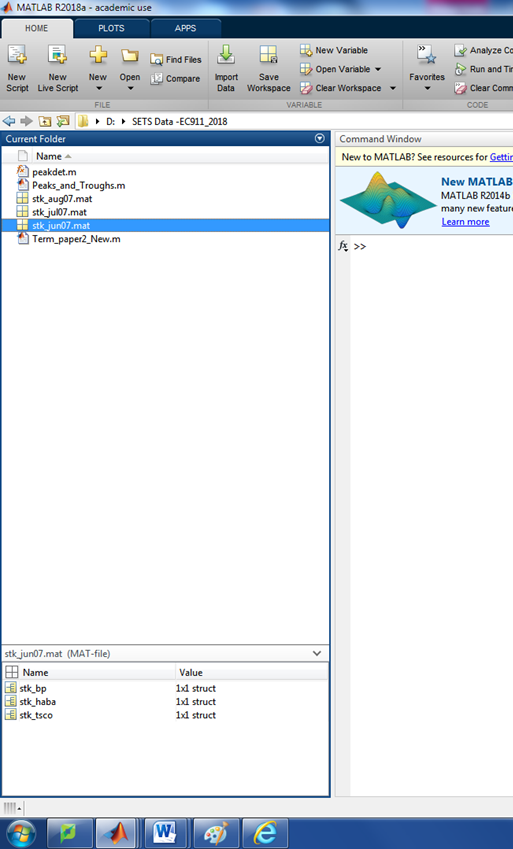
**Notes for Variables to be downloaded From MATLAB program on Shape of Order Book**

These notes shows the data, the main codes and output from MATLAB program needed in order to answer questions 2.1a and 2.1b in the Term Paper on Limited Order Book, EC911.

1. **Some Background information of the SETS Matlab Data and Code**

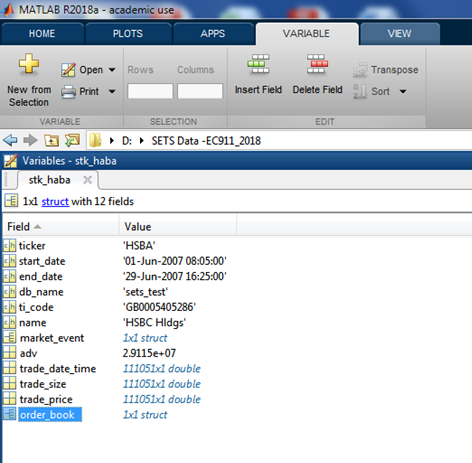
The SETS Matlab data is given for 3 stocks, BP (***stk\_bp***), HSBC (***stk\_haba***) and Tesco (***stk\_tsco***), over 3 months June (***stk\_jun07***), July (***stk\_jul07***) and August (***stk\_aug07***). As shown in Figure 1, by double-clicking each month we get the data on the three stocks on that month.

Figure 1: The three stocks over the 3 months (June-August 2007)



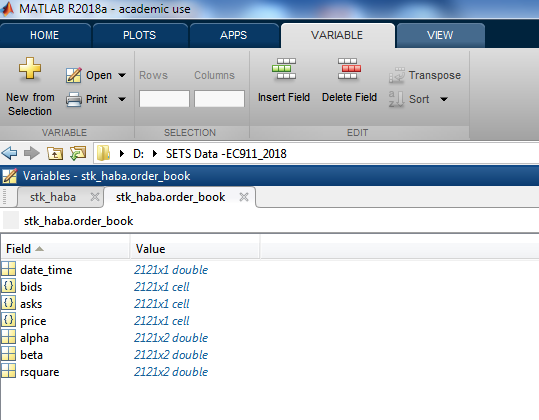
To examine the data set for a stock for a month (for example, ***stk\_haba*** for June 2007) we double-click on ***stk\_haba***(in Figure 1). This opens a window illustrated by Figure 2.

Figure 2: The HBSC Stock Data in *stk\_haba* folder for June 2007



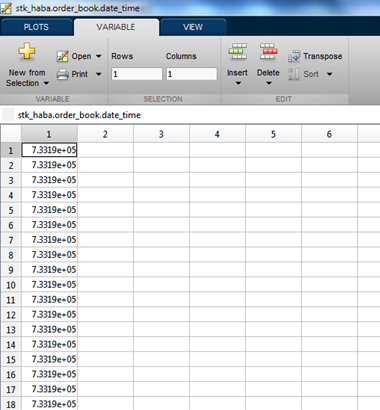
However, for our analysis we focus on the data in the Order\_book. To open Order book we double-click on ***order\_book***(Figure 2), and open the window shown in Figure 3,

Figure 3: The HBSC Stock Data in *order\_book* folder for June 2007



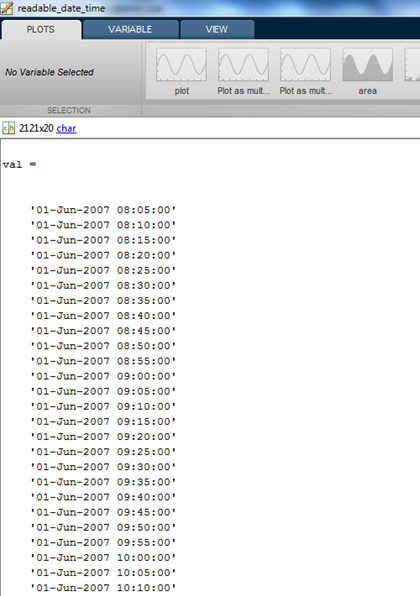
The first thing to note is that the data is organized according a time stamp Date, hour, minute, second saved in ***stk\_haba.order\_book.date\_time***. However, in this folder the ***date\_time*** is shown in “Matlab format” as we can see in Figure 4

Figure 4: The June 2007 Date Time in “Matlab Format”



To convert it into, a “readable data” we use the command ***readable\_date\_time = datestr(stk\_haba.order\_book.date\_time)***. This yields the date time presented in Figure 5

Figure 5: The June 2007 Date Time in “Readable Format”



In folder called bids and asks (Figure 3) you get a files of date organized as a tuple, shown for example in Figure 6. The first entry (***’01-Jun-2007 08:05:00’***) say ***104x 2 double*** which means that the order book in the bid side at that time stamp has 104 tuples given as first col : price and 2nd column as the quantity as shown in Figure 7.

Figure 6: The Data on Bid Side in 5 minutes time stamp in June 2007

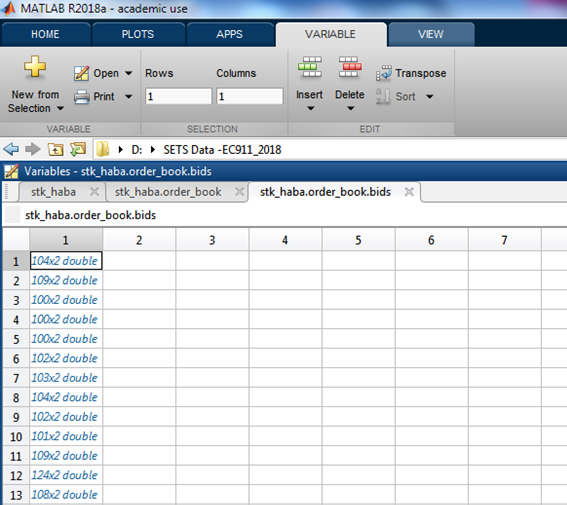
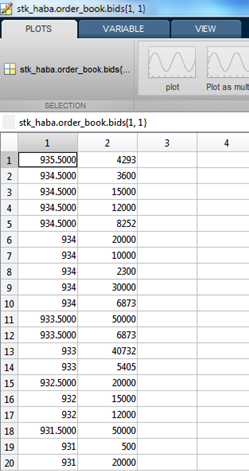
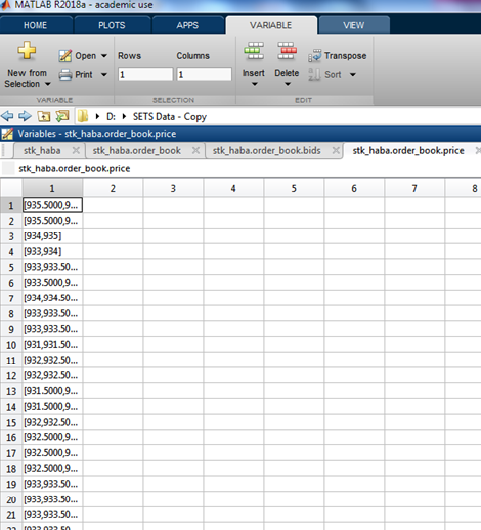


Figure 7: The Data on Bid Side (price and quantity) at 01-June-2007 08:05:00 time stamp in June 2007



Under price (in Figure 3) for each time stamp you get the best bid and the best ask. For example in the first entry we have 935.5000 and 936.

Figure 8: The Bid and Asks Best Price in 5 minutes time stamp for June 2007



The remaining data in the Matlab Data folder for a stock’s monthly data (Figure 3), gives you’re the results of regression analysis of the so called Notional VWAP Curves specified in Markose-Malik. At the end of the next section we show how its generated and its use.

1. **The Main Codes and Output from MATLAB Program Needed to for the Term Paper on Limited Order Book, EC911.**

In the data analysis section of the papers, question 2.1a states: “***Extract the best bids and asks for three stocks, HSBC, Glaxo and Tesco (one stock is allocated to each student) and plot on a 5 minutely basis for 3 months and give sample statistics***”. For this question we need:

* a time series of the best bids and asks prices,
* the mid-price,
* return

%% Calculate and Plot mid-price and Stock returns for August %% \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

k = stk\_haba.order\_book.date\_time;

The best bids and ask prices are given as “price” in the subfolder “order\_book” located in the folder stk\_haba. To call for the prices we use the following command

price =cell2mat(stk\_haba.order\_book.price); -

Then the mid-price and the price return are calculated, respectively, as

%% mid price calculation

mid\_price=mean(price,2);

%calculate return

r=[0; diff(log(mid\_price))];

Now we can plot the mid-price and return and calculate the sample statistics. This can be done in MATLAB or in Excel.

Question 2.1b states: “***Build the demand/bid and supply ask curves using the Malik-Markose method Pick 3 days and plot the 5 minutely best bids and asks for these three days. Discuss, the choice of specific peaks and troughs at which the price trend detection analysis will be conducted. Specify the conditions for a local price rise and those for a price fall. Backtest your results against the actual data that you gave for the 3 days***”.

To answer this question we start by identifying the 3 days to be analysed. Then:

* Extract the time series of the best bids and asks for that day;
* Calculate the mid-price
* Calculate the price return and cumulative price return
* Identify the peaks and troughs
* Estimate and plot the bid and ask side NVWAP curves, and
* Estimate the slope of the NVWAP to capture the steepening/flattening effect of the curves

This can be obtained using the following commands/block of command%% Calculate and Plot Stock returns 3 days and Find peaks and Troughs

%% \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

startDate = datenum('2007-08-01 8:05:00');

endDate=datenum('2007-08-31 16:25:00');

k = stk\_haba.order\_book.date\_time;

Note that 5 minutes best bid and asks for August have 2222 observations. If we count these observations, we find that the best bid and asks for 14th August (from 08:05 to 4:25) corresponds to the observations from 910 to 1010. So to get the best bid and best ask price for this day we use the code.

Day\_14\_sample = datestr(k(910:1010),'HH:MM');

price\_Day\_14 = cell2mat(stk\_haba.order\_book.price(910:1010));

Then the mid-price is calculated as

mid\_price\_Day\_14 =mean(price\_Day\_14,2);

Then the return (r\_Day\_14) is calculated

r\_Day\_14=[diff(log(mid\_price\_Day\_14))];

And the cumulative return as

cum\_r\_Day\_14 = cumsum(r\_Day\_14);

**FINDING THE PEAKS AND TROUGHS IN THE CUMULATIVE RETURN**

The Peaks and troughs of the cumulative return are identified by the command

%% find the indices of peaks and troughs that are at least r=25bps in return

[maxtab14, mintab14] = peakdet(cum\_r\_Day\_14,0.0025);

Here MAXTAB gives the peaks and MINITAB the troughs (a pair, the number of the observation and the respective return value). Then, we need identify the exact time stamp for each observation number of that day. We can do it in excel. But can do it straight away in MATLAB. Then the cumulative return with peaks and troughs are plotted using the following command.

figure

%subplot(221)

hold on; plot(mintab14(:,1), mintab14(:,2), 'g\*');

plot(maxtab14(:,1), maxtab14(:,2), 'r\*' );

plot(cum\_r\_Day\_14,'b','lineWidth',2);

xlabel('Time')

ylabel ('Cumulative Return(%)')

legend('Trough','Peak', 'Cumulative Return','Location','Northwest')

title('Peaks and Troughs in 14th August 2007')

xdate = Day\_14\_sample;

datetick('x','HHPM')

The same can be applied for full month as

startDate = datenum('2007-08-01 8:05:00');

endDate=datenum('2007-08-01 16:25:00');

k = stk\_haba.order\_book.date\_time;

Day\_14\_sample = datestr(k(910:1010),'HH:MM')

%First5dayaug07 = datestr(k(1:505),'DD:HH:MM');

price =cell2mat(stk\_haba.order\_book.price);

price\_Day\_14 = cell2mat(stk\_haba.order\_book.price(910:1010));

%% mid-price calculation

mid\_price=mean(price,2);

%calculate cumulative return

r=[0; diff(log(mid\_price))];

cum\_r = cumsum(r);

[maxtab, mintab] = peakdet(cum\_r,0.0025);

figure

%subplot(221)

hold on; plot(mintab(:,1), mintab(:,2), 'g\*');

plot(maxtab(:,1), maxtab(:,2), 'r\*' );

plot(cum\_r,'b','lineWidth',2);

xlabel('Time')

ylabel ('Cumulative Return(%)')

legend('Trough','Peak','Cumulative Return','Location','Northwest')

title('Peaks and Troughs in 14th August 2007')

xdate = Day\_14\_sample;

datetick('x','HHPM')

**ESTIMATION OF THE BIDS AND ASK SID NVWAP CURVES**

Once we find the peaks and troughs and the exact time stamp of the day, we calculate the bids and ask side NVWAP curves. As the



We need the bids and ask prices and the respective volume to each time stamp. This

%% \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 14th August 2007\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*%%

%% Time Stamp 1 14th August 2007

%% BIDS SIDE

t1=datenum('2007-08-14 11:10:00');

idx1=find(stk\_haba.order\_book.date\_time>=t1,1,'first');

bids1=stk\_haba.order\_book.bids{947};

**tval1=cumsum(bids1(:,1).\*bids1(:,2));**

**tvol1=cumsum(bids1(:,2))**

**NVWAP1=tval1./tvol1;**

figure

subplot(421)

plot(tvol1,NVWAP1, 'r','lineWidth',2);

hold

%% ASKS SIDE

asks1=stk\_haba.order\_book.asks{947};

**tval1=cumsum(asks1(:,1).\*asks1(:,2));**

**tvol1=cumsum(asks1(:,2));**

**NVWAP1=tval1./tvol1;**

subplot(421)

plot(tvol1,NVWAP1, 'b','lineWidth',2);

xlabel('Cumulative Volume')

ylabel ('NVWAP')

legend('birds','asks','Location','Northwest')

title('Start Downtrend')

**REGRESSION OF NOTIONAL VWAP CURVES**

To estimate the slope (β1) the NVWAP curve that capture steepening/flattening effect of the curves, we run a regression in equation 3.7 pag. 6 of the Inacio Term Paper such that

%% Find the Index of the timestamp

stock=stk\_haba;

order\_book = stk\_haba.order\_book;

n=length(order\_book.date\_time);

adv=stock.adv;

b0=NaN(n,2);

b1=NaN(n,2);

rsquare=NaN(n,2);

q=NaN(n,2);

min\_pos=5;

for k = 1:length(order\_book.date\_time);

best\_limit = order\_book.price{k};

if numel(best\_limit) ~= 0 && best\_limit(2) > best\_limit(1)

bid = order\_book.bids{k};

ask = order\_book.asks{k};

date\_time\_vec = datevec(order\_book.date\_time(k,:));

time = date\_time\_vec(4) + date\_time\_vec(5)/60;

if (~isempty(bid) && ~isempty(ask) && time >= (8+1/60) && time <= 16.5)

bids\_cumvol = cumsum(bid(:,2));

bids\_cumval = cumsum(bid(:,1) .\* bid(:,2));

bids\_nvwap = bids\_cumval ./ bids\_cumvol;

asks\_cumvol = cumsum(ask(:,2));

asks\_cumval = cumsum(ask(:,1) .\* ask(:,2));

asks\_nvwap = asks\_cumval ./ asks\_cumvol;

q(k,:) = [sum(bid(:,2)) sum(ask(:,2))];

if length(best\_limit)==2

x = bids\_cumvol./adv\*100;

y = (bids\_nvwap-mean(best\_limit))/mean(best\_limit)\*10000;

if length(x) >= min\_pos

try

[b,bint,r,rint,gof] = regress(log(-y),[ones(size(x)) (x)]);

b0(k,1) = b(1);

b1(k,1) = b(2);

rsquare(k,1) = gof(1);

catch ME

ME.message;

end

end

x = asks\_cumvol./adv\*100;

y = (asks\_nvwap-mean(best\_limit))/mean(best\_limit)\*10000;

if length(x) >= min\_pos

try

[b,bint,r,rint,gof] = regress(log(y),[ones(size(x)) (x)]);

b0(k,2) = b(1);

b1(k,2) = b(2);

rsquare(k,2) = gof(1);

catch ME

ME.message;

end

end

end

end

end

end

return

This code will gives you time series of the intercepts (b0), the bid and ask side NVWAP slopes (b1) and total volume available on both sides of limit order book (q). Then we calculate the changes in the bids and ask side NVWAP slopes and changes in the both side volumes between the peaks and troughs. This calculation can be made in excel file and can yield a table below.

Hypothesis:

Uptrend 14 August 2007. Evaluate the start point (trough) and the end point (peak) of the length of time when cumulative returns increased continuously without falling by more than 25 bps:

The bid side is lengthening (0.06 = dlogQtyBid > 0 check first row of Table ) and slope of bid NVWAP curve is flattening (- 0.16 , dlog b <0)

The ask side is shortening (-0.19 = dlog QTyAsk <0 ) and slope of ask NWAP curve is steepening (0.19 = dlog b >0)

