Thesis
Documentation

Mahdi Mir

June 26, 2021

Contents

1 Scripts

1.1 a

1.1.1 Indices Data

1. Indeices data got from Hamid-Reza Amiri.

1. I Clean data, add gregorian date to it.
2. I make it consistent with the notation and convention of date in the adj prices data.

Indices data saved in CleanData directory.

1.1.2 Adjusted Prices Data

- The adjusted prices data set got from Hamid-Reza Amiri in the associated folder.
- ID-Tickers map got from Hamid-Reza again.

1. The data got prices only for traded days. Holidays are absent. When Trading is halted on a security there is no entry in data.
2. I constructed all unique IDs X Dates pairs. By Cartesian multiplication of unique IDs and unique days.
3. I sort data by dates.
4. I forwardly fill close prices for each ID (that's why I sort by dates).i.e. If security is halted for trading then it would have NAN entries in the associated (ID, Date) pair, by forward filling the close prices if the security has a price in a prior date then for all NAN values from the prior date to next trading day prices is set to the prior date.
5. Then I remove remained NAN close price rows. Those are ones before their IPO (Has no prior not-NA close price). 6. For the filled days, I define a new column named 'TradeHalt' which is True for halted days and false otherwise. These are the days that security cannot be traded.
6. Date column in the data is Gregorian date, I create a new column named 'JDate' which is the Jalali counter part Date.
7. I added ticker names by the other dataset which maps 'IDs' to 'Tickers'.
8. I remove IDs without Tickers. These are not stocks of firms.
9. I create a new column called 'JYearMonth', which contains year and month of each row as an integer.e.g. 138001, for the first month of year 1380.
10. Finally I keep these columns: [ID, Ticker, Date, JDate, Close, TradeHalt]

1.1.3 Outputs

1. py/cleanData/a-indices.parquet
2. py/cleanData/a-adjprices.parquet

Adjusted Price data saved in CleanData directory.

1.2 b

1.2.1 Adjusted Prices Metadata