3rd set SAS assignments

Another help site for sas:

`http://en.wikiversity.org/wiki/Data_Analysis_using_the_SAS_Language`

Application of the Glosten/Harris (1988) model

1. Take the dataset `aby_t_ass3` and `aby_ass3` and read them in as temporary SAS datasets. `aby_t_ass3` contains transaction data of a stock traded on the New York Stock Exchange and `aby_ass3` contains the corresponding quote data. Open the datasets and make yourself familiar with their structure.

2a. Plot the transaction price in a SAS graph. Label your axis and interpret the graph.
2b. Create descriptive statistics for both datasets. Use `proc means` to get the average, minimum, maximum and standard deviation for the bid and ask price in the quote dataset and the transaction price in the trade data set. Use the `ODS pdf` statement and the `proc print` procedure (see SAS help) to write the table containing the decriptives into a pdf file in your home folder.

3. Use the date and time variables to create a datetime variable in each dataset. Hint: Use the `dhms` function. Format the datetime variable. In the transaction dataset substract 5 seconds from the timestamp (datetime variable) of each transaction in order to incorporate reporting delays. Create a trade indicator in the transaction data set (simply generate a new variable that equals "T").
In the quote dataset calculate the midquote (name it `midquote`). Keep only the datetime variable and the midquote in the dataset.

4. Use `proc sql` to merge both datasets and keep all (!) observations. Have a look at the dataset you created.

5. We need to match the prevailing midquote to each transaction. Therefore use the following

commands in a data step:

```
retain prev_midq;
if midquote ne . THEN do;
prev_midq=midquote;
end;
```

Try to figure out what these statements do.

6. Use the indicator for a trade and keep only the timestamps when there is a transaction in the dataset (Hint: Use a data step and a `where` statement).

7. Create an indicator for the direction of trade that equals 1 for a buy and -1 for a sell. Use the same procedure as in the last assignment, i.e. first compare the transaction price to the midquotes, check how many trades are classified. Then use the tick test for those trades that have not yet been classified as buy or sells (Delete those trades that have not been classified after having used the tick test and going back three lags).

8. The Glosten and Harris model is based on the following equation

$$P_t - P_{t-1} = c_0(Q_t - Q_{t-1}) + c_1(Q_t V_t - Q_{t-1}V_{t-1}) + z_0 Q_t + z_1 Q_t V_t + e_t$$

where Q gives the trading indicator, V the trading volume in number of shares traded, and P is the transaction price.

i) Generate the missing variables and use `proc reg` to estimate the parameters in the equation above by OLS.

ii) Merge the estimated parameters back to the dataset (Hint: Use a data step and the `merge` statement instead of the `set` statement and then list the datasets you want to merge. You also need the `by` statement (the by variable is the indicator from Task 3)).

iii) Then create the following variables:

Implied spread

$$2(C_t + Z_t) = 2(c_0 + c_1 V_t) + 2(z_0 + z_1 V_t) \tag{1}$$

Share of implied spread attributable to adverse selection costs

$$\alpha = \frac{z_0 + z_1 V_t}{z_0 + z_1 V_t + c_0 + c_1 V_t} \tag{2}$$

Share of implied spread attributable to order processing costs

$$\gamma = \frac{c_0 + c_1 V_t}{z_0 + z_1 V_t + c_0 + c_1 V_t} \tag{3}$$

iv) Finally, calculate the effective spread from the data directly (see last assignment) and calculate the mean for all four variables. Interpret the results.

/* Additional Task:*/

Turn the whole program into a macro, which is flexible concerning the input data sets and the final output data set in 8iv). Call you macro using the data for the second stock provided on the course homepage.