Tone of MD&A and Its Impact on Firm Achievements

Team members:

黄海昀 尤姝婉 齐月 赵依婷 朱星宇

Team name: One

1 Background

2 Approaches

3 Results

4 Reflection & Conclusion

Background

01

part

Our aim

Text information is an important part of a company's disclosure.

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

BASIC CONCEPT

MD&A

Management discussion and analysis (MD&A) is the section of a company's annual report in which management provides an overview of the previous year's operations and how the company performed financially.

Basic introduction

Positive emotional tendency has a positive effect on the firm’s performance.

?

Approaches

02
part

Approaches

Data collecting

Machine learning

Data collecting

借鉴 Davis et al. （2012）、Tetlock et al. （2008）、Davis and Tran （2012）、陈小悦和徐晓东（2001）等文献，建立了模型（1）－（2）⑧：

$$ROE_{iT+1} = \alpha_0 + \alpha_1 TONE_{iT} + \alpha_2 ROE_{iT} + \alpha_3 DROE_{iT} + \alpha_4 YRET_{iT} + \alpha_5 SIZE_{iT} + \alpha_6 AGE_{iT} + \alpha_7 FSHR_{iT} + \alpha_8 HFD5_{iT} + \alpha_9 MB_{iT} + \alpha_{10} LEV_{iT} + \alpha_{11} MRETSTD_{iT} + \sum IND + \sum YEAR + \varepsilon_{iT} \quad (1)$$

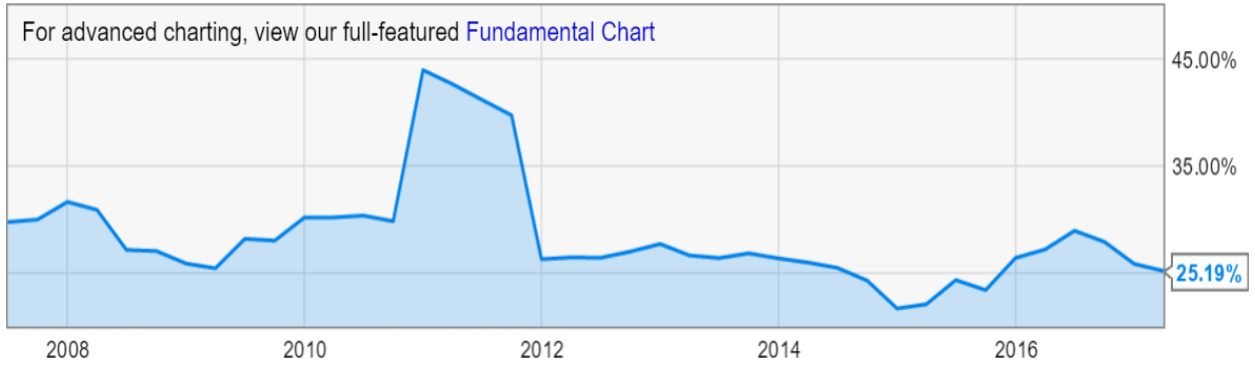
$$ROE_{iT+1} = \alpha_0 + \alpha_1 TONE_{iT} + \alpha_2 ROE_{iT} + \alpha_3 P/E\ ratio_{iT}$$

Reference:
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

Data collecting

Coca-Cola Return on Equity (TTM) Chart View Full Chart

1d | 5d | 1m | 3m | 6m | YTD | 1y | 5y | 10y | Max — Export Data | Save Image | Print Image



Coca-Cola Historical Return on Equity (TTM) Data

Date Range: 03/31/1984 to 03/31/2017 Get Data — Export Data

Viewing 1 of 3 First Prev Next Last

Data for this Date Range

March 31, 2017	25.19%	Dec. 31, 2010	43.95%
Dec. 31, 2016	25.85%	Sept. 30, 2010	29.83%
Sept. 30, 2016	27.91%	June 30, 2010	30.36%
June 30, 2016	28.95%	March 31, 2010	30.19%
March 31, 2016	27.20%	Dec. 31, 2009	30.19%
Dec. 31, 2015	26.41%	Sept. 30, 2009	28.03%
Sept. 30, 2015	23.41%	June 30, 2009	28.20%
June 30, 2015	24.34%	March 31, 2009	25.44%
March 31, 2015	22.08%	Dec. 31, 2008	25.88%

Coca-Cola PE Ratio (TTM): 32.32 for June 2, 2017

View 4,000+ financial data types

Search Add Browse...

Coca-Cola PE Ratio (TTM) Chart View Full Ch

1d | 5d | 1m | 3m | 6m | YTD | 1y | 5y | 10y | Max — Export Data | Save Image | Print Imag

From: https://ycharts.com/

Data collecting

SEC Form 10-Q is the Securities and Exchange Commission’s form for the quarterly performance filings of public companies.

Information in a 10-Q filing includes:

Financial statements,

Management discussion and analysis (MD&A) of any material changes in revenues and expenses in the context of previous quarterly comparisons,

...

SEC 编辑

美国证券交易委员会 Securities and Exchange Commission。缩写为：SEC。根据《1934年证券交易法》于当年成立的美国联邦政府专门委员会，旨在监督证券法规的实施。委员会由五名委员组成，主席每五年更换一次，由美国总统任命。

EDGAR 编辑

本词条缺少名片图，补充相关内容使词条更完整，还能快速升级，赶紧来编辑吧！

EDGAR（Electronic Data Gathering, Analysis, and Retrieval System）,即电子化数据收集、分析及检索系统。1996年，美国SEC规定所有的信息披露义务人（美国上市公司）都必须进行电子化入档。

Data collecting

How to quantify the tone?

-Deciding the tone of certain words in the text:

the Loughran and McDonald dictionary (LMD)

127	ABLE
336	ABUNDANCE
338	ABUNDANT
437	ACCLAIMED
476	ACCOMPLISH
477	ACCOMPLISHED
478	ACCOMPLISHES
479	ACCOMPLISHING
480	ACCOMPLISHMENT
481	ACCOMPLISHMENTS
619	ACHIEVE
620	ACHIEVED
621	ACHIEVEMENT
622	ACHIEVEMENTS
625	ACHIEVES
626	ACHIEVING
900	ADEQUATELY
1130	ADVANCEMENT
1131	ADVANCEMENTS
1132	ADVANCES
1133	ADVANCING

11	ABANDON
12	ABANDONED
13	ABANDONING
14	ABANDONMENT
15	ABANDONMENTS
16	ABANDONS
53	ABDICATED
54	ABDICATES
55	ABDICATING
56	ABDICATION
57	ABDICATIONS
72	ABERRANT
73	ABERRATION
74	ABERRATIONAL
75	ABERRATIONS
81	ABETTING
140	ABNORMAL
141	ABNORMALITIES
142	ABNORMALITY
143	ABNORMALLY

Reference:
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.

Data collecting – codes part 1

```
# -*- coding: utf-8 -*
text = '''
'''
# import textblob
from textblob import TextBlob

textTB = TextBlob(text)
# tokenize text into words
words = textTB.words
```

Data collecting – codes part 2

```
10  f=open("positive.txt","r")
11  ff=open("negative.txt","r")
12  s=0
13  n1=0
14  n2=0
15  positivelist=dict()
16  negativelist=dict()
```

```
18  for line in f:
19      positivelist[line[:-1]]=1
20  for line in ff:
21      negativelist[line[:-1]]=1
22  for word in words:
23      if word in negativelist:
24          n1=n1+1
25  for word in words:
26      if word in positivelist:
27          n2=n2+1
28  print("总字数：",len(words))
29  print("neg:",n1)
30  print("pos:",n2)
31  print("tone:",(n2-n1)/(n2+n1))
32  f.close()
33  ff.close()
```

Data collecting

How to quantify the tone?

TF-IDF to quantify the tone

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

tf-idf [编辑]

维基百科，自由的百科全书

tf-idf（英语：term frequency－inverse document frequency）是一种用于信息检索与文本挖掘的常用加权技术。tf-idf是一种统计方法，用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。tf-idf加权的各种形式常被搜索引擎应用，作为文件与用户查询之间相关程度的度量或评级。除了tf-idf以外，互联网上的搜索引擎还会使用基于链接分析的评级方法，以确定文件在搜索结果中出现的顺序。

Reference:
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.

Data collecting

	A	B	C	D	E	F
1	company	year	ROE(T+1)	ROE(T)	P/E(T)	tone(T)
2	Cocacola	2007, 1	0. 3056	0. 2995	16. 6	0. 12139
3	Cocacola	2007, 2	0. 2976	0. 3056	16. 11	0. 0086
4	Cocacola	2007, 3	0. 2999	0. 2976	17. 52	0. 14793
5	Cocacola	2007, 4	0. 3164	0. 2999	18. 62	0. 15337
6	Cocacola	2008, 1	0. 309	0. 3164	18. 2	0. 03916
7	Cocacola	2008, 2	0. 2716	0. 309	17. 49	0. 10417
8	Cocacola	2008, 3	0. 2705	0. 2716	16. 18	-0. 06667
9	Cocacola	2008, 4	0. 2588	0. 2705	15. 94	-0. 10112
10	Cocacola	2009, 1	0. 2544	0. 2588	14. 25	-0. 12097
11	Cocacola	2009, 2	0. 282	0. 2544	14. 33	-0. 38776
12	Cocacola	2009, 3	0. 2903	0. 282	14. 2	-0. 36402
13	Cocacola	2009, 4	0. 3019	0. 2903	16. 01	-0. 26623
14	Cocacola	2010, 1	0. 3019	0. 3019	15. 78	-0. 16058
15	Cocacola	2010, 2	0. 3036	0. 3019	14. 79	-0. 16098
16	Cocacola	2010, 3	0. 2983	0. 3036	13	-0. 11419
17	Cocacola	2010, 4	0. 4395	0. 2983	14. 96	-0. 05263
18	Cocacola	2011. 1	0. 4265	0. 3019	10. 9	-0. 09325
19	Cocacola	2011. 2	0. 4118	0. 4265	10. 79	-0. 15615

tone-cocacola

Machine learning - codes

```
16  # read tone data
17  data = pd.read_csv('tone-cocacola.csv')
```

```
24  # specify feature columns (i.e., variables that may impact the house price)
25  feature_columns = [col for col in data.columns if col not in ['company','year','ROE(T+1)']]
26  # get feature matrix X
27  X = data[feature_columns]
28  # get label vector y (house price)
29  y = data['ROE(T+1)']
30  # split X and y into training (90%) and testing sets (10%)
31  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=0)
```

```
40  #MSE of knn
41  model_knn.fit(X_train, y_train)
42  y_predict = model_knn.predict(X_test)
43  print(y_predict.tolist()[:10]) # first 10 predicted prices
44  print(y_test.tolist()[:10]) # first 10 true prices
45
46  mse = mean_squared_error(y_test, y_predict)
47  print('Mean Square Error (KNN):',mse)
48  print('Root Mean Square Error (KNN):',mse**0.5)
```

Results

03

part

Mean Square Error

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

Reflection And Conclusion

04
part

Reflection

Why is the MSE of Ford Motor slightly larger than the other two companies?

ROE of Ford Motor,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%

Some errors?

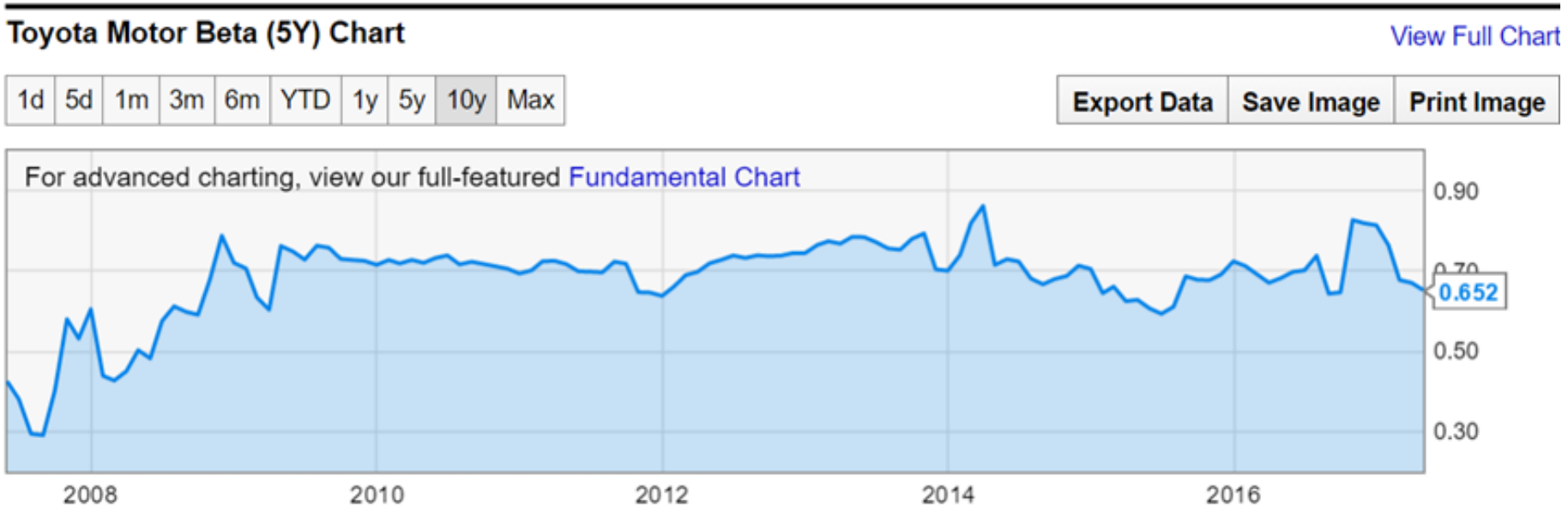
Ford Motor’s Beta

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.

beta < 1:the investment is less volatile than the market

beta >1 indicates that the investment is more volatile than the market.

The comparsion between Ford and Toyota



Conclusion

- text analysis
- machine learning

- using the tone of MD&A in 10-Q forms, ROE and PE ratio in the current period to predict firm’s future performance

- providing a new way to investors when considering which firm to invest in

Division of labor

- Coding:　赵依婷　黄海昀
- PPT making: part 1 尤姝婉; part 2 朱星宇 & 黄海昀; part 3 赵依婷; part 4 齐月
- Report writing: part 1 尤姝婉; part 2 朱星宇 & 黄海昀; part 3 赵依婷; part 4 齐月
- Presentation:　黄海昀

References

- Special thanks to our teacher for your patience and useful suggestions!

- Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, Journal of Accounting and Economics 45, 221-247.
- 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.
- 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.
- 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.
- LOUGHRAN, T., AND B. MCDONALD. Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research, 2016 , 54 (4) : 1187-1230
- 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

Thank you for your attention!