Group Project 1

This file uses the same data on bid-ask spreads (tradecost) and candidate determinants used in lab2 HW2.

We are interested in the relation between the average bid-ask spreads on stocks and the characteristics of the corresponding companies.

The data file lab2.xls contains information for the 100 stocks in the S$\&$P 100 index.

Our variable of interest (the Y variable) is the bid-ask spread (constructed as an average over the day) - or tradecost - of the

S\&P100 stocks. The explanatory, or X, variables are:

1. log volatility - The log of the daily return standard deviation
2. log size - The log of the size of the stock. Size is total outstanding number of shares multiplied by share price. Size is measured in thousands of dollars
3. log trades - This is the log of the average number of trades per day
4. log turn - This is the log of the ratio of the average number of shares traded per day (in dollars) over the number of shares outstanding (in dollars
5. NumberAnalysts - This is the number of analysts following the stock

BACKGROUND INFORMATION

Consider the following theories of the determinants of bid-ask spreads.

1. {Asymmetric information}. Stocks with greater degrees of asymmetry in information (regarding their fundamental value) tend to have wider bid-ask spreads. The number of analysts following a stock is viewed as an asymmetric information proxy. The larger it is, the lower private information, the smaller the spreads. Log turn-over is, also, seen as an asymmetric information proxy. The larger it is, the larger private information, the larger the spreads. Trading rates in excess of this proportion should be associated with informed trading.)

2. {Liquidity}. Stocks that trade more frequently and have larger market capitalization (i.e., more liquid stocks) tend to have lower bid-ask spreads. The larger log trades and log size, the larger liquidity, the smaller the spreads. Log turn-over is, also, sometimes seen as a liquidity proxy. The larger it is, the larger liquidity, the smaller the spreads.

3. {Fundamental volatility}. Stocks that have a higher volatility of fundamental values tend to have larger bid-ask spreads. Higher uncertainty about the underlying stock's value implies higher potential for adverse price moves and, hence, higher inventory risk, mostly in the presence of large imbalances to offset.

The greater the degree of asymmetric information, the wider the spreads should be as the market makers (who are not fully informed) charge a higher price when selling (raising the ask) and a lower price when buying(lowering the bid) to insulate themselves from losing money to potentially etter informed traders. We do not get to see the degree of asymmetric information between market participants but we do know the following:

1) The larger the number of analysts following a stock, the lower asymmetric information we expect to be for that stock. Analysts provide informationabout the stock, thereby uncovering its fundamental value.

2) The greater the turn-over of the stock, the higher we expect asymmetric information to be for that stock. Intuitively, the greater the turnover, the faster individuals are getting in and out of investment positions in the stock. They do this more when they believe the current price does not accurately reflect the fundamental value of the stock - that is, when they believe theypossess asymmetric (superior) information.

As indicated, turn-over is also viewed, by some, as a liquidity proxy.

****

. Import the data set and set up variable names

```
% import dataset lab2.xls (using Import Dataset)

% rename variables
tradecost = lab2{ : ,1};
logvolatility= lab2{ : ,2};
logsize = lab2{ : ,3};
logtrades = lab2{ : ,4};
logturn = lab2{ : ,5};
numberanalysts = lab2{ : ,6};
logtradecost = log(tradecost);
```

****

Run a regression of the log of the bid-ask spread on the 5 explanatory variables.

Q1. Are all independent variables significant at a 1% significance level? Explain.

```
% INSERT THE CODE HERE
x1 = table(logtradecost,logvolatility,logsize,logtrades,logturn,numberanalysts);
lm1 =fitlm(x1,'logtradecost ~ logvolatility + logsize + logtrades + logturn + numberanalysts')
```

```
lm1 =
Linear regression model:
    logtradecost ~ 1 + logvolatility + logsize + logtrades + logturn + numberanalysts

Estimated Coefficients:
                      Estimate        SE         tStat       pValue
                      _________    _________    ________    __________

    (Intercept)        -0.82932      0.44277      -1.873      0.064173
    logvolatility         1.023     0.053245      19.213     2.574e-34
    logsize            -0.14029      0.02391     -5.8673    6.5918e-08
    logtrades          -0.16936     0.035596      -4.758    7.0564e-06
    logturn           -0.098714     0.032502     -3.0371     0.0030911
    numberanalysts    5.044e-05    0.0029135    0.017313       0.98622


Number of observations: 100, Error degrees of freedom: 94
Root Mean Squared Error: 0.142
R-squared: 0.882,  Adjusted R-Squared: 0.876
F-statistic vs. constant model: 141, p-value = 4.16e-42
```

Answer:

****

Q2. Test the assumption that the coefficient associated with logvolatility is equal to 1. If this is the case, how would you interpret hte relation between daily volatility and bid-ask srepads?

```
% INSERT CODE HERE
H = [0 0 0 0 0 0; 0 1 0 0 0 0; 0 0 0 0 0 0; 0 0 0 0 0 0; 0 0 0 0 0 0; 0 0 0 0 0 0];
C = [0; 1; 0; 0; 0; 0];
[p1,F1] = coefTest(lm1,H,C)
```

```
p1 = 0.6668
F1 = 0.1866
```

Answer:

****

****

Q3. Test the assumption that the coefficients associated with logsize and logtrades are equal to each other. Interpret the result.

```
% INSERT CODE HERE
H = [0 0 1 -1 0 0];
[p2,F2] = coefTest(lm1, H)
```

```
p2 = 0.5032
F2 = 0.4516
```

Answer:

****

****

Q5. Would you exclude any variable from the regression? Explain.

```
% INSERT CODE HERE, if necessary
H = [0 0 0 0 0 1];
[p3,F3] = coefTest(lm1, H)
```

```
p3 = 0.9862
F3 = 2.9973e-04
```

Answer: We can exclude numberanalysts since its p-value is very high.

****

Q6. Let us use this model to predict what the spread will look like tomorrow.

For prediction, use a regression which does not include the number of analysts to predict.

Consider a stock which has a log size of 10.5. Suppose that for this stock we expect that for the following day the log turnover will be

-1.1, the log of the number of trades will be 7.6, and the log of the standard deviation will be -3.5.

Predict what the spread will be for this stock tomorrow. (Note that since the regression is run with log spreads you

will have to make a transformation to convert your prediction for the log spread into a prediction for the actual spread).

```
% INSERT CODE HERE
x2 = table(logtradecost,logvolatility,logsize,logtrades,logturn);
lm2 =fitlm(x2,'logtradecost ~ logvolatility + logsize + logtrades + logturn')
```

```
lm2 = 
Linear regression model:
    logtradecost ~ 1 + logvolatility + logsize + logtrades + logturn

Estimated Coefficients:
                     Estimate       SE        tStat       pValue  
                     ________    ________    _______    __________

    (Intercept)      -0.82892     0.43985    -1.8846      0.062546
    logvolatility      1.0231    0.052606     19.449    6.8322e-35
    logsize          -0.14028    0.023775    -5.9002    5.5666e-08
    logtrades         -0.1693    0.035201    -4.8094    5.6751e-06
    logturn         -0.098864    0.031164    -3.1724     0.0020368


Number of observations: 100, Error degrees of freedom: 95
Root Mean Squared Error: 0.141
R-squared: 0.882,  Adjusted R-Squared: 0.878
F-statistic vs. constant model: 178, p-value = 2.92e-43
```

```
x = [-3.5 10.5 7.6 -1.1];
% predicted value given x
y = predict(lm2, x)
```

```
y = -7.0606
```

```
% predicted values of hte regression
y1 = predict(lm2)
```

```
y1 = 100×1
   -7.4734
   -7.6471
   -5.8577
   -7.6804
   -7.3853
   -7.5895
   -7.2114
   -6.9149
   -7.7800
   -7.3886
      ⋮
```

```
% residuals (from the log specfication)
lres = logtradecost - y;
% mean of residuals
z = mean(lres)
```

```
z = -0.3912
```

```
prediction = exp(y)*exp(z)
```

```
prediction = 5.8043e-04
```

```
% compare with mean tradecost
mtradecost = mean(tradecost)
```

```
mtradecost = 6.4303e-04
```

Answer: