User Guide

**How did we get the data?**

The TAQ data is supported by Vertex, the raw data is placed in the Raw Data folder. Folder **CME commodities 2018 with 10 levels** contains the completed data, it is the aggregation of (limit order book) Event and (1-10 levels of book) Depth information. You may also do the raw data to get the completed TAQ data on your own.

**What is the data set?**

The completed data covers 6 commodities traded in CME in calendar year 2018: Crude Oil (CL), E-mini S&P 500 (ES), Eurodollar (GE), Gold (GC), Corn (ZC) and Soybean (ZS); nearby and deferred futures contracts are included (named them rank 1 and 2 of each futures contracts), the **DatesAndFutures.csv** file includes all daily contracts we should have in 2018 and more information including expiration date, days to expire, weekday/weekend, holiday (not all of them are available in the database, when the raw daily data includes no actions then the completed daily data is empty).

**Data Cleaning** (Read only if you want to generate similar completed data from the raw data by yourself, then this is helpful):

1. Unzip Raw Data: Use **Automate unzip vertex data using python.py** to unzip raw data files (depth and event) you want in python;
2. Data Organizing: Use **Data Clean Macro then Delete File (20200108).sas** to merge daily depth and event for a final data you want (make changes on the code, my code will automatically delete imported data files during the macro);
3. You may have to read **vertex\_rest\_api\_raw\_data.pdf** from vertex if you really want to do data cleaning on your own.

**Variable List:**

**Date**: date from 20180101 – 20181231 (20171231 and 20190101 are included when it’s available just in case);

**Mtime**: time in nanoseconds;

**NS**: numerical time in nanoseconds;

**OriginalID:** the system defined identity of each action;

**Token**: System allocated ID to each limit order as first 12-13 digits, with additional 6 digits sub ID as private assigned ID for each original ID, + 1 to the sub ID only when there is a CHANGE in Xcomment, indicating the limit order has changed and generated a new limit order;

**Token\_new**: it is always missing unless there is a CHANGE in Xcomment, it means new limit order has been generated after a CHANGE;

**Xcomment**: action of the limit order,

NEW = new limit order (OVERLAY, pure NEW order of derived from CHANGE orders);

DELETE = limit order disappears without executions (canceled) or the market is closed in end of trading day ();

TRADE = the execution clears all contracts in the limit order;

**Side**: side of the limit order, S = sell, B = buy;

**Volume**: the number of contracts available in the limit order at the current status (it was called Quantity in my old version data);

**Cvolume**: it is missing when it is execution or it’s a pure new limit order, otherwise it means how the limit order changes of volume from its last status;

**Price**: price of the limit order at the current status;

**Cprice**: it is missing when it is execution or it’s a pure new limit order, otherwise it means how the limit order changes of price from its last status;

**Bid1-10/Ask1-10**: bid or ask price at book level 1 to 10;

**Bidq1-10/Askq1-10**: number of bid or ask limit orders available at book level 1 to 10;

**Bidv1-10/Askv1-10**: number of bid or ask contracts available at book level 1 to 10;

**Case:** 9 Cases defined when coding raw TAQ message as below (read code for detail). Where LO=Limit Order MO=Market Order, 2 exceptions (Case=999 and . ) need further study if really exist.

/\*1.Overlay - Initialize market\*/

/\*2.New - New, MO if MO=1\*/

/\*Change - without execution\*/

/\*3.Only volume down without execution\*/

/\*4.All other cases without execution\*/

/\*Change - be executed by MO, seems lprice = price\*/

/\*5.When priority changed means more LO add when execute, so generate new limit order first\*/

/\*999.LO May escape from MO when it's executing\*/

/\*6.Partial Trade\*/

/\*7.Change - itself is a MO, i.e. price changed across spread and stayed\*/

/\*8.DELETE - without execution\*/

/\*9.DELETE - by execution\*/

/\*Else, we missed anomaly pattern of data, we should see Case = .\*/

**Rank:** If the original action in CME has been split into multiple sub-actions, then rank of them suggests the logical order of these sub-actions;

**Comment:** “I” = initialize market, OVRLAY in raw data; “M” = market order, this limit order plays a role of market order at this action; “A” = limit order side add volume (otherwise it will be traded to 0) when it’s being executed; “C” = market clearing in end of the day (if this action is not DELETE then it is meaningless).

Chao Wang

Jan 8 2020