Extracting Intraday Price Quotes for Multiple Days from TAQ WRDS Millisecond Files using SAS

5 minutes interval

Ofira Schwartz-Soicher/Todd Hines
oschwartz@princeton.edu/thines@princeton.edu

October, 2016
Notes

- The example program below may be used to extract variables from TAQ or other data files on WRDS.

- You should copy the whole code into SAS and run the whole program. Please be sure to modify the program as necessary (e.g., library name, interval length, date and sym_root.)

- Text in green or red in the document below indicates comments and will not be executed by SAS.

- This program is based upon a sample code provided by WRDS: (http://wrds.wharton.upenn.edu/)
/* Details about the PC-connection (this is always the same), it will trigger a login prompt. Use your WRDS username and password to log on*/

%let wrds = wrds.wharton.upenn.edu 4016;
   options comamid = TCP remote=WRDS;
   signon username=_prompt_;

/* Specifies where to store the datasets on your computer 
   (This should be modified according to your folder structure) */

libname myLib "H:\DSS\WRDS ";

/* Everything after 'rsubmit' will be executed on WRDS!! */

rsubmit;
/* data on WRDS is organized in libraries. The code below specifies that the data of interest is located in '/wrds/nyse/sasdata/taqms/cq' when 'taqms' is the library name. This should be modifies if your are interested in extracting data from a different folder */

libname taqms'/wrds/nyse/sasdata/taqms/cq';

/***** Input area *********************************************/
* Specifying starting time – 9:30AM;
%let start_time = '9:30:00.000't;
* Specifying interval as 5 minutes (5*60 seconds) – could be modified according to your needs by changing the 5 into the time interval of interest (e.g., 15 for 15 minutes interval) ;
%let interval_milsec =5*60;
/***** End of input area *********************************************/
* Extract data for 2 stocks from multiple days, we consider the time between 9:30am to 4:30pm, retrieving only SYMBOL, DATE, TIME and PRICE. To extract data for different/additional stocks change or add symbols in the parenthesis of the line ‘where sym_root in ('GE', 'C')’;

```
data tempx_multi;
    set taqms.cqm_20140102-taqms.cqm_20140103
taqms.cqm_20140106-taqms.cqm_20140110
    (keep= date time_m sym_root bid ask); /* Please modify files according to dates of interest; */
    format time_m time12.3;
    where sym_root in ('GE', 'C')
        and time_m between '9:30:00.000't and '16:30:00.000't;
run;
/* Sorting the data by SYMBOL, DATE and TIME */
proc sort;
    by sym_root date time_m;
run;
Title 'Print several lines of the original data for checking';
```
data milsec2014_multi;
    set tempx_multi;
        by sym_root date time_m;
        retain itime rtime iprice;  *Carry time and price values forward;
            format itime rtime time12.3;
        if first.sym_root=1 or first.date=1 then do;
            /*Initialize time and price when new symbol or date starts;*/
                rtime=time_m;
                iprice=bid;
                itime= &start_time;
        end;
        if time_m >= itime then do;  /*Interval reached;*/
            output;  /*rtime and iprice hold the last observation values;*/
            itime = itime + &interval_milsec;
            do while(time_m >= itime);  /*Need to fill in all time intervals;*/
                output;
                itime = itime + &interval_milsec;
            end;
        end;
    rtime=time_m;
    iprice=bid;
    keep sym_root date time_m itime iprice rtime;
run;
Title "Final output -- \&interval_milsec";

**proc print** data=milsec2014_multi (obs=400);
  var sym_root date time_m itime iprice rtime;
run;

/* Downloading the dataset in SAS format */
**proc download** data=milsec2014_multi out=myLib.tempx_multi;
run;

/* Stop processing statements on WRDS and continue working locally */
endrsubmit;

• Exporting the SAS dataset into a .CSV so could be read in Excel. You may want to change the name if the interval is changed or looking at a different file;

**proc export**
DATA=myLib.tempx_multi
OUTFILE="H:\DSS\WRDS \tempx_multi.csv"
DBMS=csv REPLACE;
run;

* Note: To see time as hours, minutes and seconds the time formatting may have to be modified on the Excel spreadsheet: From the “Number” tab select “Time” and “13:30:55”;}