**First-Day Returns and the Intended Use of Proceeds**

**Laura Chen**

*Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA*

1. **Introduction**

When large companies announce that they are going public, or undergo an initial public offering (IPO), their shares are often priced less than their intrinsic value. This phenomenon is commonly referred to as IPO underpricing, wherein a new stock’s first-day closing price is above the initial offering price. A NY Times article noted that IPOs in the U.S. have been underpriced by about 16.8% on average over the last 50 years, meaning that there’s an estimated $125 billion in the last twenty years that companies could have earned in addition if their shares were not underpriced. Besides the U.S., many influential countries also suffer from this worldwide phenomenon, such as China, where underpricing has been very severe, averaging 137.4% within 20 years. Previous research has shown that there is a section in the IPO prospectus that discloses a company’s “use of proceeds,” including a specification of the company’s future goals and amounts they are committing to specific assets (Wyatt 625). Investors frequently glance over this section before making decisions to invest in the company’s stock to have an idea of the company’s future financial state and intentions. However, a close examination of the “use of proceeds” section reveals that firms disclosing vague information about their long-term projections typically exhibit higher underpricing. While the majority of the research has focused on the amount of uncertainty expressed in the IPO companies’ goals and underlying sentiment of SEC filings, there are still questions of what potential factors related to these companies’ intentions are the strongest indicators of IPO underpricing. By delving into the categorization of intended use of proceeds in the U.S. IPO market, we can obtain a better understanding of how IPO underpricing is related to what a given IPO company plans to use their proceeds towards and the sentiment of their intentions in text.

1. **Literature Review**

Textual analysis is increasingly being used to examine the tone and sentiment of corporate reports, newspaper articles, investor message boards, and much more in the business realm. Many methods have been publicized for use, and some of the most recognized ways include Tim Loughran and Bill McDonald’s six sentiment word lists, described below. Another developing approach involves the categorization of IPO companies’ intentions. These procedures have been found to be significantly related to many different financial and market indicators, some of which are stock return volatility and first-day returns.

Before starting to perform complex methods utilized in the field of textual analysis, it is imperative to understand the context that this type of analysis shall be used in and what various methodologies researchers have tested. Tim Loughran and Bill McDonald’s “Textual Analysis in Accounting and Finance: A Survey”details the nuances of textual analysis and quantitative procedures that have been used in relation to the business world. In their paper, Loughran and McDonald have recommended that a regular expression processor should be used when parsing text as part of a key component in any programming language. They noted that regular expressions, or regex, provides for efficient pattern searching within text, allowing the parsing of documents to be done with ease (Loughran and Mcdonald 1220). Besides suggestions and a discussion of methods, like using the natural log of gross 10-K file sizes as a simple proxy for linkage to stock return volatility, the two authors emphasized three main takeaways: (1) maintain a focus on readability, which is an attribute of financial disclosure that’s relevant to examining information assimilation into stock prices, (2) understand that a small number of words will dominate frequency counts and can produce outliers, and (3) avoid word lists and algorithms that are derived in the context of other disciplines. This last point is especially crucial in terms of textual analysis as different words can have a positive or negative connotation depending on what context it’s referring to.

The sentiment of a word in corporate documents has been found to be related to stock returns, volatility, and even investors’ decisions. In a separate paper by Tim Loughran and Bill McDonald, called “When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks”, the two researchers developed six word lists (negative, positive, uncertainty, litigious, strong modal, and weak modal) to accurately capture words and their sentiment in a financial context. Previous researchers had created sources for word classification, a common one being Harvard Psychosociological Dictionary’s Harvard IV-4 TagNeg (H4N) file, but it misclassified almost 73.8% of negative words in business literature. Hence, Loughran and McDonald (LM) sampled all 10-Ks and 10-K405s from SEC EDGAR’s website over 1994 to 2008 and utilized a bag of words algorithm along with term frequency-inverse document frequency (tf-idf) to construct their list of words. In addition, they examined all words occurring in at least 5% of the documents to consider their most likely usage in financial documents, placing them into their respective categories of word lists (Loughran and McDonald 44). Through a successful test of the word lists by examining the market’s reaction at the time of a 10-K filing, LM’s sentiment word lists were able to perform much better compared to the H4N file. LM’s word lists are now one of the standard sources utilized in parsing financial corpuses that is recognized in textual analysis.

More research has been done over the past years in relation to LM’s sentiment word lists, such as LM’s “IPO First-day Returns, Offer Price Revisions, Volatility, and Form S-1 Language”, which analyzes the tone of information contained in firms’ Form S-1. This form is the first SEC filing in the IPO process, therefore if an uncertain tone is prevalent throughout the document, it becomes more challenging for investors to accurately value the IPO. Based on a sample of 1,887 completed U.S. IPOs from 1997-2010 obtained from Thomson Financial Securities Data, LM parsed through these IPOs’ S-1 and 424 (final IPO prospectus) forms and measured document tone using their six word lists. From their findings, firms with higher uncertainty experienced higher first-day returns and the percentages of uncertain, weak, and negative words in the S-1 forms were much more powerful in explaining levels of underpricing compared to positive percentages (Loughran and McDonald 308). Furthermore, word frequencies in S-1 forms were just as good at predicting post-IPO return volatility as the tone of 424 forms, providing evidence that initial uncertainty is long-lived. While word sentiment demonstrates an impact on IPOs’ returns, there remains the question of whether other factors also have significant relationships with financial indicators.

Although previous studies have focused on word sentiment and how they relate to the market, another study has branched out to categorizing different purposes in what an IPO company intends to do with the firm after going public and finding connections between this categorization and predicting firm value. Anne Wyatt’s “Is There Useful Information in the ‘Use of Proceeds’ Disclosures in IPO Prospectuses”uncovers evidence of investors using information from the ‘use of proceeds’ section of IPO companies’ documents to distinguish IPOs with differing uncertainties and to predict firm survival. In a sample of 241 Australian IPOs between 1994-2000, Wyatt collected their use of proceeds disclosures along with financial characteristics of each firm at the time of the IPO. The most important step was classifying the use of the proceeds disclosures into different categories based on purpose (growth, production, financing activities, etc.) and whether specific assets were identified for growth or production investments. Through various linear regression models and significance testing, every category except for capital expenditures contained incremental information that was useful for predicting firm survival (Wyatt 653). The new approach of categorization brings an opportunity for further questions to be developed that are centered around whether it could be applied towards other market indicators and investor intentions.

1. **Data and Sample Construction**
   1. *Data*

This study is based on a sample of 2,462 IPO companies in the United States with an issue date between 1997 and 2019. The primary source of data from which the list of IPO companies was collected is Nasdaq’s IPO list. In addition, each IPO company’s ‘Use of Proceeds’ text was web scraped from Form 424B in the same source. Stock price information was retrieved from CRSP (Center for Research in Security Prices, LLC), which maintains a database of security prices, returns, and volume data for the NYSE, AMEX, and NASDAQ stock markets. Some of the main characteristics of the IPO companies include Filing Date, Issue Date, Company Name, main SIC code, ticker symbol, offer price, shares offered, stock price at close of offer, and 9-digit CUSIP number. These variables come from SDC Platinum, which offers a variety of financial databases.

* 1. *Variables*

In this study, our target variable is the amount of IPO underpricing, which we define as the percentage difference between the offer price and closing price of the first trading day. In addition, we control for the U.S. market using VIX Index, which measures the stock market’s expectation of volatility, and NASDAQ Composite Index, which is one of the most followed stock market indices in the United States.

Below, we describe several of the most relevant independent variables we intend to investigate to relate to the phenomenon of IPO underpricing, including sentiment of the ‘Use of Proceeds’ texts, categorization of ‘Use of Proceeds’, relation to the technology industry, and Google Trends interest volume.

* **Sentiment of ‘Use of Proceeds’:** Within each IPO company’s ‘Use of Proceeds’, we compute the proportion of the amount of sentiment represented by the text. There are six sentiment categories, which include positive, negative, litigious, strong modal, weak modal, and constraining as defined by Tim Loughran and Bill McDonald’s six sentiment word lists. Therefore, we calculate six values for each ‘Use of Proceeds.’
* **Categorization of ‘Use of Proceeds’:** Four categories are created to categorize IPO company intentions in the ‘Use of Proceeds’ text: acquisitions, financing transactions, production investment, and growth investment. These categories include a variety of words that best capture its activity category (See Appendix 1A for category dictionary).
* **Relation to technology industry:** The “Internet” variable represents a binary variable with the encoding 0=Not a tech company and 1=Is a tech company. A company is determined to be related to the technology industry using the internet identifications of Thomson Financial Securities Data, Dealogic, and IPOMonitor.com.
* **Google Trends interest volume:** We define an IPO company’s Google Trends interest volume as the average historical interest volume on Google Trends over 30 days before the IPO company’s issue date. Each day’s interest volume is an integer ranging between 0 and 100, with higher values representing more interest over time.
  1. *Methodology*

*3.3.1. Data Extraction*

To extract the ‘Use of Proceeds’ text for our data sample, we built a web scraping algorithm that would parse through the HTML document of Form 424B (the final SEC form submitted before going public) and extract the specified section for each IPO company on NASDAQ’s “IPO Calendar” list [8] between the years 1997-2019. Next, we accessed CRSP data on WRDS (Wharton Research Data Services) and extracted stock price information such as first-day closing prices and offer prices using SQL.

A Python dictionary was made for the four categories investigated, where relevant keywords were listed for each category (Appendix A). Category proportion is then determined for each IPO’s ‘Use of Proceeds’ text using a bag-of-words model.

To account for the sentiment of the ‘Use of Proceeds’ texts, Tim Loughran and Bill McDonald’s six sentiment word lists [4] are used to classify financial words in the text into their respective sentiment category. Sentiment proportion is then determined based on the amount of words in each category over the total words in the corpus.

* 1. *Data Summary Statistics*

Table 1 reports the summary statistics of sentiment category proportion for each of the six sentiment categories.

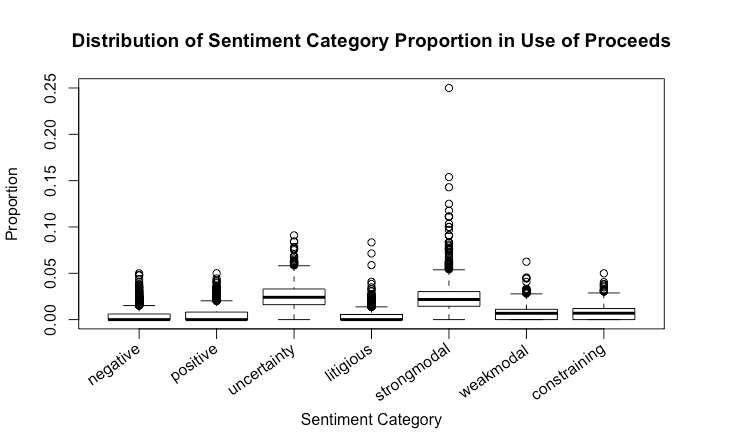
**Table 1.**

Figure 1 reports the distributions of category proportion of our data sample. All appear to be right-skewed except for Financing Transactions, which has a bell-shaped curve.

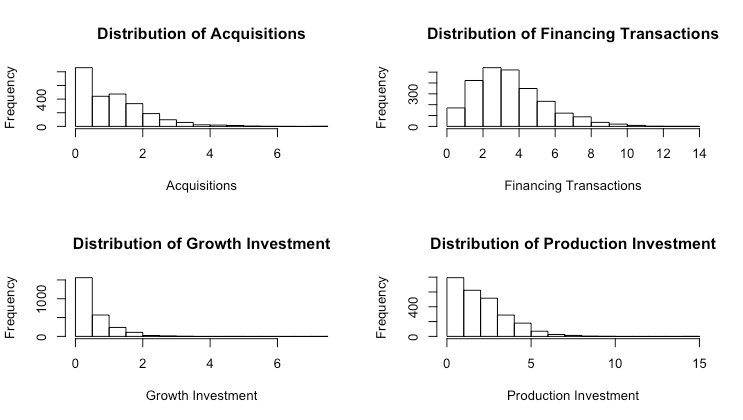
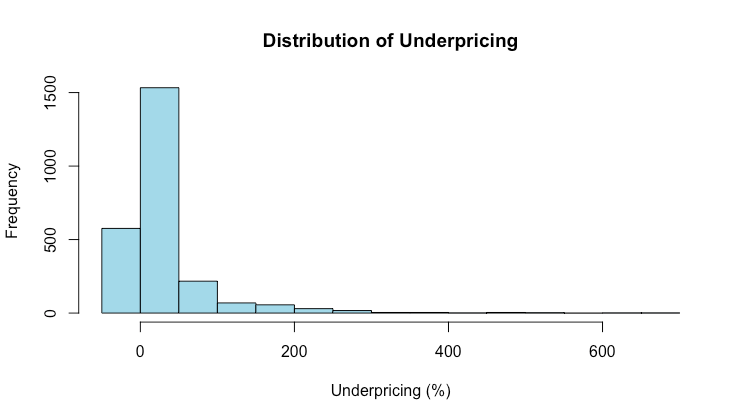
**Figure 1.**

Figure 2 reports a histogram distribution of IPO underpricing of our data sample. The distribution is skewed right, with an average underpricing of 30%.

**Figure 2.**



**Appendix A.**

In this appendix, we provide a dictionary of words related to each of the four categories investigated in the ‘Use of Proceeds’ text of IPO companies. As we parse through each ‘Use of Proceeds’ text, a frequency count is obtained for each relevant word for each category and normalized using the word count of the document corpus.

|  |  |
| --- | --- |
| **Category** | **Relevant Words** |
| Growth Investment | Joint venture, research and development, R&D, exploration, explore, growth, create, creation, partnership |
| Production Investment | Capital expenditures, expenditure, equipment, product, construction, office, store, facilities, licensing, expansion, commercialization |
| Acquisitions | Acquire, acquisitions, merge, acquisition |
| Financing Transactions | Working capital, repay, repayment, debt, loan, indebtedness, cashout, invest, investment, purchase, repurchase, securities, pay, paid, borrow, refinance, remittance |

**References**

1. FRED. “CBOE Volatility Index: VIX.” *FRED Economic Data*, https://fred.stlouisfed.org/series/VIXCLS. Accessed November 2020.
2. FRED. “NASDAQ Composite Index.” *FRED Economic Data*, https://fred.stlouisfed.org/series/NASDAQCOM. Accessed November 2020.
3. Loughran, Tim and Bill McDonald. “IPO first-day returns, offer price revisions, volatility, and form S-1 language.” *Journal of Financial Economics*, vol. 109, 2013, 307-326.
4. Loughran, Tim, and Bill McDonald. “Loughran-McDonald Sentiment Word Lists.” *Software Repository for Accounting and Finance*, University of Notre Dame, https://sraf.nd.edu/textual-analysis/resources/.
5. Loughran, Tim and Bill McDonald. “Textual Analysis in Accounting and Finance: A Survey.” *Journal of Accounting Research*, vol. 54, no. 4, 2016, 1187-1230.
6. Loughran, Tim and Jay Ritter. “Why Has IPO Underpricing Changed Over Time?” *Financial Management*, 2004, 5-37.
7. Loughran, Tim and Bill McDonald. “When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.” *The Journal of Finance,* vol. 66, no. 1, 2011, 35-65.
8. Nasdaq. “IPO Calendar.” *Nasdaq IPO*, https://www.nasdaq.com/market-activity/ipos. Accessed 2020.
9. Ritter, Jay. “A List of Internet IPOs.” *University of Florida Warrington College of Business,* https://site.warrington.ufl.edu/ritter/ipo-data/. Accessed November 2020.
10. Solomon, Steven. “Why I.P.O.’s Get Underpriced.” *The New York Times*, 27 May 2011, https://dealbook.nytimes.com/2011/05/27/why-i-p-o-s-get-underpriced/
11. Wyatt, Anne. “Is there useful Information in the ‘use of proceeds’ disclosures in IPO prospectuses?” *Accounting and Finance,* vol. 54, 2014, 625-667.