Signal Fitness and Verifiability, Divergence of investor Opinions, and IPO Underpricing

Sayan Sarkar
London Business School
Strategy & Entrepreneurship
ssarkar@london.edu

Abstract

Most past studies on investors’ responses to signals or information during fundraising pertain to examining the antecedents of average level of investor opinions. However, there is very limited knowledge on what might cause investors to converge or diverge in their opinions. The scope to extend knowledge is underscored by the fact that IPO underpricing, the key measure of capital-raising success at IPOs, is driven by the level of divergence rather than average level of investors’ estimates of the true firm value. I adopt a screening theory perspective and postulate that divergence, i.e. whether investors form heterogenous or homogenous outlook about investing in an IPO will be sensitive to characteristics of signals produced by the IPO firm. I theorize that higher divergence will be precipitated by lower levels of a) signal fitness and b) signal verifiability. I map these constructs to dimensions of IPO prospectus content, which I analyze using a mix of Topic Modeling and dictionary-based techniques. I exploit the proliferation of user-generated content on social media to use a novel method to compute divergence from express investor opinions, bypassing limitations of traditional divergence measures. I find support for my predictions. In addition, I demonstrate how investor sensitivity to signals is contingent on uncertainty a) in the environment and b) about managers’ motives. The study enriches the micro-foundations of organizational resource-acquisition strategy, besides informing the literatures on a) investor response to signals and information, and b) IPOs.
Signal Fitness and Verifiability, Divergence of investor Opinions, and IPO Underpricing

ABSTRACT

Most past studies on investors’ responses to signals or information during fundraising pertain to examining the antecedents of average level of investor opinions. However, there is very limited knowledge on what might cause investors to converge or diverge in their opinions. The scope to extend knowledge is underscored by the fact that IPO underpricing, the key measure of capital-raising success at IPOs, is driven by the level of divergence rather than average level of investors’ estimates of the “true” firm value. I adopt a screening theory perspective and postulate that divergence, i.e. whether investors form heterogenous or homogenous outlook about investing in an IPO will be sensitive to characteristics of signals produced by the IPO firm. I theorize that higher divergence will be precipitated by lower levels of a) signal fitness and b) signal verifiability. I map these constructs to dimensions of IPO prospectus content, which I analyze using a mix of Topic Modeling and dictionary-based techniques. I exploit the proliferation of user-generated content on social media to use a novel method to compute divergence from express investor opinions, bypassing limitations of traditional divergence measures. I find support for my predictions. In addition, I demonstrate how investor sensitivity to signals is contingent on uncertainty a) in the environment and b) about managers’ motives. The study enriches the micro-foundations of organizational resource-acquisition strategy, besides informing the literatures on a) investor response to signals and information, and b) IPOs.
Signal Fitness and Verifiability, Divergence of investor Opinions, and IPO Underpricing

INTRODUCTION

The ability to acquire financial resources is important to growth and survival of organizations, big or small. External financing can not only facilitate investments into manufacturing, research and development, or distribution capacity and capabilities but also increase visibility to potential customers and suppliers, and the media in general (Cumming, 2012:1). Besides, fundraising is often an important lever for wealth-attainment by entrepreneurs and existing investors (McGrath, 1999). Organizational outcomes at capital-raising exercises are a function of how potential investors respond to the investment opportunity that the organization represents. They, in part, base their investment decision by imputing the underlying firm quality and value from the signals transmitted by the firm (Cohen & Dean, 2005). Literature provides evidence that such signals could correspond to either a) “hard” information, such as about inter-organizational ties (Gulati & Higgins, 2003), patents (Heeley, Matusik & Jain, 2007) and social capital of CEOs (Fischer & Pollock, 2004); or b) “soft” aspects information, such as tone (Loughran & McDonald, 2013; figurative language (Cornelissen, Clarke & Cienki, 2012), or optimism (Anglin et al, 2018) as evinced through disclosures, pitches, presentations, etc.

Most studies examine antecedents of average opinion of prospective investors and its implication on fundraising outcomes. Consider that all investment decisions are decision-making under uncertainty. No investor can know for sure what the future holds for the capital-seeking organization (Knight, 1921) and can at best arrive at informed conjectures about “true” firm value or share price. Therefore, it stands to reason that investors will develop heterogeneous estimates about “true” firm value or returns to investing. While some will develop optimistic predictions and invest, others will be pessimistic and abstain from investing, a phenomenon termed as divergence of investor opinions (Miller, 1977). Divergence is inseparable from investment decision under uncertainty, yet, the literature sheds little light on distributional properties of investor opinion other than the average. The importance of divergence is
underscored by the fact that for capital-raising exercises that function like auctions – notably IPOs – it is the level of divergence of opinions of first-day investors, rather than their average opinion, that drive IPO underpricing, the key metric of capital-raising performance of the IPO firm (Miller, 1977; Rock, 1984; Hong & Stein, 2007). While much literature on IPOs has concerned itself with understanding the relationship between divergence and IPO underpricing, less attention has been accorded to understanding possible relationships between signals produced by the firm and divergent opinion-formation by investors.

The purpose of this paper is to address the research opportunity at the juncture of the two aforementioned puzzles. I ask: Can characteristics of signals produced by an organization determine whether investors converge or diverge in their opinion about whether to invest? I go on to examine the economic implications of variation in level divergence of investor opinions on resource-acquisition outcomes for IPO firms (and wealth-attainment outcomes for the entrepreneur).

To this end, I take a screening theory perspective (Stiglitz, 1975) and postulate that fundraising outcomes will eventually be driven by how boundedly-rational investors screen firms for quality – as a function of their interpretation of signals produced at the time of fundraising. The key principles undergirding screening theory are twofold; a) signals are meaningless without receivers. Simply put, investors have agency and can use signals to screen for quality regardless of the strategic intent behind the production of signals (e.g. Weiss, 1995). And b) signals have no intrinsic property that render them unequivocal or unambiguous; signals necessitates interpretation and hence investors have the scope develop heterogeneous interpretation of signals (e.g. Park & Patel, 2015). Based on these assumptions I develop theorization that leads me to predict that, level of divergence will be higher the lower the level of a) signal fitness (extent to which signal correlates with true, unobservable quality), and b) signal verifiability (extent to which receivers can verify signal-quality linkage); pertaining to business strategy, business
model, and risks and uncertainties thereof. I also predict that the main effects will be strengthened under conditions of a) *high industry uncertainty* and b) *low uncertainty about entrepreneurial motives*, which enable me to demonstrate how screening for adverse selection is a) contingent on conditions the larger decision-making environment and b) operates in conjunction with screening for moral hazard.

I assess support in the empirical context of about 450 IPOs of ventures that have gone public since 2015 in NYSE and NASDAQ, the leading stock exchanges in the US. I map the abstract signal dimensions to attributes of textual content of Risk Factors section of IPO prospectuses. I use dictionary-based and Topic Modeling techniques to analyse such content. For measuring divergence of investor opinions, I exploit the proliferation of user-generated content on social media. I use a novel way to measure divergence directly from actual opinions tweeted on a leading social-media investment platform, overcoming certain limitations of a conventional, trading-volume measure where opinions could be merely imputed. I find evidence that supports my predictions.

This paper has multiple implications for management theory and practice. First, it informs the resource acquisition literature about a novel set of antecedents that determine resource-acquisition performance at IPOs – a critical juncture in the life-cycle of a growing organization. Second, it informs the literatures on investor response to information and signals about *divergence* of investor opinions – and its antecedents and consequences – as opposed to *average* level investor opinion. Third, the study could also speak with the literature on IPOs by offering theories about factors that drives divergence in the first place, counterbalancing the focus on divergence-IPO performance relationships. This is also to the best of my knowledge joins a limited body of research that applies screening theory explicitly to an organizational context. This complements most extant work which is about its mirror image – signalling theory – i.e. analysing a signaling event from the signaller perspective. Finally, to entrepreneurs and top
managers, the study can generate insights on how to strategically produce information (while staying compliant and informative) to marshal higher cash proceeds at IPOs. Relatedly, investors could be made aware of how they interpret signals and information under uncertainty, thereby enabling them to make better-informed investment decisions.

CONCEPTUAL FRAMEWORK

Uncertainty, Divergence, and Capital-Raising

Divergence of investors’ opinions ("divergence") can be termed as the phenomenon where a pool of prospective investor develops heterogeneous estimates or outlooks about the returns to investing in an investment opportunity (Bamber, Barron & Stober, 1997; Miller, 1977; Hong & Stein, 2007). For the purposes of this paper, such investment opportunities refer to equity investments into entrepreneurial ventures or large corporations, either public or private. Labels such as investor disagreement (Ferson & Lin, 2014) and belief dispersion (Dontoh & Ronen, 1993) have also been used in past research to denote the same concept. To illustrate, in a hypothetical crowdfunding round or a stock market offering, it is implausible to expect all prospective investors to either invest, or abstain, en masse. Rather, varying degree of potential investors invest and abstain, i.e., there is some divergence. Divergence can be said to be high when investors widely vary in their private estimates of firm value (Miller, 1977) or when the investor pool exhibits a somewhat even mix between optimists and pessimists, without one camp dominating the other. Divergence is low when investors either converge towards an optimistic or a pessimistic outlook (i.e. optimists vastly outnumber pessimists or vice versa), or have a narrow range of private estimates of the firm value.

At a capital-raising exercise it is impossible for investors to know the precise dimensions of uncertainty on which future returns depend on, let alone the precise distribution of future returns. This is a classic scenario of Knightian (1921) uncertainty. (This is contrast with pure risk, such as a game of dice. Admittedly the exact future state is unknown, but the underlying
distribution of possible returns is known with certainty.) The only way for investors to proceed is to use the available information at hand, and develop (ideally, informed) conjectures about the future value-creation potential of the organization’s strategy. It stands to reason the such subjectivity and divergence go hand-in-hand. If the future is so uncertain, it is unrealistic to believe that investors will somehow arrive at a singular expectation about returns to investing in the fund-seeking organization (Miller, 1977).

The importance of divergence to management theory is underscored by the fact that when a capital-raising exercise is designed or functions like an open-bid auction, it is the level of divergence, rather the average level, of investor opinions that drive capital-raising outcomes. Seminal work in financial economics show evidence towards the existence of a “winner’s curse”, whereby an auction winner is cursed in a way as they end up overpaying the most, relative to the average bid of all bidders (Thaler, 1988). This phenomenon has been directly mapped to IPOs by scholars, who have demonstrated that US IPOs function very much like auctions, given that the SEC bans short sellers from the first 30 days of an IPO (Ljungqvist, 2007). We know that IPO underpricing rises monotonically with the level of divergence among prospective investors on the first day of IPO trading, rather than the average level of their opinions (Miller, 1977).

IPO underpricing, or the ratio of a) the difference between closing price at first day of trading at an IPO and the offer price, and b) offer price is the universal metric for measuring capital-raising performance at an IPO (e.g. Pollock & Rindova, 2003; Pollock & Gulati, 2007). High underpricing implies the entrepreneur or existing investors receive less cash proceeds from equity sale than it would have been possible in the wake of the market-determined price. This is because the firm only receives proceeds as per the lower offer price, while new, first day investors hold stocks now valued at a higher closing price. IPO underpricing represents “money left on the table”, i.e. an inferior resource-acquisition or wealth-attainment outcome for the organization and the entrepreneur, respectively (Certo et al, 2011). Thus, a better understanding
of antecedents of variation in divergence levels can lead to a better understanding of resource-acquisition performance at a crucial, high-profile juncture of a growing organization.

**Puzzle 1: Average Level vs. Divergence of Investor Opinions**

An important step towards opening the black box of divergence is to review extant research on investor response to signals and information under uncertainty. Fundraising is characterized by information asymmetry between market participants – the organization has superior information about its quality than prospective investors (Cohen & Dean, 2005). To alleviate this asymmetry, fund-seeking organizations produce and disseminate information about its strategy to potential investors – either voluntarily (Kim et al, 2017; Ibrahim, 2015), or in order to ensure regulatory compliance (Daily, Certo & Dalton, 2005; Lowry & Shu, 2002). This is mainly done through business plans, online pitches, live presentations (generally by entrepreneurial ventures) or financial disclosures (generally in the public equity markets).

Boundedly rational investors interpret these signals, impute firm quality to estimate returns to investment, and make their investment decisions. The first group of literature discusses how signals may emanate from “hard” information, ranging inter-organizational ties (Gulati & Higgins, 2003), patents (Heeley, Matusik & Jain, 2007) and social capital of CEOs (Fischer & Pollock, 2004). The second group emphasizes that investors ultimately respond to subjective constructions of the investment opportunity. The focus is on signals based on “soft” attributes of the information, such as tone (Loughran & McDonald, 2013; figurative language (Cornelissen, Clarke & Cienki, 2012), or optimism (Anglin et al, 2018)

While this research provides rich insights into investor decision-making and its organizational implications, it primarily deals with the first statistical moment of investor response, namely the average optimism or pessimism level or funding propensity of the investor pool. We do not, however, know much about other distributional properties of investor response. The opportunity to extend this literature is underscored by the fact that one such property –
divergence – has a significant bearing on arguably the most important capital-raising round in the life cycle of a growing organization. Opening the black box of divergence could indeed add to knowledge about organizational resource-acquisition and investor decision-making under uncertainty. This is the first research opportunity that motivates this study.

**Puzzle 2: Divergence as an Antecedent of IPO Underpricing**

The IPO literature offers four main theories on the determinants of underpricing. First, entrepreneurs could intentionally under-price their stock in order to stimulate demand from investors by offering them the prospect of making immediate gains from holding the stock (Ljungqvist, 2007). Second, firms could also intentionally under-price their stock as an insurance against liability arising out of future litigations (Drake & Vetsuypens, 1993). Third, investors could seek higher compensation for fundamentally riskier firms (e.g. Filatotchev & Bishop, 2002) or for learning more about these firm’s true values (Sherman & Titman, 2002). Finally, the share price could rise relative to the offer price simply because investors disagree about their subjective estimates about the true value of the IPO firm (Miller, 1977; Hong & Stein, 2007). A situation akin to “winner’s curse” unfold, and greater the spread of estimates of “true” share price or rift between optimists and pessimists, the higher the first-day investors have to pay for buying the stock relative to the offer price.

While the IPO performance consequences of divergence are well-documented, possible relationships between divergence and characteristics of the very signals produced by the IPO firm remains an unanswered puzzle. At the extreme, two otherwise similar IPOs may witness different IPO underpricing simply because in one case, the investors converge towards a pooling equilibrium about their interpretation of signals produced by the IPO firm, and in the other case, investors end up developing heterogeneous interpretation of signals, resulting in a separating equilibrium between optimists and pessimists. Opening the black box of divergence – and better
understanding drivers of IPO underpricing – is the second research opportunity that motivates this study.

**Screening for Adverse Selection by Investors**

Fundraising events such as IPOs are characterized by information asymmetry between various parties. On one hand, the entrepreneur and managers of the fund-seeking organization have superior information about their strategy, business model, future prospects, and risk and uncertainty therein (e.g. Zhang & Wieserma, 2009; Heely, Matusik & Jain, 2007). On the other hand, prospective investors have incomplete information and must rely on signals produced by the organization (e.g. Certo, 2003). Signals are information or actions that can reveal unobservable characteristics of the sender. Simply out, a signal is information or action A that indicates underlying quality or phenomenon B is true (Kirsch, Goldfarb & Gera, 2009). See Figure 1 for a graphical representation of how capital-raising events such as IPOs and its related processes can be characterized as a signalling environment.

---

**Insert Figure 1 about here**

---

It comes as no surprise that the concept of signaling environment has been mobilized in management research to model various interactions between an organization and its stakeholders. However, most studies attribute much agency to the signaler, e.g. the IPO firm in the present case. Simply put, *signal production* varies according to organizations’ true type, resulting in a separating equilibrium between high-quality and low-quality organizations (Spence, 1978; Weigelt & Camerer, 1988). Implicitly, it has been assumed that signals evoke unequivocal interpretations by all receivers. Therefore, it is solely the heterogeneity in *signaler type* that drives the nature of the resultant equilibrium of receiver opinions.

Limited theoretical attention has been accorded to the other party in such transactions, the signal receiver, e.g. prospective IPO investors in the present case. Screening theory is the less-
examined mirror of signaling theory, that postulates that ultimately, receiver response to a signal is a function of their interpretation of the signal (Stiglitz, 1975; Weiss, 1995; Kang, 2008; Sanders & Boivie, 2004). The central actor or agent is the receiver, who, in this case, are prospective investors. Consider two simple principles that undergird “screening theory” – a theoretical lens that I will use to peer into the black box of divergence. First, signals are incomplete without the receiver. In reality, the outcome of any signaling exercise is eventually driven by investor response to signals (Connely et al, 2001). Receivers can exert agency while interpreting signals, regardless of the signaler’s strategic intent underlying signal production. Second, signals do not have intrinsic properties that interpret themselves, or render them unequivocally or unambiguously. Signals necessitate interpretation (Heil & Robertson, 1991) by receivers. In a fundraising context the receivers, i.e. investors are boundedly rational who in reality, develop subjective interpretations of signals (March, 1978). It is thus implausible to expect a uniform level of converge or divergence from one signaling event (i.e. IPO) to another.

I postulate that divergence of investor opinions – or variation to the extents to which investors converge or diverge about their outlook on an IPO could be sensitive to the very characteristics of the signals. I define the scope of the study in terms of two important but less examined signal dimensions from the information economics school. The first property is signal fitness (Park & Patel, 2015; Park & Mezias, 2005). On one hand, investors are more likely to be able to clearly impute an IPO firm’s future value-creation potential from certain signals. On the other hand, certain other signals can be unreliable, i.e., boundedly-rational investors may not be able to parse out implications for firm future value from “noise”. The second property is signal verifiability (Hermalin & Katz, 1991). On one hand, certain signals can be more standard or commonplace in nature, i.e. conforming to established “types”. Signal receivers (i.e. IPO investors) have an established base of knowledge and heuristics to draw upon to aid their ability to verify the implications of such signals. On the other hand, other signals could be unique and not conforming to established types. Signal receivers can only access a limited or non-existent
body of prior knowledge and heuristics to interpret these and therefore, less verifiable. Note that both fitness and verifiability are distinct from the notion of signal credibility, which simply has to do with whether receivers think a signal is true or honest (Arthurs et al, 2008). The notions of fitness or verifiability can operate independent of signal credibility, thereby obviating the need for making any restrictive assumptions about signaler (i.e. IPO firm) agency.

In the following section, I will map these signal dimensions to attributes of “soft” information in IPO prospectuses in order to develop empirically-tractable hypotheses. The IPO prospectus “describes the strategic, operational, and financial status of an IPO firm” to potential investors (Park & Patel, 2015). IPO prospectuses are a mix of past track record and planned future strategy. It is also a mix of standard content following SEC guidelines and content authored at the discretion of IPO firms. These allows for a potentially meaningful variation in prospectus content, and hence variation along the two signal dimensions.

An IPO prospectus is a legal document and the Board of Directors are liable for misrepresentation and concealment of any material information (Hanley & Hoberg, 2012), which underscores the sanctity and comprehensiveness of the prospectus to prospective investors. In addition, the SEC stipulates a “silent period” between filing of the final IPO prospectus and IPO launch on the stock market, wherein the IPO firm can disseminate no new material information (Deloof et al, 2009). Thus, scholars argue that all firm-value-relevant information should eventually be ascribed to the IPO prospectus. In summary, the IPO prospectus can be considered to be the definitive, comprehensive, and original source of information and signals for prospective investors. This is substantiated by widespread empirical support that most investors depend on the prospectus to evaluate an IPO firm – both in academic research (e.g. Mather, Ramsay & Steen, 2000; Hanley & Hoberg, 2010) and business press (e.g. zacks.com; investopedia.com).

**Signal Fitness: Market- or Technology-Focus**
It is obvious that an IPO prospectus will have a mix of content that pertains to the a) market or demand side of its strategy, business model, and risks and uncertainties; as well as b) the technology or supply side of its business (Gulati & Higgins, 2003). I argue that investors are likely to process and interpret market signals in a fundamentally different way from technology signals. The logic is as follows. The entrepreneur or senior management of the IPO firm has much less control over how the market will shape in the future, than about product or technology development. In other words, the sources of uncertainty in market are largely exogenous to firm actions (Tong & Li, 2011; Kaplan & Stromberg, 2004). Imputing the impact of signals pertaining to exogenous factors on future growth of an IPO firm (and hence, returns to investing in the IPO firm’s stocks) requires a two-step estimation from a potential investors. First, not only they have to predict how shocks outside firm control will pan out in the future but also how the fund-seeking organization will respond to such shocks. Second, the prediction exercise is further complicated by the possibility that different organizations will have asymmetric responses to exogenous uncertainty. The nested nature of the prediction task makes the linkage between firm signals and quality less clearly imputable. In other words, these signals have low fitness or “noisy” in nature, which induces heightened subjectivity and conjecturing from investors.

In contrast, the entrepreneur or senior management of the IPO firm high agency over the product or technology that will be developed by them - either within firm boundaries or through contractual agreements (Kaplan & Stromberg, 2004). In other words, uncertainties emanating from product or technology are largely endogenous in nature. These uncertainties can be reduced through firm actions. In addition, investors have disciplining mechanisms – capital allocations and corporate governance – at their disposal to induce the firm to do so. Thus, imputing the linkage between endogenous uncertainty and future firm value is essentially a one-step estimation for the investors, and the strategy-performance linkage much clearer. It is paradigmatically less noisy than the exogenous uncertainty-firm value linkage.
The more the layers of uncertainty, i.e. the lower the perceived fitness of signals, the higher the scope for subjectivity and conjecturing. Thus, while interpreting market-related content in IPO prospectus, I expect investors to be more induced to draw upon idiosyncratic prior information sets, beliefs, and biases than while interpreting technology-related content, increasing the likely spread of their opinions. All else being equal, divergence of opinions should be higher when market-related signals dominate technology-related signals. Thus, even if we assume that all IPO firms are otherwise identical and are producing signals rationally, we could have a separating equilibrium – between optimistic and pessimistic investors - simply on account of low perceived fitness of such signals. In contrast, when technology-related signals are in preponderance, prospective investors gravitate towards a pooling equilibrium – either optimistic or pessimistic because the signals are perceived to be high-fitness, which facilitates clear interpretation.

The organizational manifest of high divergence will be high IPO underpricing, which translates to “money left on the table” for the IPO firm and its entrepreneurs and existing investors. In formal terms,

\[ H1: \text{The higher the prevalence of market-related content to technology-related content in IPO prospectus the higher the level of divergence of investor opinions at IPO.} \]

**Signal Verifiability: Standard or Idiosyncratic Content**

Firm strategy can be at once unique and commonplace. For any organization, its strategy, business model, etc. will have some aspects that are unique or idiosyncratic to the organization. At the same time, it will plausibly have aspects of strategy or business model that it shares with comparable organizations. Following this logic, an IPO prospectuses will have a mix of idiosyncratic content unique to itself, and “standard” or commonplace information or signals that do not vary much among comparable IPO firms (Hanley & Hoberg, 2010). (In line with past research, I define comparable IPO firms as a) rival firms, who are competing for the same
product-market opportunities, or b) contemporaneous firms, who are vying for the same cyclical financial capital flows and witnessing the same variations in macroeconomic conditions.)

Standard content conform well to established body of knowledge and heuristics that investors can draw upon for interpretation. Thus, these are likely to be perceived as more verifiable signals. Verifiable signals are likely to have been analyzed in greater detail than unverifiable signals, so investors will have at their disposal a greater array of prior knowledge and heuristics to impute firm quality from verifiable signals. Furthermore, it is likely that verifiable signals correspond to well-documented success drivers or risks to the businesses. This means that the firm can be predicted to choose from a narrow, constrained set of strategic alternatives in the future. It is reasonable to believe that the signal-quality linkage is clearer in the case of verifiable signals, thereby eliciting greater convergence from the investors.

This is less so in the case of unverifiable signals emanating from idiosyncratic IPO prospectus content. The task of interpreting them is more open-ended in nature owing to no or limited established knowledge base to refer to. Also, they could pertain to singular business success factors and risks, which are likely to elicit a wider set of possible strategic actions by the firm. Thus, the import of unverifiable signals on firm value or returns to investing is likely to be interpreted by investors a greater number of ways. Note that this will lead to a separating equilibrium between optimist and pessimist regardless of strategic intent of the signaling IPO firm.

The organizational impact is manifest in the form of higher IPO underpricing or, inferior resource-acquisition for the IPO firm. In formal terms,

\[ H2: \text{The higher the level of idiosyncratic signals relative standard signals in an IPO prospectus the higher the level of divergence of investor opinions at IPO.} \]

**Uncertainty in the Environment**
Strategy evaluation during IPO – in essence screening for adverse selection – does not take place in a vacuum where investors are immune to all other information, stimuli, or considerations. I turn my attention to two non-firm factors that contextualizes investors’ response to signals. First, consider the fact firms are nested in its environment. To that end, investor response firm signals is likely to be moderated by response to uncertainty in the broader environment in which the fund-seeking firm resides in. I pick one salient level where firms are nested – an industry.

Consider that not all industries are perceived to be equally uncertain. In some industries, the linkage between strategy and performance is easy to decipher, while in other industries the linkage is less clear (Heeley, Matusik & Jain, 2007). One important dimension in which industries differ is knowledge-intensiveness. Knowledge intensive assets are difficult to value for outsiders. There is greater scope for information asymmetry both in terms of asset or strategy quality and entrepreneurial motives (Barth et al, 2001). This engenders a closer scrutiny of firm signals at IPO, relative to IPO firms from more physical asset-intensive industries. After all, physical assets are easier to value, and therefore terminal cash flows to firm are more easily estimable. This means in knowledge-intensive industries, investors are likely to weigh firm signals and firm-specific information more heavily than other extraneous considerations.

The heightened sensitivity to IPO prospectus based signals by the firm is expected to amplify the baseline effects of signal reliability and conformity on divergence of investor opinions. In formal terms,

\( H3a: \) The effect of \( H1 \) will be stronger in knowledge-intensive industries than in less knowledge-intensive industries.

\( H3b: \) The effect of \( H2 \) will be stronger in knowledge-intensive industries than in less knowledge-intensive industries.
Uncertainty about Managerial Motives

When investors evaluate an IPO for potential investments, it is easy to assume that they evaluate the firm strategy in conjunction with its managers. It is well-documented that investors jointly evaluate the “jockey” and the “horse” while making investment decisions (Kaplan, Sensoy & Stromberg, 2009). In the case of an IPO, this translates to prospective investors not only a) imputing the impact of strategy and business details as disclosed in an IPO prospectus on firm value (i.e. screening for adverse selection) but also b) imputing whether the management can be relied on to deliver or deviate from the stated strategy (i.e. screening for moral hazard).

I argue that investors will place greater emphasis on the IPO prospectus when the payoffs of senior management is highly dependent on firm performance. When incentives are aligned, investors are likely to assume that the management will stick with the strategy stated in the IPO prospectus (Werner & Tosi, 1995). This means that investors are likely to weigh firm signals and firm-specific information more heavily than other extraneous considerations. The heightened sensitivity to IPO prospectus based signals by the firm is expected to amplify the baseline effects of signal reliability and conformity on divergence of investor opinions. On the other hand, when incentives are delinked, the management is assumed to be less likely to stick with the stated strategy. Investors become less likely to weigh heavily information from IPO prospectus, thereby attenuating the baseline effects of signal reliability and conformity. In formal terms,

\[ H4a: \text{The effect of } H1 \text{ will be when managerial incentives are strongly linked with firm performance than when the linkage is weak.} \]

\[ H4b: \text{The effect of } H2 \text{ will be when managerial incentives are strongly linked with firm performance than when the linkage is weak} \]

METHODOLOGY

Context and Data Sources
I choose the empirical setting of previously private-held organizations going public at NYSE and NASDAQ stock exchanges in the United States. The window of observations in this study spans from 1st January, 2015 to 31st October, 2018, and covers a sample of 495 IPOs. It may be worthwhile to observe that business press and social media can at times exhibit differing views of what constitutes an IPO. Thus in line with much past work (e.g. Pollock & Rindova, 2003) I plan to restrict my sample only to “true” venture IPOs by disregarding the following six categories of equity issuance. a) IPOs by real estate investment trusts, b) IPOs by hedge or mutual funds, c) floating of American depository receipts (except where the ADR is the first time the venture is issuing public equity globally), d) Follow-on public offers (FPO/ SEO) and e) exchange offers for other financial instruments.

I create a proprietary dataset on US IPOs by combining data from several public and paid databases, through a mix of hand-collection and automation using computer programs. I first determine the identity of all IPO firms from NASDAQ website (Grable, Lytton & O’Neill, 2004). NASDAQ maintains a comprehensive database of all US IPOs regardless of the stock exchange a firm sought listing in. Thereafter I collected daily stock price and trading volume data from Yahoo Finance website. I also validated Yahoo Finance data against CRSP database (Dezso & Ross, 2012) and found the data to be consistent across sources. I hand-collected the IPO disclosures (such as prospectuses and registration documents) from Security and Exchanges Commission’s EDGAR database (Carpenter & Pollock, 2003) and coded the relevant prospectus-based variables manually. I collected investor opinion data from StockTwits, a leading investor social-media platform (Cookson & Niessner, 2016) using a Python program. For creating prospectus-based data I variously used a) Linguistic Inquiry and word Count (LIWC) software (Pennebaker et al, 2007) for dictionary-based variables, and b) Python program that implemented parts of the Stanford Topic Modeling Toolbox (nlp.stanford.edu) for variables created using Topic Modeling analyses.
The majority of studies utilizing textual analysis of IPO prospectus content have considered the whole IPO prospectus as the source of information or signals. However, I contend that investors have limited cognitive capacity and are unlikely to read all parts of a prospectus, often running over 100,000 words or 300 pages, thoroughly. To this end, I conducted semi-structured interviews with seven seasoned professionals from the investing domain—equity investors, investment bankers, and stock market Directors. My interviews revealed that investors are likely accord importance to only select parts of the prospectus, with the Risk Factors section being the unanimous opinion of the most value-relevant section. The following interviewee quotes corroborate this finding.

“Reading the full prospectus is not possible as it can reach 300 pages. But I make sure I read the risk factors section as it contains detailed, pertinent information that the media won’t cover.”
- VP - Investment bank, with large personal investments in equities

“In the fog of optimistic news, it is the risk factors section that portrays the true colours of the company going public. What are the genuine, hidden roadbumps?”
- Director, Equity Primary Markets, Global top-10 Stock exchange

In this light, I depart from some past studies in considering the content of only the Risk Factors sections of prospectus for appropriate parts of my analysis.

Variables

Capital-raising performance. Consistent with much past work on IPOs, I opt to measure this as IPO underpricing or first-day IPO returns, which is the relative change in a venture’s firm value on the first day its shares trade on the public stock market (e.g. Pollock & Rindova, 2003). I collected daily stock price data from Yahoo Finance website. This measure is expressed as:
**IPO underpricing** = (Share price at close of first-day trade – offer price per share)/ Offer price per share

**Divergence of investor opinions.** This refers to differences in investors’ opinions about whether to invest into a venture going public. For long, researchers have imputed investor opinion from individual-level trade data (which itself is often proprietary to stock trading platforms). Even if such data is used, it is difficult to separate the empirical effect of divergence, since investors can have heterogeneous motives behind actual trades. For example, an investor can trade because of liquidity, a consideration orthogonal to his opinion on the venture’s prospects. To address this challenge, I exploit the recent proliferation of user-generated content on social media to collect investor opinion as explicitly expressed by them, bypassing the need for proxies. In line with Cookson & Niessner (2016), I collected and parsed tweets posted by investors on StockTwits, a leading social media-investing platform, using a proprietary Python program.

StockTwits was founded in 2008 as a social networking platform for investors to share their opinions about stocks. On StockTwits, users - actual and potential investors – regularly express their opinions predicting future changes to stock price/ firm valuation and categorize themselves as bullish (optimistic) or bearish (pessimistic). The website has a Twitter-like user-interface where participants post tweets of up to 140 characters. They use "cashtags" (similar to hashtags) with the relevant company’s ticker symbol (e.g. $AAPL for Apple, Inc.) to link one’s tweet to a particular company. See Exhibit 4 for a sample of tweets made by investors discussing the Snapchat IPO. Past work (e.g. Cookson & Niessner, 2016; Antweiler & Frank, 2004) has shown that opinions expressed at leading investor stock message boards are representative of opinion of all investors at large. See Figure 2 for a snapshot of StockTwits interface. For each message or “tweet” I also observe a user identifier, the tweet text, and user-supplied indicators
for sentiment (bullish, bearish, or unclassified), and the corresponding cashtag. See Figure 3 for sample investor profile.

---

**Insert Figures 2-3 about here**

---

I follow Antweiler and Frank (2004) to construct a divergence measure that reflects the optimism (bullishness) and pessimism (bearishness) at a firm level. The relevant set of tweets for each IPO is all tweets published starting seven days before the IPO, until 4PM, the equities market trading closing time, on IPO opening day. For any firm,

\[
Divergence = 1 - \frac{\text{abs}(N_{\text{bullish}} - N_{\text{bearish}})}{(N_{\text{bullish}} + N_{\text{bearish}})}
\]

Where \(N_{\text{bullish}}\) and \(N_{\text{bearish}}\) are the number of tweets with bullish and bearish sentiment tags, respectively.

In my sample, close to 70% of all tweets are unclassified, i.e. neither bullish nor bearish. One potential concern is that users explicitly classify tweets when they are more certain of their sentiment than when they leave sentiment unclassified. However, Cookson & Niessner (2016) indicated that unclassified tweets tend to have very similar distributional properties with respect to sentiment-classified tweets. This leads me to assume that many otherwise strongly-bullish or -bearish users simply leave their message unclassified because of lack of time or motivation. Furthermore, a scrutiny of a small sample of unclassified messages indicated that there exists a fair proportion of messages where it is evident the actual sentiment expressed was unequivocally bullish or bearish. I manually read all unclassified tweets and tagged appropriate tweets (a further 10% of total) from this pool as either bullish or bearish and updated the dataset. I preferred a manual approach to an algorithmic or dictionary-based approach for this step. Many
messages on StockTwits contain sarcastic or disparaging language while expressing optimism/approval or pessimism/disapproval. Prior work has shown that machine learning or dictionary based approaches perform unsatisfactorily in understanding such textual content.

**Market focus (H1).** I used a dictionary-based approach to compute this measure. I first carefully read a sample of prospectuses – spread across multiple industries and years – to come up with two bags of words. The first corresponded to *market* words, such as X and y. These are generally used to describe strategy, business plans, risks and uncertainties, future predictions, etc. in relation to the demand-side, i.e. customers and the market. The second bag-of-words correspond to *technology* words. These are generally used to describe strategy, business plans, risks and uncertainties, future predictions, etc. in relation to supply-side i.e. technology, product development, and manufacturing. I then calculated the word frequencies of both bags-of-words by inputting the two custom-made dictionaries in Linguistic Enquiry and Word Count (LIWC), a text analysis software widely used in social science application (Pennebaker, et al. 2015). Finally, I coded *market focus* as a binary variable where the difference of *market* words frequency and *technology* words frequency was over the sample median, and 0 if below sample median.

**Similarity (H2).** I use Topic Modeling techniques in order to measure similarity of prospectus content between two IPO firms. First, in line with prior research, I limit the consideration set for each firm to all firms that either a) belong to the same industry, for any given year in the sample or b) IPOed in the same year, regardless of which industry they operated in (Hanley & Hoberg, 2010). Thereafter, I used Topic Modeling technique to create a document similarity score for each pairwise combination of prospectuses. I implemented the above algorithm using a Python computer program, setting the number of topics to 1000 and
number of iterations to 1,000. The document similarity metric used was Jensen-Shannon (JS) Divergence, which takes a value between 0 and 1 depending on distance between the distribution of topics of two documents (Lin, 1991). The final measure for a firm was the simple average of all JS scores for all pairs of prospectuses where a) one firm was the focal firm and b) the other firm belonged to the consideration set defined above.

Knowledge-intensiveness (H3). Synthesizing prior research on knowledge-intensive (versus physical-assets intensive) industries, I label industries within 2-digit SIC codes 28 (chemical and allied products) and 73 (business services) as knowledge-intensive. These SIC codes subsume all biotechnology and information technology firms in my sample, respectively. Collectively they account for 52% of firms in my sample. This is coded as a binary variable which equals 1 if a firm is from a knowledge-intensive industry, and 0 otherwise.

Incentives alignment (H4). I compute this measure in terms of collective post-IPO shareholding by all employee-directors in the Board of Directors. These persons would typically include the CEO, the Chairman of the Board, and all other executive directors, and typically exclude independent, non-executive directors. I code this variable as 1 if the total post-IPO shareholding exceeds 50%, and 0 otherwise.

Control variables. I incorporate a host of control variables in order to account for the possibility that firms self-select into different prospectus styles depending on their true, intrinsic quality. In line with literature on IPOs, these variables account for uncertainty and agency conflicts that arise during IPOs, and also found to have significant empirical impact on IPO underpricing. Firm age. All else being equal, signals from an older venture are likely to engender a level of uncertainty lower than that of a newer venture. (Beatty & Zajac, 1990). I compute this as the logarithm of firm age, expressed in years elapsed between the date of its registration as a
business and the first date of the IPO. **Firm size.** A bigger firm may have advantages related to scale that may serve it well in the post-IPO phase. **Size** can also be construed as a signal of quality and ability to survive, leading to an easier formation of judgement around its future prospects (Certo, Daily & Dalton, 2001). I record firm size as the logarithm of latest reported annual revenue in US$ million. Past performance. Investors likely weight heavily past financial performance of a venture while forming opinions about its value creation potential in the post-IPO phase (Walter, Kroll & Knight, 2010). I record this measure as trailing return on assets (%). **Tangible assets:** Higher the proportion of tangible assets in a firm’s balance sheet, lower is the perceived uncertainty in the minds of the investors (Baker & Wurgler, 2006). After all, tangible assets are easier to value and sell than intangible assets. I record this measure as the ratio of all assets except intangibles or goodwill, and total assets. **Insider retention.** On account of this information asymmetry, investors may discount the value of a venture where the existing owners seek to sell off a significant portion of their holdings; high insider retention is less likely when the shares are "lemons" (Ehrhardt & Nowak, 2003). I record this measure as the ratio of total shares retained by pre-IPO shareholders to the total number of shares outstanding (post-IPO close). **Underwriter fees.** It is a widespread norm for the investment bank (or syndicate of banks) underwriting a US IPO to charge 7% of IPO proceeds as underwriting fees or discount. A fee higher than 7% could indicate a risky offering, and less than 7% a less risky offering. I code this variable as 1 if underwriting fees and discounts exceed 7%, and 0% otherwise. **Underwriter prestige.** In line with past work, I distinguish between reputed underwriters with a high market share of IPO deals and relatively less reputed and smaller underwriters. The former group can be supposed to deal with large and less risky IPOs than the latter group (Higgins & Gulati, 2003). I code this variable as 1 if an IPO if the solo, lead or joint-lead underwriter of the IPO is among
one of the following five investment banks; Goldman Sachs, JP Morgan, Morgan Stanley, Bank of America-Merrill Lynch, and Credit Suisse (relbanks.com).

I use an ordinary least-squares (OLS) regression equation to examine my hypotheses.

RESULTS

I report key descriptive statistics for key dependent and independent variables in Table 1. The number of observations vary as not all data are available for all firms in my sample. Furthermore, I disregarded firms where some observations clearly had outliers and I could not verify their accuracy from multiple sources. The final sample comprised 439 IPOs. However, not all IPOs attract investor tweets on StockTwits from the first day. Hence, I have investor opinion data for 169 IPOs in the sample. I report the pairwise correlations between independent variables in Table 2. There is no evidence of systematic high correlation among the variables. Finally, I report the industry-wise composition of the 439 IPO firms in the original sample in Table 3.

Before proceeding to hypothesis-testing, I establish if divergence of investor opinion is indeed positively associated with IPO underpricing. This relation has been well-established in finance literature, and I check if the data in my sample shows a pattern consistent with this. Regressing IPO underpricing on divergence shows the relationship is large in magnitude and statistically significant (coeff. = 0.33; p<0.005). I report the results in Table 4.

Having demonstrated the consequences of divergence of investor opinions on IPO outcomes, I turn my attention to examine the hypotheses pertaining to the antecedents of divergence; I report the regression output in Table 5. In Model 1, the effect of market focus on divergence is large and significant (coeff. = 0.184; p<0.012). This implies the higher the
reference to market relative to reference to technology in the prospectus the higher the level of divergence of investor opinions, lending support to H1. (which, in turn, is expected to lead to higher IPO underpricing). In Model 2, the coefficient of similarity on divergence is again large and significant (coeff. = -0.462; p<0.071). This means that the more similar the content of a prospectus is to that of similar IPO firms, the lower is the level of divergence (and hence, IPO underpricing) for the focal firm. This lends support to H2.

Insert Table 5 about here

The interactions effects pertaining to H3a-b can be seen in Models 3-6. Model 3 and 4 collectively indicate that the positive effect of market focus on divergence is larger and significant for knowledge-intensive industries (2-digit SIC 28 or 73), but smaller and not significant than for other industries, indicating support for H3a. Similarly, Models 5-6 demonstrate that the similarity effect is more negative and significant for IPOs from knowledge-intensive industries, thereby supporting H3b.

The interactions effects pertaining to H4a-b can be seen in Models 7-10. Model 7 and 8 collectively indicate that the positive effect of market focus on divergence is larger and significant for where equity retention by insiders exceed 50% than for firms where retention is less than 50%, lending support to H4a. Similarly, Models 9-10 demonstrate that the similarity effect is also as per the predictions of H4b.

I thereafter move on to report certain robustness checks. Consider that the main analyses of the paper are based on a social-media based measure of divergence; the ability to directly measure investor opinions enables me to address some shortcomings of the traditional, trading-volume based measure, from which divergence could at best be imputed. In light of the ubiquity
of the conventional measure, I examine the relationships between the *social media divergence* and *trading volume divergence*. *Trading volume divergence* is expressed as the ratio of total number of shares traded on IPO first day stock trading to the total number of shares on offer in the IPO. First, I observe that *social media divergence* is a strong predictor of *trading volume divergence* (coeff. = 1.665; p=0.000). In turn, *trading volume divergence* is found to be strong predictor of IPO underpricing (coeff. = 0.128; p<0.019). I report core results in Table 6. In other words, *social media* and *trading volume* divergence measures are strongly correlated, and both divergence measures strongly predict IPO underpricing.

Second, I examine if the empirical patterns holds if I consider the full IPO prospectus instead of the Risk Factors section. For H1, the coefficient of *market focus* is positive and large, but not statistically significant at p<0.10 level. For H2, however, the coefficient of *similarity* continues to be negative and statistically significant (coeff. = -0.5; p<0.059). Core results are reported in Table 9. I infer that the empirical patterns with the Risk Factors section continue to hold to a moderate extent for the full IPO prospectus; the differences in investor response to full IPO prospectus and the Risk Factors section represent an avenue for future research.

-------------------------------------

**Insert Tables 6-7 about here**

-------------------------------------

Finally, H3a-b, some may question the appropriateness of an equity-retention (by CEO and employee directors) measure to capture incentives alignment between the firm and managers. The underlying assumption is that it is ultimately the CEO and other executive directors who is responsible for authoring and implementing the strategy laid out in an IPO prospectus. Three observations support this assumption. First, the IPO is a legally-binding document, wherein the authors of the prospectus are liable for any possible misrepresentations
and omissions of material fact in the prospectus. Since it is the Board of Directors who hold ultimately fiduciary responsibility, it is the directors who have the final agency and responsibility for the prospectus content. Second, within the Board of Directors, employee directors often have greater day-to-day say over firm strategy than independent directors, whose role, in reality, is to provide sporadic guidance and governance. Finally, since it is investment banks who underwrite the equity sale at the IPO, it is plausible that it is the underwriters who help write part of the prospectuses. However, an ANOVA analysis indicates that the underwriter-effect accounts for only 18-38% of total variation in relevant textual attributes (market focus and similarity). The final agency for the IPO prospectus, can thus be deemed to lie with the firm. The detailed results are available on request from the author.

DISCUSSION

Through empirical examinations I find support for my hypotheses. First, divergence of investor opinions and in turn, IPO underpricing, is a) positively-associated the relative occurrence of market (as opposed to technology) related content in risk factors section of a prospectus and b) negatively associated with similarity in content of the risk factors section of the prospectuses to that of IPOs by rival and contemporaneous firms. In addition, the main effects are amplified for a) knowledge-intensive industries and b) high equity retention by employee directors. These findings can inform extant knowledge of management theory and practice in several ways.

First, it informs the organizational resource-acquisition literature, specifically the literature on entrepreneurial finance. The paper examines a novel set of antecedents of resource-acquisition performance, namely – investor response to various fundamental characteristics of signals produced by the resource-seeking organization. To the best of my knowledge, this joins one of the few papers in management to explicitly consider the fact that firm- signals can be
equivocal, and possibly the first that goes on to establish the impact of the heterogeneous interpretation of signals to capital-raising contexts. The second anticipated contribution is to literatures on investor response to strategy communicated by the firm. Almost all studies examine *average* level of opinion of potential investors as they are exposed to fundraising pitches, business plans, strategy or earnings announcements, or IPO prospectuses. This study informs the above studies by examining a novel distributional property of investor opinions that can drivee fundraising outcomes, namely, *divergence* (variously referred to as dispersion or disagreement).

Finally, the study speaks to the literature on IPOs – across management, finance, and accounting disciplines. So far, the role of divergence on IPO outcomes had been well documented. However, there existed little insights in terms of how divergence emerges in first place, and what role information or signals might have to play in causing heterogeneous interpretation of an investment opportunity. This study opens the black box of divergence, so to speak, and offers one set of possible explanations of why investors diverge or disagree. The study also makes an empirical contribution, specifically in terms of computing investors sentiment measures directly from opinions expressly communicated by investors. In doing so, I demonstrate the rich potential and relevance user-generated content on social media platforms holds for management research.

The study has clear business policy implications. To the extent that IPO underpricing means the firm receives less cash proceeds on account of pricing their shares at a price below the market-determined level at the end of first-day trading, IPO underpricing tantamount inferior resource-acquisition for the firm (and wealth-attainment for existing investors, such as entrepreneurs and VCs). The implication is that, if investors are indeed sensitive to the market-vs technology balance of prospectus content, and similarity of prospectus to that of other prospectuses, firms should try to author their prospectus strategically. This could increase their
expected proceeds at an IPO (while of course ensuring they remain compliant in terms of reporting regulations).

I conclude by arguing that this study offers multiple future research opportunities. First, we could examine if the determinants of divergence of investor opinions at IPOs are generalizable to market reactions to other strategy communication events, such as M&A or earnings announcements. Second, I compute investor sentiment from StockTwits, a leading investor social-media platform. It would be interesting to check if the same empirical patterns hold for other social media platforms, such as Twitter or Yahoo Finance, and develop a theoretical reasoning if for any systematic differences among various platforms. Finally, researchers could explore the impact of other textual dimensions of the IPO prospectus on divergence of investor opinions, and if the effects are consistent or different across various prospectus sections.

REFERENCES


https://finance.zacks.com/importance-prospectus-1022.html

https://nlp.stanford.edu/software/

https://www.investopedia.com/terms/p/prospectus.asp

https://www.relbanks.com/worlds-top-banks/top-investment-banks


FIGURES & TABLES

Figure 1: Fundraising as a signaling exercise

![Signaling Exercise Diagram]

Figure 2: Investor tweets on StockTwits (web interface)

![StockTwits Interface]

$SNAP Cramer says now is not the time to buy SNAP. He said that a week or so ago too. Gotta love him...

SirCheddarMan
$SNAP Anybody else notice that the new flop goggles come in tide pod colors?

CapitalistWrite
$SNAP waiting for 20 my puts will be so big league. This is a $2.50 stock. Bye

joehelpick
$SNAP is in positive sentiments, with sudden increase in downloads of Snapchat took place. agrud.com/markets/nyse-snap

JorelLaraKalei
Jim Cramer Mad Money Lightning Round cnbc.com/2018/03/02/cameras...
Figure 3: User (investor) profile on StockTwits

Table 1: Key descriptive statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underpricing</td>
<td>439</td>
<td>.167</td>
<td>.285</td>
<td>-.373</td>
<td>1.412</td>
</tr>
<tr>
<td>Divergence (tweets-based)</td>
<td>165</td>
<td>.104</td>
<td>.208</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Market focus (0/1)</td>
<td>439</td>
<td>.487</td>
<td>.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Topic similarity</td>
<td>439</td>
<td>.947</td>
<td>.040</td>
<td>.773</td>
<td>.990</td>
</tr>
<tr>
<td>Log (age)</td>
<td>439</td>
<td>1.881</td>
<td>.857</td>
<td>0</td>
<td>3.04</td>
</tr>
<tr>
<td>Log (revenue)</td>
<td>439</td>
<td>3.397</td>
<td>2.676</td>
<td>0</td>
<td>9.524</td>
</tr>
<tr>
<td>Equity retention, %</td>
<td>439</td>
<td>49.604</td>
<td>25.806</td>
<td>0</td>
<td>94.400</td>
</tr>
<tr>
<td>Asset tangibility, %</td>
<td>439</td>
<td>.573</td>
<td>.446</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Underwriter fees, %</td>
<td>439</td>
<td>.6475</td>
<td>1.181</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Underwriter prestige (0/1)</td>
<td>439</td>
<td>.572</td>
<td>.495</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Pairwise correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-.30</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.19</td>
<td>.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.50</td>
<td>-.32</td>
<td>.35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-.04</td>
<td>-.00</td>
<td>.34</td>
<td>.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-.09</td>
<td>.07</td>
<td>.05</td>
<td>-.29</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-.05</td>
<td>-.03</td>
<td>-.03</td>
<td>.03</td>
<td>.03</td>
<td>-.05</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.16</td>
<td>-.09</td>
<td>.13</td>
<td>.30</td>
<td>.11</td>
<td>-.00</td>
<td>-.06</td>
<td>1</td>
</tr>
</tbody>
</table>

Variable List: 1 = Market focus. 2 = Similarity. 3 = log (age). 4 = log(revenue). 5 = Equity retention. 6 = Asset tangibility. 7 = Underwriter fees. 8 = Underwriter prestige.
Table 3: Industry-wise breakdown of IPOs

No. of IPOs

- Chemicals
- Business Services = Holding Co.
- Instruments
- Electronic
- Others

Table 4: Divergence on IPO underpricing

<table>
<thead>
<tr>
<th>DV= IPO underpricing</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divergence</td>
<td>0.337***</td>
</tr>
<tr>
<td>R²</td>
<td>.06</td>
</tr>
<tr>
<td>N</td>
<td>165</td>
</tr>
</tbody>
</table>

*p < 0.10, ** p < 0.05, *** p < 0.01

Table 5: Examining Hypotheses 1-4b

<table>
<thead>
<tr>
<th>DV= Divergence</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market focus</td>
<td>.18**</td>
<td>.17**</td>
<td>.09</td>
<td>.33***</td>
<td>- .04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic similarity</td>
<td>-.46*</td>
<td>-.72**</td>
<td>-.96</td>
<td>-1.25*</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log age</td>
<td>-.08**</td>
<td>-.06**</td>
<td>-.05*</td>
<td>.14**</td>
<td>-.05*</td>
<td>-.14</td>
<td>-.07**</td>
<td>-.07*</td>
<td>.02</td>
<td>-.10**</td>
</tr>
<tr>
<td>Log revenue</td>
<td>.01</td>
<td>.02*</td>
<td>.00</td>
<td>.03</td>
<td>.02</td>
<td>.07</td>
<td>.02</td>
<td>.03</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Equity retention</td>
<td>-.00</td>
<td>-.00*</td>
<td>-.00</td>
<td>-.00</td>
<td>.00</td>
<td>.82</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Asset tangibility</td>
<td>.07</td>
<td>.21*</td>
<td>.04</td>
<td>.12</td>
<td>.23*</td>
<td>.16</td>
<td>.16</td>
<td>.16</td>
<td>.43**</td>
<td>-.20</td>
</tr>
<tr>
<td>Underwriter fees</td>
<td>-.30</td>
<td>-.10</td>
<td>-.31</td>
<td>.10</td>
<td>.09</td>
<td>12.16</td>
<td>-.7.8*</td>
<td>.15</td>
<td>5.13</td>
<td>.01</td>
</tr>
<tr>
<td>Underwriter prestige</td>
<td>.00</td>
<td>.014</td>
<td>.05</td>
<td>-.13</td>
<td>.05</td>
<td>-.11</td>
<td>-.03</td>
<td>.07</td>
<td>.04</td>
<td>.10</td>
</tr>
<tr>
<td>Industry effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>169</td>
<td>152</td>
<td>102</td>
<td>67</td>
<td>102</td>
<td>50</td>
<td>101</td>
<td>67</td>
<td>80</td>
<td>81</td>
</tr>
<tr>
<td>R-squared</td>
<td>.66</td>
<td>.58</td>
<td>.40</td>
<td>.93</td>
<td>.41</td>
<td>.91</td>
<td>.73</td>
<td>.78</td>
<td>.61</td>
<td>.79</td>
</tr>
</tbody>
</table>

*p < 0.10, ** p < 0.05, *** p < 0.01
Table 6. Social-media and conventional divergence measures

Panel A: Tweet divergence and trading-volume divergence

<table>
<thead>
<tr>
<th>DV= Trading-volume divergence</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweet divergence</td>
<td>1.66***</td>
</tr>
<tr>
<td>R²</td>
<td>.19</td>
</tr>
<tr>
<td>N</td>
<td>169</td>
</tr>
</tbody>
</table>

*p < 0.10, ** p < 0.05, *** p < 0.01

Panel B: Trading-volume divergence and IPO underpricing

<table>
<thead>
<tr>
<th>DV= IPO underpricing</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trading-volume divergence</td>
<td>.13**</td>
</tr>
<tr>
<td>R²</td>
<td>.02</td>
</tr>
<tr>
<td>N</td>
<td>439</td>
</tr>
</tbody>
</table>

*p < 0.10, ** p < 0.05, *** p < 0.01

Table 7: Full IPO prospectus content

<table>
<thead>
<tr>
<th>DV= Divergence</th>
<th>H1</th>
<th>H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market focus</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Topic similarity</td>
<td>- .46**</td>
<td></td>
</tr>
<tr>
<td>Log age</td>
<td>-.08**</td>
<td>-.06**</td>
</tr>
<tr>
<td>Log revenue</td>
<td>.01</td>
<td>.02*</td>
</tr>
<tr>
<td>Equity retention</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Asset tangibility</td>
<td>.09</td>
<td>.20*</td>
</tr>
<tr>
<td>Underwriter fees</td>
<td>-.22</td>
<td>.08</td>
</tr>
<tr>
<td>Underwriter prestige</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Industry effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>169</td>
<td>169</td>
</tr>
<tr>
<td>R-squared</td>
<td>.64</td>
<td>.66</td>
</tr>
</tbody>
</table>

*p < 0.10, ** p < 0.05, *** p < 0.01