Analysis of Competitive Speech in Annual Reports: 1998-2017

June 15, 2018

Background

The Securities and Exchange Commission (SEC) requires that all major companies produce a 10-K report, an annual filing that details the financial performance and operations of a firm.\(^1\) A recent article in *The Economist* summarized the findings of a study by AXA Investment Managers Rosenberg Equities, which examined the language used in 10-K filings (“What Annual Reports Say,” 2017). The researchers found that the frequency of occurrence for the word “competition” and other related terms declined significantly over the past twenty years.

I assess the validity of these findings. The data used throughout this paper is constructed with the 10-K filings of the largest 3000 American firms over the twenty-year period from 1998 through 2017. These companies are collected using the membership lists of the RUSSELL 3000 market index, which are adjusted annually.\(^2\) The 10-K disclosures are obtained from the SEC’s online database, EDGAR.

Data

I automate the textual analysis of the 10-K filings using the computer language R. In addition to calculating the length of the reports, the program sums the total number of occurrences of the word “competition” and its related counterparts. This includes “competitive,” “competitiveness,” “compete,” and “competitor,” as well as plural forms. I eliminate filings with under 3,000 words when performing the analysis. After accounting for missing reports, the data set is found to consist of 55,817 firm-years.

I examine the use of competitive language at the market, industry, and firm levels. Industry classifications are obtained from the Bloomberg Terminal, and include Basic Materials, Consumer Goods, Consumer Services, Financials, Health Care, Industrials, Oil & Gas, Technology, Telecommunications, and Utilities.

The collection of historical annual reports is restricted by their availability. The EDGAR database contains few filings before 1995. In order to have the most robust data set possible, I restrict attention to a 20-year period.

Statistical and Graphical Summary

A topical statistical analysis of the data reveals that while the structure of a 10-K is standardized, its length across firms is highly variant. A histogram of word counts exhibits a right skew, with few firms completing annual reports of 100,000 words and greater. The grand mean for word count is 41,563 and the grand median is 39,483. In all of the following graphics, I do not exclude companies whose filings are greater than 100,000 words, but note that the results do not change significantly when these outliers are removed.

The grand mean for the use of competitive language is 37.34 words per document. During the twenty year period, the yearly mean rose from around 21 words to 48, with a similar trend for the median. The inclusion of the word “competition”, (and its capitalized form), rose from an average of approximately 7 times per filing to just over 11 in 2017.

---

\(^1\)See https://www.sec.gov/fastanswers/answersreada10-khtm.html for a detailed description of the 10-K filing.

\(^2\)The RUSSELL 3000 index consists of the largest 3000 American companies, with size calculated based on total market capitalization. ("RUSSELL U.S. Equity Indexes", 2018)
The average number of instances of “competition” and its related counterparts per 10,000 words fell from 12.5 in 1998 to just under 9 in 2017. Constructing the best linear predictor for competitive language frequency reveals that the year of reporting is highly statistically significant, and its coefficient is negative. I assume the following linear model:

\[ \text{CompFreq} = \beta_0 + \beta_1 \text{Year} \]

Table 1: Competitive Language Frequency over Time

<table>
<thead>
<tr>
<th>Year</th>
<th>CompFreq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>-0.216***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Constant</td>
<td>444.751***</td>
</tr>
<tr>
<td></td>
<td>(8.932)</td>
</tr>
<tr>
<td>N</td>
<td>55,817</td>
</tr>
<tr>
<td>R^2</td>
<td>0.041</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.041</td>
</tr>
<tr>
<td>Residual Std. Error</td>
<td>6.036 (df = 55815)</td>
</tr>
<tr>
<td>F Statistic</td>
<td>2,367.811*** (df = 1; 55815)</td>
</tr>
</tbody>
</table>

Notes: ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Even after controlling for industry, the year of reporting proves to be statistically significant at the 99% confidence level. This fact is unsurprising when considering the following graph. It is apparent that the
Overall decreasing trend is prevalent across most industries, particularly in the Technology sector.

The negative relationship identified in the above analyses raises the following question: Are the companies that compose the RUSSELL 3000 in recent years simply less inclined to discuss competition? To answer this, I consider the amount of time firms have remained in the index, summarized below.

While a large portion of companies in the RUSSELL 3000 maintained their membership for only a short period of time, many have remained in the index for the entire duration of the 20-year period. From the
completed data set, I construct a panel data set including only the companies in the latter group. This generates 832 observations. I perform a firm-level analysis.

I apply a fixed effects model to the panel observations. Again, reporting year proves to be statistically significant with >99% confidence, and its coefficient is negative. The same result is obtained when controlling for industry.

Frequency of Competitive Language as a Valid Metric

Several articles have been written concerning the growing length of 10-K filings in the past few years. The author can corroborate these findings; in fact, the increase in average length is surprisingly consistent over time.

One may pose the criticism, then, that pure word counts of competitive language serve as better metrics than the ratios considered extensively by this paper. It is true that these raw measures have increased over time across most industries. However, the author argues that the significant increase in word count should have been accompanied by a commensurate or greater increase in the use of competitive descriptive language. Many of the sections of the 10-K that were modified during the 20-year period were those that pertain most to the discussion of competition, including the formalization of Section 1.A Risk Factors and the addition of Section 7.A Quantitative and Qualitative Disclosures about Market Risk. Available literature supports this assumption as well: Dyer, Lang, and Stice-Lawrence (2017) find that risk factor disclosure is one of the three topics that contributed most to the lengthening of annual reports, along with fair value and internal controls. Thus, I have chosen ratios as the preferred metric.

---

3Over the past 10 years, the mean competitive word counts for companies in the Industrials category fell from around 100 to just under 80. It should be noted, however, that the sample size obtained for this industry is comparatively small.

4While a discussion of risk factors has always been required for 10-K filers, the SEC mandated that it become its own section in 2005. The addition of Section 7.A came after the passage of Sarbanes-Oxley (SOX), a reactionary piece of legislation following the Enron scandal.
Technical Appendix

The membership list for the RUSSELL 3000 index and industry classifications were obtained from the Bloomberg Terminal. All filings were obtained from SEC’s online database, EDGAR. When the 10-K by itself was not available, a complete submission text was taken and truncated to include only the annual report. Words are defined to be any string of alphanumeric characters preceded by a space. All filings were edited to remove HTML tags (through automation with R). Amendments to the annual reports were removed, and only 10-K and 10-K405 forms were incorporated into the data set.

Bibliography

