The positive similarity of company fillings and the cross-section of stock returns

Matúš Padyšák

Abstract

It is already well-documented that textual analysis of 10-K & 10-Qs can be largely profitable. This research studies the similarity of language used in the filings using data which enables to analyze what type of language is similar. Results show that the similarity of the positive language is the most profitable option. From a practical point of view, the positive similarity effect is examined. Results show that the lowest positive similarity stocks significantly outperform the highest positive similarity stocks. The effect cannot be explained by the common asset pricing models, nor by the change of sentiment in the financial reports. Therefore, the positive similarity effect could be considered as a distinct anomaly in the financial markets. In the long-only implementation, the strategy is highly profitable, and in the long-short implementation, the strategy has impressive consistency and risk-adjusted return (0.84).

Authors:
Matúš Padyšák, Senior Quant Analyst, Quantpedia.com

Introduction

A 10-K or a 10-Q are periodically filled reports by publicly-traded companies. Those reports consist of relevant information about financial performance, and rightfully reports receive great interest from investors, analysts but also academics as well. No doubt, financial reports should be informative. After all, they should report the financial prospects of companies. However, in the recent period, 10-K and 10-Q reports have changed.

Firstly, there is a high and increasing volume of published 10-K&10-Q reports. There is also a gradual shift to non-numerical information, which means that the reports are consisting of an increasing amount of text-based information that is harder to analyze. This allows management to include a lot of white noise into reports or "manipulate" the choice of the language. Mandatory filings could be written using a speculative language to present the information in a better light, while still being in line with rules. There is a possibility to create a picture of the company that is favourable for the image of the company; however, it does not tell the absolute truth. According to the Pulliza (2015), a document could include an overemphasis of positive opportunities while undermining the plausibility of possible risks, such as using the stronger "likely" versus, the weaker "may." There is also an opportunity for firms to add distance between pieces of related information, increasing the amount of cognitive effort to deduce truthfulness, such as projecting a large gain or loss in a table but communicating a decreased or increased likelihood of that figure in another section.

Secondly, the increasing volume of text in the fillings and technological advances has led to a textual analysis of the reports. Advanced machine learning methods, and compared to the last century, the sufficient computing power for advanced algorithms offer us a new data-driven view on the increasingly bigger data of 10-K &10-Q reports. Therefore, it is possible to identify the type of the language used, compare the reports with the previous, identify changers/non-changers or the similarity of langue used compared to the earlier reports. One of the most significant results is the work of Cohen et al. (2019). According to this research that is commonly known as "the Lazy prices", firms that do change reports outperform non-changers.

In this novel research, with the help of the Brain Company, which provided the data of textual analysis of reports, I study the similarity of language. Analyzing preliminary results, the similarity of the reports yields different results based on the type of language used. There are two significant differences across types of language similarities recognized by the Brain Company (all, positive, negative, uncertainty, litigious, constraining and interesting language). The decile sorts of stocks do not have a unified pattern across all types of language. The highest (lowest) decile is not always the most profitable (unprofitable). Moreover, the degree of profitability changes from a kind of language to another.

The aim of this paper is to precisely analyze possible usages of the similarity data in a trading strategy. Moreover, the objective is to find the most profitable type of language that could bring the most economically significant results. For example, by examining portfolio sorts based on the similarity of uncertainty language, I can confirm the results of Cohen et al. (2019), where the lowest similarity (changers) decile is outperformed by the highest similarity decile (non-changers). However, the results reverse when positive language is considered. For a positive language, the lowest similarity decile outperforms the highest similarity decile by 0.41% monthly. Furthermore, the spread is much larger than each type of language which is identified by the Brain. As a result, this paper is focused on the usability of a positive similarity score in a trading strategy. The most profitable strategy implementation would be a long-only variant, with an average return of 1.14% monthly.

Additionally, the results show that the returns of similarity deciles are robust to the sentiment extracted from the financial reports. Through a double-sorts, where stocks are firstly sorted based on their similarity score and secondly by their changes in sentiment extracted from reports, it can be shown that the performance does not depend on the change of sentiment. Such a result suggest that the effect could be considered as an anomaly in the financial markets, where the similarity of positive language negatively affects subsequent performance.

Naturally, the research paper Lazy prices is very close to this research, but there are three main differences. Firstly, as it was previously mentioned, this paper is focused on the similarity of positive language only, motivated by the search for the most profitable strategy. Secondly, the holding period is shorter (one month compared to three months), and stocks are sorted into deciles based on their most recent 10-K or 10-Q report. Each stock is ranked every month and not only after the new report is released. Stock enters the portfolio based on the most recent report. For contrast, in the Lazy prices, stock enters the portfolio in the month after the public release of either 10-K or 10-Q. The approach used in this paper brings two main benefits: simplicity and diversification. Because the most recent reports ensure that the investment universe is larger. Tests have showed that the strategy could be formed with the average similarity based on the past 200 days, holding stocks for another 200 days (which is much longer than three months). Therefore, there is no need to always wait for the newest information if the portfolio can be much more diversified. Lastly, the investment universe is different. The Lazy prices examine the effect on approximately 4000 stocks (based on the count of reports, backtesting length and quarterly period), which has to include smaller capitalization stocks and possible liquidity issues. I study the effect on the stocks with large market capitalization since the Brain analyzes company reports for approximately the largest 1000 US stocks. As a result, the investment universe includes mostly large caps with better liquidity and lower slippage costs and spreads.

The main contribution of this paper to the literature is twofold. Firstly, it successfully shows the possible usage of modern machine learning methods (and alternative data) in finance. Secondly, I show that the new (dissimilar) is always better, at least if it is connected with the positive language and financial reports. There seems to be a clear pattern, where the firms with the lowest similarity of 10-K and 10-Q reports outperform firms with the highest similarity. Additionally, the effect seems to be an anomaly, since the strategy has a low correlation to the common equity risk factors (value, size, momentum, investments or profitability) and is uncorrelated to the market factor. Moreover, the alpha from common asset pricing models is both economically and statistically significant.

Data

The Brain Company kindly provided dataset "The Brain Language Metrics on company filings". The dataset monitors several language metrics on 10-Ks and 10-Qs company reports for the largest stocks and consists of calculated language metrics like financial sentiment, percentage of words belonging to financial domain classified by language types, differences between last filings or similarity of language between the most actual report and previous one. The similarity of positive language is in the main interest of this paper. The similarity says how similar filings are, however, the dissimilarity could be caused by a different wording, where the exact synonyms are used, or there could be used less (or more) positive language compared to the previous report. Let's consider a short and simple example.

We have a dictionary of only three positive words: positive, credible and adaptable. Words can be interpreted in the vector, where zero means that the word is missing and one that the word is obtained. For example, the sentence: "Firm is credible and adaptable", can be represented by the vector [0,1,1]. Each report is made by only one short sentence:

$REPORT_1$: The firm is credible.

$REPORT_2$: The firm is adaptable.

In this case, the sentiment is the same (one positive word in each sentence). However, the words are different, and the cosine similarity is zero.

$REPORT_3$: The company is credible.

Comparing the third and the first report, reports are similar and have the same sentiment.

$REPORT_4$: The company is heading to bankruptcy.

Comparing the fourth report with the first, reports are not similar (different language is used). Furthermore, the change of sentiment (from first to fourth) is negative.

$REPORT_5$: The company is both credible and adaptable.

Comparing the fifth to the first, reports are to some extent similar, but the sentiment is more positive.

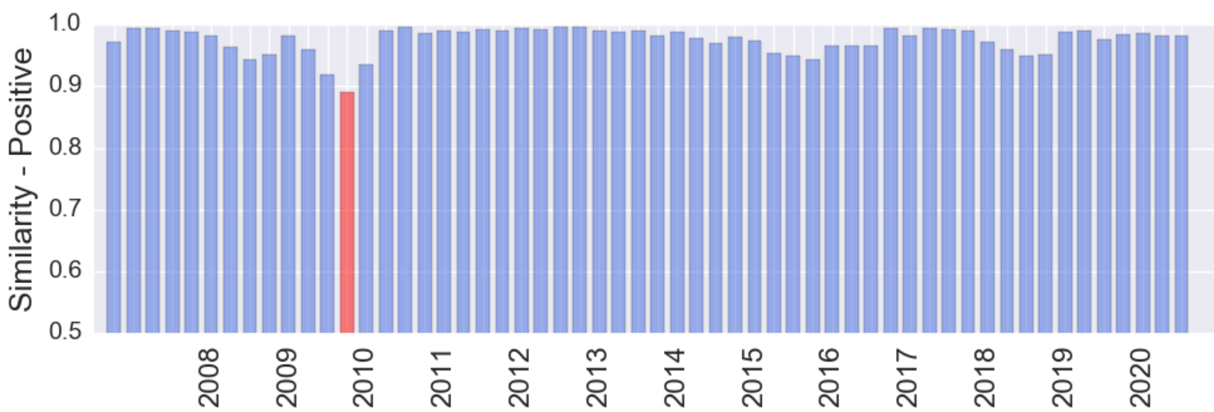
This short example aimed to show that low similarity could be caused by different words used with the same sentiment or completely different sentiment. Therefore, I also utilize the Brain data on sentiment change (delta) to double sort stocks. At first, by the similarity of positive language and secondly by the sentiment delta.

The Brain data are matched with stock prices obtained from Yahoo Finance and filtered to a monthly frequency since the Brain Language Metrics are updated daily. The matching of Yahoo Finance data (stocks available at Yahoo dataset) with Brain data leaves us with 637 stocks with a history spanning from 28.2.2007 to 29.5.2020. This includes two major financial crises, the global financial crisis during 2007-2008 and recent COVID-19 market crash. However, it allows examining the performance also during large downturns. The factor returns are obtained from Kenneth R. French's data library.

Positive similarity as a trading strategy

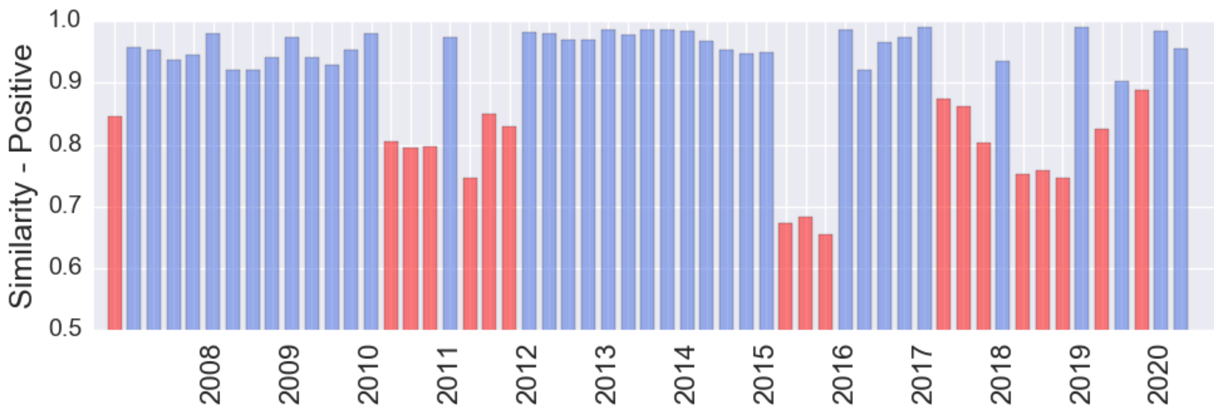
Since the aim is to explore the economic significance of the similarity data, the usability is explored through a classical portfolio sorting approach. At the end of the month, stocks are ranked according to their most recent positive similarity scores.

Figure 1 Positive similarity score for ACADIA Pharmaceuticals



ACADIA is an excellent example of a stock that has a similar positive language in the filings. It is expected that it will have consistently high ranks.

Figure 2 Positive similarity score for NVIDIA



On the other hand, NVIDIA frequently changes the positive language and filings are often dissimilar. It is natural to expect that NVIDIA will frequently have a low rank.

While some firms have similar positive language, some change it frequently. In general, the effect of the positive similarity can be examined by sorting stocks into deciles and examine the performance.

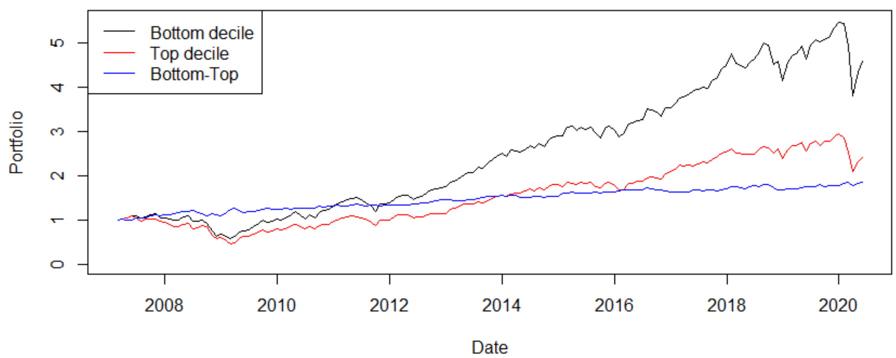
Table 1 Portfolio sorts for positive similarity. Stocks are ranked each month according to their positive similarity and sorted into deciles. Table present average monthly return (Ret) and standard deviation (Sd) in percentage points.

Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Ret	1.14	1.13	1.16	0.93	1.11	0.97	1.03	0.82	0.81	0.73
Sd	5.60	5.64	5.46	5.31	5.67	5.61	5.69	5.34	5.23	5.68

While the differences are not extreme, they are present, and in general, the return declines with raising decile. Therefore, low positive similarity stocks tend to outperform high similarity stocks. Additionally, while the first three deciles are highly statistically significant, the last three deciles are statistically insignificant on the 5% significance level (in other words, the p-value is higher than 0.05).

One viable strategy could be going long the first or even the first three deciles. Either used as a standalone strategy or in a portfolio as a building block, the investor could profit from the edge that low similarity stocks have over high similar. Furthermore, the positive similarity score can be utilized in a typical anomaly-based long-short strategy and compared to the common equity risk factors.

Figure 3 Performance of top, bottom and bottom minus top portfolios



A dollar invested in the bottom decile 28.2.2007 would result in 4.58 USD on 29.5.2020. For top decile, it would be 2.40 USD and 1.84 USD for bottom minus top portfolio. The bottom-top portfolio seems to be flat, but it is, in fact, very consistent and non-risky, yet not that profitable.

Figure 4 Performance of Bottom minus Top portfolio

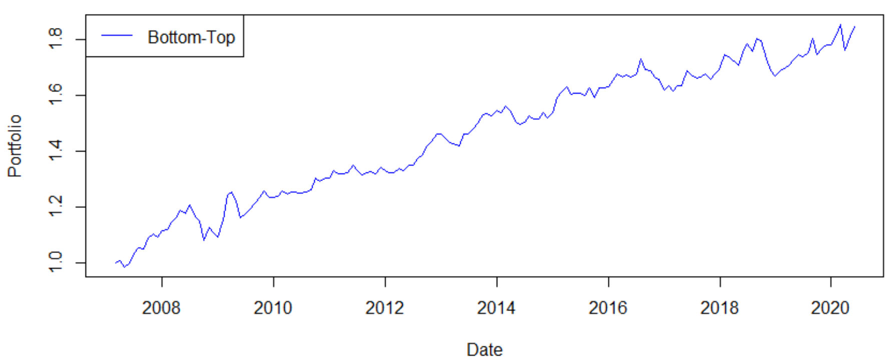


Table 2 Performance metrics for Top, Bottom and Bottom-Top portfolios. Return and volatility are annualized. Risk-adjusted return is return divided by volatility.

	Return	**Volatility**	**Risk-adjusted return**	**Max Drawdown**
Bottom	**14.58%**	19.40%	0.75	-49.50%
Top	9.11%	19.70%	0.46	-58.81%
Bottom-Top	5.47%	**6.48%**	**0.84**	**-10.43%**

Table 3 Correlation of the long-short portfolio with common equity factors

	Long-Short	**MKT**	**HML**	**SMB**	**RMW**	**CMA**	**MOM**
Long-Short	1	-0.01105046	-0.3260071	0.07559683	0.02888553	-0.22780154	0.2472834
MKT	-0.01105046	1	0.3147639	0.45432595	-0.32191713	-0.11677026	-0.4014346
HML	-0.3260071	0.31476395	1	0.37936967	-0.12789744	0.50586284	-0.479986
SMB	0.07559683	0.45432595	0.3793697	1	-0.30724505	0.0804875	-0.2853258
RMW	0.02888553	-0.32191713	-0.1278974	-0.30724505	1	0.06721442	0.1637133
CMA	-0.22780154	-0.11677026	0.5058628	0.0804875	0.06721442	1	-0.1090429
MOM	0.24728338	-0.4014346	-0.479986	-0.2853258	0.16371326	-0.10904288	1

Interestingly, the strategy is uncorrelated with the market factor, which can also be indirectly observed from the consistency of the performance from figures 3 and 4. It seems that the long-short strategy could be utilized also during market downturns. The strategy is, to a small extent, negatively correlated with value and investment factors, and positively correlated with the momentum factor.

It is already clear that the strategy is economically significant, especially for a risk-averse investor, given the consistency and low volatility. To some extent, the strategy is correlated with common equity factors, but they probably cannot explain the performance.

The worth of consideration is also the attribution of stocks in the extreme deciles (bottom and top) into sectors. While stocks from the top decile are most often attributed to the Financial

Services sector, stocks from the bottom decile are most often attributed to the Consumer cyclical, Industrials and Technology sectors. Attribution plots are presented in Appendix A.

Table 4 Asset pricing models. Alphas are monthly, t-statistics are in parentheses. *** denotes 0.1%, ** denotes 1%, * denotes 5% and $^{\cdots}$ denotes 10% significance levels.

CAPM					
$\boldsymbol{\alpha}$	$\boldsymbol{\beta_{Mkt-Rf}}$				
0.414** (2.721)	-0.004 (-0.137)				
Carhart					
$\boldsymbol{\alpha}$	$\boldsymbol{\beta_{Mkt-Rf}}$	$\boldsymbol{\beta_{HML}}$	$\boldsymbol{\beta_{SMB}}$	$\boldsymbol{\beta_{UMD}}$	
0.3* (2.097)	0.027 (0.779)	-0.195*** (-3.648)	0.175** (2.690)	0.065$^{\cdots}$ (1.877)	
Five-Factor					
$\boldsymbol{\alpha}$	$\boldsymbol{\beta_{Mkt-Rf}}$	$\boldsymbol{\beta_{HML}}$	$\boldsymbol{\beta_{SMB}}$	$\boldsymbol{\beta_{RMW}}$	$\boldsymbol{\beta_{CMA}}$
0.288393$^{\cdots}$ (1.935)	0.006411 (0.173)	-0.241695*** (-3.885)	0.178060** (2.669)	0.073251 (0.759)	-0.066632 (-0.583)

Robustness

Results show that alphas are both statistically and economically significant. Therefore, it seems to be that the similarity of positive language in company filings can be a distinct anomaly that cannot be fully explained by common equity factors.

With a short example in the data section, it was outlined that dissimilarity can be caused by either different sentiment or simply by using different positive words or even synonyms. In this section, I check whether the performance can be attributed to the positive (negative) change in the sentiment of 10-K & 10-Qs. Similar to the previous examination, results are obtained by checking the economic significance by double-sorts. I firstly sort stocks into quintiles based on their positive similarity score, and secondly into quintiles again by sorting the stocks by their change of sentiment (delta).

Table 5 Double-sorts, Sim1 is the lowest positive similarity quintile, Delta1 is the lowest delta sentiment quintile. Returns presented in the table are average monthly returns.

	Delta1	**Delta2**	**Delta3**	**Delta4**	**Delta5**
Sim1	1.1131722	1.1158128	1.363432	0.8110353	1.281859
Sim2	0.9682288	0.9993955	1.0935296	1.2509164	0.918202
Sim3	1.0680012	0.8904987	1.0708121	1.190584	1.003792
Sim4	0.7010845	0.908613	1.0665435	0.9586686	1.020037
Sim5	0.6634943	0.8669145	0.8748371	0.6853864	0.745307

Looking on the extreme portfolios, Sim1-Delta5, Sim5-Delta1, and comparing it to the Bottom-Top strategy in the previous section, there seems to be a minor improvement of using another sorting based on the delta of sentiment. However, the differences in the rows of Table 5 are statistically insignificant. Kruskal-Wallis test (commonly known as non-parametric ANOVA) does not reject the null hypothesis that differences in medians across different deltas in each individual row (Sim1, Sim2, …) are equal. Additionally, a sensible comparison is to set side by side returns of Decile 1 from Table 1 and Sim1-Delta5 returns, and also returns of Decile 10 and Sim5-Delta1. For the first case, the Wilcoxon test is used, and the difference in medians is statistically insignificant (p-value 0.284). For the latter case, the difference is also statistically insignificant (p-value 0.512). To sum it up, there seems to be a small improvement of double-sorting on the difference of sentiment, but the differences to a simple sort are not statistically significant. Also, the differences across different deltas in similarity quintiles are not significant. The results suggest that the positive similarity is a distinct anomaly and is not caused by the different sentiment of reports, but rather by the similarity alone.

Possible explanations

To my best knowledge, this paper is the first that documents the usage of positive language similarity in financial reports. Feldman et al. (2009), to some extent, studied tone change and numbers of positive and negative words. However, this is closer to the change of sentiment extracted by the Brain from the filings of companies. As it was previously mentioned, the most related research is Cohen et al. (2019). In this article, the conclusion is that companies have tendencies to repeat the most recent filings. Subsequently, the repetition creates similarity. A change creates dissimilarity, and according to the results of this paper, it is profitable.

On the other hand, in the Lazy prices, authors, focusing on overall changes, state that the difference is terrible news and stock returns subsequently deteriorate. I hypothesize that the dissimilarity of positive language is profitable because the companies have a greater motivation to change the filings if they think that it would influence investor positively. Therefore, the hypothesis is that the effort to change the positive language should positively impact subsequent returns because management does not have the motivation to change report if it would harm the company significantly. If there are possibilities to change as little as possible, it is natural to expect that the changes in positive language should be beneficial.

The literature recognizes the efforts of managers to plot the situation in a positive or misty rather than realistic way. As an example, Li (2006), has measured the readability of public company annual reports by examining the length of the report and the Fog index. Findings point to the tendencies to manipulate readers. Firstly, companies with lower earnings produce harder to read reports – reports are longer and have higher Fog. The conclusion is that management may be opportunistically choosing the readability to hide unfavourable information.

To sum it up, there are at least a few possibilities. Firstly, it can either signal a new piece of positive information. Secondly, the positive performance can be attributed to the better governance of the company, which can be proxied by the effort to change the report. Lastly, it can be a pure manipulation of the management, and there is a reversal in the longer term. However, results for 200 days sorting and rebalancing period were almost the same as for one month period. The more extended period includes two new reports (a piece of new information) and many chances for a reversal if the report would be to some extent manipulated to attract capital, which is the reason that this option is not very probable. Examination of the aforementioned hypotheses could be an aim of the following research in the future.

Conclusion

Textual analysis of company 10-K & 10-Qs can be a good addition in the portfolios. Past research has identified that the change (similarity) of reports could be utilized as a trading signal. According to the results, different types of language have distinct implications for the subsequent stock returns. Preliminary results have revealed that the similarity of positive language provided by the Brain Company can be a base of the most economically significant strategy among all languages considered. Such a strategy is significantly profitable as a long-only: the investor goes long the lowest similarity stocks. The difference of returns between the lowest similarity stocks and the highest similarity stocks is around 5% (yearly, with Sharpe ratio 0.84). Therefore the long-short portfolio may not seem to be as profitable compared to the long-only. However, it has significant risk-adjusted return and very consistent performance and can be very attractive for risk-averse investors. Moreover, the results are based on stocks with large market capitalization since the Brain analyzes company reports for approximately the largest 1000 US stocks, which ensures better liquidity and lower slippage costs and spreads.

Results also suggest that the low positive similarity effect is a distinct anomaly in the financial markets. Asset pricing models cannot wholly explain the performance of the strategy. Additionally, the change of sentiment extracted from fillings also cannot explain the positive similarity effect.

The exact mechanism that drives the low positive similarity effect is foggy. The hypothesis is that the companies have a greater motivation or are willing to change the filings if they think that it would influence investor positively. The effort to change the positive language should positively influence subsequent returns because management does not have the motivation to change report if it would harm the company significantly. A deeper examination of the underlying mechanism is an idea for further research.

Related literature

Carhart, M.M. (1997), On Persistence in Mutual Fund Performance. The Journal of Finance, 52: 57-82. doi:10.1111/j.1540-6261.1997.tb03808.x

Cohen, Lauren and Malloy, Christopher J. and Nguyen, Quoc, Lazy Prices (March 7, 2019). 2019 Academic Research Colloquium for Financial Planning and Related Disciplines, Available at SSRN: https://ssrn.com/abstract=1658471 or http://dx.doi.org/10.2139/ssrn.1658471

Davis, Angela K. and Tama-Sweet, Isho, Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases Versus MD&A (June 17, 2011). Available at SSRN: https://ssrn.com/abstract=1866369 or http://dx.doi.org/10.2139/ssrn.1866369

Fama, Eugene F. and French, Kenneth R., A Five-Factor Asset Pricing Model (September 2014). Fama-Miller Working Paper, Available at SSRN: https://ssrn.com/abstract=2287202 or http://dx.doi.org/10.2139/ssrn.2287202

Feldman, Ronen and Govindaraj, Suresh and Livnat, Joshua and Segal, Benjamin, Management's Tone Change, Post Earnings Announcement Drift and Accruals (May 11, 2009). Available at SSRN: https://ssrn.com/abstract=1287083 or http://dx.doi.org/10.2139/ssrn.1287083

Li, Feng, Annual Report Readability, Current Earnings, and Earnings Persistence (September 15, 2006). Ross School of Business Paper No. 1028, Available at SSRN: https://ssrn.com/abstract=887382 or http://dx.doi.org/10.2139/ssrn.887382

Pulliza, J. (2015). An Analysis of Speculative Language in SEC 10-K Filings. https://doi.org/10.17615/6kk9-xe26

Appendix A

Figure A1 The lowest positive similarity sector attribution plot

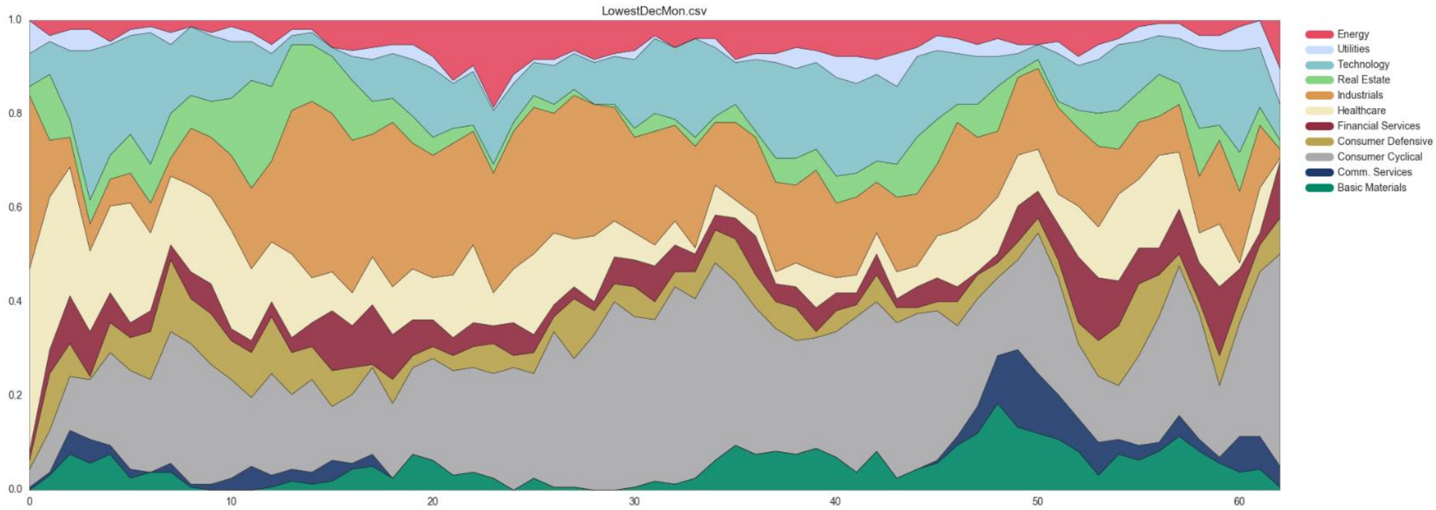


Figure A2 The highest positive similarity sector attribution plot