Tone of MD&A and Its Impact on Firm Achievements

Team members:

516120910137 尤姝婉;

516120910151 黄海昀;

516120910155 齐月;

516120910160 赵依婷;

516120910161 朱星宇

目录

Chapter 1　Background

1. problem

Text information is an important part of a company's disclosure, providing information to the investors, helping them get a more comprehensive assessment of the company's performance by describing the company's activities.

Text is different from data as the information it contains is often difficult to quantify and be expressed accurately.

2. aim of the project and our assumption

Our project aims to research on the emotional tone of information contained in the MD&A part of quarter report and its relationship between the performance of listed companies and to　predict the firm's future performance with available information and data.

Based on the text data form the American listed companies, In this project, we use the tone of 10-Q form MD&A part of the current period and some other data(will be talked about in detail in the following parts) to predict the ROE of the next period and see whether it will work.

In our project, we worked on three giant firms: Apple, Ford Motor and Coca-Cola. The reason why we chose these 3 companies is that they are from 3 different industries so that we can get a better view of how well our idea works on an overall basis.

We assume that the tone of MD&A has a positive relationship with the ROE of the next period; which is to say, the larger the value of tone, the more likely company will do better in the next period.

Chapter 2 Approaches

1. data collecting

We use the dictionary-based word segmentation method (also called "bag of words method") to calculate the number of words that can represent the tone of the text. There are two important elements of this method: dictionary and matching principle. The basic idea is: if a word in text is found in the dictionary, the word is identified and the program takes it into count. The advantage of the method is that the process is very easy to achieve. But the tricky part is that we have to find an appropriate dictionary to identify the words we want in the text.

In our project, we apply the Loughran and McDonald dictionary to find the words with special tones. The Loughran and McDonald dictionary, also known as the LMD dictionary, which contains the analysis of tones of millions of words, specially applied in the field of business analysis. Some of the "positive" and "negative" words are listed below.

127	ABLE	11	ABANDON
336	ABUNDANCE	12	ABANDONED
338	ABUNDANT	13	ABANDONING
437	ACCLAIMED	14	ABANDONMENT
476	ACCOMPLISH	15	ABANDONMENTS
477	ACCOMPLISHED	16	ABANDONS
478	ACCOMPLISHES	53	ABDICATED
479	ACCOMPLISHING	54	ABDICATES
480	ACCOMPLISHMENT	55	ABDICATING
481	ACCOMPLISHMENTS	56	ABDICATION
619	ACHIEVE	57	ABDICATIONS
620	ACHIEVED	72	ABERRANT
621	ACHIEVEMENT	73	ABERRATION
622	ACHIEVEMENTS	74	ABERRATIONAL
625	ACHIEVES	75	ABERRATIONS
626	ACHIEVING	81	ABETTING
900	ADEQUATELY	140	ABNORMAL
1130	ADVANCEMENT	141	ABNORMALITIES
1131	ADVANCEMENTS	142	ABNORMALITY
1132	ADVANCES	143	ABNORMALLY
1133	ADVANCING		

We use the code to go through the MD&A part of 10-Q forms of 3 different companies (Apple, Coca-Cola and Ford Motor) in 10 years (10 years*4 quarters = 40 for each company). The reason why we chose these 3 companies is that they are from 3 different industries so that we can get a better view of how well our idea works on an overall basis.

In the code we import the TextBlob package to split the text into separated words and remove the spaces and punctuations between them. We can count the number of positive and negative words in each 10-Q form so that we can calculate the tone of each MD&A part.

To quantify the tone of MD&A parts so that we can apply it in the regression model, we use a classical approach called TF-IDF (term frequency–inverse document frequency) algorithm.

By calculating the weighted proportion of emotional vocabulary, we can get the positive and negative emotional tendencies of the report, and by applying the formula below, we can quantify the tone value of the text.

$$\text{TONE} = \frac{POSTONE - NEGTONE}{POSTONE + NEGTONE}$$

In this formula, POSTONE represents the positive tendencies of the report (=(the number of positive words)/(the total number of words of the text)), while NEGTONE represents the negative tendencies (=(the number of negative words)/(the total number of words of the text)). The value of TONE indicates the overall emotional value of the MD&A part. The larger the value, the more positive the MD&A tends to be.

Now that we have the value of 40 tone values of one company during a 10-year period of time, we save them in 3 Excel forms (each for one company) for the use in the machine learning process.

2. machine learning

First of all, the python program reads the new tone value data using Pandas package. We divide the data into two parts. One is features variables, including company name, year, ROE(T), PE(T) and tone value. The other is ROE(T+1). We split the data into 90% of training data and 10% of testing data. We chose these as feature variables because there is continuity of the ROE in a company, so the ROE in the previous period do matter. And PE ratio partly reflects the growth potential of a firm

4

$$ROE_{iT+1} = \alpha_0 + \alpha_1 TONE_{iT} + \alpha_2 ROE_{iT} + \alpha_3 P/E\ ratio_{iT}$$

Then, we prepared feature matrix X and label vector y, and use holdout validation to split X and y into two parts: training (90%) and testing (10%).

As the value X and Y are prepared, it is time to train a model .We used the algorithms (KNeighborsRegressor,LinearRegressor,DecisionTreeRegressor,RandomForestRegressor) in turn and computed the MSE of each algorithm and compared them to decide which one is most suitable for our project. The one with the smallest MSE can be regarded as the best way. If the MSE is extremely small, it indicates that our assumption proves to be true, that the positive emotional tendency has a positive effect on the firm's performance.

Chapter 3   Results

The results of machine learning prove to be quite favorable. (see the table below)

Apple	KNeighbors Regressor	Linear Regressor	DecisionTree Regressor	RandomForest Regressor
MSE	0.0048403412	0.000869849531182	0.0035967275	0.00229037107856

Ford	KNeighbors Regressor	Linear Regressor	DecisionTree Regressor	RandomForest Regressor
MSE	2.38750786428	0.705674478239	2.825898082	2.0793695911

Coca-cola	KNeighbors Regressor	Linear Regressor	DecisionTree Regressor	RandomForest Regressor
MSE	0.00031257984	0.000216178912362	0.0004017125	0.000376583789722

With regression analysis, for Apple and Coca-Cola, the MSE of linear regression are both smaller than 0.001. The MSE of Ford Motor is slightly larger, about 0.7, but is also favorable, indicating that our project can successfully and precisely predict the future performance of a firm with available data like the current ROE, PE ratio and tone of the MD&A part in the 10-Q form of the current quarter.

Chapter 4　Problems we met & solution/explanation

1. Data NaN?

When working on Ford, we found out that some of their PE ratios were not available due to negative EPS. So in the excel, these data are NaN and couldn't be processed with the program.

To solve this problem, we add a line of code:

$$data = data.fillna(method='ffill')$$

This line of code deals with the unavailable data automatically so that our programs can run smoothly.

2. Why is Ford's result not so good?

As is mentioned in Part 3, the results of machine learning prove to be quite favorable. However, there are still some questions worth thinking. Why is the MSE of Ford Motor slightly larger than the other two companies?

After we checked its historical data, we found a little clue. Here is its historical ROE data, as you can see, the data varied intensely between year 2006-2012.

Dec. 31, 2012	33.63%
Sept. 30, 2012	120.30%
June.30,2012	147.60%
March.31,2012	210.10%
Dec. 31, 2011	360.00%
Sept. 30, 2011	301.10%
June.30,2011	1.99K%
March.31,2011	387.40%
Dec. 31, 2010	169.80%
Sept. 30, 2010	132.80%
June.30,2010	-90.49%
March.31,2010	-61.93%
Dec. 31, 2009	-22.44%
Sept. 30, 2009	37.90%
June.30,2009	55.59%
March.31,2009	272.30%
Dec. 31, 2008	1.11K%
Sept. 30, 2008	566.90%
June.30,2008	575.20%
March.31,2008	143.80%
Dec. 31, 2007	584.30%
Sept. 30, 2007	-2.30K%
June.30,2007	354.90%
March.31,2007	200.20%
Dec. 31, 2006	137.70%
Sept. 30, 2006	-55.82%
June.30,2006	-17.50%
March.31,2006	-8.13%

We don't really understand what happened during these years although we have done a lot of research. Maybe the data on the website contain some mistakes.

Another possible reason accounting for this is that the beta of Ford Motor is relatively large. In finance, the beta (β or beta coefficient) of an investment indicates whether the investment is more or less volatile than the market as a whole. In general, a beta less than 1 indicates that the investment is less volatile than the market, while a beta more than 1 indicates that the investment is more volatile than the market. We compared the beta of Ford and TOYOTA and found that Ford's beta is much larger.
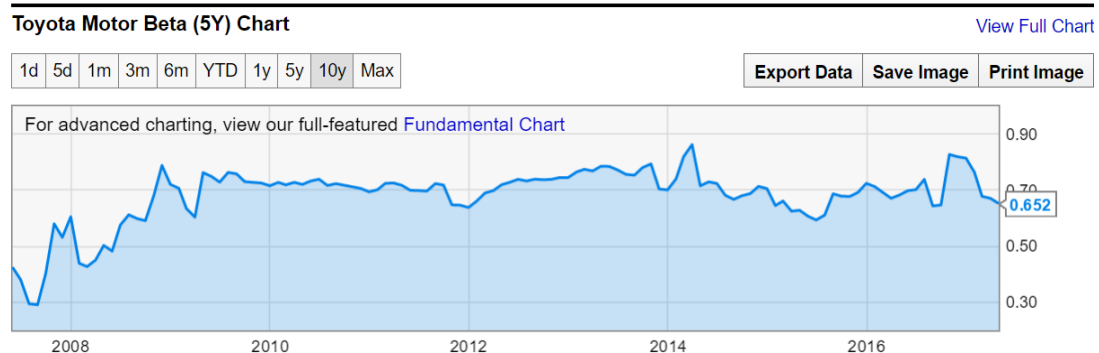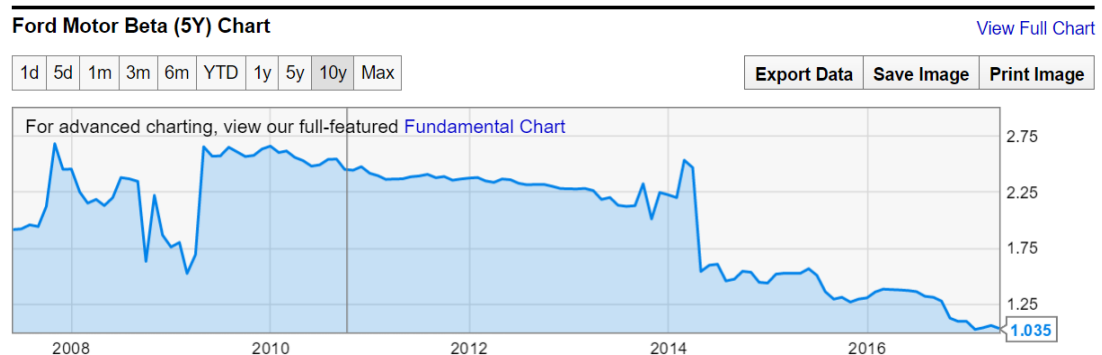
The comparison between Ford Motor Beta and Toyota Motor Beta indicates that the risk of Ford Motor is relatively high, and the investment is rather volatile. The managers' discussion and analysis may not be able to keep pace with the change of the market.

If this is true, then using MD&A to predict the ROE of a company is not applicable for companies with a large beta.

Chapter 5   Reflection & conclusion

There are still a lot of things need to be improved in our project. For example, we copied the 10-Q forms manually to our program, which is low-efficient. Maybe the better way is that we should use some Python crawler technique to get the text we want on the Internet.

Also, we only took 3 feature variables into account, which may affect the precision of the prediction. We can add more variables such as the leverage rate and the industry index to the program to get a more precise prediction.

In conclusion, our project used the text analysis and machine learning technique and succeeded in using the tone of MD&A in 10-Q forms, ROE and PE ratio in the current period to predict firm's future performance, thus providing a new way to investors when considering which firm to invest in.

Appendix   References

【1】 Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, Journal of Accounting and Economics 45, 221-247.

【2】 Li, Feng, 2009, The determinants and information content of the forward-looking statements in corporate filings - a Naïve Bayesian machine learning approach, Working paper, University of Michigan.

【3】 Tetlock, Paul C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, Journal of Finance 63, 1437-1467.

【4】 LOUGHRAN, T., AND B. MCDONALD. "When is a Liability not a Liability? Textual analysis, Dictionaries, and 10-Ks." Journal of Finance 66 (2011): 35-65.

【5】 LOUGHRAN, T., AND B. MCDONALD. Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research, 2016 , 54 (4) : 1187-1230

【6】 谢德仁，林乐，《管理层语调能预示公司未来业绩吗？——基于我国上市公司年度业绩说明会的文本分析》，《会计研究》，2015,2