.0006402037	5.92e-8	1.60e-7	-0.009148842	-0.00180183
0709	0.0005541411	0.0003907624	3.65e-8	6.26e-8	-5.40e-3	-5.82e-5
0710	0.0004284492	0.0004234418	3.98e-8	7.56e-8	0.005295621	0.003162072
0713	0.0004292954	0.0003748533	4.57e-8	6.30e-8	0.02600775	0.009292702
0714	0.000427353	0.000421939	5.62e-8	8.80e-8	0.02563317	0.02369774
0715	0.0004847082	0.0004588641	3.33e-8	9.10e-8	-0.0004462891	0.001161786
0716	0.0003836711	0.0002751401	3.61e-8	5.85e-8	0.0258462	0.03480529
0717	0.0003539351	0.000397179	1.25e-7	9.38e-8	0.03570511	0.05018066
0720	0.0004515131	0.0005285777	1.21e-7	1.43e-7	0.003608613	0.002555863
0721	0.0004533309	0.0004568103	6.38e-8	1.16e-7	0.01081075	0.03179945
0722	0.0005517805	0.0006323116	9.04e-8	1.97e-7	0.002086137	0.021157756
0723	0.000523238	0.0003976895	8.64e-8	7.48e-8	-0.025839806	-0.007773562
0724	0.0005433408	0.000498606	9.61e-8	1.14e-7	-0.037799776	-0.009001335

Table 2. Market-observed and simulated spread, short-term variance and daily return.

6 Conclusion

In this paper, we propose a simple Markovian model for the limit order book. The model includes all types of market activities, such as limit order placement, market order execution and limit order cancellation. Even with many market features, the model is simple enough to be described by less than ten parameters.

With normal distribution for limit order placement, the model bridges the gap between short-term non-i.i.d. jumps and long-term diffusion. This process is much closer to the real market than the Black-Scholes model, or any other SDE-based stochastic processes. Spread and market depth emerge naturally from simulation, without further artificial assumptions.

A specific limit order placement distribution, point distribution $\delta(x)$, can reproduce spread and short-term variance very accurately. The simple model is also able to roughly predict the direction of the daily movement.

Limit order book models are real "dynamical" models of the market, because the price is formed as a result of the real processes of order placement, execution and cancellation, rather than being determined by hand-written SDEs. Though SDE models are superior in term of tractability, order book models have the ability to relate the real world via order operations. The model we propose in this paper is deliberately made as simple as possible, by assuming time-homogeneity, uniform order size, fundamental independence, uniform cancellation rate and Markov property. In future research, more complex models can be proposed to model the real market, by releasing those restrictions.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under the grant agreement no. 289032 (HPCFinance).

References

[1] M. B. Garman, *Market microstructure*, *Journal of financial Economics* **3** (1976), no. 3 257–275.

[2] M. O'Hara, *Market microstructure theory*, vol. 108. Blackwell Cambridge, 1995.

[3] D. B. Keim and A. Madhavan, *The cost of institutional equity trades*, *Financial Analysts Journal* (1998) 50–69.

[4] J. Coughenour and K. Shastri, *Symposium on market microstructure: A review of empirical research*, *Financial Review* **34** (1999), no. 4 1–27.

[5] B. Biais, P. Hillion, and C. Spatt, *An empirical analysis of the limit order book and the order flow in the paris bourse*, *the Journal of Finance* **50** (1995), no. 5 1655–1689.

[6] L. R. Glosten, *Is the electronic open limit order book inevitable?*, *The Journal of Finance* **49** (1994), no. 4 1127–1161.

[7] C. A. Parlour, *Price dynamics in limit order markets*, *Review of Financial Studies* **11** (1998), no. 4 789–816.

[8] I. Roşu, *A dynamic model of the limit order book*, *Review of Financial Studies* **22** (2009), no. 11 4601–4641.

[9] S. Maslov, *Simple model of a limit order-driven market*, *Physica A: Statistical Mechanics and its Applications* **278** (2000), no. 3 571–578.

[10] S. Maslov and M. Mills, *Price fluctuations from the order book perspectiveempirical facts and a simple model*, *Physica A: Statistical Mechanics and its Applications* **299** (2001), no. 1 234–246.

[11] E. Smith, J. D. Farmer, L. s. Gillemot, and S. Krishnamurthy, *Statistical theory of the continuous double auction*, *Quantitative finance* **3** (2003), no. 6 481–514.

[12] H. Luckock, *A steady-state model of the continuous double auction*, *Quantitative Finance* **3** (2003), no. 5 385–404.

[13] R. Cont, S. Stoikov, and R. Talreja, *A stochastic model for order book dynamics*, *Operations research* **58** (2010), no. 3 549–563.

[14] J.-P. Bouchaud, J. Farmer, and F. Lillo, *How markets slowly digest changes in supply and demand*, *Fabrizio, How Markets Slowly Digest Changes in Supply and Demand (September 11, 2008)* (2008).

[15] A. Bovier, J. Černỳ, and O. Hryniv, *The opinion game: Stock price evolution from microscopic market modeling*, *International Journal of Theoretical and Applied Finance* **9** (2006), no. 01 91–111.

[16] D. Challet and R. Stinchcombe, *Analyzing and modeling 1+ 1d markets*, *Physica A: Statistical Mechanics and its Applications* **300** (2001), no. 1 285–299.

[17] M. G. Daniels, J. D. Farmer, G. Iori, and E. Smith, *How storing supply and demand affects price diffusion*, *Preprint* (2001).

[18] M. G. Daniels, J. D. Farmer, L. Gillemot, G. Iori, and E. Smith, *Quantitative model of price diffusion and market friction based on trading as a mechanistic random process*, *Physical review letters* **90** (2003), no. 10 108102.

[19] F. Slanina, *Mean-field approximation for a limit order driven market model*, *Physical Review E* **64** (2001), no. 5 056136.

[20] P. Malo and T. Pennanen, *Reduced form modeling of limit order markets*, *Quantitative Finance* **12** (2012), no. 7 1025–1036.