Title: **Quantifying Trading Behavior in Financial Markets Using Google Trends**

Authors: Tobias Preis, Helen Susannah Moat, and H. Eugene Stanley

Commentator: Sangno Lee

The purpose of this comment is not to criticize the paper, but to clearly understand their findings. They found that Google Trends data can be a proxy for investors’ sentiments. Using Google trends data, they obtained high returns for some search terms. For instance, ‘debt’ term produces 326% accumulative return from January 1, 2004 to February 22, 2011. We were so surprised to see the remarkably highest returns that we asked Dr. Preis about their data and code. He provided us with their data and code that used for implementing the paper. We are deeply appreciating their supports. Thanks to their help, we are able to figure out some mistakes in the paper. We would like to describe them in terms of implementation and logic issues.

**A. Implementation Issues**

They implemented the code with R language. The paper was silent when the current volume of search is the same with the average of the last three weeks. From the source code, we found that they just copy the previous week return (line 43-44). *Value* variable stands for the difference between the current volume and the average of volume of the previous three weeks. In the trading algorithm (line 75-82), if *Value* variable is greater than zero, they take short position, and if *Value* variable is lower than zero, they take long position.

For instance, from 2004-08-22 to 2004-08-22, the search volume of ‘debt’ term is 0.18, and the previous three weeks data are 0.19, 0.173333, and 0.176667, respectively. If we calculate the average for the last three week data, we have (0.19 + 0.173333 + 0.176667) / 3 = 0.1799999, almost 0.18. Thus, if we take 0.18 for the average value, *Value* is 0, but if we take 0.17999999, *Value* is not zero. The value has a very small positive (0.18 – 0.179999) due to *double* precision presumed in R. Although the difference is very small, because of double precision issue, the trading will take short position. There were 15 cases for zero value in the data. When we take very small difference as zero, we had *297*% cumulated return for ‘debt’ term (i.e. value<-round((now-previous), digits=6) line 61).

Another issue for implementing the trading algorithms is that they mismatched trading strategy and obtained returns. For instance, let’s look at the first few lines of the data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No | Goog\_SD | Goog\_ED | debt | DJIADate | DJIA | Return |
| 1 | 1/4/2004 | 1/10/2004 | 0.21 | 1/12/2004 | 10485.18 | 1 |
| 2 | 1/11/2004 | 1/17/2004 | 0.21 | 1/20/2004 | 10528.66 | 1 |
| 3 | 1/18/2004 | 1/24/2004 | 0.21 | 1/26/2004 | 10702.51 | 1 |
| 4 | 1/25/2004 | 1/31/2004 | 0.213333 | 2/2/2004 | 10499.18 | 0.992452049 |
| 5 | 2/1/2004 | 2/7/2004 | 0.2 | 2/9/2004 | 10579.03 | 1.005196564 |

When No = 4, Value has 0.00333 (0.021333 – [(0.21 + 0.21 + 0.21) / 3]). Because *Value* is positive, trading strategy takes short position and the return of trading is 0.992 (1 / (10579.03 / 10499.18)). However, the match of Google search volume and DJIA is not appropriate. The paper describes that the trading algorithm takes the following logic.

1. At Sunday night (Jan 25, 2004), an investor figures out whether the search volume of ‘debt’ term increases or decease. In this case, an investor found the search volume of ‘debt’ term increase by 0.00333.
2. Then, the investor set a plan to take short position.
3. On Monday (Jan 26, 2004), the investor borrows DJIA ETF from a broker and sells them on stock market.
4. On the next week Monday (Feb 2, 2004), the investor buys DJIA ETF and returns them to the broker.
5. Then, the investor only knows the outcome from short position on Feb 2, 2004.
6. Thus, the code has to record the return of 0.9924 at the *fifth* row, not *fourth* row.

Although the final outcome is the same with the recoding in the current row in this case, this implementation produces another problem in the logic issues.

**B. Logic Issue**

Let return *r*, stock price *p*, time *t*. The return in a period can be defined as *total return* and *rate of return*.

Total Return: *R* = *p*(t+1) / *p*(t). (1)

Rate of Return: *r* = (*p*(t+1) – *p*(t)) / *p*(t). (2)

They use total return for obtaining compounding return during the sample period (Jan 2004 – Feb 2011). Note that 1 + *r* = *R*, and *p*(t+1) = (1+*r*)*p*(t).

And for short and long position, they calculate accumulative return (AR) as

AR = *R*(t-1) / [*p*(t+1) / *p*(t)] for short position (line 78) (3)

AR = *R*(t-1) × [*p*(t+1) / *p*(t)] for long position (line 82). (4)

I think that the formula for calculating the accumulative return for short position has a problem. Alternatively, Equation (3) can be expressed as AR = *R*(t-1) × [*p*(t) / *p*(t+1)]. In this case, Total Return, *R*, changes as *R* = *p*(t) / *p*(t+1). Then, the role of input and output changed.

For example, suppose that an investor makes short selling $5 at *t*, and the stock prices $4 at *t+1*. According to Equation (2), Rate of Return from this short position is -[(4 – 5) / 5 ]= 20%, and Total Return 120% (i.e. 1 + *r* = *R*). But the total return from Equation (3) is 5/4 = 125%, and the rate of return is 25%. As the stock price falls sharply at *t+1*, total return from Equation (3) is larger. For example, suppose that the stock price at *t+1* becomes $1. In this case, Total Return and Rate of Return from Equation (3) is 5/1 = 500%, 400%, respectively. However, Rate of Return and Total Return from Equation (1) and (2) is -(1-5)/5 = 80%, 180%, respectively.

The calculation for the return from short position should keep [*p*(t+1) / *p*(t)] formula by adding negative sign. Regardless short or long position, the absolution value of return from short and long position is the same but only the sign is reverse. There are two reasons for keeping *p*(t+1) / *p*(t) formula. First, when stock price is falling, if we use Equation (3), short position produces higher return than long position. For instance, suppose that a stock price falls from $5 to $1. From a long position, an investor losses 80%, but from a short position, another investor gains 400%. Thus, the return from short and long is not symmetric and this short position cannot be used for hedging risks. Second, in terms of input and output, *p*(t) should be denominator for calculating return. To short, investors need a margin account (i.e. borrowing). For instance, Charles Schwab requires investors to keep a minimum of $5,000 to trade on margin. When an investor shorts a stock, he borrows a stock from a broker and sells it. Then, investors need to buy back the stock for the future. Before investors buy back the stock, they need to keep cash in their accounts for the case of the stock price rise. If the stock rises but investors think it will drop and keep short position, they have to add more money to their account to cover greater potential losses. Thus, *p*(t) should be used as denominator for calculating returns of short selling. By changing the return from short position, I obtained only 6% profit using ‘debt’ term.

I think this result comes from inaccurate implementation of return calculation in the paper. They described return from short and long position as follows (p. 2).

“If we take a short position – selling at the closing price p(t) and buying back at price p(t+1) – then the cumulative return *R* changes by log(*p*(t)) – log(*p*(t+1)). If we take a long position – buying at the closing price *p*(t) and selling at price *p*(t+1) – then the cumulative return R changes by log(*p*(t+1)) – log(*p*(t)).”

If we take logarithms return instead of simple return, the return from short position is the negative return from long position. There is no the denominator issue in calculating logarithms return. They might implement log calculation with division of simple return without considering the denominator issue. In the logarithms return, accumulative return is simply sum of all returns.

**Conclusion**

When evaluating performance from investments, we need to use proper concept of return and carefully implement it. Given that there are many concepts of returns, including compound and simple interest returns, arithmetic and geometric return, and dollar-weight and time-weight return, we need to very cautious to use and implement those concepts.