Using Earnings Call Transcripts to Predict Stock Performance

Jonathan Khalfayan, Justin Kahl, Santiago Rodriguez, Matthias Schmitz, Sam Sklar, Juan Pablo Villarreal
Roadmap

1. Obtain Earnings Call Transcripts
2. Apply Natural Language Processing algorithm
3. Run regressions to weight keywords
4. Develop scoring algorithm to determine when to buy/sell
5. Run backtest evaluation of algorithm
Earnings Calls

3Q Fiscal 2017
ADP Earnings Call & Webcast

May 3, 2017

KLONDEX MINES LTD

2017 FIRST QUARTER EARNINGS
MAY 5, 2017
Stock Criteria

- **Small**
  - Between the 25th and 75th percentile based on their market capitalization

- **Cheap**
  - Below 25th percentile EV to Ebitda ratio

- **Highly Leveraged**
  - Above mean LT Debt/EV
Binary Variables

LT debt year 1 > LT debt year 2

Prior year returns below median:

% Revenue Growth > % Asset Growth

High Gross Profit/Assets relative to LT Debt/Assets
Benchmark and Relative Performance

<table>
<thead>
<tr>
<th>Total Returns</th>
<th>Benchmark Returns</th>
<th>Alpha</th>
<th>Beta</th>
<th>Sharpe</th>
<th>Sortino</th>
<th>Volatility</th>
<th>Max Drawdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.4%</td>
<td>111.5%</td>
<td>0.01</td>
<td>0.53</td>
<td>0.57</td>
<td>0.85</td>
<td>0.13</td>
<td>-30.4%</td>
</tr>
</tbody>
</table>

Cumulative performance:
- Algorithm **52.52%**
- Benchmark (SPY) **109.02%**

Weekly returns **$16,262**

Transactions

Week of Mar 26, 2017
Data Collection

- Earnings call transcripts from THOMSON REUTERS, manually scraped
Data Collection

Next Steps:

1. Incorporate a larger number of equities in the dataset.
   a. Available extensive dataset still pending exploitation

2. Enhance current dataset
   a. Fill in gaps of non-existent data
Roadmap

1. Obtain Earnings Call Transcripts
2. Apply Natural Language Processing algorithm
3. Run regressions to weight keywords
4. Develop scoring algorithm to determine when to buy/sell
5. Run backtest evaluation of algorithm
Language Processing

Methodology:

1. Categorize words from .txt file into their corresponding fiscal quarter
2. Account for discrepancies in length of transcript, case, and punctuality
3. Count the frequency of specific words in each quarter
4. Write csv files of organized data for regression analysis
Next Steps:

1. Account for different tenses of words.
   a. Potentially use global vectors for word representation (GloVe research group at Stanford)
   b. Multi-Dimensional Matrixes with similar words mapped at neighboring locations within the Matrix, word frequency calculated for set of words.

2. Investigate two word keyphrases such as “Pay Down”
Word Selection Criteria

- We started analyzing the frequencies of 26 keywords that could possibly predict future growth.
- The presence of words like “cost-cutting,” “deleveraging,” and “growth” that we believe will lead to these small, highly-leveraged, companies to produce higher returns than the benchmark

- For every company’s earnings call transcript:
  - Frequency = # of appearances of keyword within specified fiscal quarter
<table>
<thead>
<tr>
<th>Quarter</th>
<th>cost</th>
<th>prudent</th>
<th>long-term</th>
<th>short-term</th>
<th>capital</th>
<th>reduce</th>
<th>revenue</th>
<th>efficient</th>
<th>synergies</th>
<th>growth</th>
<th>sales</th>
<th>marketing</th>
<th>save</th>
<th>structure</th>
<th>negotiate</th>
<th>pay</th>
<th>tax</th>
<th>rates</th>
<th>demand</th>
<th>supply</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 2017</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2016</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2015</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2014</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2013</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2012</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2011</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01 2010</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Data

Screenshot table containing frequency values for one company for desired fiscal quarters and keywords:
Roadmap

1. Obtain Earnings Call Transcripts
2. Apply Natural Language Processing algorithm
3. **Run analysis to weigh keywords**
4. Develop scoring algorithm to determine when to buy/sell
5. Run backtest evaluation of algorithm
## Key Word Analysis

<table>
<thead>
<tr>
<th>Word</th>
<th>Score</th>
<th>Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Save</td>
<td>1.37</td>
<td>122</td>
</tr>
<tr>
<td>Dividend</td>
<td>.894</td>
<td>739</td>
</tr>
<tr>
<td>cutting</td>
<td>.805</td>
<td>92</td>
</tr>
<tr>
<td>cut</td>
<td>.772</td>
<td>473</td>
</tr>
<tr>
<td>short-term</td>
<td>.754</td>
<td>333</td>
</tr>
<tr>
<td>pay</td>
<td>.720</td>
<td>594</td>
</tr>
<tr>
<td>demand</td>
<td>.685</td>
<td>2053</td>
</tr>
<tr>
<td>prudent</td>
<td>.682</td>
<td>195</td>
</tr>
<tr>
<td>reduction</td>
<td>.657</td>
<td>1107</td>
</tr>
</tbody>
</table>
Next Steps
Roadmap

1. Obtain Earnings Call Transcripts
2. Apply Natural Language Processing algorithm
3. Run regressions to weight keywords
4. Develop scoring algorithm to determine when to buy/sell
5. Run backtest evaluation of algorithm
Algorithm

**Insight:** By using the correlations found in the Earnings Call Transcript analysis, we can score companies by quarter as new transcripts are released.

**Methodology:**

- When high frequencies of key terms occur (cost cutting, dividend, reduce, save) increment score.
- Add additional binary values for whether long term debt has decreased from last year (signalling debt paydown) as well as increasing Gross Profit/Assets ratio.
- Buy companies with highest scores that also meet binary metrics. Try on Quarterly vs Yearly holding period
Roadmap

1. Obtain Earnings Call Transcripts
2. Apply Natural Language Processing algorithm
3. Run regressions to weight keywords
4. Develop scoring algorithm to determine when to buy/sell
5. Run backtest evaluation of algorithm
Stock Performance

- To prevent look-ahead bias, we will trade with one quarter lag.
- For instance, we will act upon data obtained from Q1 2010 at the beginning of Q3 2010.
Hypothesis and Predictions

- Investing in small, cheap, highly-leveraged stocks with improving asset turnover and estimated debt paydown that we filter by keywords such as “cost-cutting” will lead to a strategy that will outperform the S&P 500 Index.
Thank You!