cme.mdp: Clean and Analyze Chicago Mercantile Exchange Market Data in R

Richie Ma, Brian Peterson

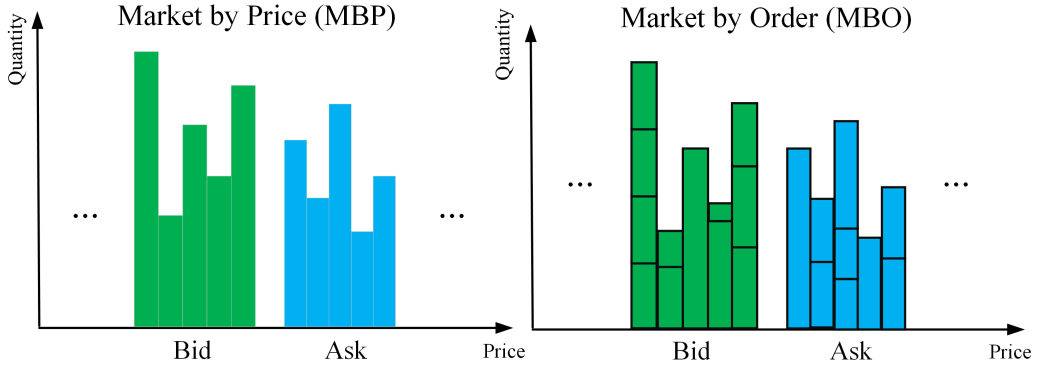
University of Illinois at Urbana-Champaign

April 12, 2025

Introduction

- Financial markets have become more transparent and exchanges can provide high-frequency data for better monitoring market in real-time.
  - Creating more demand about the high-frequency data usage (academia and industry).
- Most exchanges do not disseminate tabulated market data to non-member market participants, and almost all market data are specially coded to enhance the communication efficiency.
  - e.g., binary coding, for both inbound and outbound messages
- Thus, financial economists need to know how to clean these untabulated data at first, which is a substantially time-consuming task.
- `R cme.mdp` package focuses on how to clean and analyze the CME market data under the FIX protocol.
  - Market by Price (MBP) and Market by Order (MBO)

CME market data overview



cme.mdp package

Installation

- The `cme.mdp` package can be installed through Github by `devtools`.
- All dependent packages will be installed automatically.
- All data-related process is done by `data.table` package.

```
1 install.packages("devtools") #Install the 'devtools' package if you haven't
2 devtools::install_github("richie-ma/cme.mdp")
```

cme.mdp package

Trade summary

```
library(cme.mdp)
gold_trades <- trade_summary(
  mbp_input = "R:/xcec_md_gc_fut_20200102-r-00390.gz",
  date = '2020-01-02',
  price_displayformat = 0.1
)
#> CME MDP 3.0 Trade Summary
#>  contracts: GCF0 GCG0 GCG1 GCH0 GCJ0 GCM0 GCQ0 GCV0 GCZ0
head(gold_trades[["GCF0"]])
#> Key: <Code, Seq>
#>          Date  MsgSeq              SendingTime             TransactTime   Code
#>        <Date>   <num>                   <char>                   <char> <char>
#> 1: 2020-01-02 7672654 20200102020146175426919 20200102020146174657851   GCF0
#> 2: 2020-01-02 7676350 20200102020218681738774 20200102020218678765327   GCF0
#> 3: 2020-01-02 7676361 20200102020218682147202 20200102020218678773403   GCF0
#> 4: 2020-01-02 7676364 20200102020218682531825 20200102020218678775243   GCF0
#> 5: 2020-01-02 7676366 20200102020218682721907 20200102020218678777285   GCF0
#> 6: 2020-01-02 7676370 20200102020218682975735 20200102020218678781159   GCF0
#>         Seq      PX  Size   Ord   agg
#>       <num>   <num> <num> <num> <num>
#> 1: 2275838 1518.1     5     1     0
#> 2: 2276877 1518.0     2     1     0
#> 3: 2276886 1518.0     2     1     0
#> 4: 2276890 1518.0     1     1     0
#> 5: 2276894 1518.0     1     1     0
#> 6: 2276898 1518.0     1     1     0
```

cme.mdp package

Trade summary with matching details

```
## The sample data is COMEX gold futures on October 25, 2017
gold_trades_orders <- mbo_match_details(
  mbo_input = "R:/xcec_md_gc_fut.gz",
  price_displayformat = 0.1
)
#> CME MDP 3.0 Trade Summary with Matching Details
#>  contracts: GCZ7
head(gold_trades_orders[["GCZ7"]])
#> Key: <MsgSeq>
#>         Date    MsgSeq             SendingTime             TransactTime    Code
#>       <char>     <num>                  <char>                   <char> <char>
#> 1: 20171025 14899367 20171025123000003425726 20171025123000000215181   GCZ7
#> 2: 20171025 14899367 20171025123000003425726 20171025123000000215181   GCZ7
#> 3: 20171025 14899504 20171025123000007263052 20171025123000007035151   GCZ7
#> 4: 20171025 14899504 20171025123000007263052 20171025123000007035151   GCZ7
#> 5: 20171025 14899513 20171025123000008425450 20171025123000008189723   GCZ7
#> 6: 20171025 14899513 20171025123000008425450 20171025123000008189723   GCZ7
#>         Seq     PX   Qty   Ord   agg trade_id     order_id matched_qty
#>       <num>  <num> <num> <num> <num>     <num>       <char>       <num>
#> 1: 5207583 1273.6     1     2     1  15508475 762981983336           1
#> 2: 5207583 1273.6     1     2     1  15508475 762981983304           1
#> 3: 5207611 1273.5     1     2     1  15508484 762981983442           1
#> 4: 5207611 1273.5     1     2     1  15508484 762981983435           1
#> 5: 5207616 1273.6     3     3     1  15508485 762981983447           3
#> 6: 5207616 1273.6     3     3     1  15508485 762981983367           2
```

cme.mdp package

Quote messages with order details

```
## The sample data is COMEX gold futures on October 25, 2017

sp500_quotes_orders <- mbo_quote_queue(
  mbo_input = "R:/xcme_md_es_fut.gz",
  price_displayformat = 0.01
)
#> CME MDP 3.0 Quote Messages with Queue Information
#>  contracts: ESZ7 ESH8 ESM8
head(sp500_quotes_orders[["ESZ7"]])
#>        Date  MsgSeq             SendingTime          TransactTime Update
#>      <char>  <char>                  <char>                <char>  <num>
#> 1: 20171025 4504335 20171025123000001200661 20171025123000001087823      0
#> 2: 20171025 4504336 20171025123000001208956 20171025123000001096167      2
#> 3: 20171025 4504355 20171025123000089527420 20171025123000089296549      2
#> 4: 20171025 4504355 20171025123000089527420 20171025123000089296549      2
#> 5: 20171025 4504356 20171025123000089681329 20171025123000089555777      1
#> 6: 20171025 4504357 20171025123000089698822 20171025123000089558125      1
#>       Side    Code      PX    Qty Order_priority     Order_id    Seq
#>     <char> <char>   <num> <num>          <num>       <char> <num>
#> 1:       1   ESZ7 2579.5     1     5535661496 644490560830    NA
#> 2:       0   ESZ7 2549.5     1     5535661193 644490560657    NA
#> 3:       0   ESZ7 2564.5     2     5535661466 644490560815    NA
#> 4:       0   ESZ7 2564.5     2     5535661495 644490560829    NA
#> 5:       0   ESZ7 2523.5    41     5535661510 644490542185    NA
#> 6:       0   ESZ7 2528.0    22     5535661511 644490429553    NA
```

cme.mdp package

order_book()

- `order_book()` function provides an integrated method to reconstruct outright, implied, and consolidated limit order books easily.
  - The (maximum) book depth `level` is given by users. Currently, `cme.mdp` will check whether `level` is equal to the maximum book depth stipulated by the CME.
  - Consolidated book will be processed when users set `consolidate = TRUE` (default).
  - Also support to process only a single security. e.g., `security = 'ESH9'`.
  - A whole trading week's quote message file should be used though now the package does not check this.

```
1 book <- order_book(mdp_quote_msgs_list, "2019-01-07", level = 10,
     consolidated = TRUE, security = 'ESH9')
```

Visualization

e.g., book_bar()

- cme.mdp provides a set of easy-to-use visualization functions for limit order books.
  - e.g., book_bar() is used to either plot static or animated limit order book.

```
book_bar <- function(mbp_order_book, implied_book=NULL, level=NULL,
    animation=FALSE, anim_fps=10){...}
```

  - An animated example from Github: https://github.com/richie-ma/cme.mdp

- Thank you!
- Github repo: `https://github.com/richie-ma/cme.mdp`
- Personal website: `www.richiema.com`