News Sentiment and Equity Returns

Institute of Management Science
Financial Management & Controlling

Ao.Univ.-Prof. Mag. DDr. Thomas Dangl
Dipl. Ing. Stefan Salbrechter
Overview

- We study the influence of financial news on equity returns.
- Therefore we perform a Sentiment Analysis of company related news articles.
- We train a state-of-the art Natural Language Model (BERT) to capture the news sentiment.
- Fundamental Hypothesis:
  - Positive Sentiment is associated with positive future asset returns
  - Negative Sentiment is associated with negative future asset returns
- News sentiment trading strategy
- 100% data driven approach

What we find

- Our sentiment measure shows predictive quality
- It takes up to two days until financial news are fully incorporated into asset prices
- Markets are very efficient

<table>
<thead>
<tr>
<th></th>
<th>Fama-French Reg. (FF5+MOM)</th>
<th>Long Portfolio</th>
<th>Long/Short Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha /Month</td>
<td>2.25%</td>
<td>6.46%</td>
<td></td>
</tr>
</tbody>
</table>
Natural Language Processing (NLP)

NLP is a growing subfield of computer science and artificial intelligence that gives computers the ability to process, analyze and understand human language.

Methods for generating features from text

- Dictionary approaches
- Bag-of-words
- Word embeddings (Word2Vec)
- Contextual word embeddings (BERT)

Use cases of NLP

- Machine translation
- Spam detection
- Text summarization
- Virtual agents and chatbots
- Sentiment analysis

Fig 1: Bag-of-words approach

Fig 2: Word embeddings approach
Data

News Data
- Thomson Reuters news dataset
- Time period: 1996 to 2020
- Selection of English news, related to S&P 500 companies
- 372,438 cleaned news articles
- We further consider:
  - News novelty („fresh“ vs. „stale“)
  - News topics (analyst forecast, earnings reports)

Price data\([1]\)
- In Total 1130 S&P 500 Companies
- Adjusted Open/Close Prices
- Frequency: Daily

[1] Data source: Thomson Reuters DataStream
BERT Model *(Bidirectional Encoder Representations for Transformers)*

- Developed by Google in 2018
- Pre-trained on a book corpus with 800 million words and the English Wikipedia with 2,500 million words in 2018.[1]
- Uses contextual word embeddings
- BERT is part of the Google search algorithm [2]
- Multiple use cases:
  - Question Answering
  - Named Entity Recognition
  - Text Similarity
  - Classification

---

BERT\textsubscript{MINI} vs. BERT\textsubscript{BASE}

<table>
<thead>
<tr>
<th></th>
<th>BERT\textsubscript{MINI}\textsuperscript{[1]}</th>
<th>BERT\textsubscript{BASE}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer layers</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Hidden embedding size</td>
<td>256</td>
<td>768</td>
</tr>
<tr>
<td>Max. Input Sequence</td>
<td>256</td>
<td>512</td>
</tr>
<tr>
<td># Parameters</td>
<td>\textbf{18.95 Mio.}</td>
<td>110 Mio.</td>
</tr>
<tr>
<td>Pre-training data</td>
<td>Thomson News (4 GB)</td>
<td>Book corpus &amp; Wikipedia</td>
</tr>
<tr>
<td></td>
<td>Reuters News (4 GB)</td>
<td></td>
</tr>
</tbody>
</table>

Advantages of BERT-Mini

- Focus on financial domain
- Less computation power required
- Training cheaper compared to BERT\textsubscript{BASE}
- No look-ahead bias

Disadvantages of BERT-Mini

- Limited input sequence length of 256 words

\textsuperscript{[1]} Turc, Iulia; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2019): Well-Read Students Learn Better: On the Importance of Pre-training Compact Models.

Model architecture

Fig. 5: Model architecture

**Model architecture**

**FinNewsBERT**

```
CLS Token

[CLS] T1 T2 ... Tn
```

**Text2Topic**

```
T1 T2 ... Tn
```

News Article

research alert texas instruments raised to buy, first boston raised texas instruments inc to a buy from a hold based on share price, a source said.

```
Input layer
Hidden layer 1
Hidden layer 2
Hidden layer 3
Output layer
```

```
0.04
0.14
0.11
0.04
0.24
0.26
0.02
0.01
0.04
0.21
0.02
0.81
0.02
0.01
```

```
\hat{y}
```

```
0.97
0.01
0.02
```

Fig. 5: Model architecture
Annotation of Financial News

- Artificial neural networks are trained on datasets that contain large numbers of example input-output pairs (Supervised Machine Learning).
- We use the joint behaviour of news articles and stock returns to derive the sentiment labels [1]

1. Calculation of daily idiosyncratic returns
2. Calculation of the arithmetic mean values between $t-1$, $t$, and $t+1$
3. Calculation of z-values over a 2-year rolling window
4. Annotation:
   - $z$-value > 1.4: positive label
   - $z$-value < -1.4: negative label
   - $-1 < z$-value < 1: neutral label

Training of FinNewsBERT

1. Pre-training
   - Unsupervised learning
   - Masked Language Modelling (MLM)

   "lincoln national corp deutsche bank starts with <mask> rating price target $84."

<table>
<thead>
<tr>
<th>Predictions for the &lt;mask&gt; Token</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>buy</td>
<td>0.56</td>
</tr>
<tr>
<td>hold</td>
<td>0.24</td>
</tr>
<tr>
<td>neutral</td>
<td>0.08</td>
</tr>
<tr>
<td>outperform</td>
<td>0.05</td>
</tr>
<tr>
<td>sell</td>
<td>0.03</td>
</tr>
</tbody>
</table>

2. Fine-tuning
   - Supervised learning
   - Fine-tuning on the training dataset with annotated financial news
   - Iterative Training
   - Training over 3 epochs
   - Loss function: Symmetric Cross Entropy Loss \[1\]

![Training and validation accuracy](image)

Fig. 7: Iterative pre-training and fine-tuning. The models are re-trained every two years.

Fig. 8: Training and validation accuracy (The model is trained with data from 1996 to 2017)

---

Event Study

Fig. 9: Abnormal CAPM-Returns (in %) of different sectors.
News are released between market close at 4:00pm on day t-1 and market opening at 9.30am on day t. Time period: 01-2002 to 01-2020.
Forecast of the price movement

Prediction whether the share price of company XY will rise or fall on the following trading day in response to financial news.

- News article classified as positive & $r(t+1) > 0$ $\rightarrow$ True Positive
- News article classified as negative & $r(t+1) < 0$ $\rightarrow$ True Negative

<table>
<thead>
<tr>
<th>Prediction</th>
<th>All news</th>
<th>Filter: Topic: Analyst Forecast</th>
<th>Filter: Topic: Analyst Forecast &amp; $z$-value$(t) &lt; -1.96$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>51.40 %</td>
<td>52.03 %</td>
<td>52.36 %</td>
</tr>
<tr>
<td>Avg. return p.a. (arithm.)</td>
<td>-13.77 %</td>
<td>33.03 %</td>
<td>-31.16 %</td>
</tr>
<tr>
<td>Support</td>
<td>61916 (48.44 %)</td>
<td>65878 (51.55 %)</td>
<td>19746 (49.09 %)</td>
</tr>
</tbody>
</table>

Table 1: Precision and average returns of the predictions at time $t+1$

Fresh & stale news published during market closing hours, 4:00pm (day $t-1$) to 9:30am (day $t$)
Short-term Momentum Effect? (1)

- Stale news show larger abnormal returns than fresh news at time t+1
- Strongly negative returns at time t (z-value < -1.96) tend to be followed by large negative returns at time t+1 (and also at t+2 in the case of fresh news)

Short-term Momentum Effect? (2)

- This effect can only be observed in combination with negative news.
- The effect can always be observed except for the period 2008-2009 (financial crisis).
- Without taking the news into account, a short term reversal can be observed in the markets.

→ Effect is not driven by recessions.

Fig. 11: fresh & stale News  
\[ z(t) < -1.96 \]

Fig. 12: No consideration of news sentiment  
\[ z(t) < -1.96 \]
Backtest (Settings B)

Trading strategy:
- Positive sentiment → LONG
- Negative sentiment → SHORT
Buy: MOO(t)
Sell: MOO(t+1)

Fig. 16: Backtest
Fresh news, analyst forecast, time-window: 4pm - 9:30am (17.5h), z-value(t, neg. news) < -1.96

Table 3: Performance metrics of the long and the long/short portfolio (Settings B)

<table>
<thead>
<tr>
<th></th>
<th>Long portfolio</th>
<th>Long/short portfolio</th>
<th>Market portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. weight</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>News</td>
<td>Fresh news, Topic: Analyst Forecast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z-value(t) neg./pos. news</td>
<td>(−∞, -1.96)/([∞/∞])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>1.44</td>
<td>2.26</td>
<td>0.36</td>
</tr>
<tr>
<td>CAGR</td>
<td>44.04% p.a.</td>
<td>61.22% p.a.</td>
<td>6.56% p.a.</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>30.58% p.a.</td>
<td>44.00% p.a.</td>
<td>18.33% p.a.</td>
</tr>
<tr>
<td>Beta</td>
<td>0.84</td>
<td>0.49</td>
<td>1</td>
</tr>
<tr>
<td>Alpha</td>
<td>13.43 bps/day</td>
<td>29.94 bps/day</td>
<td></td>
</tr>
<tr>
<td>Avg. Portfolio-size</td>
<td>1.93</td>
<td>1.92 (long)</td>
<td>0.42 (short)</td>
</tr>
<tr>
<td>Trades p.a.</td>
<td>454</td>
<td>553</td>
<td></td>
</tr>
<tr>
<td>Return/trade</td>
<td>18.31 bps</td>
<td>24.06 bps</td>
<td>(50.28 bps im Short leg)</td>
</tr>
</tbody>
</table>
Backtest (Settings B)

**Fig. 17:** Investment ratio of the long and the long/short portfolio

**Fig. 18:** Ex-ante beta of the long and the long/short portfolio
Table 4: Fama-French factor analysis of the two backtests with monthly alphas in basis points. **, *** corresponds to significance levels of 5% and 1%.
Conclusion

- The presented models are able to extract a sentiment measure from financial news.
- Financial news contains predictive information about stock returns. Positive correlation between news classification (sentiment) and stock returns.
- Information from financial news is priced in very quickly, usually within a day.
- We observe abnormal returns of up to 2 days after strongly negative news events.
- The additional news topics are valuable features that significantly improve the performance of the model.
- Analyst forecast news contains more predictive power than general financial news.
- Training of BERT on financial text corpora is beneficial
Outlook

- By considering companies with lower market capitalization, higher achievable returns can be expected. According to Kelly et. al. (2019), the price reaction is about four times as large for small companies as for companies with high market capitalization.

- By combining news sentiment with other signals (e.g. historical price data, volatility, market trends, etc.) the return per trade could be further improved.

- The news sentiment could be used as an additional signal for other (longer-term) trading strategies.

- A higher number of additional topic features might further improve prediction accuracy.

References


Cong, Lin; Liang, Tengyuan; Zhang, Xiao (2018): Textual Factors: A Scalable, Interpretable, and Data-driven Approach to Analyzing Unstructured Information.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Gentzkow, Matthew; Kelly, Bryan T.; Taddy, Matt (2017): Text as Data


Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey (2013): Efficient Estimation of Word Representations in Vector Space.


Turc, Iulia; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2019): Well-Read Students Learn Better: On the Importance of Pre-training Compact Models.
