Explaining Underpricing Through the Textual Analysis of IPO Registration Statements

Tom Lee; University of Pennsylvania
Sarah Rice; University of Connecticut School
Steve Rock; University of Colorado at Boulder
Mike Willenborg; University of Connecticut

Contact: srice@business.uconn.edu

November 2009

This study focuses on the relationship between IPO underpricing and disclosure; specifically, disclosure related to the Risk Factors section of an IPO registration statement. Filing an IPO registration statement (e.g., S-1!) with the Securities and Exchange Commission (SEC) is required of all companies intending to go public. Within the registration statement, the intent of the Risk Factors section is to discuss any known risks relevant to the business, industry, and long term financial performance of the IPO; that is, to discuss any factors that make the offering speculative or risky to prospective investors. The Risk Factors section is among the first sections presented in an IPO registration statement, and thus it seems it would be a particular point of focus for investors. However it is unclear what, if any, effect the risk factors disclosure section has on IPO pricing.

When a company is preparing to go public, it often sets its initial (or IPO) price by estimating investor demand via a process termed book building and by going public in what is termed a ‘firm commitment’ offering¹. When a company’s shares are ‘underpriced’ this means the price at which shares trade in the aftermarket (typically proxied by the closing price on the first trading day) is higher than the the initial (or IPO) price at which the company sells its shares to the underwriter. The result is “money left on the table”, or IPO proceeds that could have gone to the company had offered its shares at a higher price. Theoretical models (e.g., Beatty and Ritter, JFE 1986) document conditions under which an inverse relationship exists between ex-ante uncertainty and IPO underpricing, thus it seems intuitive that risks disclosed in the IPO prospectus (as a proxy for ex-ante uncertainty) should be associated with the underpricing of the offering. Along these lines, certain papers in the literature document an inverse association between a count of the sheer number of risk factors disclosed in the IPO registration statement and IPO underpricing (e.g., Beatty and Welch, JLE 1996). However, to date the literature is silent regarding the association between textual analysis of risk factors disclosure in the IPO registration statement and IPO underpricing, thereby providing motivation for our study. We address this void by developing quantifiable measures to characterize the risk factors section of an IPO registration statement, and use these measures to test for an association with IPO underpricing. In particular, we are interested in whether some facet of IPO underpricing is seemingly associated with the market ignoring the information disclosed in the risk factors section of the registration statement, and therefore overvaluing shares above the offered price. With this in mind our first research question asks:

¹ A process where the underwriter takes orders from fund managers indicating the number of shares they wish to purchase and the price they are willing to pay for those shares.
1) Is there an association between the information disclosed in Risk Factors section of the IPO registration statement and IPO underpricing?

We develop additional tests to investigate underpricing in ‘hot’ IPO markets. Our goal is to determine if external market conditions impact the level to which the market impounds risk factor information disclosed in the IPO registration statement. For example, in a hot IPO market where excess underpricing is often a reality, might this be due to investor’s optimism, as evidenced by their willingness to ignore disclosed information about risk? To address this question we identify a market where IPO underpricing was clearly extreme, namely the time period from the fourth quarter of 1998 through third quarter of 2000, known largely as the dotcom bubble. During this era, average first day returns exceeded 60%, and the incidence of excessive returns (e.g., instances wherein a company’s shares doubled on the first trading day) among firms that went public was high. Moreover, unlike other time periods, the average three year return on those underpriced companies in the dotcom era show losses of up to 64% on average (Ritter and Welch 2002), indicating the initially high market value of these firms often vanished. This leads to the conjecture that while first-day prices closed much higher, there was no justification for favorable long term performance. This provides further evidence of mispricing, suggesting the market may have ignored vital information when pricing the IPO. Thus, our second research question asks:

2) Does the association between the information disclosed in the Risk Factor section of the IPO registration statement and IPO underpricing differ (decrease) during ‘hot’ IPO markets?

Text Analysis Methodology

To answer our research questions we extracted the risk factors section from each IPO prospectus of firms that went public from 1998 through 2002. We construct three sets of discrete and continuous variables to characterize the text within each risk factors section using a combination of syntactic and semantic automated text processing algorithms. The first set of variables is generated on each document independently. We use word and sentence counts to derive measures of length and complexity including the Flesch-Kincaid readability score suggested by former SEC Chairman Christopher Cox as a possible standard for financial regulatory filings. The second set of variables is comparative. Borrowing from information theory, we measure the Kullback-Leibler divergence between the word distribution of a particular filing and the expected distribution for all filings in the same year or within the same industry. The third set of variables are constructed using semantic analysis. We model a reference set of filings using Latent Dirichlet Allocation to extract word clusters that characterize the common topics raised in the risk factors section. Individual filings are classified based upon their inclusion or exclusion of these common topics. We control for firm specific characteristics, including size, age, and industry while also considering the underwriter, the audit firm and access to venture capital in our empirical analysis.

Conclusion
This work aids in explaining the relationship between risk factors disclosure and IPO underpricing, providing additional insight into how this association might be influenced by the type of IPO market (hot vs cold). Our work provides an important contribution to the finance and accounting literature striving to shed light on the anomaly of IPO underpricing. This study may also aid practitioners in developing ways to more accurately and efficiently set IPO prices.

Finally, we contribute to the academic disclosure literature by demonstrating an approach by which to characterize a section of text in the IPO prospectus. While this work is limited to just the risk factors section, we believe our methodology can be applied to other areas of the IPO prospectus. The ability to create a reliable and quantifiable characterization of the text within an IPO prospectus opens the door to a host of other questions, relating different elements of IPO disclosure to economic outcomes.

**Status**

We are in the process of finalizing our analysis and will have conclusive results to report at the Utah Winter Workshop in March 2010.