Accessing TAQ data from D-CAF Servers at Aarhus University

In order to access and download Trade And Quote (TAQ) data from the servers at Aarhus University you need:

1. Access to the networks at Aarhus University - Either by being on the premises or by VPN

2. Access to the network drive is obtained by contacting dbsupport@econ.au.dk with an email containing your AU login

    (a) If not affiliated with AU, contact dbsupport@econ.au.dk with your information, including department, university and birthday to obtain a username to the VPN.

3. A working copy of MATLAB

4. Template files available on the server

Access

The TAQ data is saved on a local server in HDF5[1] format for the years 1993 to 2013 consolidated in seconds. Data will also be available for 2014 and forward, but with a timestamp in miliseconds. This highly compressed format enables us to store large amount of data in small footprint. The HDF5 format is widely supported by most major programming languages[2]), and has libraries independently developed[3].

After ensuring that you are connected to Aarhus University's network either wired or wireless through Eduroam, you have to add the network drive located at "\\cafmaster.econ.au.dk\data\". On Windows this is done by right-clicking "Computer" and choosing "Map Network Drive". After this you are presented with a window like in Figure (1). You need to click of "Connect using different credentials" and input the login credentials.

Figure 1 – Network Drive

---
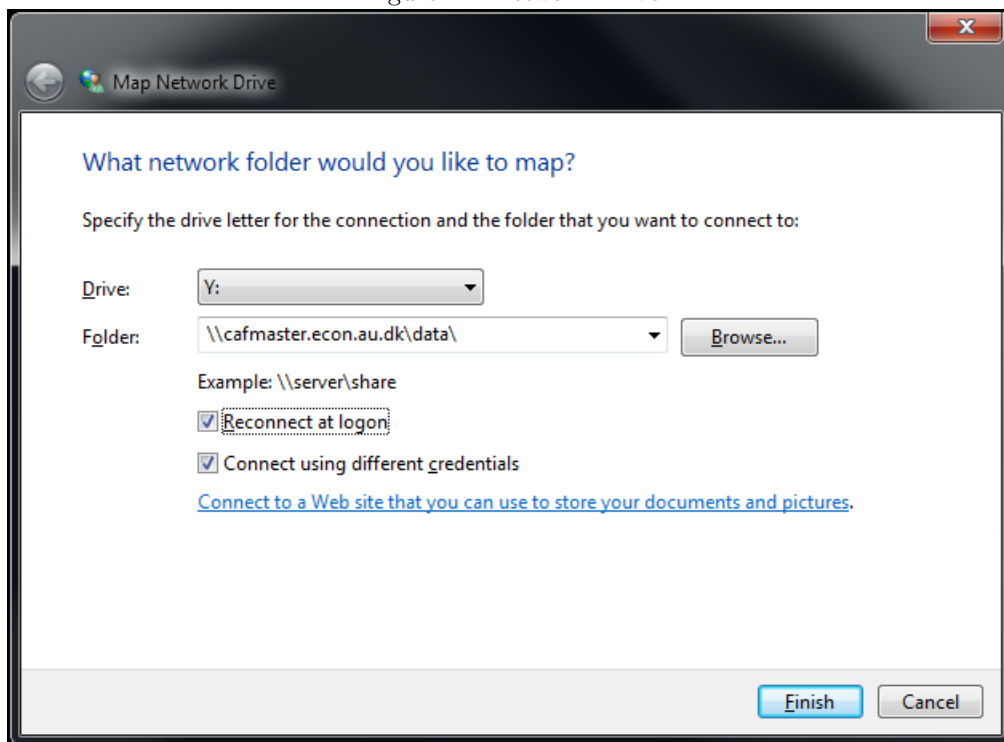
[1] http://www.hdfgroup.org/
[2] http://www.hdfgroup.org/products/hdf5_tools/index.html
[3] http://www.hdfgroup.org/products/hdf5_tools/SWSummarybyName.htm

After successfully connecting to the network drive, you can access the drive and navigate to the folder called "TAQ".

If you cannot gain access to the network drive ensure and check if you have fullfilled Step 1 above. Otherwise, contact dbsupport@econ.au.dk

Downloading data

In the TAQ folder we have stored the raw TAQ files in HDF5 format, as well as several program files for MATLAB. In "TAQ\matlab\windows" you will find two folders (named hdf5 and templates) which you can copy to your own computer. In the following I have copied the two folders to a folder named "TAQ" on the C: drive of my computer, and this will be my working directory. You can put it anywhere you have access to.

In "hdf5" there are the three functions we need to get access to the TAQ data, each defined of their own .m files. .m files are MATLAB files.

convert_to_escaped.m is a helper function that cleans up any issues with characters.

taq_h5_filename_daily.m is helper function that tells the **taq_h5_read_daily.m** file where to look for our files.

Figure 2 – taq_h5_filename_daily.m

```
1      function [shortFileName,fullFileName] = taq_h5_filename_daily(date,base)
2
3 -    if nargin==1
4 -        base = [];
5 -    end
6
7 -    shortFileName = ['taq_' num2str(date) '.h5'];
8
9 -    if isempty(base)
10 -       base = '//cafmaster.econ.au.dk/data/TAQ/TAQHDF5/';
11 -   end
12 -   fullFileName = fullfile(base,shortFileName);
13
```

The function **taq_h5_filename_daily.m** (Shown in Figure (2)) takes 2 arguments, date and base. The date argument is used to construct the filename to look for in our database, while base is the location of the actual data. If base is not specified, the function will assume that the base is "//cafmaster.econ.au.dk/TAQ/TAQHDF5/". We will specify the base correctly when we use the **taq_h5_read_daily.m function**.

taq_h5_read_daily.m contains the function that does the actual reading of the data. If you are used to loading in .txt and .csv files in MATLAB, consider this a similar function, where instead of .csv/.txt files we are loading in a special type of file, called HDF5. The taq_h5_read_daily function takes 5 arguments: type, ticker, date, base, metaData. These are explained in Table (1).

Table 1 – Arguments

Argument	Value	Mandatory
type	Character - "Quotes", "QuoteIndex", "Trades" or "TradeIndex"	Yes
ticker	Character - Ticker name (e.g. "IBM", "MSFT", or "SPY")	Yes
date	Date - Must be open trading day	Yes
base	file path - Location of the HDF5 files	No
metaData	Helper-data used when loading indeces	No

Finally, we have a file called **daily.m** in the folder "templates". This MATLAB file puts together all the functions explained above, and shows you how to access the data. Open up the **daily.m** file in MATLAB and you will see something like Figure (3).

Figure 3 – daily.m

```
1 -     clear;
2 -     clc
3 -     close all;
4
5       % matlab hdf5 utilities for TAQ
6 -     addpath('//cafmaster.econ.au.dk/TAQ/matlab/windows/hdf5/');
7
8
9
10 -    TAQTr = taq_h5_read_daily('Trades', 'SPY', 20020521);
11
12      % You can also use 'Quotes', and change 'SPY' to another ticker.
13
14 -    TAQQu = taq_h5_read_daily('Quotes', 'SPY', 20020521);
15
16      % To make it faster with repeated stocks on the same day
17
18 -    metadata = taq_h5_read_daily('TradeIndex', [], 20020521);
```

First we use "addpath" to add the path where the functions we need to use are located to our current MATLAB installation.

When you first open **daily.m** you will see it uses "addpath" on //cafmaster.econ.au.dk/TAQ/matlab/windows/hdf5. However, we want to change this to the location where we saved the "hdf5" folder. In my case, "C:\TAQ\hdf5". Remember to change back slashes to forward slashes. So now we should have something similar to Figure (4), but with your specific path in "addpath".

Figure 4 – daily.m

```
1 -     clear;
2 -     clc
3 -     close all;
4
5       % matlab hdf5 utilities for TAQ
6 -     addpath('C:/TAQ/hdf5');
7
8
9
10 -    TAQTr = taq_h5_read_daily('Trades', 'SPY', 20020521);
11
12      % You can also use 'Quotes', and change 'SPY' to another ticker.
13
14 -    TAQQu = taq_h5_read_daily('Quotes', 'SPY', 20020521);
15
16      % To make it faster with repeated stocks on the same day
17
18 -    metadata = taq_h5_read_daily('TradeIndex', [], 20020521);
19
```

Next we consider the first example of reading in trades, to the object called TAQTr.

Here we call the function **taq_h5_read_daily** using the three mandatory arguments as specified earlier. We want "Trades", we want it for the S&P500 tracker, which has the ticker name "SPY" and we want it for the 21th May, 2002 - 20020521. Running the first ten lines of code, should now return a 1x1 struct in your workspace called TAQTr, which holds the data requested as shown in Figure (5). The different variables are explained in Table (2).
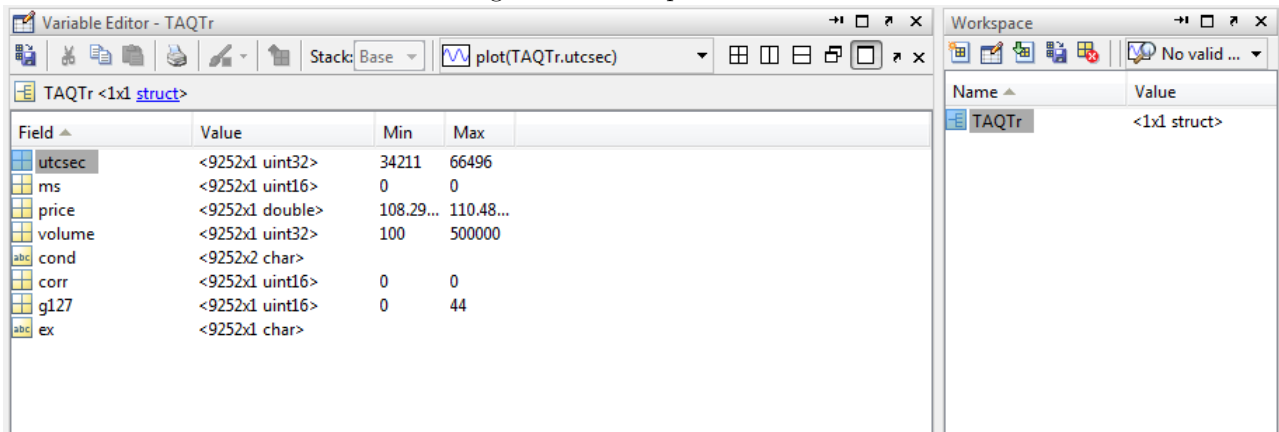
Figure 5 – Workspace - Trades

Table 2 – TAQ Variables - Trades

utcsec	Number of seconds (measured from EST)
ms	Milliseconds (Only valid for data from 2014)
price	Price
volume	Volume
cond	Condition of trade
corr	Correction
ex	Exchange abbreviations (See Table (4))

Similarly, for we can request quotes by running line 14 in figure 6:

Figure 6 – daily.m

```
1 -    clear;
2 -    clc
3 -    close all;
4
5      % matlab hdf5 utilities for TAQ
6 -    addpath('C:/TAQ/hdf5');
7
8
9
10 -   TAQTr = taq_h5_read_daily('Trades', 'SPY', 20020521);
11
12     % You can also use 'Quotes', and change 'SPY' to another ticker.
13
14 -   TAQQu = taq_h5_read_daily('Quotes', 'SPY', 20020521);
15
16     % To make it faster with repeated stocks on the same day
17
18 -   metadata = taq_h5_read_daily('TradeIndex', [], 20020521);
19
```

In this case, we request the quotes for the same day and ticker as with the trades. Now we have a TAQQu 1x1 struct in our workspace with the data requested. The different variables are explained in Table (3).
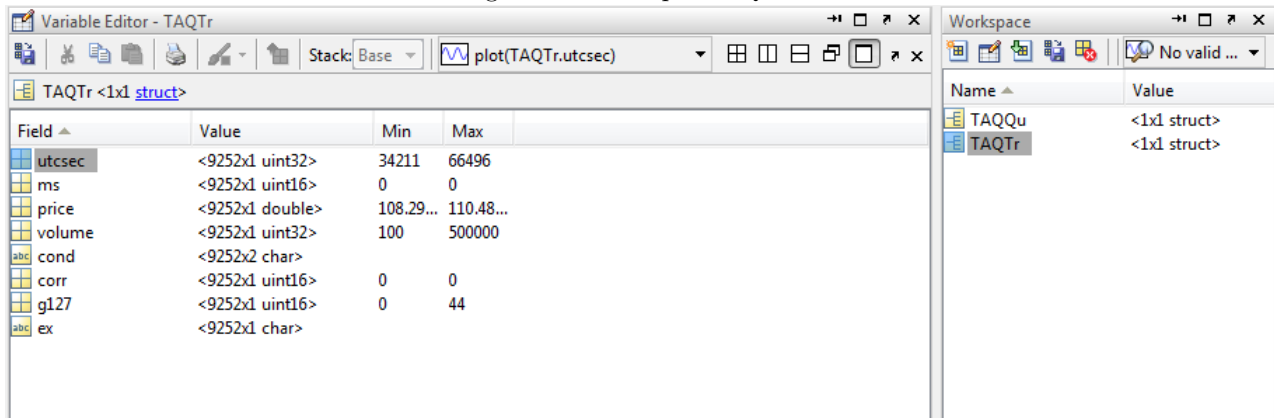
Figure 7 – Workspace - Quotes

Table 3 – TAQ Variables - Quotes

utcsec	Number of seconds (measured from EST)
ms	Milliseconds (Only valid for data from 2015)
bid	Bid price
ofr	Offer price
bidsize	Bid size in number of round lots (Counted in 100s)
ofrsize	Offer size in number of round lots (Counted in 100s)
mode	Quote condition
ex	Exchange on which the quote occurred (See Table (4))

Table 4 – Exchanges and abbreviations

Exchange name	Abbreviation
NYSE	N
Boston	B
Arca	P
INSX	C
NASDAQ	T
NASD ADF	D
Philadelphia	X
ISE	I
Chicago	M
CBOE	W
BATS	Z

Troubleshooting

1. Are you currently connected by wire or VPN to AU's network?

2. Have you properly copied the template files to your own computer?

3. Have you added the correct folders, with the MATLAB files containing functions, to your MATLAB Path?

4. Have you correctly defined the location of the HDF5 TAQ Data?

5. Is the ticker you are using correct?

6. Is the date provided a trading day on the exchanges? `http://www.nyx.com/en/holidays-and-hours/nyse`