Management Science and Financial Engineering
Vol 19, No 1, May 2013, pp.37-41
ISSN 2287-2043 | EISSN 2287-2361 |
http://dx.doi.org/10.7737/MSFE.2013.19.1.037

A Stochastic Model for Order Book Dynamics: An Application to Korean Stock Index Futures

Yongjae Lee
Industrial and Systems Engineering, KAIST
Woo Chang Kim*
Industrial and Systems Engineering, KAIST

(Received: December 21, 2012 / Revised: March 20, 2013 / Accepted: March 22, 2013)

ABSTRACT

This study presents an application of stochastic model for limit order book (LOB) dynamics to Korean Stock Index Futures (KOSPI 200 Futures). Since KOSPI 200 futures market is widely known as one of the most liquid markets in the world, direct application of an existing model is hardly possible. Therefore, we modified an existing model to successfully model and predict the dynamics of extremely liquid KOSPI 200 futures market.

Keywords: Limit Order Book (LOB), KOSPI 200 Futures, Stochastic Modeling, Markov Chain, Financial Engineering

* Corresponding Author, E-mail: wkim@kaist.ac.kr

1. INTRODUCTION

In an order-driven market, the limit order book (LOB) is known to embed important information about short-term price movement. A lot of researchers have studied the nature of LOB to understand and capture the information about price dynamics embedded in LOB.

Bouchaud *et al*. (2002), Farmer *et al*. (2004), Hollifield *et al*. (2004) studied about the statistical characteristics of LOB. Parlour (1998), Foucault *et al*. (2005), Rosu (2009) presented Equilibrium models for LOB, but due to the difficulties in the estimation of some parameters, the models were inappropriate for applications. Bouchaud *et al*. (2008), Smith *et al*. (2003), Bovier *et al*. (2006), Luckock (2003), Maslov and Mills (2001), Cont *et al*. (2010) proposed stochastic models that explain the dynamics of LOB. While all other researchers focused on the steady-state of LOB, Cont *et al*. (2010) focused on conditional probabilities of future dynamics given the current LOB state. Obviously, the conditional probabilities are very easy to interpret and utilize in the investment processes.

Following the enforcement of Integration law of capital market, the demands for advanced investment techniques like High-Frequency trading are growing in Korean financial market. However, little is known about LOBs of Korean stock markets. Therefore, in this study, we present an application of the LOB model of Cont *et al*. (2010) (CST model) to KOSPI 200 futures market, which is the symbolic market in Korea, to provide a theoretical prediction of price dynamics of KOSPI 200 futures.

However, according to ‘Jan-Jun 2012 Volume Trends’ by Futures Industry Association (2012), KOSPI 200 futures is one of the most heavily traded equity index futures and options in the world. Due to this extreme liquidity, direct application of existing models is not obvious. Hence, we propose a modification of CST model to deal with the liquidity of KOSPI 200 futures, and finally present a verification of the model through the theoretical prediction of price dynamics of KOSPI 200 futures.

2. MODEL DESCRIPTION

2.1 Limit Order Book

CST considered LOB as a continuous-time process X(t).

$$X(t) \equiv \left(X_1(t), \cdots, X_n(t)\right)_{t \geq 0}$$

Here, {1, ⋯, n} is a price grid representing multiples of a price tick covering appropriate range of possible prices. Each $|X_p(t)|$ is the number of outstanding limit orders at price p. If $X_p(t) > 0$, then there are $X_p(t)$ number of ask orders at price p, and if $X_p(t) < 0$, then there are $-X_p(t)$ number of bid orders at price p.

2.2 Dynamics of LOB

In CST model, they considered three types of price dynamics: (1) Limit order, (2) Market order, and (3) Cancellation.

Limit orders are the orders that specify price and quantity. These orders go in to LOB and wait to be matched by market orders. Limit orders are the only way to increase the order quantity in LOB. Market orders are the orders that specify quantity only. Therefore, they are matched by existing limit orders in LOB which gives the best price, and therefore decreases the number of orders at matched price. Finally, cancellation is the cancellation of limit orders, which decreases the number of orders at specified price.

CST modeled these dynamics as independent Poisson processes. For limit orders and cancellations, the arrival rates differ with respect to the distance from opposite best quote. However, since market orders do not specify the price, the market order arrival rate is a constant.

To simplify the model, they made an assumption that every orders must be of 'unit size.' They have chosen the average size of limit orders as the unit size.

Under these settings, CST have proved that this continuous-time process becomes a continuous-time Markov chain.

2.3 Model Calibration

The calibration of CST model is mere a counting of different orders, which makes the model convenient for real-world application.

The arrival rate for limit orders at distance i is as follows.

$$\hat{\lambda}(i) = \frac{N_l(i)}{T}$$

$N_l(i)$ is just the total number of limit orders arrived at a distance i from the opposite best quote and T is the total sample time in minutes.

Similarly, the arrival rate for market orders is

$$\hat{\mu} = \frac{N_m}{T}\frac{S_m}{S_l},$$

where N_m is the total number of market orders. Here, S_m is the average size of market orders and S_l is the average size of limit orders. Recall that CST have assumed every orders must to be of unit size, which is chosen as the average size of limit orders S_l. Therefore, the second term makes the market orders to be of unit size.

Finally, the cancellation rate is given by

$$\hat{\theta}(i) = \frac{N_c(i)}{TQ_i}\frac{S_c}{S_l},$$

where $N_c(i)$ is the total number of limit orders arrived at a distance i from the opposite best quote and S_c is the average size of cancellations. One difference here is Q_i term which is the average number of orders at a distance i from the opposite best quote. Since CST assumed cancellation rate to be proportional to the number of existing orders at the price level, they divided it by the average number of orders at a distance i. In calculation, it is used after multiplied by the number of existing orders at i.

3. APPLICATION TO KOSPI 200 FUTURES

3.1 Modification on CST Model

In CST paper, the conditional probabilities are calculated using two-sided Laplace transform and inverse Laplace transform. Also, two other ways were suggested, one was by defining an appropriate transient discrete-time Markov chain, and the other was by Monte-Carlo simulation.

In this paper, we have chosen the transient discrete-time Markov chain method, because it was more intuitive and convenient for us to identify the problem of applying CST model directly to KOSPI 200 futures and modify the model to deal with the difficulties arose from the extreme liquidity of KOSPI 200.

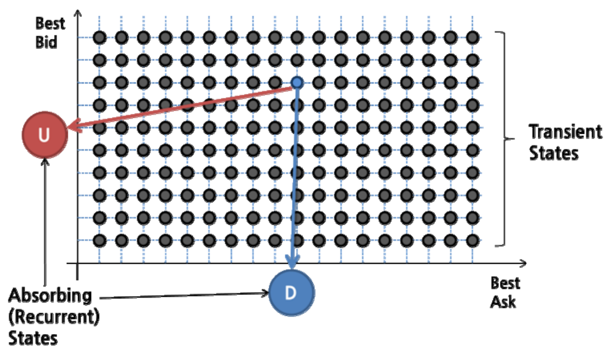

Figure 1. Transient Discrete-Time Markov Chain for LOB

First, as in the CST model, we only considered the orders at best ask and best bid prices. Transient states were defined as the states having positive number of orders at both best ask and best bid. We have two absorbing states, one is the mid-price going up, and the other is the mid-price going down. Therefore, this Markov chain travels through the transient states until it reaches either 'Mid-price Up' or 'Mid-price Down.'

If the bid-ask spread is just one tick, the process ends up in 'Mid-price Up' only when the number of orders at best ask becomes zero. However, if the bid-ask spread is more than two ticks, then if there comes a limit buy order inside the spread, the mid-price also goes up. Therefore, calculation is little different depending on the bid-ask spread.

Once the model calibration is done and the Markov chain is defined as above, the conditional probabilities are easy to compute. Ross (1996) introduced a one-step conditioning equation for the set of conditional probabilities.

$$f_{ij} = \sum\nolimits_{k\in T} P_{ik} f_{ij} + \sum\nolimits_{k\in T} P_{ik}, \quad i \in T$$

Here, f_{ij} is the probability of the process entering the recurrent state j given the current state i. In our context, this can be interpreted as the conditional probability of price going up (or down) given the current numbers of limit orders at best bid and ask. T denotes the set of all transient states, R denotes the set of states communicating with j and P_{ik} is the transition probability from state i to state k.

Since the transition probabilities P_{ik} are known as we know the arrival rates of Poisson processes, this becomes a simple system of linear equations.

However, since KOSPI 200 futures market is extremely liquid, even if we set the appropriate upper bound for the number of orders, the size of matrix becomes huge. If we have a huge matrix, even a simple matrix inversion becomes unstable. Therefore, we cannot achieve meaningful results by applying this method directly to KOSPI 200 futures.

Consequently, we enlarged the unit size of orders to reduce the size of matrix. This enables us to cover the same space with less number of states. Undoubtedly, this makes our model less sensitive, but regarding the liquidity of KOSPI 200 futures, it would be better to be less sensitive to small movements to successfully model the overall dynamics. As a consequence, there exist a trade-off between the sensitivity of model and the accuracy of calculation. If we pick too large unit size, then the model only counts for very big movements. On the other hand, if the unit size is small, we cannot achieve accurate and meaningful results. We can determine the position between two by choosing an appropriate unit size.

3.2 Data and Experiment Results

In this study, we used bid and ask sides' 5 best quotes tick data of KOSPI 200 futures from February 21, 2012 to April 30, 2012.

Figure 2 and Figure 3 present the conditional probabilities of mid-price going up calculated under different settings. For both figures, the number of orders to be covered is fixed as 600. However, for Figure 2, we used the unit size as 10, and for Figure 3, we used 20. Therefore, the number of states is 60 in Figure 2 and 30 in Figure 3.

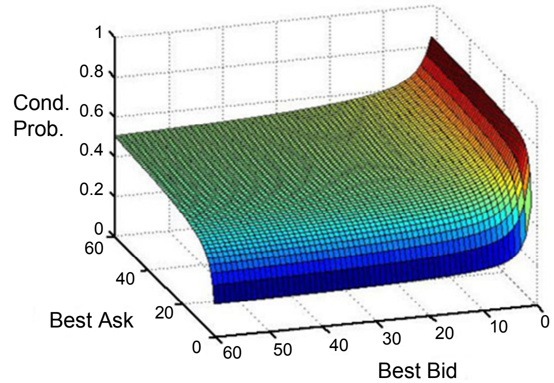

Figure 2. Conditional Probabilities of Mid-Price Going Up (60×10)

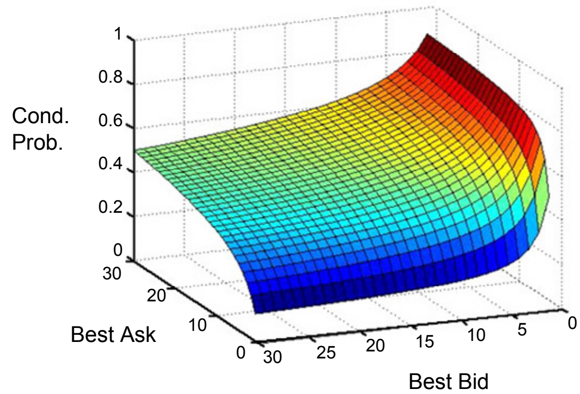

Figure 3. Conditional probabilities of mid-price going up (30×20)

By increasing the unit size, the number of states is decreased. In most of the states in Figure 2, the conditional probabilities are very close to 0.5. However, in Figure 3, we have much more information of future price movement. Therefore, we can see from the figures that the accuracy of model can be improved by selecting an appropriate unit size.

Table 1 provides the predictability comparison between our revised model (unit size: 20, # of states: 30) and direct application of CST model. For instance, if we achieved 52% probability of mid-price going up from the model, then we watch if the next move is actually going up. By aggregating such experiments, we attained the numbers in the empirical test column. As can be seen in Table 1, the prediction of our model is much more accurate than the direct application of CST model.

Table 1. Predictability Comparison of CST Model and Revised Model

Mid-price going Up				
Theoretical Prediction	CST Model		Revised Model	
	Empirical Test	Occurrence	Empirical Test	Occurrence
0.51~0.53	0.529	552657	0.508	207360
0.53~0.55	0.564	459979	0.535	342776
0.55~0.57	0.599	349464	0.558	286470
0.57~0.60	0.657	296803	0.576	384211
0.60~0.63	0.722	187625	0.617	247058
0.63~0.65	0.759	104775	0.662	175024
0.65~0.67	0.785	83547	0.712	134437
0.67~0.7	0.808	69056	0.723	172778
Mid-price going Down				
Theoretical Prediction	CST Model		Revised Model	
	Empirical Test	Occurrence	Empirical Test	Occurrence
0.51~0.53	0.547	565744	0.538	210916
0.53~0.55	0.591	469346	0.554	358392
0.55~0.57	0.640	353720	0.562	288200
0.57~0.60	0.682	289907	0.609	383071
0.60~0.63	0.743	183176	0.655	251073
0.63~0.65	0.771	108364	0.695	176188
0.65~0.67	0.797	80712	0.728	132616
0.67~0.7	0.817	70977	0.741	175123

However, as the theoretical prediction goes over 0.65, our empirical test resultstend to show higher numbers. This is probably due to the upper bound in the number of orders. If we choose very large upper bound, then we have unstable matrix calculation problem again. Therefore, it is inevitable to compromise between large upper bound and stable matrix calculation. Thus, there exist some cases that the number of orders goes over the upper bound. In such cases, the model predicts as if there exists only the upper bound number of orders, but actual probability of mid-price going up is little higher. However, if we already have high expectation of mid-price going up, then the errors in actual probability in such cases cost less than the errors near 50:50 cases.

4. CONCLUSION

This paper presented a way to successfully model the limit order book of an extremely liquid financial product like KOSPI 200 futures. Since the conditional probability analysis gives more useful information than the steady-state analysis, we have chosen CST model as our guideline. However, due to the highly liquid nature of KOSPI 200 futures, some problems occurred when we tried to apply the model directly. Therefore, we have suggested a remedy of increasing the unit size and we have shown that applying this approach gives much more accurate and meaningful results. However, it should be noted that there was a trade-off between the accuracy and the sensitivity. When applying this method, the position between two must be determined depending on the purpose of using this model.

Domestic demand for advanced investment techniques is growing fast and the stochastic modeling of LOB can play a very crucial role in high-frequency trading. Hopefully, this study would be helpful for people in financial industry looking for high-frequency trading opportunities in Korea.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A1011 157).

REFERENCES

Bouchaud, J.-P., D. Farmer, and F. Lillo, "How markets slowly digest changes in supply and demand. T. Hens, K. Schenk-Hoppe, eds," *Handbook of Financial Markets: Dynamics and Evolution*, Elsevier: Academic Press, 2008.

Bouchaud, J.-P., M. Mezard, M. Potters, "Statistical properties of stock order books: empirical results and models," *Quantitative Finance* 2 (2002), 251-256.

Bovier, A., J. Cerny, O. Hryniv, "The opinion game: Stock price evolution from microscopic market modeling," *International Journal of Theoretical and Applied Finance* 9 (2006), 91-111.

Cont, R., S. Stoikov, R. Talreja, "A stochastic model for order book dynamics," *Operations Research* 58, 3 (2010), 549-563.

Farmer, J. D., L. Gillemot, F. Lillo, S. Mike, and A. Sen, "What really causes large price changes?," *Quantitative Finance* 4 (2004), 383-397.

Foucault, T., O. Kadan, and E. Kandel, "Limit order book as a market for liquidity," *Review of Financial Studies* 18, 4 (2005), 1171-1217.

Futures Industry Association, *Jan-Jun 2012 Volume Trends*, Retrieved December 18, 2012, from Futures Industry Association website: http://www.futuresindustry.org/volume-.asp, 2012.

Hollifield, B., R. A. Miller, and P. Sandas, "Empirical analysis of limit order markets," *Review of Economic Studies* 71, 4 (2004), 1027-1063.

Luckock, H., "A steady-state model of the continuous-double auction," *Quantitative Finance* 3 (2003), 385-404.

Maslov, S. and M. Mills, "Price fluctuations from the order book perspective-empirical facts and a simple model," *PHYSICA A* 299, 234 (2001).

Parlour, C. A., "Price dynamics in limit order markets," *Review of Financial Studies* 11, 4 (1998), 789-816.

Ross, S., *Stochastic Process*, John Wiley and Sons, 1996.

Rosu, I., "A dynamic model of the limit order book," *Review of Financial Studies* 22 (2009), 4601-4641.

Smith, E., J. D. Farmer, L. Gillemot, and S. Krishnamurthy, "Statistical theory of the continuous double auction," *Quantitative Finance* 3, 6 (2003), 481-514.