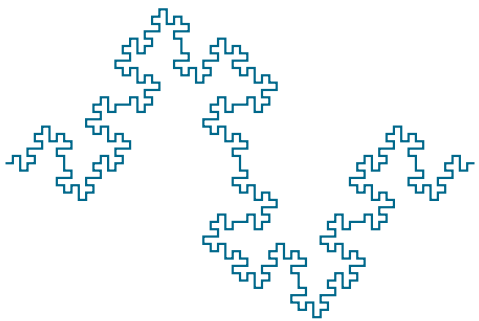
Modelling Global Security Markets

Dustin Johnson
Applied Quantitative Methods



March 17, 2015

What is the Efficient Market Hypothesis (EMH)?

- The debate of market efficiency was started in 1900 by Bachelier who showed that commodity prices follow a "random walk" or "drunkard's walk".
- If a drunk were lost in a vacant field, what would be the most efficient search pattern to find the drunk at some time later?

The EMH essentially claims:

- it is impossible to ”beat the market” because stock market efficiency causes share prices to always incorporate and reflect all relevant information.
- it is impossible for investors to either purchase undervalued stocks or sell stocks for inflated prices.
- it is impossible to outperform the overall market through expert stock selection or market timing.

There are arguments for and against.

- Believers argue it is pointless to search for undervalued stocks or to try to predict trends in the market through either fundamental or technical analysis.
- Warren Buffett has consistently beaten the market over long periods of time, which by definition is impossible according to the EMH.

Is what you are learning pointless? Does market analysis have any justification at all?

Some evidence for either claim:

- Equity markets indeed react quickly to new information
- Authors just "assume... the usual assumptions of the linear regression are satisfied", therefore justifying weak form market efficiency
- Recently, researchers aim to analyze how mistakes in valuation in one market can be transmitted to other nation's security exchanges

HYPOTHESES

H_0: The price movement of a particular stock exchange index is not correlated with any other foreign exchange indices.

H_A: The price movement of a particular stock exchange index is correlated with some other foreign exchange indices (argument against the semi-strong form of the EMH in global security markets).

ACTIVITY PART 1

A research paper makes the following claim:

> *the Hong Kong Stock Exchange (HKSE) is highly vulnerable to major foreign market innovation transmission with the largest influence from the United States.*

Our job:
Regress the one-day lagged Hang Seng Index daily returns to the returns of three other major indices. This means that the HSI column vector will be shifted down one spot relative to the other indices (an additional NA you must deal with). Why conduct regression in this manner? What do we hope to achieve?

RETRIEVE DATA

If you haven't done so already, retrieve the adjusted closing prices of following data and compile into a single data frame:

- Hong Kongs Hang Seng Index (HSI)
- Japans Nikkei Average 225 Index (N225)
- United States S&P 500 Stock Index (GSPC)
- United Kingdoms FTSE 100 Index (FTSE)

You will then convert them to daily rates of return and shift the HSI by lag while keeping other data fixed.

VISUALISE

Once your data has been combined into a neat data.frame, examine the following:

1. summary information of the data (min, max, median, etc)
2. pairs plots for collinearity effects amongst the variables
3. evidence of anything abnormal

THE MODEL

Construct the following multiple regression model in R:

$$HIS = \beta_0 + \beta_1 FTSE + \beta_2 JAP + \beta_3 S\_P + \epsilon$$

Examine your results.

1. Are the coefficients significant?
2. Conduct residual analysis.
3. Interpret results

CROSS-CHECK

Do results coincide with that of the statement made in the research paper?

CONCLUDE PART 1

1. We have now found that HKSE is highly vulnerable to major foreign market innovation transmission

2. However, all the empirical tests so far violate most of the assumptions behind multiple regression analysis.

3. The results still do not tell us when and how to make day-to-day investment decisions even if we turned a blind eye on these requirements.

CAN I USE LOGISTIC?

Can we create a model that will tell us if the Hang Seng Index will close 0.5% or more above the level of the day before?

Perhaps we would also like to be able to make money when the index moves down by shorting the index?

Let's modify the data for the dependent variable into a binary format - 0 for daily return less than 0.5 percent (triggering threshold), and 1 for daily return of 0.5 percent or more.

Use this to support the intuition gained from the regression!

ACTIVITY PART 2

Transform or "mutate" the dependent variable into two separate binary vectors. Call them long and short, as below:

$$HSI_{long} = \begin{cases} 1 & \text{if } R \geq 0.5\% \\ 0 & otherwise \end{cases}$$

$$HSI_{short} = \begin{cases} 1 & \text{if } R \leq -0.5\% \\ 0 & otherwise \end{cases}$$

CONSTRUCT LOGISTIC MODEL

Run the following two models in R, which the first corresponding to a long position, and the second corresponding to a short position.

$$P(HSI_{long} \geq 0.5\%) = \frac{1}{(1 + e^{\beta_0 + \beta_1 FTSE + \beta_2 JAP + \beta_3 S\_P})}$$

$$P(HSI_{short} \leq -0.5\%) = \frac{1}{(1 + e^{\beta_0 + \beta_1 FTSE + \beta_2 JAP + \beta_3 S\_P})}$$

the models predict whether or not the Hang Seng Index will close 0.5% or more above the level of the day before, or contrarily for the short position.

RESULTS

Examine the following for your two logistic regression fits:

1. Are the coefficients significant?
2. Identify the classification error
3. Identify the success rate
4. Interpret the results
5. BONUS: Try to build a "confusion matrix" (matrix of predictions vs observed)

CONCLUSION

Let's answer the following questions in groups!

1. Can we reject H_0?
2. Is there an arbitrage opportunity for a knowledgeable investor to make meaningful profits?
3. Is the semi-strong form of EMH supported in the global security market?

EXTRAS

If you have time, I strongly encourage you to take a deeper look into the following and test them out yourself:

1. Can you identify other issues with this model?
2. You may remember the issue with reusing the training set for predictive purposes. Conduct a cross-validation routine to determine the true classification prediction error. Would your results change?
3. Can you improve on the results of this paper?