MAFS5270: Mathematical Market Microstructure – Theory and Applications, 2019

Assignment 2 – Analyzing Response Function of Market Impact at Microstructure Level

Due Date: 11:59pm, Mar. 31, 2019

It is well known that a transaction can generate market impact to the price of a traded instrument. In this assignment, we consider the case of commodity futures trading and a delicate impact process at a tick-by-tick level.

1. **Kernel Functions**: Assume that trading of a commodity futures contract commenced at a time point that is far away from the current time point: $t > 0$. Further assume that at exactly the same time point t, a transaction happened with a price of $p_t > 0$, size of $V_t > 0$, and a sign of ϵ_t which is +1, -1 or 0. +1 indicates that the trade is buyer-initiated, -1 indicating seller-initiated, and 0 indicating no specific direction can be determined.

   To understand how all the trades that happened before the trade at time point t (or, the "current trade") influenced the market to attain a trade price of p_t, we consider the following assumptions:

   (1) First of all, we assume that all the trades that happened before the trade at time point t should have impact on the current trade, even though we are quite certain that the most recent ones should have higher, more direct impact on the current trade than "older" trades that happened at time points t' with $t' \ll t$;

   (2) Secondly, we assume that for a trade that happened at t', the direction of its impact on the prices of future trades after it will be the same as the direction of this trade, $\epsilon_{t'}$; that is, if the trade at t' is buyer-initiated, then it inserts upward pressure on the prices of all trades that are after it, and vice versa if the trade at t' is seller-initiated;

   (3) Thirdly, we assume that the magnitude of the impact of trade at t' on future trades will be in proportion to the size of this trade, but in a non-linear fashion.

   With the above assumptions, we contemplate that the trade price at time point t will be a combination of an independent innovation term and the impacts of all trades that happened before t, with older trades that have more decayed impact than new ones. Mathematically, we can express the trade price, p_t, as:

$$p_t = \sum_{t'<t}\left[G(t-t')V_{t'}^{\alpha}\epsilon_{t'}\right] + \varepsilon_t. \qquad (1)$$

   In equation (1), $G(t)$ is called a kernel function that is instrument-specific, but not trade-specific. Figure 1 shows the shape of a typical kernel function, which typically increases with its argument $t - t'$ while it is small, then decays to 0 when $t - t' \to +\infty$. A kernel function reflects the dynamics of order books, including limit order insertion, limit order deletion and market order interaction with order books. Parameter α in equation (1) is positive and less than 1, which has been observed frequently in markets across different kinds of assets. In addition, we often assume that the innovation term ε_t for $-\infty < t < +\infty$ in equation (1) follows an independent random process from the impact process that we focus on. ε_t is often driven by stock-specific processes such as news, etc.

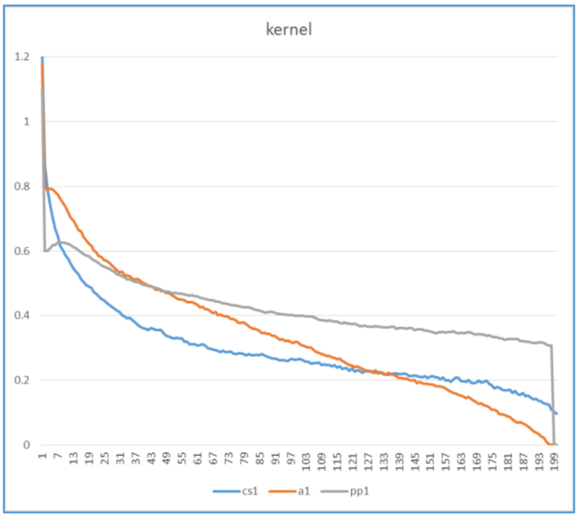

Figure 1: A few sample kernel functions extracted from tick data.

2. **Response Functions**: A response function is defined as follows:

$$R_l = \langle (p_{t+l} - p_t)\epsilon_t \rangle_{over\,t}. \tag{2}$$

The response function gives the average price change over a time lag of l, conditioned on the direction of the trade at time t, then averaged over all the trades that one can gather in his/her sample. In practice, especially for instruments that one can only get "sub-samples" of market data (such as one snapshot every a few hundred milliseconds or a few seconds), (2) is often re-written as

$$\widetilde{R}_l = \langle (\hat{p}_{t+l} - m_t)\epsilon_t \rangle_{over\,t} \tag{3}$$

where m_t is the mid-quote of the snapshot at time t, and $\hat{p}_t$ is the VWAP of all the trades that happened between t and $t + \Delta t$, two adjacent snapshots with $\Delta t \sim 250$ milliseconds for commodity futures that we consider in this assignment. $\hat{p}_{t+l}$'s are subsequent VWAP prices for trades the fall between $t + l\Delta t$ and $t + (l+1)\Delta t$ for $l = 1,2,\ldots$ In this set-up, m_t is also called the "prevailing mid-quote" for $\hat{p}_t$ because it appears right before the trades between t and $t + \Delta t$. Figure 2 shows an example response function with its definition in equation (3). It shows that $R_l\ (l = 0, 1, 2, \ldots)$ is often statistically significantly positive for relatively small l (such as $l \lesssim 200$) and can extend to as large as $l \sim 1000$, which is a clear sign of market impact: for trades with $\epsilon_t = +1$, a <u>positive</u> response function indicates that, on average, $\hat{p}_{t+l}$ in equation (3) is <u>higher</u> than m_t; for trades with $\epsilon_t = -1$, a <u>positive</u> response function indicates that, on average, $\hat{p}_{t+l}$ in equation (3) is <u>lower</u> than m_t. In both cases, the trade at t has "directional" impact on subsequent trade prices. Note that, because we assume that the innovation term ε_t for $-\infty < t < +\infty$ follows an independent random process, it does not affect the response function. For most of the time, one often divides the right-hand side of equation (3) by the bid-ask spread of the prevailing quote at t

so as to smooth out the effect of fluctuating bid-ask spread.

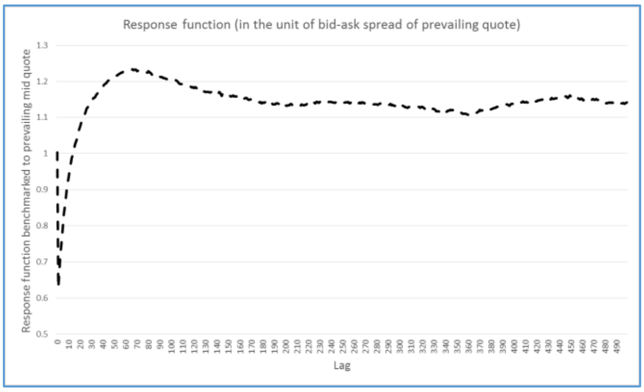

Figure 2: A sample response function in the unit of bid-ask spread of prevailing quote.

3. **<u>Sign and Size Auto-correlation</u>**: Let's define

$$C(l) \equiv \langle \epsilon_t \epsilon_{t+l} V_{t+l}^{\alpha} \rangle \tag{4}$$

over all of t's. For financial instruments that have relatively constant trade sizes, we often further simplify matters and include trade size in equation (4) as follows:

$$C(l) \sim \bar{V}^{\alpha} \langle \epsilon_t \epsilon_{t+l} \rangle, \text{ or } \; C(l) \sim \bar{V}^{\alpha} c(l) \;\text{ where }\; c(l) \equiv \langle \epsilon_t \epsilon_{t+l} \rangle. \tag{5}$$

where $\bar{V}$ indicates an average trade size. One of the rationales behind equation (5) is that, when $\alpha < 1$, marginal increase in trade size has relatively subdued influence on the response function calculations in (2) and (3).

It is often found from empirical studies that, for commodity futures contracts that we focus on in this assignment, $\langle \epsilon_t \epsilon_{t+l} \rangle$ follows a power law as follows: $\langle \epsilon_t \epsilon_{t+l} \rangle \sim |l|^{-\gamma}$ for $0 < \gamma < 1$. Note that, at least empirically, $\langle \epsilon_t \epsilon_{t+l} \rangle \sim \langle \epsilon_t \epsilon_{t-l} \rangle$, which is often called "time homogeneity". Therefore, we often assume that $c(l) \sim c(-l) \sim c(|l|)$.

4. **<u>Question 1 (25 points)</u>**: With (1), (4) and (5), prove that the response function defined in (2) can be written as

$$R_l \sim \bar{V}^{\alpha} [\textstyle\sum_{0<t' \le l} G(t')\, c(t'-l) \; + \sum_{t'>l} G(t')\, c(t'-l) - \sum_{0<t'} G(t')\, c(t')].$$

5. **<u>Question 2 (25 points)</u>**: With the data provided with this assignment (see appendix of this assignment on column definitions in the data file), construct $\widetilde{R}_l$ for $0 \le l \le 500$ as defined in equation (3) using all the available trades provided.

6. **<u>Question 3 (25 points)</u>**: With the data provided with this assignment, construct $\widetilde{R}_l\big|_V$ for $0 \le l \le 500$ as defined in equation (3) for trades in different groups of trade sizes. That is, if we label all trades that have sizes that $v_i < V_i \le v_{i+1}$ as group i, calculate $\widetilde{R}_l\big|_{v_i < V_i < v_{i+1}}$

for $0 \le l \le 500$ as defined in equation (3) for all trades within group i. Comment on your findings from this analysis, especially on how the response function depends on trade sizes. In this assignment, we define: $v_1 = 0, v_2 = 2, v_3 = 5, v_4 = 10, v_5 = 15, v_6 = 20, v_7 = 30, v_8 = 40, v_9 = 55, v_{10} = 90, v_{11} = 100000$.

7. **<u>Question 4 (25 points)</u>**: For $l = 10, 20, 30, 40, 50, 75, 100, 125, 150, 175, 200, 250$, plot $\log(\widetilde{R}_l\big|_{v_i<V_i<v_{i+1}})$ as a function of $\log(\langle V_i \rangle)$ and fit the data into a straight line. Compare the slopes of different straight lines for different l. $\langle V_i \rangle$ is the average of trade sizes of all trades in group i.

Appendix: <u>Explanation of the format of provided data</u>

Attached to this assignment are two tick data files for the commodity futures contract of dlpp (Polypropylene) for the months of July and August of 2016, respectively. The two files are of the same data format. Each file has 9 columns (see Figure 3 as a snapshot of a section of one of the files). The meanings of columns are:

- <u>Column A</u>: This is an index column for sequential trades of each trading day;
- <u>Column B</u>: This column indicates the trading date in the format of YYYYMMDD;
- <u>Column C</u>: This column indicates the time of each individual snapshot in the format of HHMMSSxxx where "xxx" indicates milli-seconds;
- <u>Column D</u>: This column indicates the total number of contracts traded between the current row and the immediate next row; that is, it is the total number of contracts traded from the Time of the current row to the Time of the immediate next row;
- <u>Column E</u>: This column indicates the VWAP of all the trades that happened between the current row and the immediate next row;
- <u>Column F</u>: This column indicates the sign of the VWAP as compared to the prevailing midQuote indicated in Column G;
- <u>Column G</u>: This column indicates the mid-quote of the bid-ask quote snapped at the Time indicated in column C; that is, this is the mid-quote of the prevailing quote of the VWAP in column E;
- <u>Columns H/I</u>: These two columns indicate the bid price and ask price of the prevailing quote snapped at the Time indicated in Column C; the average of these two columns gives "midQ" in column G.

	A	B	C	D	E	F	G	H	I
1		Date	Time	Size	VWAP	Sign	midQ	BP1	SP1
22212	49401	20160701	145851602	2	5224	-1	5225	5224	5226
22213	49402	20160701	145851842	6	5226	1	5225	5224	5226
22214	49403	20160701	145852103	2	5226	1	5225	5224	5226
22215	49407	20160701	145853006	2	5224	-1	5225	5224	5226
22216	49409	20160701	145853635	2	5224	-1	5225	5224	5226
22217	49413	20160701	145854613	2	5224	-1	5225	5224	5226
22218	49415	20160701	145855105	8	5225	0	5225	5224	5226
22219	49416	20160701	145855444	10	5224	-1	5225	5224	5226
22220	49417	20160701	145855665	24	5224	-1	5225	5224	5226
22221	49418	20160701	145855890	2	5226	1	5225	5224	5226
22222	49419	20160701	145856196	14	5225.429	1	5225	5224	5226
22223	49421	20160701	145856591	6	5224.667	-1	5225	5224	5226
22224	49423	20160701	145857196	2	5224	-1	5225	5224	5226
22225	49424	20160701	145857426	6	5226	1	5225	5224	5226
22226	49425	20160701	145857697	12	5224	-1	5225	5224	5226
22227	49426	20160701	145857740	32	5224.125	-1	5225	5224	5226
22228	49433	20160701	145859882				5225	5224	5226
22229	0	20160704	90100004	78	5252	1	5251	5250	5252
22230	1	20160704	90100278	34	5252	1	5251	5250	5252
22231	2	20160704	90100479	6	5252	1	5251	5250	5252
22232	3	20160704	90100688	2	5250	-1	5251	5250	5252
22233	4	20160704	90101025	40	5250.4	-1	5251	5250	5252
22234	5	20160704	90101231	2	5252	1	5251	5250	5252
22235	6	20160704	90101492	2	5250	-1	5251	5250	5252
22236	7	20160704	90101769	6	5252	1	5251	5250	5252
22237	9	20160704	90102285	4	5252	1	5251	5250	5252
22238	10	20160704	90102536	4	5252	1	5251	5250	5252
22239	11	20160704	90102787	42	5250.095	-1	5251	5250	5252

Figure 3: A section of one of the attached data file as an example.

This is the end of assignment 3.