Imperial College London

Market Microstructure
Problem Set 1

Tuesday 28 January 2020
Johannes Muhle-Karbe, j.muhle-karbe@imperial.ac.uk
Claudio Bellani, c.bellani17@imperial.ac.uk

Exercise 1 ("Roll's model with geometric mid-price and uncorrelated unbalanced order flows")**.** We would like to modify Roll's model simultaneously in two directions. The first modification concerns the dynamic of the mid-price m_t: rather than assuming this dynamic to be an arithmetic random walk, assume that it follows a geometric random walk, namely that there exist strictly positive square-integrable i.i.d. random variables ϵ_t such that $\mathbb{E}[\epsilon_t] = 1$ and $m_{t+1} = \epsilon_{t+1} m_t$ for all t. The second modification is to relax the assumption that $\mathbb{P}(d_t = 1) = 1 - \mathbb{P}(d_t = -1) = 1/2$ and account for the possibility of unbalanced order flow, i.e. that $\mathbb{P}(d_t = 1) = 1 - \mathbb{P}(d_t = -1) = \theta$ for some $0 < \theta < 1$ and all t. We keep however in place the assumption of uncorrelated flows, namely $\mathrm{cov}(d_t, d_s) = 0$ for all $s \neq t$, and the independence of d_t from any of the ϵ_s. Under these two modifications, derive an expression for the spread S in terms of (the parameters of the model and) the covariance $\mathrm{Cov}(p_{t+1} - p_t, p_t - p_{t-1})$.

Exercise 2 ("Roll's model with correlated order flows and non-transaction prices")**.** We would like to modify Roll's model for measuring spreads in order to take two phenomena into account. The first phenomenon is the possibility that at some time the reported price p_t is not a transaction price at the bid or at the ask price (i.e. it is not $m_t \pm S/2$) but it is actually the mid-price m_t itself. This may happen if the trading venue, having observed a period (e.g. a time window $[t-1, t]$ of one minute) where no transaction happens decides to report $p_t = m_t$. Alternatively, this may happen if a transaction is executed inside the spread because of the presence of hidden orders or because of favourable treatments that some market makers might grant to some of their clients. The second phenomenon is the possibility that the direction of the $(t+1)$-th trade is actually dependent on the direction of the t-th trade.

Using the notation introduced in the lectures, we model the first phenomenon by saying that $\mathbb{P}(d_t = 0) = q_0 \arctan(S)$ for some $0 \leq q_0 \leq 2/\pi$ and all t. This captures the fact that the higher the spread the higher the probability of non-occurrence of trades or the possibility of a hidden order inside it. We model the second phenomenon by saying that

1. the event $d_{t+1} = 0$ is independent of d_t for every t;
2. there exist coefficients $0 \leq q_{-1,1},\ q_{0,1},\ q_{1,1} \leq 1$ independent of S and such that for every t the probability that $d_{t+1} = 1$ conditioned on $\{d_t = k,\ d_{t+1} \neq 0\}$ is $q_{k,1}$, where $k = -1, 0, 1$.

In this Exercise 2 we keep in place the linear structure $m_{t+1} = m_t + \epsilon_{t+1}$ for a sequence of square-integrable centred i.i.d. random variables ϵ_t, independent of any other variable in the model.

Assume stationarity of the chain (d_t) and derive an expression for the spread S in terms of (the parameters of the model and) the covariance $\mathrm{Cov}(p_{t+1} - p_t, p_t - p_{t-1})$.

[Hint: You might observe
$$2\mathbb{P}(d_{t+1} = 1 | d_{t+1} \neq 0, d_t) = (d_t + 1) q_{1,1} + (1 - d_t) q_{-1,1} + (1 - d_t^2)\big(2q_{0,1} - q_{1,1} - q_{-1,1}\big).]$$

Exercise 3. From the Blackboard folder `Assignment_1`, you can download the datasets

```
MSFT_2012-06-21_34200000-57600000_message_10.csv
MSFT_2012-06-21_34200000-57600000_orderbook_10.csv
```

The first file reports the trading activity for the symbol MSFT on Thursday 21 June 2012. Every row in the table represents an event, i.e. an order sent to the exchange. In the first column you can find the time stamp of the event, in the second column you can find the event type (identified by an integer from 1 to 7), in the fourth, fifth and sixth columns you can find the size, price and direction of the order. The event identifiers are as follows

1: Submission of a new limit order
2: Cancellation (partial deletion of a limit order)
3: Deletion (total deletion of a limit order)
4: Execution of a visible limit order
5: Execution of a hidden limit order
6: Indicates a cross trade, e.g. auction trade
7: Trading halt indicator.

The second file provides an online reconstruction of the LOB that results form the orders reported in the first dataset. A more detailed description of these two datasets is available at https://lobsterdata.com/info/DataStructure.php.

This exercise requires you to perform some measurements of spread and liquidity based on the provided datasets, by implementing Roll's model and Amihud's liquidity measure in python. More specifically your task is threefold:

1. Focus first on the events of type 4 in the MSFT message file that happened between 10.30am and 2.30pm. From this series, estimate the average spread S using the classical Roll's model and its extension of Exercise 1. Evaluate your estimation by comparing your result to the average spread readily readable from the MSFT orderbook.
2. Extract now the series of transaction prices (and their direction) that correspond to events of type either 4 or 5 that happened between 10.30am and 2.30pm. From this series, estimate the average spread S by using the extension of Roll's model introduced in Exercise 2. Again evaluate your estimation by direct comparison with the true value derived from the MSFT orderbook.
3. Using transaction prices and volumes that correspond to events of type 4, perform Amihud's ILLIQ measure for the time windows 9am-10am, 11.30am-1.30pm, and 3pm-4pm. Comment on the intraday variation.

Solutions

Solution to Exercise 1

Notice that from our assumptions it follows in particular that $\mathbb{E}[m_t]$ is constant in t, and that

$$\mathbb{E}[d_t d_s] = \mathbb{1}(s = t) + (2\theta - 1)^2 \mathbb{1}(s \neq t).$$

Therefore,

$$\begin{aligned}
\mathrm{Cov}(p_{t+1} - p_t, p_t - p_{t-1}) =& \mathbb{E}[(p_{t+1} - p_t)(p_t - p_{t-1})] \\
=& \mathbb{E}[m_t m_{t-1}(\epsilon_{t+1} - 1)(\epsilon_t - 1)] \\
&+ \frac{S^2}{4}\mathbb{E}[(d_{t+1} - d_t)(d_t - d_{t-1})] \\
=& \frac{S^2}{4}\left((2\theta - 1)^2 - 1\right).
\end{aligned}$$

Hence we observe in particular that whether the mid-price follows an arithmetic or a geometric random walk has no influence on Roll's estimation.

Solution to Exercise 2

Let $\pi = \{\pi_{-1}, \pi_0, \pi_1\}$ be the stationary distribution of d_t. We know already that $\pi_0 = q_0 \arctan(S)$. By taking expectation on both sides of the equation in the hint, we get

$$2\frac{\pi_1}{1 - \pi_0} = (2\pi_1 + \pi_0)q_{1,1} + (2\pi_{-1} + \pi_0)q_{-1,1} + \pi_0(2q_{0,1} - q_{1,1} - q_{-1,1}).$$

From this equation and $1 = \pi_{-1} + \pi_0 + \pi_1$ we can derive π_1 and π_{-1}.

By stationarity of d_t and $\mathbb{E}[\epsilon_t] = 0$ we have that $\mathbb{E}[p_{t+1} - p_t] = 0$. Hence,

$$\mathrm{Cov}(p_{t+1} - p_t, p_t - p_{t-1}) = \frac{S^2}{4}\mathbb{E}[(d_{t+1} - d_t)(d_t - d_{t-1})].$$

We readily have that $\mathbb{E}[d_t^2] = \pi_1 + \pi_{-1}$. Hence, the exercise is concluded once we establish $\mathbb{E}[d_t d_{t+1}]$ and $\mathbb{E}[d_{t-1} d_{t+1}]$.

Notice that $d_t(1+d_t) = 2\mathbb{1}(d_t = 1)$, that $d_t(1-d_t) = -2\mathbb{1}(d_t = -1)$, and that $d_t(1-d_t^2) \equiv 0$. Hence,

$$\mathbb{E}\left[d_t \mathbb{P}(d_{t+1} = 1 | d_{t+1} \neq 0, d_t)\right] = \pi_1 q_{1,1} - \pi_{-1} q_{-1,1}.$$

With this formula we can compute

$$\begin{aligned} \mathbb{E}[d_t(1+d_{t+1})] =& \mathbb{E}[d_t \mathbb{E}[1+d_{t+1}|d_t]] \\ =& \mathbb{E}[d_t \Big(\pi_0 + \mathbb{E}[1+d_{t+1}|d_{t+1} \neq 0, d_t](1-\pi_0)\Big)] \\ =& \mathbb{E}[d_t \Big(\pi_0 + 2(1-\pi_0)\mathbb{P}(d_{t+1} = 1|d_{t+1} \neq 0, d_t)\Big)] \\ =& (\pi_1 - \pi_{-1})\pi_0 + 2(\pi_1 q_{1,1} - \pi_{-1} q_{-1,1})(1-\pi_0), \end{aligned} \tag{1}$$

and similarly

$$\mathbb{E}[d_t(1-d_{t+1})] = (\pi_1 - \pi_{-1})(2-\pi_0) - 2(\pi_1 q_{1,1} - \pi_{-1} q_{-1,1})(1-\pi_0). \tag{2}$$

Finally, $1 - d_{t+1}^2 = \mathbb{1}(d_{t+1} = 0)$ and so

$$\mathbb{E}[d_t(1-d_{t+1}^2)] = (\pi_1 - \pi_{-1})\pi_0. \tag{3}$$

With the preliminary formulas (1), (2) and (3) above, we will now compute $\mathbb{E}[d_t d_{t+1}]$ and $\mathbb{E}[d_{t-1} d_{t+1}]$. As for the former, it is already contained in (1), which we rewrite as

$$\mathbb{E}[d_t d_{t+1}] = (1-\pi_0)\Big((2q_{1,1} - 1)\pi_1 - (2q_{-1,1} - 1)\pi_{-1}\Big).$$

As for the latter,

$$\begin{aligned} \mathbb{E}[d_{t-1} d_{t+1}] =& \mathbb{E}[d_{t-1} \mathbb{E}[d_{t+1}|d_t]] \\ =& (1-\pi_0)\mathbb{E}[d_{t-1}\Big(2\mathbb{P}(d_{t+1} = 1|d_{t+1} \neq 0, d_t) - 1\Big)] \\ =& (1-\pi_0)(\pi_1 - \pi_{-1})\Big(2(1-\pi_0)q_{-1,1} + \pi_0 q_{0,1} - 1\Big) \\ & + 2(1-\pi_0)^2(q_{1,1} + q_{-1,1})\Big(\pi_1 q_{1,1} - \pi_{-1} q_{-1,1}\Big). \end{aligned}$$

This concludes the exercise.