Simulating Limit Order Book Models

Samuel Watts

sam@samueldwatts.com

University of Oxford

April 2016

Contents

1 Introduction

Most of the world's exchanges operate an order-driven trading system. These systems provide all market participants with a limit order book (LOB) that contains all currently outstanding limit orders. As described in the lecture notes[4] and Gould *et al.*[5], traders submit *orders* to the exchange at time t_x with price p_x and size $\omega_x > 0$ (respectively $\omega_x < 0$). This order is a commitment to sell (respectively buy) up to $|\omega_x|$ units of the traded asset at a price no less than (respectively no greater than) p_x. Prices must be a multiple of the *tick size*, π, the smallest price interval between different orders. *Market orders* are those that immediately match, causing a trade to occur. If the order cannot be immediately fulfilled it is termed a *limit order* and stored in the limit order book until either fulfilled later, or cancelled by the original trader.

The highest buy limit order price in the LOB at time t is called the *bid price*. This is the best price that a trader can sell at at time t. Analogously, the *ask price* at time t is the lowest price of the offer limit orders and the best price one can buy at. Market orders are fulfilled at these prices in order to guarantee the best possible price at time t. The mid-price is the mean of the two prices:

$$m(t) = \frac{b(t) - a(t)}{2}$$

When making trading decisions, many traders assess the LOB according to the *depth profile*. The depth is the maximum amount that can be immediately traded at a specific price and time. It is the total outstanding size of limit orders at price p and time t, denoted $n^b(p,t)$ on the bid-side of the LOB and $n^a(p,t)$ on the ask side. The bid-depth profile at time t is the set of all ordered pairs $(p, n^b(p,t))$ and ask-depth is defined similarly.

As discussed by Gould *et al.*[5], there are well-established statistical regularities common to a wide range of markets. It is proposed that LOB dynamics are integral to the price-discovery process and that some of these statistical regularities emerge as a result of LOB structure and dynamics.

In order to investigate this possibility we can use simplified models of a LOB. These toy models capture key dynamics, while remaining mathematically tractable. There are two main approaches to modelling LOBs: perfect-rationality and zero-intelligence.

Perfect-rationality models are trader-centric, attempting to recreate the trader's decision process. These models assume the trader is perfectly rational and acts optimally to maximise their utility by following a specified trading strategy based on the information available in the LOB.

Zero-intelligence models, however, are exogenous of the trading decision. Instead they assume trading decisions are random and model the order flows as stochastic processes. Order flows are dependent on the state of the LOB, as well as its recent history in some models.

In this paper we discuss two related zero-intelligence LOB models. In section 2 we describe the two models, then in section 3 we explain how they were simulated. In section 4 we explain how the model parameters were estimated from real data and input into the simulations. In section 5 we discuss the results of the simulations and compare them with the behaviour of real LOBs. Finally we conclude.

2 Two zero-intelligence order-flow LOB models

The two models discussed in this paper were developed by Smith *et al.* [8] and by Cont, Stoikov, and Talreja [3]. These papers use slightly different notation so for clarity the notation from Cont *et al.* [3] will be used with some personal adjustments.

Both models are zero-intelligence models that model all order flows as independent Poisson processes. The model in Smith *et al.* is designed to capture the long-run statistical properties of the LOB, whereas the model in Cont *et al.* is designed to capture the short-run temporal evolution. The only differences between the two models are the order arrival rates. For both models:

- All orders are for unit size σ.
- Buy market orders arrive with fixed rate μ.
- Sell market orders arrive with fixed rate μ.

For Smith *et al.*:

- Buy limit orders arrive with fixed rate λ at prices $-\infty < p < a(t)$.
- Sell limit orders arrive with fixed rate λ at prices $b(t) < p < \infty$.
- Each active limit order is cancelled with fixed rate θ.

Whereas, for Cont *et al.*:

- Buy limit orders arrive with rate $\lambda(i)$.
- Sell limit orders arrive with rate $\lambda(i)$.
- Each active limit order is cancelled with rate $\theta(i)$.

where i is the distance, in ticks, of the price from the opposite best quote, i.e.

$$i = \begin{cases} (a(t) - p)/\pi, & \text{Bids: } p < a(t) \\ (p - b(t))/\pi, & \text{Offers: } p > b(t) \end{cases}$$

Cont *et al.*[3] modelled the rate function $\lambda(i)$ as a power law:

$$\lambda(i) = \frac{k}{i^{\alpha}}$$

whereas $\theta(i)$ is modelled as a step function with fixed rates for $1 \le i \le 5$ and $\theta(i) = \theta(5)$ for $i > 5$.

3 Simulating the Models

In order to simulate these two models, a program consisting of two parts was written. One part of the program maintains the LOB and processes orders and could be used to simulate any zero-intelligence LOB model. The orders are generated by the second part, according to the selected LOB model, using standard methods for simulating Poisson processes.

For the LOB part, the market in Cont *et al.* [3] was used as the more straightforward option to implement. This is one where limit orders can be placed on a price grid $\{m, ..., n\}$ representing multiples of a price tick, π. Log prices were not used as in Smith[8] to avoid unnecessary complication. The lower boundary m and upper boundary n are chosen small enough and large enough, respectively, so that it is highly unlikely that market orders for the stock in question are placed outside the grid within the time frame of the simulation.

In the unlikely event that there are no ask orders in the book, an ask price of $n + 1$ is forced and when there are no bid orders, a bid price of $m - 1$.

Smith's model [8] assumes limit orders appear across an infinite price interval. Our simulation just considers a sample of this interval since we are only investigating behaviour near the mid-price and limit orders placed outside this interval are far enough away that they do not affect the ask or bid prices.[1]

The orders are generated order-by-order along with the time between them, rather than in continuous time or fixed time steps. Thus we consider event time, here denoted as t, i.e. $t = 5$ means after 5 events and $t + 1$ is one event after the event at t.

To generate the orders, at each step the total order rate, $r(t)$, is calculated by summing all the current rates.

$$r(t) = 2\mu + \begin{cases} \lambda(a(t) - m + n - b(t)) + \theta N_l(t) & \text{for Smith's model} \\ \sum_{i=1}^{a(t)-m} \lambda(i) + \sum_{i=1}^{n-b(t)} \lambda(i) + \sum_{i=1}^{\max(i)} \theta(i) N_l(i,t) & \text{for Cont's model} \end{cases}$$

where $N_l(t)$ is the number of outstanding limit orders at t and $N_l(i, t)$ is the number at a distance of i from the opposite best quote.

Since the individual rates depend on the LOB state they therefore change as the LOB develops, but are constant between orders. Thus they are recalculated at each step. From this, the time til the next order is generated:

$$\tau(t) = -\frac{\log(\rho)}{r(t)}$$

where ρ is a random number uniformly chosen between 0 and 1. This algorithm works due to the memoryless nature of Poisson processes.

The next order is selected at random. It is chosen with probability according to its rate proportional to the total rate. For example, a buy market order would be selected

[1] If required, the area outside this interval could be approximated by steady state behaviour where arrival and cancellation rates at each price balance out. As discussed in Section 5 this behaviour appeared even within the limited interval simulated.

with probability $\mu/r(t)$. This process was repeated until the total time passed the desired simulation period, e.g. one trading day.

4 LOB data and Model Parameters

To run meaningful simulations of the two models, values for the parameters are required. These were approximated from real LOB data. LOB data for one trading day of Microsoft's stock on NASDAQ were used, taken from LOBSTER Academic Data's website[2]. These data consist of a time-stamped sequence of market events as well as quotes (prices and quantities of outstanding limit order) for the ten best price levels on each side of the order book. The prices are given in US dollar price times 10000, with a tick size of $\pi = 100$, i.e. one cent.

These data cover too small a time span to be truly representative. We do not know whether the day in question was a typical trading day or not. However they are sufficient to give a meaningful simulation.

To estimate the parameters a procedure similar to the one described by Cont *et al.*[3] was utilised. This is one that uses mean averages to estimate parameters.

Both models assume orders arrive in chunks of σ orders. An argument could be made to choose $\sigma = 100$, as this was the mode order size in the LOBSTER data sample. This is most likely due to traders preferring to work in round order sizes for psychological reasons. However, following Cont *et al.*[3], $\sigma = S_l$, the average size of a limit order, was chosen. Using S_l as the order size makes sense since limit orders add to the LOB while market and cancellation orders are removals of existing limit orders so it is intuitive to operate with a unit representative of limit order size. Choosing $\sigma = 100$ would scale the rates below by $S_l/100$.

According to Smith *et al.*[8], the order size is an important determinant of both liquidity and volatility. More investigation is required to see if this choice similarly effects the simulations of the Cont model.

Since $\sigma = S_l$, market and cancellation order rates are scaled to σ according to their average order sizes, S_m and S_c respectively.

The arrival rate of buy and sell market orders for both models is estimated by:

$$\hat{\mu} = \frac{1}{2} \frac{N_m}{T_*} \frac{S_m}{S_l}$$

where T_* is the total trading time in the sample (in seconds) and N_m is the number of market orders (both buy and sell). The rate is halved since it is assumed that buy and sell orders arrive at equal rates, an assumption borne out by the data. Market orders matched by hidden limit orders are ignored since these do not affect the best prices.

For Smith's model[8] we estimate the limit order arrival rate per unit price interval by:

$$\hat{\lambda} = \frac{1}{PI_l} \frac{N_l}{T_*}$$

[2] `https://lobsterdata.com/info/DataSamples.php`

where N_l is the number of limit orders in the sample and PI_l is the average size of the price interval in which limit orders could arrive in. In this model an infinite number of limit orders arrive in each time interval, since they appear across an infinite price interval. The data sample used does not cover an infinite interval of course but only contains orders with prices within the ten best prices on each side. Thus it only contains limit orders that arrive within a limited price interval. Therefore, if M is the number of quote rows, i.e. the number of events, and $a_{10}(T), b_{10}(t)$ are the tenth best ask and bid prices respectively, we can estimate the price interval size by:

$$PI_l = \frac{1}{M}\sum_{t=0}^{M}(a(t) - b_{10}(t) + a_{10}(t) - b(t))$$

This produces a reasonable limit order rate near the mid-price but is not representative for orders further away. Comparison between the total limit order rates for LOBSTER data containing orders within ten and fifty price levels showed that over 90% of limit orders arrive within the best ten prices so therefore this rate estimation is meaningful for the purposes of this paper.

To estimate the cancellation rate for the model in Smith *et al.*[8] the steady-state shape of the order book, Q, needs to be estimated. This is the average number of outstanding limit orders. This value is required since the cancellation rate at each event time is directly proportional to the number of outstanding limit orders. If $S_l(t)$ is the total size of outstanding limit orders at t:

$$Q = \frac{1}{S_l}\frac{1}{M}\sum_{t=0}^{M} S_l(t)$$

This gives the estimate of the cancellation rate as:

$$\hat{\theta} = \frac{1}{Q}\frac{N_c}{T_*}\frac{S_c}{S_l}$$

where N_c is the number of cancellation and deletion orders in the sample.

For the model in Cont *et al.*[3] the limit order arrival rate function can be estimated by:

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

where $N_l(i)$ is the total number of limit orders that arrived at a distance of i price ticks from the opposite best quote. Values for $N_l(i)$ were only taken for $1 \leq i \leq 10$. This is because all limit orders within this distance must be included in our data sample, whereas, if there are any gaps in the depth profile, it cannot be guaranteed that all orders for a given $i > 10$ would be included. Then the function was extrapolated by fitting a power law function of the form:

$$\lambda(i) = \frac{k}{i^\alpha}$$

	i									
	1	2	3	4	5	6	7	8	9	10
$\hat{\mu}$	0.7953									
$\hat{\lambda}$	0.5030									
$\hat{\theta}$	0.0235									
$\hat{\lambda}(i)$	5.6014	4.4890	1.2115	0.6304	0.5723	0.4058	0.2613	0.2516	0.3744	0.2432
k	6.8587									
α	1.515									
$\hat{\theta}(i)$	0.2461	0.1313	0.0313	0.0188	0.0130	0.0124	0.0099	0.0135	0.0178	0.0164

Table 1: Simulation parameters

as in [3] which follows the suggestion of Zovko and Farmer[11] and Bouchaud et al[1]. The power law parameters k and α were obtained by using least-squares fit.

For the cancellation rate for Cont's model[3] we again need to estimate the steady-state shape of the order book. However we need to estimate it at each tick distance from the opposite best quote, denoted Q_i. To do this the average number of orders bid at a distance of i ticks from the ask price, $S_l^B(i,t)$, was calculated:

$$Q_i^B = \frac{1}{S_l}\frac{1}{M}\sum_{t=0}^{M} S_l^B(i,t)$$

Q_i^A was calculated analogously and the average of the two taken as Q_i. This gives us an estimator for the cancellation rate function:

$$\hat{\theta}(i) = \begin{cases} \frac{1}{Q_i}\frac{N_c(i)}{T_*}\frac{S_c}{S_l}, & 1 \leq i \leq 5 \\ \hat{\theta}(5), & i > 5 \end{cases}$$

where $N_c(i)$ is the number of cancellation (partial deletions) and deletion orders in the sample at a distance of $1 \leq i \leq 5$ ticks from the opposite best quote. A preliminary analysis appears to show this to be a representative rate function as the estimated cancellation rates remain reasonably steady at a distance of more than 5 ticks. This can be seen in the calculated values for $\hat{\theta}(i)$, $i > 5$ in Table 1.

Estimated parameter values for Microsoft stock are given in Table 1.

5 Simulation Results

Simulations were run using the program described in Section 3 with the input parameters from Section 4. The outcome of these simulations was be investigated and compared with real limit order book behaviour.

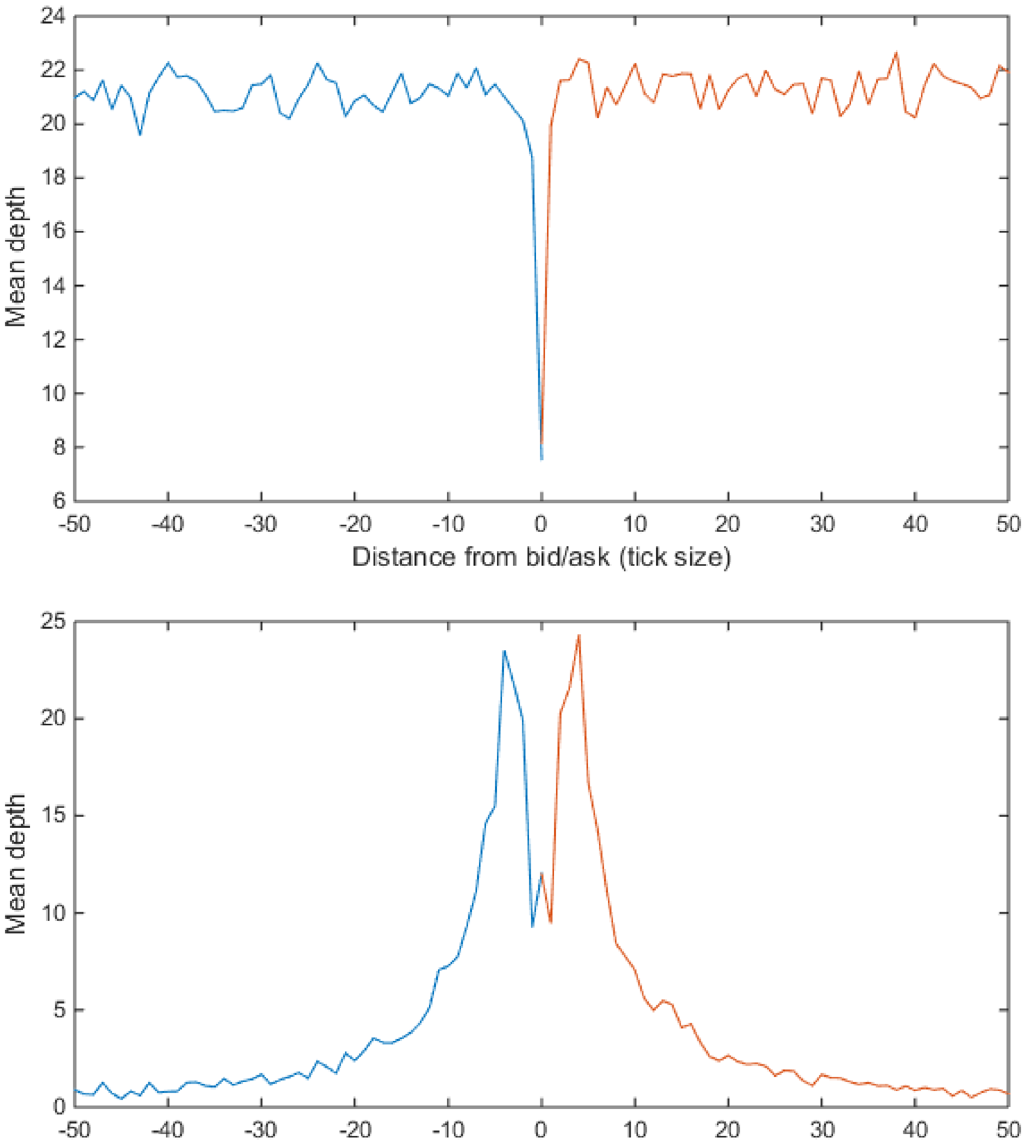

Figure 1: Mean relative depth profile generated by simulations of the Smith (top) and Cont (bottom) models. The buy-side is in blue on the left and offer-side in red on the right.

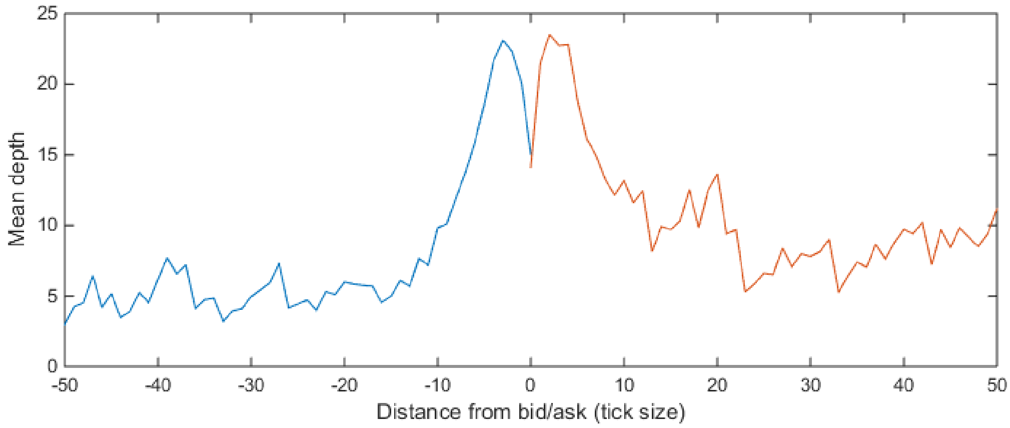

Figure 2: Mean relative depth profile of MSFT stock LOBSTER data sample. The buy-side is in blue on the left and offer-side in red on the right.

First we look at the depth profiles. In Figure 1 you can see graphs displaying the mean relative depth profiles produced by simulations of the two models. This is the mean number of limit orders outstanding at a each number of ticks away from the bid price for bids and ask price for offers. This is an indicator for the amount of liquidity available to an impatient trader and gives an indication of how much the price will change if the trader 'walks up the book'.

As you can see in both cases the depth is quite symmetrical, which is expected since both models have the same rates for both buy and sell orders. This matches western exchanges which also exhibit symmetrical relative depth profiles[5]. Compare this with the sample data (see Figure 2) which does not exhibit a symmetrical mean relative depth profile. This is most likely due to the short time scale of the sample (only one trading day) and the clear downward direction the price takes during the sample (see Figure 3).

Both models also have lower depths at the bid and ask prices due to market orders eating into the orders there.

For the Smith model (top of Figure 1) the mean depth is relatively stable and high for prices more than a few ticks away from the best prices. This is due to these prices being sufficiently far away that they are only affected by limit and cancellation orders. From the values estimated for $\hat{\lambda}$ and $\hat{\theta}$ (see Table 1) these two rates become equal at 21.4 outstanding orders at a price tick, which is where we see the depth profile levelling out at.

For the Cont model (bottom of Figure 1) we see an 'm' shape. This is due to the interplay between the high limit order and cancellation rates adjacent to the opposite best price as well as the market order rate. Away from the mid-price the limit order arrival rates tail off following the power-law distribution, as cancellation rate becomes constant.

A wide range of markets have been reported to exhibit an 'm' or 'hump' shape[5]. This is where the absolute value of the mean depth available increases of the first few relative prices and subsequently decreases. This shape is clearly visible in the sample

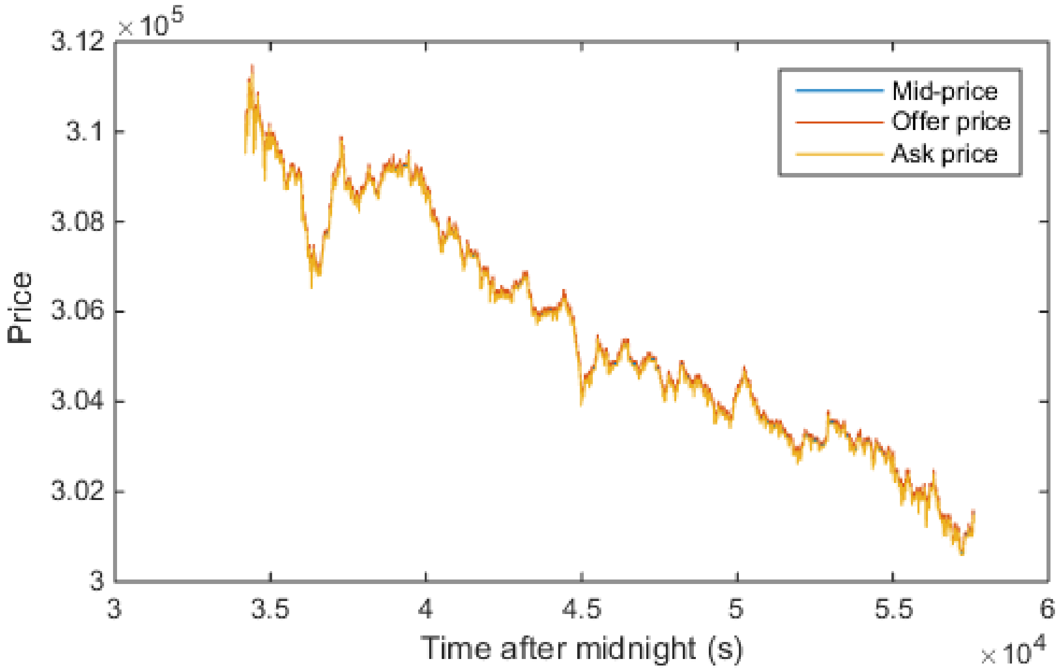

Figure 3: One trading day of MSFT stock price data used to estimate parameters.

data (Figure 2) as well as the simulation of the Cont model but is absent for the Smith model simulation.

As discussed in Smith *et al.*[8] this tail depth behaviour is quite unrealistic, resulting from the uniform order placement process. However, when they considered only orders that eventually get executed humps also appeared in the depth profile.

Cont *et al.*[3] mentioned that this feature does not result from fine-tuning of the model parameters or additional elements such as correlation between order flow and past price moves. In our simulation of Cont's model the hump peaks at four ticks whereas in the sample data it peaks at three. In their simulation[3], however, they had a hump at two ticks from the opposite best quote, which, assuming a spread of only one tick, means an equivalent hump at only one tick in our definition. Thus it appears that the position of the hump is dependent on the parameters passed into the model. This matches real LOB behaviour since the location of the hump varies across markets[5].

This 'hump' feature arises in the model due to the relation between market order rates and limit order rates. Thus it seems likely this will appear in most markets where traders can choose between submitting a market order (be impatient) or a limit order (be patient). Rosu[7] conjectures this is the case when large market orders are sufficiently likely.

Now we turn to the mid-price. Figure 3 displays the actual mid-price process from the sample LOBSTER data. Figure 4 displays the mid-price processes produced by the model simulations.

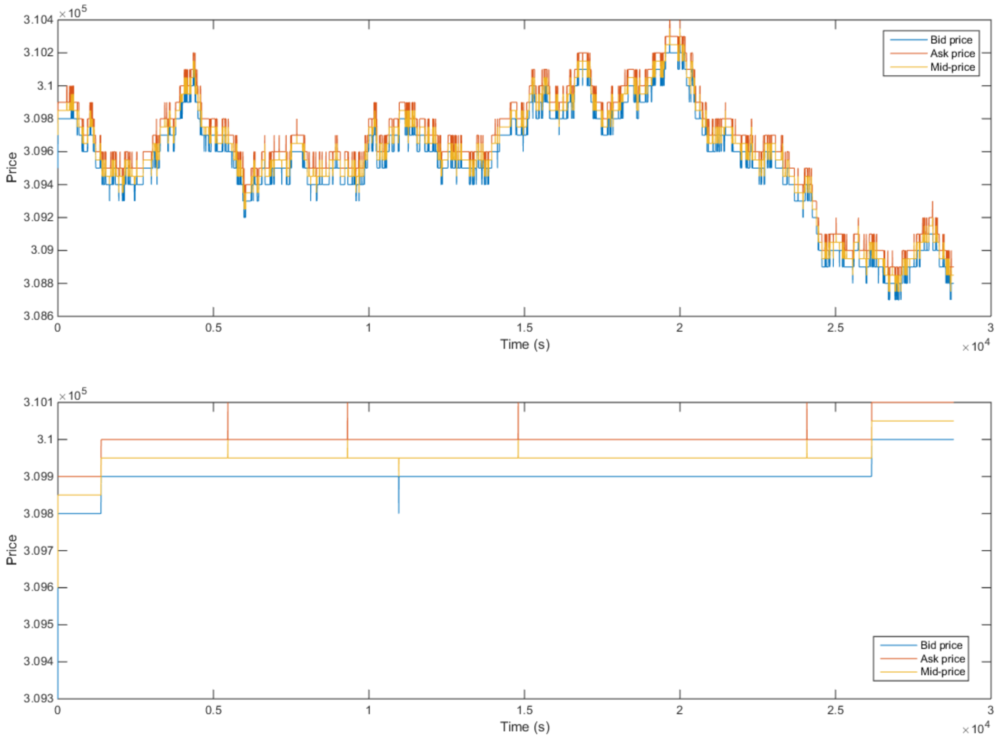

Figure 4: Bid, mid-, and ask prices over a trading day from simulations of the Smith (top) and Cont (bottom) models.

For the Cont model simulation it was discovered that the rate parameters, estimated from the data, produced low volativity and a relatively unchanging mid-price. This was due to the depths at the bid price and offer price quickly reaching an equilibrium sufficiently high that only rarely were all the orders matched or cancelled, thus changing the mid-price. This meant there were few datapoints for analysis of mid-price behaviour.

First the mid-price returns were investigated. The logarithmic mid-price return over timescale of Δt is defined as:

$$R_{\Delta t}(t) = \log\left(\frac{m(t)}{m(t-\Delta t)}\right)$$

Empirical probability density functions were estimated for both models over the timescale of seconds, minutes and hours (see Figures 5 and 6).

For returns over seconds and minutes there were insufficient data to confidently make any comment on the distribution. Over the timescale of hours however we start to see heavy tails appear. Mid-price returns have been reported to display tails heavier than a normal distribution[5]. Preliminary analysis implies this to be the case for the Smith model simulation. This could be related to the inverse cubic law of returns as reported by Gu *et al.*[6] amongst others. However, without a larger data sample the decay of the tails of the distribution cannot be accurately assessed.

Next we turn to autocorrelation of the mid-price returns. The autocorrelation function A of a time series X is given by:

$$A_X(l) := \frac{1}{k-l}\sum_{i=1}^{k-l}(X(t_i) - \langle X\rangle)(X(t_{i+l}) - \langle X\rangle)$$

where $\langle X\rangle$ is the mean of the series.

Again we did not have sufficient mid-price data from the Cont model simulation to produce anything meaningful, unfortunately. Nevertheless, for the Smith model simulation we found that the autocorrelation function did decay like a power law, $O(l^{-\alpha})$. A curve fitted to the simulation data by least-squares had exponent of $\alpha = 0.0003212$. Since this is within $(0, 1)$ the logarithmic mid-price of the Smith model exhibits long memory[5].

This result is in accordance with the findings in Smith *et al.*[8] who also found non-zero autocorrelations of the mid-point price. They found this autocorrelation persisted for timescales up to approximately 50 market order arrival times.

For various markets, studies have reported that they exhibit no long memory except on the shortest timescales[2]. This implies that there is a 'perfect rationality' effect in real markets that eliminates the arbitrage opportunity that arises with autocorrelation in the Smith model.

6 Conclusion

Simulations of two key zero-intelligence LOB models displayed that these types of models can capture significant distinguishing features of LOB behaviour while remaining ana-

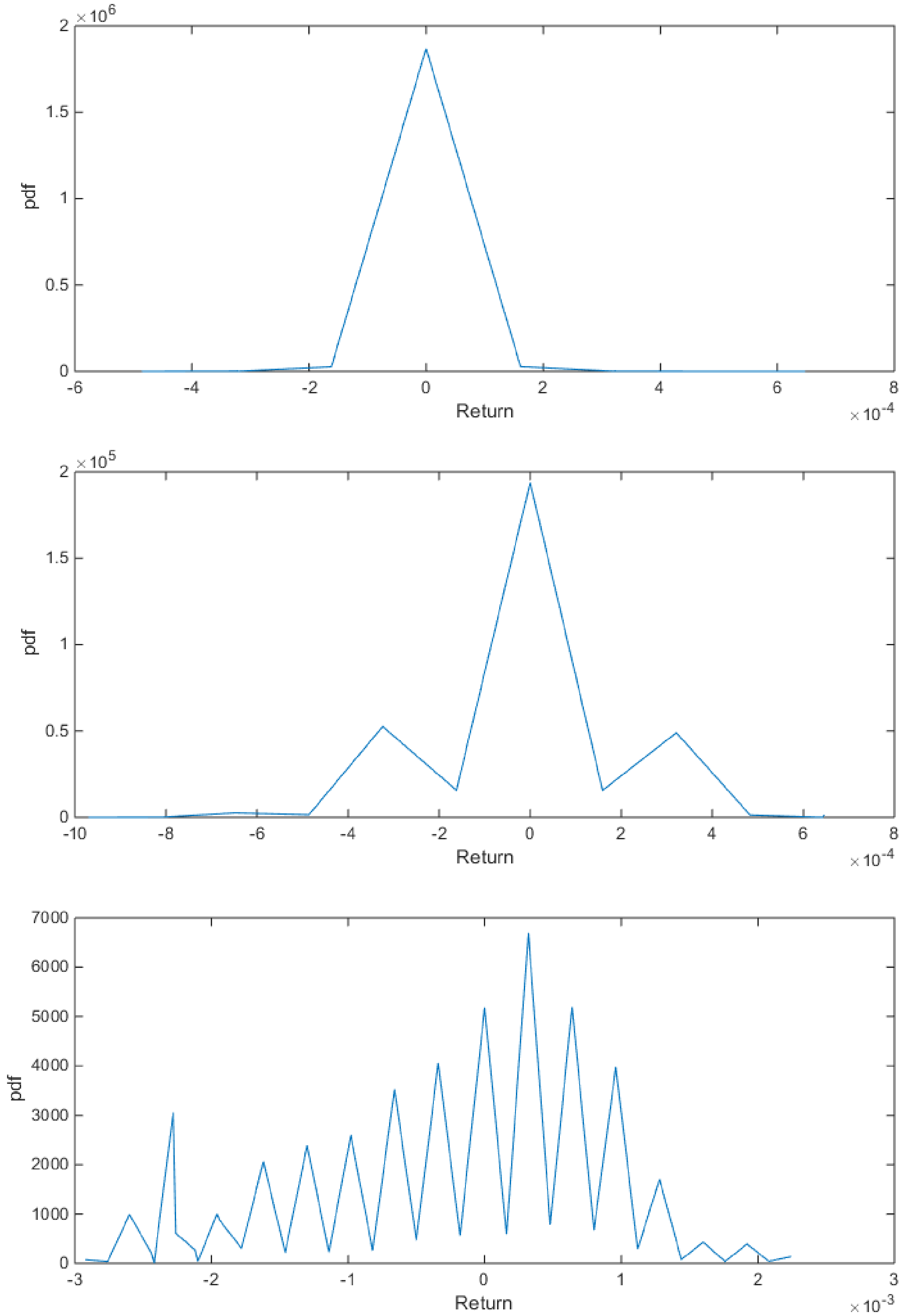

Figure 5: Distribution of logarithmic mid-price returns from simulations of the Smith model at timescales of seconds (top), minutes (middle) and hours (bottom).

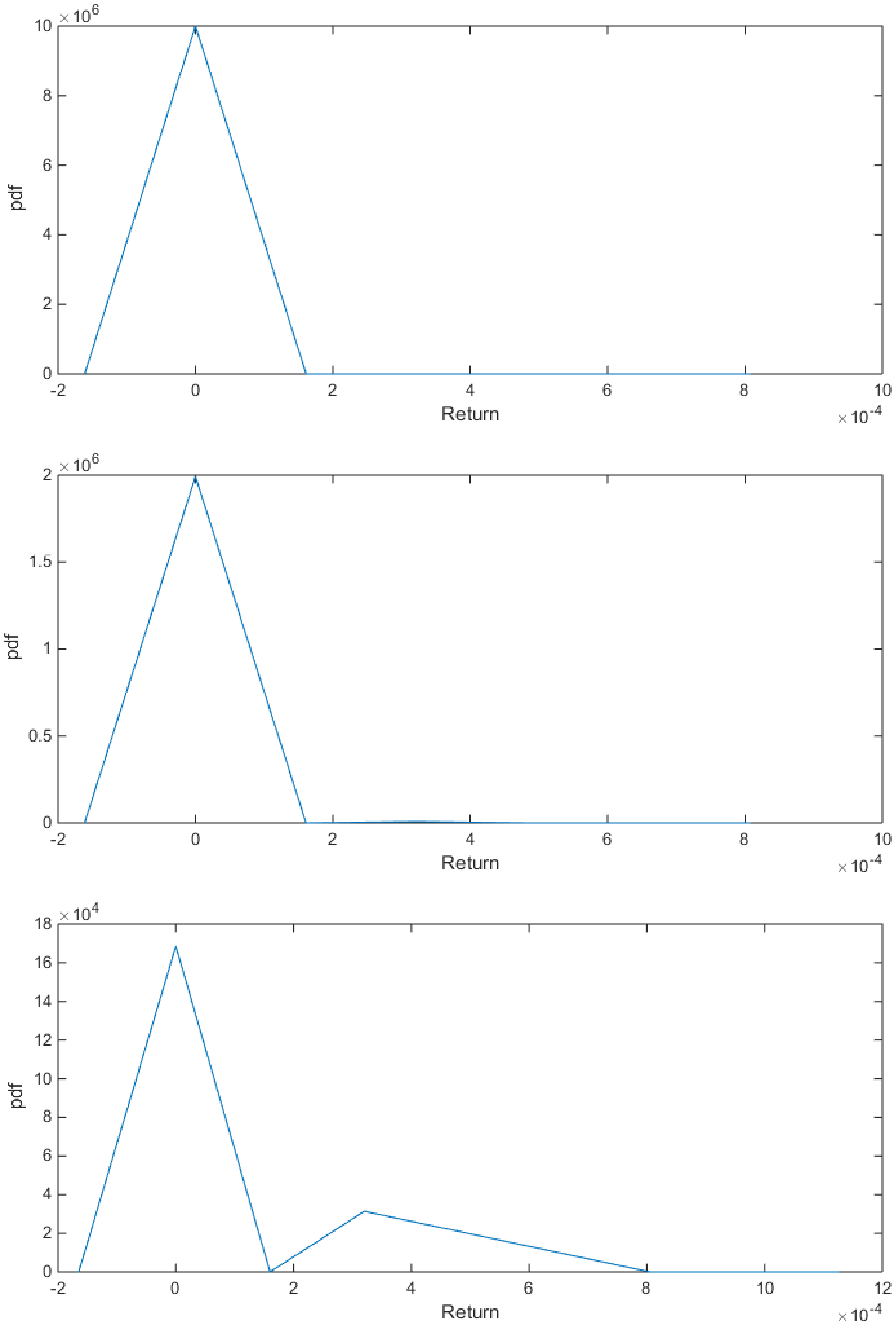

Figure 6: Distribution of logarithmic mid-price returns from simulations of the Cont model at timescales of seconds (top), minutes (middle) and hours (bottom).

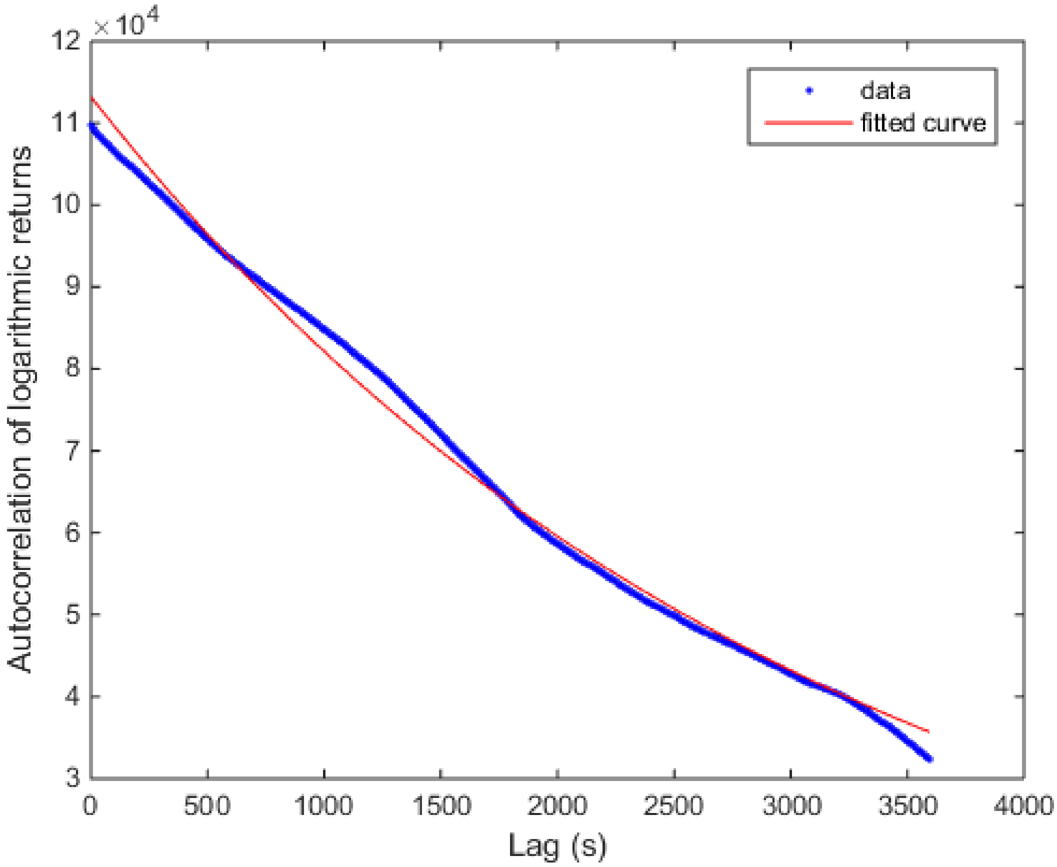

Figure 7: Autocorrelation function of logarithmic mid-price returns generated from simulation of the Smith model. The data is displayed along with a fitted power-law function.

lytically tractable and easy to implement. In particular they exhibit appropriate mean relative depth profiles.

Research into mid-price behaviour such as by Gu *et al.*[6] used data samples containing millions of trades. My simulation results only contain tens of thousands. This meant there were not enough datapoints to provide meaningful statistical analysis on the mid-price distribution for these models. Nevertheless, they appear to exhibit some heavy-tailed distributions but further investigation is required to say this with certainty.

The Smith model simulation displays autocorrelation on short timescales in accordance with real markets, however these correlations persist for longer in the model than in real LOBs.

One significant stylized fact not displayed by either model is volatility clustering which is displayed in real mid-price returns[5]. This is where time series of absolute or square mid-price returns exhibit long memory. It indicates that large price changes tend to follow other large price changes.

One proposed explanation for volatility clustering is the strategic splitting of orders by traders[5]. This could be easily incorporated into either model discussed in this paper. For example we could make market orders arrive in closely spaced chunks or have an additional 'large market order' rate that is split into order chunks.

One way volatility clustering could also be incorporated into the model is by randomly "switching" regime between ones of high and low activity, with long range dependence as suggested by Cont[2]. This would significantly increase the difficulty of parameter estimation however.

Another potential avenue is to have arrival rates as a function of recent order arrival rates and the number of recent order arrivals; say for example modelling the market rate as a decaying function of the time since the last market order. Zhao[10] and Toke[9] both extended the Cont model following this approach, producing simulated order flows and spread dynamics that more closely matched empirical data.

In summary, these two zero-intelligence models provide explanation for some fundamental LOB behaviour and serve as a solid foundation to extend to more realistic models.

References

[1] Jean-Philippe Bouchaud, Marc Mézard, and Marc Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.

[2] Rama Cont. Long range dependence in financial markets. In Jacques Lévy-Véhel and Evelyne Lutton, editors, *Fractals in Engineering*, pages 159–179. Springer London, 2005.

[3] Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.

[4] Martin Gould. LOBs and market microstructure. *MSc Mathematical Finance Lectures, Module 7*, 2016.

[5] Martin D. Gould, Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

[6] Gao-Feng Gu, Wei Chen, and Wei-Xing Zhou. Empirical distributions of chinese stock returns at different microscopic timescales. *Physica A: Statistical Mechanics and its Applications*, 387(2-3):495–502, January 2008.

[7] Ioanid Rosu. Liquidity and information in order driven markets. *Social Science Research Network Working Paper Series*, October 2008.

[8] Eric Smith, J. Doyne Farmer, Laszlo Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, October 2002.

[9] IoaneMuni Toke. market making in an order book model and its impact on the spread. In Frédéric Abergel, BikasK Chakrabarti, Anirban Chakraborti, and Manipushpak Mitra, editors, *Econophysics of Order-driven Markets*, New Economic Windows, pages 49–64. Springer Milan, 2011.

[10] Linqiao Zhao. *A MODEL OF LIMIT-ORDER BOOK DYNAMICSAND A CONSISTENT ESTIMATION PROCEDURE*. PhD thesis, Diss. Carnegie Mellon University Pittsburgh, 2010.

[11] Ilija Zovko and J. Doyne Farmer. The power of patience: a behavioural regularity in limit-order placement. *Quantitative finance*, 2(5):387–392, 2002.