# Measuring the Sentiment indicator for earnings calls

25 Apr, student number: 21090478

# Overview

In this work, we built a sentiment indicator for companies in the FTSE100 and then test its predictive ability.

The work contains the following four parts:

## 1. Prepare data

- In this part, we show the description of our data and how we collected the data.

## 2. Build sentiment indicator

- This part demonstrates how we created the sentiment indicator.
- Besides, this part will discuss the choice of lexicon and the designing of the indicator.

## 3. Test the predictive ability

- We tested the predictive ability of our sentiment indicator via spearman correlation and linear model.

## 4. Conclusion, Limitation and improvements

- In this part, we show the description of our data and how we collected the data.

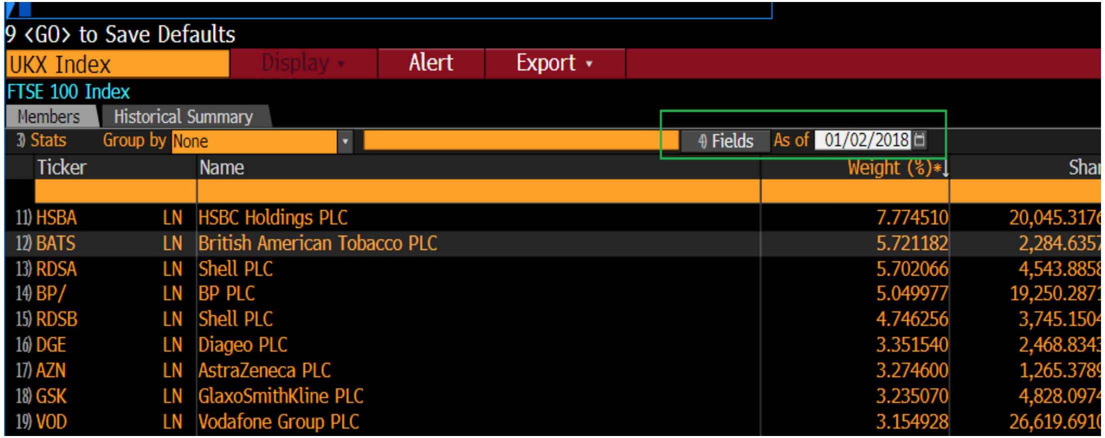
# Prepare Data — Description

- We collected 15 companies' earnings call for building sentiment indicator.
- And then test its predictive ability on fundamentals (the growth rate of the net income, shown as the right chart) .
- We filled the missing value by using average values of the same year.

| | 2018Q2 | 2018Q3 | 2018Q4 | 2019Q1 |
|---|---|---|---|---|
| AHT | 0.255667 | -0.208255 | -0.223760 | 0.542840 |
| AZN | 0.673380 | -0.783894 | 1.508308 | -0.303938 |
| BARC | 0.291988 | -0.038081 | -2.049859 | 2.006252 |
| BP | 0.199607 | 0.131285 | -0.805620 | 0.564445 |
| GSK | 0.014401 | 0.273149 | -0.110208 | -0.038099 |
| HSBA | 0.002145 | -0.005762 | -0.789011 | 0.831934 |
| IAG | 1.400756 | 0.534903 | -0.852565 | -1.956990 |
| IMB | 0.241984 | 0.016095 | -0.213782 | 0.258560 |
| LLOY | -0.056829 | 0.164216 | -0.339206 | 0.258317 |
| NWG | -0.144187 | 0.075412 | -0.371735 | 0.259776 |
| RDSA | -0.005269 | 0.071019 | -0.401570 | 0.421063 |
| RRS | -0.100541 | 0.166166 | -0.213782 | 0.258560 |
| SHP | 0.032545 | -0.187017 | -0.213782 | 0.258560 |
| SKY | 0.241984 | 0.016095 | -0.213782 | 0.258560 |
| SN | 0.241984 | 0.016095 | -0.213782 | 0.258560 |
| TUI | 0.582135 | 0.016095 | 2.083622 | 0.258560 |

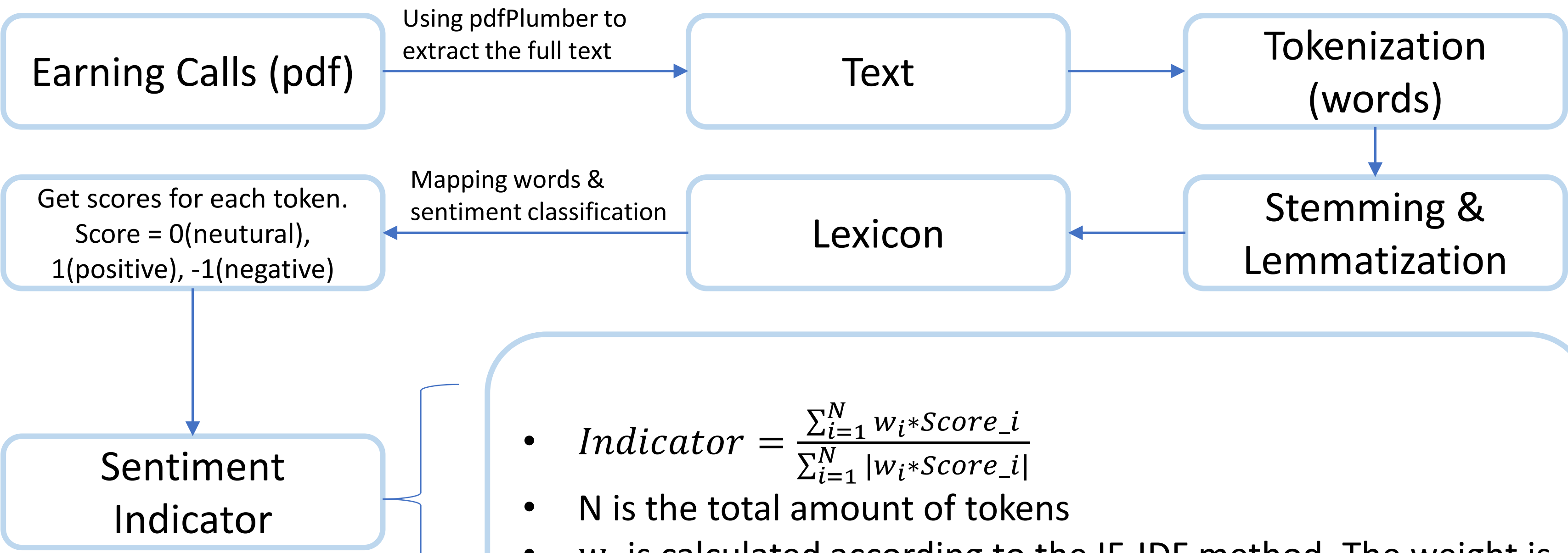
# Prepare Data — Description

- We used the transcripts of the earning calls based on the 2018 Q1
- We used the FTSE100 membership list on 02 Jan 2018 to avoid supervisor bias.




# Build Sentiment Indicator — Methodology (1)

**We followed the following process to build the lexicon-based sentiment indicator:**




- $Indicator = \frac{\sum_{i=1}^{N} w_i * Score\_i}{\sum_{i=1}^{N} |w_i * Score\_i|}$
- N is the total amount of tokens
- $w_i$ is calculated according to the IF-IDF method. The weight is used for measuring the importance of a single word.

# Build Sentiment Indicator — Methodology (2)

## The choice of lexicon:

| Lexicon | The total number of Positive Words | The total number of Negative Words | Comments |
|---|---|---|---|
| GI (General Inquirer) | 1915 | 2291 | General dictionary. We choose this lexicon as a benchmark. |
| LM (Loughran-McDonald) | 354 | 2355 | General dictionary. We choose this lexicon as a benchmark. |
| HW (Henry Word list [1]) | 106 | 106 | The lexicon is specific to the domain of financial disclosure. We choose this lexicon because its more powerful that general lexicon ([1] Henry 2010). Because general lexicon likely omit words that would be considered positive of negative in the context of financial disclosure and include words that would not. |
| GI+LM+HW | 3428 | 4648 | We choose this lexicon to test if the aggregation is better. |

[1]: Henry E, Leone AJ. Measuring qualitative information in capital markets research. Papel de trabajo. 2010 Nov.

# Build Sentiment Indicator — Results

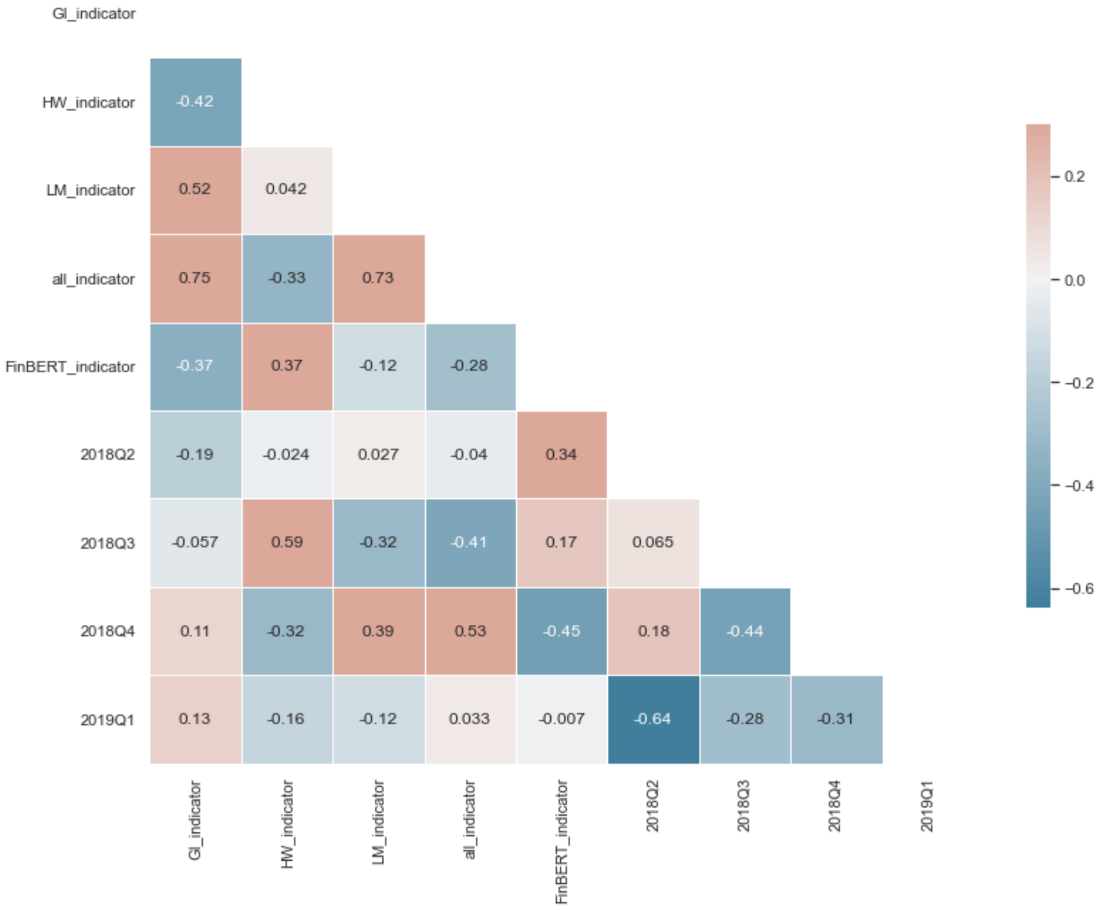
## This is the result of sentiment indicators

- According to the result, we can find that the HW_indicator can’t work well individually.
- However, when combined with other general lexicons, the HW lexicon can still contribute to the sentimental classification.

| | GI_indicator | HW_indicator | LM_indicator | all_indicator | FinBERT_indicator |
|---|---|---|---|---|---|
| AHT | 0.007041 | -0.037037 | -0.280095 | 0.267057 | 0.793154 |
| AZN | 0.102525 | -1.000000 | -0.037069 | 0.417137 | 0.001990 |
| BARC | 0.161701 | -0.084011 | -0.580018 | 0.028644 | 0.696333 |
| BP | -0.276162 | NaN | -0.748861 | -0.137148 | -0.000534 |
| GSK | 0.664433 | 1.000000 | 0.317251 | 0.543742 | 0.007323 |
| HSBA | -0.193567 | 1.000000 | -0.583392 | 0.145945 | 0.876217 |
| IAG | -0.285631 | 1.000000 | -0.609506 | -0.126754 | 0.893922 |
| IMB | -0.706026 | 1.000000 | -0.532175 | -0.120275 | 0.545867 |
| LLOY | 0.459951 | -1.000000 | -0.649864 | 0.100593 | 0.023318 |
| NWG | -0.444763 | 1.000000 | -0.585233 | -0.183766 | 0.150081 |
| RDSA | 0.179671 | 1.000000 | -0.341220 | 0.159918 | 0.052938 |
| RDSB | 0.179671 | 1.000000 | -0.341220 | 0.159918 | 0.052938 |
| RRS | -0.072092 | NaN | -0.781551 | 0.034104 | 0.166713 |
| SHP | 0.105015 | -0.037037 | -0.220538 | 0.122063 | 0.012169 |
| SKY | -0.028981 | 1.000000 | 0.067982 | 0.041272 | 0.828924 |
| SN | -0.007237 | NaN | -0.285665 | 0.101019 | 0.009866 |
| Tesco | 0.110168 | 1.000000 | 0.022899 | 0.225074 | 0.799676 |
| TUI | 0.052117 | NaN | -0.338428 | 0.305283 | 0.018070 |

# Test Predictive Ability (1)

## The Correlation Heatmap:




- The FinBERT indicator has positive correlation with T+1 and T+2 fundamentals. It means the indicator could have predictive ability in the short term (T+1, T+2), but the predictive ability will diminish in the long term (T+3, T+4)

- The LM indicator and LM+GI+HW indicator has negative correlation with T+1 fundamentals while positive correlation with T2 fundamentals. It means the indicators predictive ability is time-lagged, and it will diminish in the long term.

# Test Predictive Ability (2)

## Comparison on the average growth rate of net income among different groups

| Lexicon | index | 2018Q2 | 2018Q3 | 2018Q4 | 2019Q1 |
|---|---|---|---|---|---|
| LM | postive | 0.128193 | 0.144622 | -0.161995 | 0.110230 |
| | negtive | 0.258240 | -0.002266 | -0.221180 | 0.279750 |
| GI | postive | 0.223502 | -0.086596 | 0.031693 | 0.425444 |
| | negtive | 0.260467 | 0.118786 | -0.459257 | 0.091676 |
| LM+GI+HW | postive | 0.181132 | -0.041681 | -0.098068 | 0.417597 |
| | negtive | 0.424540 | 0.189424 | -0.560925 | -0.218552 |
| FinBERT | postive | 0.244809 | 0.008416 | -0.174326 | 0.238168 |
| | negtive | 0.199607 | 0.131285 | -0.805620 | 0.564445 |

- The positive group and the negative group of LM+GI+HW and FinBERT have significant differences in 2018Q3 and 2018Q4. It means the indicator has predictive ability on T+2 and T+3 fundamentals.

# Test Predictive Ability (3)

UCL

## Results from the univariate Linear Regression

$$Foundmtens_{i,T+j} = a + \beta * sentimental\ indicator_{i,T} + \varepsilon_{i,T}$$

- The results show that the FinBERT indicator has significant coefficients in the four regression. And the LM+GI+HW indicator has significant coefficients in the T+2 and T+3 regressions.

| Lexicon | index | 2018Q2 | 2018Q3 | 2018Q4 | 2019Q1 |
|---|---|---|---|---|---|
| LM | linear intercept | -0.595601 | -0.543121 | -0.520605 | -0.553533 |
| | linear Coef | -0.188346 | -0.218893 | -0.197272 | -0.194687 |
| | p_value of coef | 0.001023 | 0.000180 | 0.000304 | 0.000527 |
| | linear_R2 | -0.070656 | 0.038398 | 0.092106 | -0.055333 |
| GI | linear intercept | -0.192335 | -0.199428 | -0.196076 | -0.223939 |
| | linear Coef | 0.235890 | 0.166386 | 0.177680 | 0.159874 |
| | p_value of coef | 0.830463 | 0.849176 | 0.917425 | 0.725685 |
| | linear_R2 | -0.032718 | -0.067912 | -0.058352 | -0.053255 |
| LM+GI+HW | linear intercept | -0.023469 | 0.007919 | 0.032138 | -0.015721 |
| | linear Coef | 0.246034 | 0.214282 | 0.229867 | 0.223522 |
| | p_value of coef | 0.098278 | 0.036690 | 0.013054 | 0.083589 |
| | linear_R2 | -0.069729 | 0.111234 | 0.225353 | -0.070242 |
| FinBERT | linear intercept | 0.003386 | 0.108090 | 0.087145 | 0.097859 |
| | linear Coef | 0.469973 | 0.518878 | 0.468776 | 0.538449 |
| | p_value of coef | 0.047180 | 0.005547 | 0.007463 | 0.007871 |
| | linear_R2 | 0.053583 | -0.039516 | 0.148230 | -0.071376 |

# Conclusion, Limitation and improvements (1)

UCL

## Conclusion:

- We found that the three methods (the correlation heatmap, the comparison among groups, the univariate regression) show conflicting results. It's may be because the sample is too small, which leads to the unreal and confusing results.
- We can only see that the indicator does have predictive power for fundamentals, but this predictive power cannot be accurately quantified.

## Limitation and improvements:

Our work could the following limitations and room for improvements:

| Limitation | Improvements |
|---|---|
| The sample is too small. | Expand dataset by doing panel analysis instead of cross-sectional analysis |
| Lexicon-based method | Use deep learning method like FinBERT (Actually, we did this in our work) |
| Predictive method | Do sensitive analysis |

We will explain them one by one in the following slides

# Conclusion, Limitation and improvements (2)

UCL

## Limitation: the small sample

- The dataset size is too small, including only 15 samples. The problem comes from that we limited our cross-sectional analysis to be based on earning calls of Q12018. More specifically, most companies did not disclose the earnings call in 2018 Q1. And some companies disclosed the transcripts in 2018 Q1 but did not disclose the fundamentals during the 2018 Q1 - 2019 Q1 because they were acquired.

- The small sample may make our analysis results unreliable. Because, compared with large samples, small samples have higher requirements for data quality and are more difficult to reflect the real patterns.

## Improvements: Panel Analysis

- We can use panel analysis to replace cross-sectional analysis. Once we remove the restrictions (the earning call's transcript should be based on 2018 Q1), we can expand our dataset. More clearly, we use the following to formula to express our idea:

$$Foundmtens_{i,T+j} = a + \gamma_i + \beta_1 Year\&Season + \beta_2 * sentimental\ indicator_{i,T} + \varepsilon_{i,T}$$

where $j = \{1,2,3,4\}$; $i = \{1,2, \ldots N\}$

$Year\&Season$ is a one-hot variable (or dummy variable). The variable is used for capture the fixed effects from time.

and $\gamma_i$ is used for capture the fixed effects among different companies.

Then we can test the sentimental indicator's predictive ability according to the estimation of $\beta_2$.

# Conclusion, Limitation and improvements (3)

## Limitation: lexicon-based method

- Lexicon-based methods likely ignore contextual information because they can only provide single-dimensional data when mapping text to data vectors.
- Furthermore, lexicon-based methods are highly dependent on the quality of the lexicon. In other words, reliable sentiment classification is difficult without a dictionary specific to the financial domain. But even if we have domain-specific lexicons like the Henry word list, the objectivity of the lexicon itself can lead to debate.

## Improvements: Deep leaning method like FinBERT

- Deep learning method like FinBERT can be a alternative of the lexicon-based method. They can provide multiple dimensional information, thus enabling the codes to read and extract contextual information.

# Conclusion, Limitation and improvements (4)

## Limitation: the predictive method

- The predictive ability may depend on the time window. In other words, we only tested the sentimental indicator's predictive ability in Q1 2018 to the fundamentals in Q2 2018-Q1 2019. In this specific time window, the indicator may have a strong predictive ability. However , if we change the time window and test the indicator's ability at time T for fundamentals from time T+1 to T+4 again, we may get different results. And currently, our predictive method can not capture this kind of information.

## Improvements: Sensitive analysis

- We can test the sensitivity of the sentimental indicator's predictive ability by rolling the time window and testing indicators' predictive ability. And then analyzing the change of indicator's predictive ability.

# References

[1] Doran, James S., David R. Peterson, and S. McKay Price. "Earnings conference call content and stock price: the case of REITs." *The Journal of Real Estate Finance and Economics* 45.2 (2012): 402-434.

[2] Henry E, Leone AJ. Measuring qualitative information in capital markets research. Papel de trabajo. 2010 Nov.

[3] Jegadeesh, Narasimhan, and Sheridan Titman. "Cross-sectional and time-series determinants of momentum returns." *The Review of Financial Studies* 15.1 (2002): 143-157.

[4] Price, S. McKay, et al. "Earnings conference calls and stock returns: The incremental informativeness of textual tone." *Journal of Banking & Finance* 36.4 (2012): 992-1011.

[5] Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson. "Measuring news sentiment." *Journal of econometrics* (2020).