Disruptive Innovation and IPO Outcomes:

Evidence from Machine Learning

Hoang Nguyen and Anup Agrawal

July 2024

Comments welcome

* Both authors: University of Alabama, Culverhouse College of Business, Tuscaloosa, AL 35487-0224. Nguyen: hxnguyen@ua.edu; Agrawal: aagrawal@ua.edu, (205) 348-8970. We thank Mehran Azimi, Mark Chen, Doug Cook, Tony Cookson, Jerry Hoberg, Lei Kong, Kyle Lee, Josh Lerner, Baruch Lev, Kai Li, Shawn Mobbs, Sandra Mortal, Josh Pierce, N. R. Prabhala, Tian Qiu, and seminar participants at the University of Alabama for helpful comments and discussions. Agrawal acknowledges research support from the William A. Powell, Jr. Chair in Finance and Banking.

Disruptive Innovation and IPO Outcomes: Evidence from Machine Learning

Abstract

We develop a new text-based measure of a firm's engagement in disruptive innovation that does not require data on R&D or patents. We compute a disruptive innovation score (DIS) for firms doing IPOs using textual analysis of prospectuses and a semi-supervised machine learning method. DIS strongly and positively predicts firms' pre-IPO observable innovation activities, which validates our measure. We find that DIS positively predicts IPO outcomes such as initial return, trading volume, bid-ask spread, and price revision, consistent with disruptive innovation entailing high uncertainty and information asymmetry. These results are robust to a variety of controls, a propensity score matching approach, and the exclusion of technology stocks, the technology bubble period and the financial crisis period. DIS positively predicts one-year post-IPO abnormal return, and the initial returns of high DIS IPOs do not reverse over the next year, contradicting the hype hypothesis about technology stocks. Finally, DIS predicts post-IPO firm policies (such as lower leverage, higher cash holdings, and higher innovation activities) and higher firm valuation. Overall, our findings imply that disruptive innovation is not only a risky but also valuable activity for firms, and our text-based DIS measure captures disruptive innovation activities of IPO firms beyond R&D and patenting.

Keywords: Disruptive Innovation, Innovation, IPO, Machine Learning, Textual Analysis

Disruptive Innovation and IPO Outcomes: Evidence from Machine Learning

1. Introduction

Disruptive innovation (DI) refers to the application of advanced technologies to replace existing products or business models, and create new customers, new competitors, and new approaches to doing business (e.g., Christensen 1997, Christensen and Raynor 2013, Downes and Nunes 2013, Chen, Wu and Yang 2019, and Bloom et al. 2021). DI from breakthrough technologies created by young firms destroys existing industries, creates new industries, and reorders market leaders. Despite the large impact that disruptive technologies have on U.S. labor markets and firms' innovation activities (see, e.g., Bloom et al. 2021), such innovation has received scant attention in the finance literature.

Chen, Wu and Yang (2019) use patent filings by Fintech firms to measure disruptiveness of innovations by the stock price reactions to patent announcements. However, patenting does not fully capture firm innovation activities, and R&D spending is often unreported in financial data. For example, in our sample of 3,440 firms doing initial public offerings (IPO) during 1994-2021, only 52% reported R&D, 28% had filed for patents that were eventually granted, and 25% both reported R&D and filed for patents before going public. Therefore, studies that rely on R&D and patents only partially measure firms' innovation activities (see, e.g., Koh and Reeb 2015, Chen, Wu and Yang 2019, and Bellstam, Bhagat and Cookson 2021).

One strand of literature studies the impact of innovation or a culture of innovation on firm performance using textual analysis (e.g., Guiso et al. 2015, Bellstam, Bhagat and Cookson 2021, Li et al. 2021a, Li et al. 2021b, and Pacelli, Shi and Zou 2022). Without relying on R&D spending and patent data, these studies emphasize the measurement of innovation by using textual information from sources such as firm websites, analyst reports, and earning calls. Most of this research on text-based innovation focuses on well-established public firms. However, to our knowledge, no prior study has analyzed the effect of DI on private firms doing IPO, most of which are quite young. This issue is particularly important because a great deal of breakthrough innovation is carried out by young firms (see, e.g., Christensen 1997), and startups in highly disruptive technology fields tend to exit via IPO rather than by selling out (see Bowen, Frésard and Hoberg 2023), but the effect of such innovation on early firm valuation at the IPO is unknown. We try to fill this gap in the literature by measuring a firm's involvement in DI using machine learning methods on the text of IPO prospectuses, which make presentations of the firm's business, strategy and plans to investors.[1] We create a disruptive innovation score (DIS) for each IPO firm and relate it to various IPO outcomes such as first-day return, trading volume, bid-ask spread, price revision, and one-year post-IPO stock return.

The vast majority of firms do not have any patents before doing an IPO. Some of these firms may have chosen to protect their innovation as trade secrets in order to avoid disclosing the technology or business process that filing a patent entails (see, e.g., Saidi and Zaldokas 2021). Therefore, Chen, Wu and Yang's (2019) measure of DI, which relies on patents, cannot be used for these firms. Even among firms with patents, if a firm has many patents, it is computationally cumbersome to measure whether each patent represents a DI by using our textual analysis and machine learning approach to analyze the text of patent filings and then assign an average innovation score for every IPO firm. Moreover, such an approach would ignore the complementarity among a set of patents that individually represent only marginal discoveries, but collectively represent a breakthrough innovation.[2]

For IPO firms, textual analysis of IPO prospectuses is an attractive way to measure their involvement in DI. Why? Well, the prospectus is basically a written report to potential investors about the firm's business, technology, operations, financials, strategy, and prospects. If a firm is engaged in breakthrough innovation, managers have an incentive to tell investors about it in the prospectus because it is partly a marketing document. However, as the only official offering document, it is constrained by the legal requirement to tell 'the whole truth, and nothing but the truth' about the firm's business, aimed at protecting potential investors from grandiose and rosy marketing pitches. Therefore, our measure is likely to comprehensively capture firms' involvement in DI via any means, either by patenting, trade secrets, and non-disclosure agreements related to

[1] While we focus on IPO firms, our approach is quite general and can also be used to measure disruptive innovation in established public companies, e.g., using the text of earnings calls or analyst reports.

[2] We use the terms 'disruptive innovation' and 'breakthrough innovation' interchangeably.

proprietary technologies they have developed or by licensing or other agreements to use technologies developed by others.

We use keywords for 29 disruptive technologies identified by Bloom et al. (2021) and a semi-supervised machine learning method proposed by Li et al. (2021a) to quantitatively measure the discussion of DI in 3,440 IPO prospectuses between 1994 and 2021. These advanced technologies have dominated discussions between firm executives and investors in earning calls over the last twenty years (see Bloom et al. 2021), so it is reasonable to expect them to be discussed in the prospectus. We train a machine learning model, word2vec, on the text of IPO prospectuses and use Bloom et al.'s 29 technology keywords as seed words. We compute the cosine similarities between the seed word vector and each unique word in IPO prospectuses. We use the 500 words that have the highest cosine similarity scores with the seed words to construct our context-based dictionary of DI. Our expanded dictionary, shown in Table A.3 in the Appendix, identifies additional technologies beyond the 29 key technologies identified by Bloom et al. (2021), e.g., 128bit, airbag, analog converter, and laser.

We then compute a DIS for an IPO firm as the weighted count of the words from our DI dictionary that appear in the IPO prospectus divided by the total number of words in the prospectus, expressed as a percentage. The weights of words in the numerator of DIS are computed using the WF.IDF method, which gives lower weight to words that appear more frequently in the prospectus, as described in section 3. DIS captures firms' prior and future adoptions of the disruptive technologies discussed in IPO prospectuses.

As validation for our measure, we find that DIS strongly and positively predicts firms' pre-IPO observable innovation inputs and outputs such as R&D intensity, the number of patents, average citations per patent, and average value per patent. The positive relation between DIS and R&D intensity is even stronger in IPOs without patents than in IPOs with patents, implying that patenting does not fully capture firms' innovation activities. In addition, in Table A.4, we provide textual evidence that some firms that do not file patents before IPO produce DI and extensively discuss it in the prospectus. Consistent with prior findings that DI entails extreme uncertainty and information asymmetry, we find that IPOs with high DIS have higher first-day returns, trading volumes, bid-ask spreads, and price revisions. A one standard deviation increase in DIS predicts a 3.2% (1.36%) increase in first-day return (price revision).

These results are robust to inclusion of additional controls (e.g., R&D intensity, the number of patents, citations per patent, and textual sentiments in the IPO prospectus) and triple-clustered standard errors on year, industry, and state of firm location. Our main results hold for both technology firms (as defined in Loughran and Ritter (2004)) and non-technology firms, alleviating a concern that high DIS is just a proxy for technology firms. Rather, DIS measures a firm's involvement in DI, regardless of whether the firm produces the technology, uses it or otherwise engages with it. The results are also robust to propensity score matching, mitigating a concern that our results are caused by omitted variables. The results remain similar when we exclude the technology bubble period of 1998-99 and the financial crisis period of 2007-2008. Moreover, consistent with Hirshleifer et al.'s (2018) finding about stock returns of highly innovative firms, we find that IPOs with higher DIS have higher one-year post-IPO abnormal stock returns, and the initial returns of high DIS IPOs do not reverse over the next one year. The latter finding suggests that these initial returns were not caused by IPO firms capitalizing on hype over the emerging disruptive technologies.

Finally, DIS predicts post-IPO firm policies (such as lower leverage, higher cash holdings, and greater innovation activities) and higher firm growth and valuation. A one standard deviation increase in DIS predicts a 1% (8.4%) increase in post-IPO cash holdings (Tobin's Q). These findings imply that (1) DI is a risky, but valuable activity for firms, and (2) our text-based DIS measure captures DI activities of IPO firms beyond R&D and patenting.

Our DIS measure relies on Bloom et al.'s (2021) keywords for technologies that have been successful and disruptive ex-post. While our approach ensures that DIS truly measures DI, as identified by Bloom, et al. based on textual analysis of influential patents and earnings calls, it introduces a potential hindsight bias in our measure. However, we use Bloom et al.'s 29 technology keywords as just a starting point (i.e., as seed words) to come up with our context-based dictionary of 500 words. Our expanded dictionary identifies not only the subset of these technologies discussed by managers of young firms in IPO prospectuses but also additional technologies beyond those found by Bloom et al. Moreover, even though the technologies identified by Bloom et al. are successful on average, there is no guarantee that a given young firm's engagement in them will lead to good outcomes for the firm. Therefore, our measure helps to answer the question whether developing products and services related to these technologies pays off for firms doing IPO and in the long-term.

Our findings contribute to several strands of the literature. First, we contribute to the growing literature that studies the effects of DI and disruptive technologies on firm performance and labor markets (e.g., Chen, Wu and Yang 2019, and Bloom et al. 2021). Second, our study extends the current research on innovation by measuring DI using machine learning to quantify textual data, and measures innovation beyond well-established public companies (e.g., Bellstam, Bhagat and Cookson 2021, Li et al. 2021a, Li et al. 2021b) and the FinTech industry (Chen, Wu and Yang 2019). Third, our DIS measure does not require data on R&D or patents, allowing us to measure DI in firms which neither report R&D spending nor have patents. Thus, our study contributes to the literature that highlights the difference between innovation and patents (e.g., Kogan et al. 2017, Mann 2018, Cohen, Bellstam, Bhagat and Cookson 2021, and Saidi and Zaldokas, 2021). Fourth, DI carries high risk and uncertainty for both firms and investors. Therefore, our research also extends the literature that uses textual analysis to quantify the risks disclosed in IPO prospectuses (e.g., Loughran and McDonald 2013, Crain, Parrino and Srinivasan 2021, and Cumming et al. 2022). Finally, our paper adds to the growing literature in economics and finance that uses machine learning methods to quantify textual data (see Gentzkow, Kelly and Taddy, 2019, and Hoang and Wiegratz, 2023 for reviews of methods and applications) and for other predictive applications (see, e.g., Erel et al. 2021).

2. Issues and hypotheses

DI refers to the application of advanced technologies to create new products, services, business processes, or business models that improve performance to a superior level through sequel development (see, e.g., Christensen 1997, Zhou, Yim and Tse 2005, and Bloom et al. 2021). DIs are designed for new or emerging markets and provide value for new market segments, so they disrupt the existing preferences of customers (Christensen 1997). Such innovation eventually overtakes existing products in mainstream markets. Examples are an autonomous car to replace the traditional car, wireless charging to replace cable charging, or mobile payment to replace the traditional payment methods.

While DI can be attractive in terms of technologies, it is riskier for firms and investors to pursue this type of innovation than incremental innovation because it lacks pre-existing markets and customers (Christensen and Bower 1996, Zhou, Yim and Tse 2005). Because the process of DI carries high uncertainty of outcomes, it is costly for firms to finance such projects (Hall and Lerner 2010). Moreover, Loughran and McDonald (2013) find that greater uncertain tone in IPO prospectuses – as measured by higher proportions of uncertain, weak modal, and negative words – predicts higher IPO first-day returns and price revisions. Given the high degree of uncertainty involved in DI, we hypothesize that *IPOs of firms that are more involved in DI, as measured by high DIS, will have higher first-day returns and price revisions.*

Kravet and Muslu (2013) find that the number of sentences related to risk in firms' 10-K filings positively predicts higher trading volume around the filing date. Moreover, Azimi and Agrawal 2021 find that uncertainty about a firm's prospects, indicated by negative sentiment in 10-K filings, increases information asymmetry among investors, leading to higher divergence of investor opinions and higher abnormal trading volume. Given the high uncertainty involved in DI, we hypothesize that *IPOs with high DIS will have higher trading volumes.*

Previous studies find that IPOs with greater uncertainty have higher bid-ask spreads due to greater information asymmetry (Hribar 2004, Guo, Lev and Zhou 2004). Therefore, we hypothesize that *IPOs with high DIS will have higher bid-ask spreads.*

Uncertainty indicated by negative textual sentiment suggests that firms will hold more cash in the future (Azimi and Agrawal 2021). Firms in product markets with uncertain structures use cash to invest in innovation (Lyandres and Palazzo 2016). Therefore, we hypothesize that *IPOs with high DIS will have higher cash holdings.*

Finally, Bellstam, Bhagat and Cookson (2021) find that more discussion of innovation in the analyst reports of S&P 500 firms predicts higher growth opportunities for firms. Therefore, we hypothesize that *IPOs with high DIS will have higher Tobin's Q.*

3. Measurement of DI

Measuring DI based on patents and citations is challenging in firms doing IPOs because many young firms do not have (or choose not to file for) any patents before going public. While firms may not file for patents before doing IPO, they often discuss their future innovation plans in IPO prospectuses. Moreover, if firms choose not to patent their innovation, then a measure of DI based on patent filings would be inappropriate. On the other hand, a firm may have many patents, only some of which relate to DI. It is also computationally challenging to measure DI based on the text of individual patents owned by a firm and then assign an average innovation score to the firm. Moreover, as discussed in the introduction, such an approach ignores complementarities among a group of patents that collectively represent breakthrough innovation. We sidestep these issues by constructing a DI score (DIS) based on the text in IPO prospectuses, which measures firms' involvement in one or more of 29 advanced technologies identified by Bloom et al. (2021). These technologies meet the definition of DI proposed by Christensen (1997) and Zhou, Yim and Tse (2005). Bloom et al. (2021) identify these technologies based on discussions between firm managers and investors during earnings conference calls. They present strong evidence that these technologies disrupt labor markets and firm innovation activities, so they potentially also affect IPO outcomes.

Our construction of DIS follows Mikolov, et al.'s (2013) pathbreaking natural language processing (NLP) method, word2vec, used recently in finance by Li et al. (2021a), Li et al. (2021b), Azimi and Agrawal (2021), and Pacelli, Shi and Zou (2022). To reduce the noise in training data, we manually remove legal language and boilerplates which usually appear at the beginning and end of prospectuses. Before the training process, we preprocess and parse the raw IPO prospectuses by using CoreNLP to obtain meaningful groups of words in sentences. After parsing with CoreNLP, we remove the redundant words that do not carry any meaning (e.g., single letter words, stop words, and sentence punctuation marks) from documents. After training the model on 3,440 IPO prospectuses, based on the prediction of the neural network, we create the word-embedding vectors, which represent the meaning of words in the text. We use the vector of seed words and their synonyms for DI from Bloom et al. (2021). Table A.1 in the Appendix shows the keywords for 29 disruptive technologies we use as seed words. We then compute the average of the vector of seed words following Li et al (2021a). We compute the cosine similarity between the unique vector in each IPO prospectus and the average of the vector of seed words. Next, we rank the cosine similarities of words and select 500 words with the highest rank to construct a dictionary of DI.[3] We then manually inspect the context-based dictionary to include only words relevant to technologies. Table A.2 lists the top 40 technologies in this dictionary.

In the last step, we score DI based on the created dictionary. We use the term frequency inverse document frequency method with log normalization (WF.IDF), in which the words that appear more frequently in the documents carry lower weights. Finally, we construct our DIS for a firm as the weighted-frequency of words related to disruptive technologies in its IPO prospectus,

[3] This 'most similar to seed words' approach has been used in several recent studies, e.g., Hanley and Hoberg (2019), Li et al. (2021a) and Li et al. (2021b).

i.e., the number of technology-related words as a percentage of the total number of words in the prospectus, where both the numerator and denominator are adjusted to reduce the weight of more frequent words. We also compute an innovation score using seed words from Li et al. (2021a) and a total innovation score (= DIS plus Innovation score) as benchmarks. The Innovation score is the weighted-frequency of innovation-related words in the IPO prospectus, i.e., the number of innovation-related words as a percentage of the total number of words in the prospectus, adjusted to reduce the weight of more frequent words.

4. Data sources, summary statistics and correlations

4.1. Data sources and summary statistics

We obtain S-1 filings of IPO firms from 1994 to 2021 from WRDS-SEC Analytics Suite and the SEC Edgar website. Following the literature on IPOs, we exclude real estate investment trusts (REITs), American depository receipts (ADRs), closed-end funds, and IPOs with offer prices less than \$5. Our final sample includes 3,440 IPO firms. We obtain the pricing information on each IPO such as offer price and price revisions from SDC Platinum and trading data such as closing price, trading volume and Bid-ask spread from CRSP. The other control variables are from Compustat. The ranking to construct the dummy variable for underwriter reputation (Top Underwriter) is collected from Professor Jay Ritter's website. The patent data to validate DIS and innovation score are from Professor Noah Stoffman's website. Table 1 reports summary statistics of our main variables, defined in Table A.6. The key explanatory variables of interest are DIS, innovation score, and total innovation score. The summary statistics of our dependent and control variables are generally consistent with prior studies on textual analysis of IPO prospectuses (see, e.g., Hanley and Hoberg 2010, Loughran and McDonald 2013, and Crain, Parrino and Srinivasan 2021).

4.2. Correlations

Table 2 presents pairwise correlations among the variables. Both DIS and innovation score are positively and significantly correlated with first day (or initial) return, price revision, and bid-ask spread, but the magnitudes of these correlations are larger with DIS. DIS is also positively and significantly correlated with patent counts, citations per patent, patent values, R&D intensity (R&D divided by total assets), while innovation score is not.

5. Empirical results

5.1. Validation of DIS

To validate our DI measure, we follow Li et al. 2021a and Kogan et al. 2017 and use the number of patents, citations per patent, patent values or R&D spending of IPO firms as the dependent variable in validation regressions. Patent data from Kogan et al. (2017) are updated until 2020. The patent values per patent are measured in both nominal value and real value (deflated to dollar value in 1981). In this step, we only use only the patents filed before IPO offer dates because the patent filing announcements are likely to affect IPO outcomes.

Panel A in Table 3 indicates that the coefficients of DI are significantly and consistently positive across all eight models. High DIS predicts higher patent counts, citations per patent, patent values, and R&D intensity in models 1 to 8. In models 9 and 10, we include a dummy variable for whether firms filed patents before the IPO offering date, and its interaction with DIS. The negative coefficient of this interaction term implies that DIS predicts R&D intensity even more strongly for firms without patents. Thus, DIS provides a broader measure of firms' innovation activities than patents.

We next use this approach to validate the innovation score constructed using seed words from Li et al. (2021a) and present it in Panel B of Table 3. The innovation score positively predicts only the number of patents in model 2, while it predicts R&D intensity *negatively*. The implication of Panels A and B of Table 3 is that in the IPO context, DIS is more related to firms' innovation activities than the innovation score. Our finding that DIS strongly predicts firms' patenting activities and patent values is consistent with Bloom et al.'s (2021) finding that the 29 disruptive technologies they identify account for a large chunk (about 30%) of all patents granted by the USPTO between 2002 and 2016.

As a robustness check of the results in Panels A and B, we next include both DIS and the innovation score as explanatory variables in each of the eight models above. In Panel C, the coefficient of DIS remains roughly similar to that in Panel A. High DIS predicts more patents and citations, higher patent values and greater R&D spending. On the contrary, the only significant coefficients of the innovation score are *negative* for R&D in models 7 and 8. Overall, the results in Table 3 show that at least in the IPO context, DIS positively predicts firms' IPO innovation activities much more consistently than the innovation score. In section 5.4, we examine whether DIS and innovation score predict firms' post-IPO innovation activities one year after the IPO.

It is worth noting that DIS and innovation score do not measure the same aspects of firms' innovation activities — they are measured using different seed words and the Pearson correlation between them is only 0.26 (see Table 2). Also, while the dictionary for DI mainly has terms related to disruptive technologies (see Table A.2 for the top 40 terms in this dictionary), the dictionary for innovation has the synonyms of the keyword "innovation." The two dictionaries have completely different terms, with no overlap. A firm with a strong innovation score can have a low DIS. Finally, we use a principal components analysis to test the meanings of the two scores on multiple dimensions. If the first principal component (PC) is significant, but the second PC is not, then the two scores are likely to capture the same attribute (see, e.g., Loughran and McDonald 2013). In Table 4, both the first and second PCs of DIS and innovation score are highly significant in predicting patents, citations, patent values and R&D across all eight models, which suggests that DIS and innovation score measure different aspects of a firm's innovation activities.

5.2 DIS, Innovation and IPO outcomes

We next investigate the relation between DIS and each of four IPO outcomes — first day IPO return, first day trading volume, first day bid-ask spread, and price revision — and present the results in in Tables 5 to 8, respectively. The main explanatory variables of interest in these regressions are DIS, innovation score, and total innovation. Control variables include upward price revision, high tech, top underwriter, venture backing, positive EPS dummy, market return, share overhang, and natural logarithms of age and sales. The control variables are the same in the regressions in Tables 5 to 8, except that regressions of the price revision in Table 8 do not include upward price revision as a control.

In columns 2 to 4 in Table 5, DIS, innovation score, and total innovation positively predict the IPO first-day return. The economic magnitudes of the effects of DIS and innovation score on IPO first day return are comparable. A one standard deviation increase in DIS (0.09) predicts a 3.2% (= 35.469 x 0.09) increase in the IPO first-day return. A corresponding one standard deviation increase in Innovation (0.13) predicts a 3.6% (= 27.932 x 0.13) increase in the IPO first-day return.

Similarly, in columns 2 to 4 in Table 6, DIS, innovation score, and total innovation positively predict IPO first-day trading volume. A one standard deviation increase in DIS (0.09) predicts about an 11% increase in the first-day trading volume (=0.09 x ($e^{0.795}$ - 1)). Our finding is consistent with previous findings that high uncertainty in firms' disclosure predicts higher trading volume around the filing date (Kravet and Muslu 2013, and Azimi and Agrawal 2021).

In Table 7, all three innovation measures positively predict the bid-ask spread on the first trading day. The effect of DIS is somewhat larger in economic magnitude than that of the innovation score. A one standard deviation increase in DIS (Innovation score) predicts a 0.44% (0.35%) increase in bid-ask spread. This result confirms the idea that wider bid-ask spreads reflect higher uncertainty associated with DI.

Table 8 shows regressions of price revision. Among the among the three innovation scores, DIS has the largest effect on price revision. Here, a one standard deviation increase in DIS (Innovation score) predicts a 1.36% (0.62%) higher price revision. Our finding that DIS positively predicts both IPO initial return and price revision is consistent with Loughran and McDonald's (2013) finding that more uncertain information in the prospectus is associated with more uncertainty of IPO valuation.[4]

5.3 Post-IPO stock performance

We next examine whether our innovation measures predict the post-IPO stock performance. We compute the cumulative abnormal return (CAR) over months (+1, +12) after the IPO,

$$CAR_{i,1\text{-}12} = \sum_{t=1}^{12} e_{it}$$

$$e_{it} = r_{it} - r_{bt}$$

where e_{it} is the abnormal return on stock i over month t, r_{it} is the return on stock i over month t, and r_{bt} is the return on the DGTW benchmark portfolio b over month t. Portfolio b is the value weighted portfolio in the same size, book to market, and momentum (i.e., the prior 12-month return) quintile portfolio as firm i, as in Daniel et al. (1997). Based on the data from CRSP, we are able to compute $CAR_{1\text{-}12}$ for 2,552 IPO firms in our sample.

We then estimate regressions of the post-IPO abnormal return ($CAR_{1\text{-}12}$). As in Tables 5 to

[4] In Table A. 5, we report the results of horse-race regressions similar to those in column (1) of Tables 5-8, after adding Innovation as an explanatory variable. The coefficient estimates of DIS are positive and statistically significant in all the regressions, while coefficient estimates of Innovation are only significant in models 1 and 3.

8, the main explanatory variables of interest are DIS, Innovation score and Total Innovation. The model includes control variables following the recent literature on IPOs and innovation (see Crain, Parrino and Srinivasan 2021, Li, et al. 2021a, and Cumming, et al. 2022). The regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors in parentheses are clustered by industry. Table 9 summarizes the results. The estimated coefficient of DIS is significantly positive in model 1, while the coefficient on innovation score is insignificant in model 2. IPOs of firms involved in DI perform well over the next one year. A one standard deviation (0.09) increase in DIS predicts a 6.04% (= 0.09 x 67.162) increase in the 1-year post-IPO abnormal return. This finding has a similar flavor to Hirshleifer et al.'s (2018) finding that firms with more original patents have higher long-run abnormal return.

We next test the hype hypothesis that investors are unduly optimistic or exuberant about high-DIS IPOs and are willing to pay more for such stocks than they are worth. This hypothesis implies that once the initial euphoria surrounding DI by the newly public firm subsides, the stock will under-perform. To test this hypothesis, we include the interaction term between DIS and first day return as an explanatory variable in the regressions of the post-IPO abnormal return. If investors initially overpay for a high-DIS stock, which results in a large first day return, the coefficient of this interaction term should be negative. In column 1 of Table 9, the estimated coefficient of this interaction term is positive and statistically insignificant, which does not support the hype story.

5.4 Post-IPO firm policies, performance, valuation and innovation activities

We next examine whether a firm's involvement in DI at the time of IPO predicts its future financial policies, performance, valuation, and innovation activities one year after the IPO. We focus on two financial policies that are likely affected by a firm's involvement in DI: leverage and liquidity. DI projects involve a great deal of uncertainty and have low collateral values. Such projects are better financed with equity or internally generated cash, rather than debt (see, e.g., Myers 1977, Titman and Wessels 1988). That implies that firms with higher DIS will use lower leverage. Firms with higher DIS also need to hold more cash to finance high-risk innovation projects (see, e.g., Lyandres and Palazzo 2016), implying that they will have higher cash holdings. Since breakthrough innovation projects tend to be long-term in nature, and R&D investments are expensed rather than capitalized and depreciated, they should reduce net income in the short-term, implying lower return on assets (ROA). Consistent with these projects being positive NPV, we expect firms with high DIS scores to have higher valuations, as measured by Tobin's Q. Finally, we expect firms with higher DIS to have greater input (measured by R&D) and higher quantity, quality, and value of innovation output (measured by patents, citations, and patent values).

We report the regressions for each dependent variable in Table 10. In Panel A, consistent with Li et al. (2021a), we find that DIS negatively predicts leverage and ROA and positively predicts cash holdings. Consistent with our expectation, DIS negatively predicts ROA and positively predicts future sales growth, Tobin's Q, and stock return volatility. Finally, consistent with our expectation, DIS positively predicts future innovation input (R&D) and the quantity, quality and value of innovation output (patents, citations and patent values).

In Panel B, similar to the pre-IPO results, firms with high innovation score have lower R&D spending post-IPO. Contrary to the results using DIS in Panel A, innovation score at the IPO negatively predicts a firm's post-IPO cash holdings, although it predicts higher post-IPO patent values as in Panel A.

6. Robustness and identification

6.1 Additional controls

In section 5.2, we find that an IPO firm's involvement in DI, measured by DIS, positively predicts its IPO first day return, trading volume, bid-ask spread, and price revision. We next examine whether these results are robust to controlling for traditional measures of firms' innovation input and output. Accordingly, we add controls for R&D intensity, number of patents and average citations per patent filed pre-IPO to regressions similar to those in column (2) of Tables 5 to 8. We also include state dummies of IPO headquarters to control for regional effects and compute triple-cluster standard errors to test the robustness of DIS coefficient estimation. Panel A of Table A.7 shows these results. The coefficient of R&D intensity is insignificant in all four regressions. Pre-IPO patenting activity negatively predicts the return and positively predicts the trading volume on the first trading day, but is unrelated to price revision or bid-ask spread. Citations positively predict only the initial return. More important, the coefficient of DIS remains robustly significant in all four regressions, implying that IPOs with high DIS are associated with risks beyond what is indicated by R&D investment and patenting activities.

In addition, to control for the sentiment expressed in IPO prospectuses, we include negative or positive sentiment measured by the word-count method (Loughran and McDonald 2011) in our regressions. Panels B and C of Table A.7 include negative and positive sentiment, respectively, as control variables. The coefficient of DIS remains positive and statistically significant across all the four models in each panel. These results implies that our DIS measure predicts firm risk that is unexplained by the sentiment in the prospectus. The results are similar when we include both positive and negative sentiment in these regressions, so we do not tabulate them for brevity.

6.2 Tests on subsamples of technology and non-technology firms

While our DIS measure is designed to capture firms' involvement in DI, one concern is that high DIS may just be a proxy for whether a firm is in the technology sector. We next address this concern by estimating separate regressions for subsamples of technology and non-technology firms. We identify technology firms using the SIC codes listed in Loughran and Ritter (2004, Appendix D), and call the remaining firms non-technology firms. Table 11 presents the results of regressions similar to those in column (2) of Tables 5 to 8 and column (1) of Table 9 for the subsample of technology firms in Panel A and non-technology firms in Panel B. As expected, in Panel A of Table 11, the sign and significance of our main explanatory variable of interest, DIS, for technology firms are quite similar to our baseline results. Perhaps surprisingly, the results for non-technology firms in Panel B are also quite similar to those for technology firms. Thus, our results cannot be attributed to high DIS simply being a proxy for firms in the technology sector. Instead, DIS measures a firm's involvement in DI, regardless of whether the firm produces the technology, uses it or otherwise engages with it.

6.3 Tests excluding technology bubble and financial crisis

Loughran and Ritter (2004) find that the IPO pricing variables experienced an extreme distribution, especially during the peak of the technology bubble during 1998-1999. Given that these observations are so extreme, we next address the concern that they drive our empirical results. Table A.8 presents results corresponding to our baseline results in column (2) of Tables 5 to 8 and column (1) of Table 9 after excluding IPOs during 1998-98. These results are generally similar to our baseline results. Untabulated results are also similar if we exclude the financial crisis years of 2008-2009 in addition to the technology bubble years.

6.4 Propensity score matching

Our findings so far indicate that DIS reliably predicts several IPO outcomes, and firms' post-IPO financial policies and innovation activities. But is this relation causal? One identification concern is that these relations may be driven by omitted variables. We use average treatment effects (ATE) computed using propensity score matching (PSM) to mitigate this concern, using two different approaches to identify treatment and control firms.

First, we consider IPOs with DIS above (below) the sample median as treated (control) firms. We next estimate probit regressions of being treated using all the control variables and calculate the propensity score. We then match each treated firm with a control firm in the same IPO year that has the closest propensity score (see, e.g., Crain et al. 2021). We then calculate the ATEs for IPO outcomes, and post-IPO firm policies and innovation activities and summarize them in Panel A of Table A.9. In columns 1 to 4, there is strong evidence that first day return, price revision, trading volume and bid-ask spread are all significantly larger in the treated group than in the control group. In columns 5 to 10, the same is true of post-IPO cash holdings, Tobin's Q, R&D, patents, citations per patent and real value of patents.

Second, we repeat this procedure after reclassifying IPOs with DIS in the top (bottom) quartile as treated (control) firms. The results, in Panel B of Table A.9, are quite similar to those in Panel A. Overall, our PSM estimates suggest that omitted variables are unlikely to drive the impact of DIS on IPO outcomes and firms' post-IPO policies and innovation activities.

6.5 Cross-sectional tests

The effect of DIS on IPO outcomes can differ by a firm's age, size, patent possession and VC-backing. Our results so far suggest that investors in IPOs with high DIS face a great deal of uncertainty, as reflected in our findings that such IPOs have higher initial return, price revision, trading volume, and bid-ask spread. One might expect the uncertainty associated with high DIS to be lower in firms that are older, larger, already have some patents or are VC-backed. To test this notion, we start by creating a binary variable *Old* that equals one if the IPO firm's age is greater than the sample median and equals zero otherwise. We then re-estimate our baseline regressions in column 2 of Tables 5 to 8 by adding *Old* and *Old*DIS* as explanatory variables and present the results in Panel A of Table 12. We find that among firms with higher DIS, older firms have lower initial returns and price revision than younger firms.

We next create a dummy variable *Large* that equals one if the IPO firm's total assets are greater than the sample median and equals zero otherwise. Similar to Panel A, we now add *Large* and *Large***DIS* as explanatory variables to our baseline regressions and present the results in Panel B of Table 12. In all the regressions, the interaction term is statistically insignificant, suggesting that the effect of DIS on small and large firms is indistinguishable.

We next create a binary variable *Has patents* that equals one if the IPO firm owns patents that were filed before its offering date and equals zero otherwise. As in Panel A, we now add *Has patents* and *Has patents* **DIS* as explanatory variables to our baseline regressions and present the results in Panel C of Table 12. Here too, the effect of DIS is statistically indistinguishable between IPO firms with and without patents.

Next, we create a binary variable VC that equals one if the IPO firm is backed by venture capital investment and equals zero otherwise. Panel C of Table 12 presents the results of regressions where we add VC and DIS*VC as explanatory variables to our baseline specification. Here, the effect of DIS in VC-backed firms is substantially larger on initial return and smaller on trading volume, although the latter is statistically significantly at the 10% level. In all four panels, the main effect of DIS continues to be significantly positive in all four regressions.

Finally, in model 5 of each panel in Table 12, firm age, size, patent possession and VC-backing do not drive the positive relation between DIS and long-run abnormal return. The main effect of DIS on long-run abnormal return continues to be significantly positive in all four panels. Interestingly, in panel D, the positive effect of VC-backing on long-run abnormal return found by Brav and Gompers (1997) is no longer statistically significant in this specification.

7. Conclusion

This study is the first to introduce a measure of disruptive innovation, DIS, that measures IPO firms' involvement in disruptive technologies from the text in IPO prospectuses using a state-of-the-art machine learning method. We validate DIS with firms' observable innovation activities and find that IPOs with high DIS have higher pre-IPO R&D intensity, patent counts, citations per patents, and dollar values per patent. Consistent with the theory that DI involves greater risk, uncertainty and information asymmetry, we find that IPOs with high DIS have higher IPO initial returns, trading volumes, bid-ask spreads, and price revisions. IPOs with higher DIS also have higher one-year post-IPO abnormal returns, measured relative to the DGTW benchmark. We also find no evidence that the first day returns of IPOs with high DIS reverse over the following 12 months. Moreover, DIS predicts future firm policies (lower leverage, and higher cash holdings and innovation activities) and higher firm valuation. Overall, our findings suggest that DI predicts several IPO outcomes and subsequent operations, performance, and valuation of firms going public. Finally, our DIS measure captures firms' DI activities that are unexplained by R&D and patents.

Do our measures of DI miss R&D endeavors that never come to fruition? Well, adoption of breakthrough technologies involves a complex interplay between technologies, consumers, marketplace, and companies. There are many technologies that succeed in the lab but don't get adopted in the marketplace. Our paper deals with technologies that young firms aim to adopt, as discussed in IPO prospectuses. What causes some technologies to get adopted in the marketplace and others to die in the lab is an interesting question that we leave for future research.

Finally, our paper focuses on the impact of DI on the firm doing the IPO, but how does the IPO affect the valuation and performance of industry competitors, market leaders, and the industry as a whole? Do competitors experience negative valuation effects around a successful IPO caused by a large infusion of cash that allows the IPO firm to leverage its technology, increasing its disruptiveness and diffusion? Does a successful IPO create an arms race by inducing competitors to raise more capital, which in turn increases the impact of the technology? These are all interesting questions that we leave for future research.

Table 1. Summary statistics for IPO sample, 1994-2021

Statistic	N	Mean	St. Dev.	Min	Median	Max
DIS	3,440	0.09	0.09	0.00	0.05	0.46
Innovation	3,440	0.13	0.13	0.00	0.10	1.43
Patents	3,440	4.99	80.07	0.00	0.00	4,349
Citations	3,440	16.25	71.74	0.00	0.00	1,974.00
Patent value (Real)	3,440	1.74	8.08	0.00	0.00	151.28
Market return	3,440	0.96	3.30	-20.95	1.24	12.94
Share overhang	3,440	1.12	3.50	0.00	0.00	129.87
Top underwriter	3,440	0.68	0.46	0	1	1
EPS+	3,440	0.97	0.18	0	1	1
VC	3,440	0.46	0.50	0	0	1
High tech	3,440	0.57	0.50	0	1	1
Ln(age)	3,440	2.33	0.99	0.00	2.20	5.12
Ln(sales)	3,440	3.62	2.39	-4.83	3.92	11.82
R&D	3,440	28.66	33.10	0.00	14.24	486.05
ROA	3,440	-11.07	142.44	-975.04	1.16	180.92
Cash	3,440	27.94	28.14	-0.82	17.92	99.76
Leverage	3,440	23.29	73.91	0.00	6.93	182.33
Tobin's Q	3,440	3.23	5.95	0.00	1.79	39.81
Tangibility	3,440	17.06	21.74	0.00	7.80	98.40
Bid-ask spread	3,440	14.54	11.99	0.00	10.93	99.51
Ln(volume)	3,440	15.08	1.26	4.62	15.20	20.06
Price revision	3,440	-0.20	13.98	-76.81	0.00	383.33
First day return	3,440	26.12	55.68	-76.60	11.01	835.00
Long run abnormal return	2,552	-6.12	78.34	-325.31	-6.82	675.61

Note: The table presents summary statistics for IPOs in the sample. The sample consists of 3,440 US IPOs during 1994-2021 with an offer price of at least $5 per share. Table A.6 defines the variables.

Table 2. Correlation Matrix

	DIS	Innovation	Patents	Citation	Patent value	Market return	Overhang	Top underwriter	EPS+	VC	High tech	Ln(age)	Ln(sales)	R&D	ROA	Cash	Leverage	Tobin's Q	Tangibility	Bid-ask spread	Ln(volume)	Price revision
DIS																						
Innovation	0.26***																					
Patents	0.04**	0.00																				
Citations	0.15***	-0.03**	0.01																			
Patent value	0.14***	0.01	0.06***	0.26***																		
Market return	-0.02	0.00	0.00	0.02	-0.01																	
Overhang	0.09***	0.00	-0.01	0.06***	0.08***	0.00																
Top underwriter	-0.02	0.02	-0.02	-0.03	-0.02	-0.04**	-0.01															
EPS+	0.09***	0.04**	0.01	0.04**	0.01	0.01	0.02	0.02														
VC	0.18***	-0.10***	0.00	0.17***	0.09***	-0.01	-0.02	-0.08***	0.12***													
High tech	0.39***	-0.07***	0.01	0.17***	0.12***	-0.02	0.03**	-0.10***	0.14***	0.53***												
Ln(age)	-0.08***	0.17***	0.05**	-0.12***	0.00	0.00	0.00	0.06***	-0.02	-0.26***	-0.27***											
Ln(sales)	0.05**	0.25***	0.08***	-0.08***	0.03*	0.01	-0.04**	0.10***	0.22***	-0.27***	-0.26***	0.36***										
R&D	0.21***	-0.14***	0.07***	0.13***	0.12***	-0.02	-0.03**	-0.04**	0.15***	0.52***	0.58***	-0.17***	-0.29***									
ROA	0.00	0.03	0.00	-0.02	0.00	0.00	0.01	0.04**	-0.02	-0.06**	-0.11***	0.09***	0.18***	-0.24***								
Cash	0.16***	-0.05**	0.00	0.12***	0.05**	-0.01	0.00	-0.10***	0.12***	0.44***	0.46***	-0.29***	-0.37***	0.44***	-0.11***							
Leverage	-0.07***	-0.01	-0.01	-0.05**	-0.03*	0.01	-0.03	0.02	0.01	-0.12***	-0.07***	0.05**	0.02	0.06***	-0.52***	-0.12***						
Tobin's Q	0.17***	0.03*	-0.01	0.17***	0.21***	-0.02	0.03	-0.03**	0.09***	0.19***	0.22***	-0.16***	-0.08***	0.21***	-0.15***	0.25***	0.17***					
Tangibility	-0.13***	-0.04**	-0.01	-0.08***	-0.04**	-0.01	-0.04**	0.07***	0.07***	-0.25***	-0.35***	0.03*	0.31***	-0.27***	0.06**	-0.34***	0.10***	-0.12***				
Bid-ask spread	0.22***	0.05**	-0.01	0.12***	0.10***	-0.02	0.04**	-0.09***	0.08***	0.31***	0.37***	-0.23***	-0.18***	0.23***	-0.07***	0.32***	-0.08***	0.24***	-0.19***			
Ln(volume)	0.14***	0.23***	0.09***	0.02	0.14***	-0.01	-0.01	0.08***	0.05**	0.07***	0.06***	0.12***	0.37***	0.00	0.08***	-0.03*	-0.03*	0.09***	0.07***	0.21***		
Price revision	0.10***	0.06***	0.00	0.04**	0.03	0.05**	0.04**	0.00	0.02	0.00	0.04**	-0.01	0.04**	0.00	0.05**	0.01	-0.07***	0.03**	-0.04**	0.08***	0.13***	
First day return	0.20***	0.08***	0.00	0.13***	0.15***	0.07***	0.10***	-0.02	0.05**	0.19***	0.18***	-0.14***	-0.03	0.12***	-0.01	0.16***	-0.06***	0.36***	-0.10***	0.42***	0.25***	0.21***

Note: The table presents Pearson pairwise correlations among the variables. * ,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 3. Validation for innovation measures

Panel A. Validation of disruptive innovation

	Dependent variable:									
					Patent value					
	Ln(Patents)	Patents	Ln(Citations)	Citations	Nominal	Real	Ln(R&D)	(R&D)/AT	Ln(R&D)	(R&D)/AT
	OLS	*Poisson*	*OLS*	*Poisson*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DIS	0.886***	3.137***	1.103***	0.744**	0.320**	0.281**	1.260***	30.966***	1.485***	40.365***
	(0.314)	(0.667)	(0.319)	(0.296)	(0.154)	(0.130)	(0.251)	(4.139)	(0.274)	(5.970)
Ln(Patents)			0.938***	0.548***	0.703***	0.530***	0.396***	5.555***		
			(0.109)	(0.050)	(0.063)	(0.052)	(0.057)	(0.658)		
Ln(Cites)							-0.040	1.029***	-0.038	0.592
							(0.027)	(0.353)	(0.023)	(0.507)
Has patents									0.715***	14.336***
									(0.088)	(1.920)
DIS*Has patents									-0.312	-21.247**
									(0.398)	(8.667)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440
Adjusted R^2	0.362		0.570		0.611	0.552	0.621	0.622	0.598	0.616

Panel B. Validation of innovation

	Dependent variable:									
					Patent value					
	Ln(Patents)	Patents	Ln(Citations)	Citations	Nominal	Real	Ln(R&D)	(R&D)/AT	Ln(R&D)	(R&D)/AT
	OLS	*Poisson*	*OLS*	*Poisson*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Innovation	0.133	1.520***	-0.062	0.366	0.138	0.123	-0.227	-13.762***	-0.359**	-10.973***
	(0.114)	(0.452)	(0.212)	(0.232)	(0.150)	(0.132)	(0.181)	(5.079)	(0.159)	(3.471)
Ln(Patents)			0.945***	0.553***	0.705***	0.532***	0.402***	5.687***		
			(0.110)	(0.051)	(0.063)	(0.051)	(0.057)	(0.661)		
Ln(Cites)							-0.034	1.187***	-0.032	0.441
							(0.028)	(0.357)	(0.023)	(0.496)
Has patents									0.599***	14.960***
									(0.095)	(2.067)
Innovation*Has patents									0.721**	-15.156**
									(0.340)	(7.409)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440
Adjusted R^2	0.359		0.567		0.611	0.552	0.618	0.619	0.595	0.613

Panel C. Validation of both innovation and disruptive innovation

	Dependent variable:							
					Patent value			
	Ln(Patents)	Patents	Ln(Citations)	Citations	Nominal	Real	Ln(R&D)	(R&D)/AT
	OLS	*Poisson*	*OLS*	*Poisson*	*OLS*	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DIS	0.574**	2.521**	2.184***	1.117***	0.933***	0.737***	2.054***	55.011***
	(0.279)	(0.995)	(0.478)	(0.419)	(0.203)	(0.166)	(0.304)	(6.830)
Innovation	-0.015	0.505	-0.440	-0.152	0.017	0.024	-0.689**	-26.289***
	(0.095)	(0.730)	(0.314)	(0.523)	(0.163)	(0.141)	(0.273)	(8.240)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440
Adjusted R^2	0.270		0.333		0.280	0.248	0.582	0.600

Note: Regressions of traditional measures of innovation. The dependent variables are the number of patents (or its natural log), the number of citations per patent (or its natural log), natural log of patent values per patent in nominal or real terms, natural log of R&D, and R&D divided by total assets. The main explanatory variable[s] in the regressions in Panel A (B) [C] is [are] DIS (Innovation score) [DIS, Innovation score]. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors in parentheses are clustered by industry. The real value of patents equals their nominal value deflated to 1982 (million) dollars using the CPI. The sample includes only eventually granted patents filed before the IPO. Control variables are High-tech, EPS+, VC, Share overhang, Top underwriter, Ln(Age), Ln(Sales), Tangibility, Leverage, ROA, R&D intensity, Cash, and Ln(Tobin's Q). * ,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4. Principal components models

	Dependent variable:							
					Patent value			
	Patents		Citations		Nominal	Real	Ln(R&D)	(R&D)/AT
	negative binomial	*Poisson*	*negative binomial*	*Poisson*	*OLS*	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First principal component	0.115***	0.177***	0.149***	0.060***	0.063***	0.051***	0.074***	1.279***
	(0.017)	(0.007)	(0.025)	(0.005)	(0.015)	(0.012)	(0.018)	(0.384)
Second principal component	0.062**	0.126***	0.230***	0.088***	0.060***	0.047**	0.198***	6.019***
	(0.024)	(0.010)	(0.037)	(0.006)	(0.023)	(0.018)	(0.025)	(0.550)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	3,440	3,440	3,440	3,440
Adjusted R^2					0.280	0.248	0.582	0.600

Note: Regressions with the first and second principal components of disruptive innovation and innovation scores as the main explanatory variables. All regressions also include an intercept, Fama and French (1997) 48 industry dummies, and calendar year dummies. The real value of patents equals its nominal value deflated to 1982 (million) dollars using the CPI. The sample includes only eventually granted patents filed before the IPO. Control variables include High-tech, EPS+, VC, Share overhang, Top underwriter, Ln(Age), Ln(Sales), Tangibility, Leverage, ROA, R&D intensity, Cash, and Ln(Tobin's Q). Standard errors clustered by industry are shown in parentheses. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 5. Innovation and first day return

	Dependent variable: First day return			
	(1)	(2)	(3)	(4)
DIS		35.469***		
		(9.146)		
Innovation			27.932***	
			(5.231)	
Total innovation				22.861***
				(3.649)
Up revision	0.842*	0.826*	0.834**	0.825*
	(0.434)	(0.435)	(0.389)	(0.436)
High tech	2.043	0.208	1.688	0.569
	(2.315)	(2.381)	(2.239)	(2.159)
VC	12.764***	12.599***	13.221***	13.032***
	(4.220)	(4.173)	(2.619)	(4.299)
EPS+	2.082	1.814	2.207	2.011
	(1.845)	(1.633)	(2.403)	(1.813)
Market return	1.396***	1.417***	1.400***	1.413***
	(0.531)	(0.535)	(0.307)	(0.531)
Share overhang	0.916***	0.888**	0.902***	0.886**
	(0.346)	(0.346)	(0.336)	(0.347)
Top underwriter	2.365	2.211	2.379	2.278
	(1.899)	(1.851)	(2.116)	(1.901)
Ln(age)	-3.979**	-3.968**	-4.196***	-4.149**
	(1.959)	(1.960)	(1.276)	(1.959)
Ln(sales)	1.075**	1.038**	0.849**	0.867**
	(0.448)	(0.409)	(0.394)	(0.355)
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.205	0.206	0.208	0.208

Note: Regressions with the first day return as the dependent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 6. Innovation and first day trading volume (Ln(volume))

	Dependent variable: Natural logarithm of trading volume			
	(1)	(2)	(3)	(4)
DIS		0.795***		
		(0.232)		
Innovation			1.171***	
			(0.212)	
Total innovation				0.805***
				(0.149)
Up revision	0.0003	-0.0001	-0.0001	-0.0003
	(0.002)	(0.002)	(0.002)	(0.002)
High tech	0.309***	0.268***	0.294***	0.257***
	(0.072)	(0.066)	(0.072)	(0.067)
VC	0.266***	0.262***	0.285***	0.276***
	(0.040)	(0.042)	(0.040)	(0.041)
EPS+	-0.404*	-0.410*	-0.399*	-0.407*
	(0.220)	(0.223)	(0.216)	(0.220)
Market return	0.016***	0.016***	0.016***	0.016***
	(0.006)	(0.006)	(0.006)	(0.006)
Share overhang	0.008*	0.007	0.007	0.007
	(0.004)	(0.005)	(0.005)	(0.005)
Top underwriter	0.158***	0.155***	0.159***	0.155***
	(0.048)	(0.048)	(0.047)	(0.048)
Ln(age)	0.058	0.058	0.049	0.052
	(0.039)	(0.039)	(0.036)	(0.037)
Ln(sales)	0.219***	0.219***	0.210***	0.212***
	(0.026)	(0.025)	(0.024)	(0.024)
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.396	0.398	0.406	0.405

Note: Regressions with natural logarithm of first day trading volume as the dependent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 7. Innovation and first day bid-ask spread

	Dependent variable: Bid-ask spread			
	(1)	(2)	(3)	(4)
DIS		4.873^{**}		
		(2.263)		
Innovation			2.720^{*}	
			(1.451)	
Total innovation				2.540^{**}
				(1.064)
Up revision	0.024^{*}	0.021	0.023	0.022
	(0.014)	(0.014)	(0.014)	(0.014)
High tech	3.149^{***}	2.897^{***}	3.114^{***}	2.985^{***}
	(0.572)	(0.584)	(0.572)	(0.576)
VC	1.818^{***}	1.795^{***}	1.862^{***}	1.848^{***}
	(0.406)	(0.406)	(0.407)	(0.406)
EPS+	2.331^{**}	2.294^{**}	2.344^{**}	2.323^{**}
	(0.996)	(0.995)	(0.995)	(0.995)
Market return	0.074	0.077	0.074	0.076
	(0.050)	(0.050)	(0.050)	(0.050)
Share overhang	0.037	0.033	0.036	0.034
	(0.048)	(0.048)	(0.048)	(0.048)
Top underwriter	-0.615^{*}	-0.636^{*}	-0.613^{*}	-0.624^{*}
	(0.356)	(0.356)	(0.356)	(0.356)
Ln(age)	-1.062^{***}	-1.061^{***}	-1.083^{***}	-1.081^{***}
	(0.192)	(0.192)	(0.192)	(0.192)
Ln(sales)	-0.497^{***}	-0.502^{***}	-0.519^{***}	-0.520^{***}
	(0.089)	(0.089)	(0.090)	(0.089)
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.384	0.384	0.384	0.385

Note: Regressions with the bid-ask spread on the first trading day as the dependent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 8. Innovation and price revision

	Dependent variable: Price revision			
	(1)	(2)	(3)	(4)
DIS		15.072***		
		(4.725)		
Innovation			4.806***	
			(1.572)	
Total innovation				5.917***
				(1.317)
High tech	0.223	-0.555	0.162	-0.158
	(0.646)	(0.696)	(0.657)	(0.678)
VC	-0.958	-1.023	-0.878	-0.885
	(0.653)	(0.675)	(0.657)	(0.652)
EPS+	0.995	0.871	1.015	0.970
	(0.918)	(0.882)	(0.920)	(0.908)
Market return	0.265***	0.272***	0.265***	0.268***
	(0.097)	(0.096)	(0.098)	(0.097)
Share overhang	0.175	0.162	0.172	0.166
	(0.125)	(0.122)	(0.123)	(0.121)
Top underwriter	0.139	0.072	0.141	0.116
	(0.639)	(0.625)	(0.649)	(0.644)
Ln(age)	-0.406	-0.399	-0.443	-0.449
	(0.341)	(0.335)	(0.334)	(0.330)
Ln(sales)	0.244	0.228*	0.205	0.190
	(0.150)	(0.136)	(0.137)	(0.129)
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.019	0.024	0.020	0.023

Note: Regressions with price revision as the dependent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 9. Innovation and post-IPO stock performance

	Dependent variable: One-year post-IPO abnormal return		
	(1)	(2)	(3)
DIS	67.162***		
	(23.467)		
Innovation		20.523	
		(17.089)	
Total innovation			25.249**
			(11.983)
First day return	-0.077	-0.035	-0.049
	(0.053)	(0.054)	(0.065)
DIS*First day return	0.066		
	(0.306)		
Innovation *First day return		-0.257	
		(0.287)	
Total innovation *First day return			-0.084
			(0.198)
Year dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	2,552	2,552	2,552
Adjusted R^2	0.048	0.044	0.045

Note: Regressions with the one-year abnormal return as the dependent variable. Control variables include high-tech, EPS+, VC, first day return, Top underwriter, ln(age), ln(total assets), tangibility, leverage, ROA, and Tobin's Q. The regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 10. Post-IPO firm policies, performance, valuation and innovation activities

Panel A. Disruptive innovation (DIS)

	Dependent variable									
	Leverage	ROA	Tobin's Q	Cash	Growth	R&D/AT	Patents	Citations	Patent value	Volatility
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DIS (t-1)	-12.023**	-11.424**	0.660***	11.045***	73.403*	20.113***	1.945***	1.378***	0.570***	0.111***
	(5.290)	(5.013)	(0.198)	(2.694)	(39.571)	(2.451)	(0.301)	(0.227)	(0.161)	0.025
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,384	3,384	3,384	3,384	3,384	3,384	3,440	3,440	3,440	2,552
Adjusted R^2	0.264	0.394	0.391	0.609	0.304	0.699	0.237	0.242	0.159	0.328

Panel B. Innovation

	Dependent variable									
	Leverage	ROA	Tobin's Q	Cash	Growth	R&D/AT	Patents	Citations	Patent value	Volatility
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Innovation (t-1)	-7.622	1.291	-0.111	-6.037**	14.063	-9.118***	0.163	-0.182	0.260**	0.053***
	(5.848)	(5.545)	(0.219)	(2.980)	(24.460)	(2.730)	(0.194)	(0.146)	(0.103)	(0.019)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,384	3,384	3,384	3,384	3,384	3,384	3,440	3,440	3,440	2,552
Adjusted R^2	0.263	0.393	0.389	0.608	0.305	0.694	0.227	0.234	0.157	0.326

Note: Regressions with post-IPO firm policies or innovation activities or return volatility as the dependent variable. Panel A presents the regressions with DIS as independent variable. Panel B present the regressions with innovation score as independent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. The real value of patent is equal to the nominal value deflated to 1982 (million) dollars using the CPI. Control variables include high-tech, EPS+, VC, first day return, Top underwriter, ln(age), ln(total assets), Tangibility, leverage, ROA, and Tobin's Q. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 11. Regressions of IPO outcomes for subsamples of technology and non-technology firms

Panel A. Technology firms

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	49.679**	15.784***	1.137***	6.792**	62.572**
	(20.611)	(5.194)	(0.221)	(2.665)	(31.683)
Year dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	1,109	1,109	1,109	1,109	858
Adjusted R^2	0.222	0.023	0.225	0.313	0.109

Panel B. Non-technology firms

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	30.794**	13.143***	0.610***	4.155**	54.523*
	(15.005)	(4.795)	(0.132)	(1.672)	(32.289)
Year dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	2,331	2,331	2,331	2,331	1,694
Adjusted R^2	0.179	0.016	0.358	0.328	0.023

Note: Regressions with the first-day return, price revision, natural logarithm of trading volume, bid-ask spread, and one-year post-IPO abnormal return as the dependent variables. We define Technology firms as in Loughran and Ritter (2004, Appendix D). All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors in parentheses are clustered by industry. Control variables include up revision (except for model 2), EPS+, VC, market return, share overhang, Top underwriter, ln(age), and ln(sales). *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table 12. Heterogeneity in IPO outcomes

Panel A. Firm age

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	76.179***	22.457***	1.074***	5.254*	66.888**
	(14.683)	(4.083)	(0.290)	(2.788)	(26.288)
Old	6.824**	0.747	-0.154**	-1.582***	-4.850
	(3.401)	(0.948)	(0.067)	(0.484)	(6.044)
DIS*Old	-88.505***	-16.103***	-0.603	-0.765	5.693
	(18.709)	(5.215)	(0.370)	(3.530)	(33.763)
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	2,552
Adjusted R^2	0.211	0.027	0.401	0.383	0.048

Panel B. Firm size

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	38.988***	17.488***	0.894***	5.412**	93.382***
	(14.372)	(4.002)	(0.275)	(2.736)	(26.046)
Large	11.037***	2.289***	0.655***	-1.902***	3.577
	(2.807)	(0.783)	(0.054)	(0.534)	(5.840)
DIS*Large	-4.866	-5.108	-0.030	-1.827	-54.581
	(18.891)	(5.270)	(0.361)	(3.596)	(33.862)
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	2,552
Adjusted R^2	0.212	0.027	0.438	0.383	0.049

Panel C. Patent possession

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	28.660**	18.387***	0.798***	4.817*	44.803*
	(14.413)	(3.993)	(0.285)	(2.783)	(26.308)
Has patents	-0.788	-1.086	0.093	0.368	0.404
	(3.388)	(0.942)	(0.067)	(0.651)	(5.896)
DIS*Has patents	15.712	-5.625	-0.144	1.133	52.040
	(20.432)	(5.677)	(0.404)	(3.938)	(36.286)
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	2,552
Adjusted R^2	0.206	0.026	0.396	0.361	0.049

Panel D. VC-backing

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	21.499***	22.643***	0.694*	7.484**	52.834*
	(7.433)	(8.732)	(0.422)	(3.243)	(27.534)
VC	6.266*	0.571	0.242***	2.276***	8.294
	(3.393)	(0.789)	(0.050)	(0.608)	(5.354)
DIS*VC	56.461**	-14.957	0.197	-4.665*	-9.293
	(23.600)	(9.683)	(0.463)	(2.439)	(34.945)
Year dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	2,552
Adjusted R^2	0.184	0.026	0.398	0.384	0.042

Note: Regressions with the first-day return, price revision, natural logarithm of trading volume, bid-ask spread, and one-year post-IPO abnormal return as the dependent variables. Panel A (B) examines whether the effect of DIS varies for subsamples partitioned at the sample median by IPO age (firm size). Panel C (D) examines whether the effect of DIS varies by patent possession (venture capital investment). All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors in parentheses are clustered by industry. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Appendix

Table A.1. Keywords for top 29 disruptive technologies (Bloom et al. 2021)

- 3D printing
- Autonomous car
- Bispecific antibody
- Cloud computing
- Computer vision
- Drug conjugates
- Electronic gaming
- Millimeter wave
- Machine learning/AI
- Fingerprint sensor
- Fracking
- GPS
- Hybrid vehicle
- Lane departure warning
- Lithium battery
- OLED display
- Online streaming
- RFID
- Search Engine
- Smart devices
- Social networking
- Software defined radio
- Solar power
- Stent graft
- Touch screen
- Virtual reality
- Wi-fi
- Wireless charging
- Mobile payment

Table A.2. Top forty disruptive technologies in IPO prospectuses ranked by cosine similarity

1. Wifi	11. Gaming platform	21. Real time communication	31. Laser
2. Navigation system	12. Millimeter wave	22. Satellite communication	32. Machine learning/AI
3. Bidirectional	13. 3D printing	23. Video conferencing	33. Bluetooth
4. Computing platform	14. Fracking	24. Smart phone	34. Solar technology
5. Microprocessor	15. Streaming video	25. Semiconductor device	35. Flash memory
6. RFID	16. Battery system	26. Visualization system	36. Ultrasonic technology
7. Touch screen	17. Fingerprint	27. Cloud computing	37. Xray technology
8. Lithium	18. GPS technology	28. Search technology	38. Radar
9. OLED	19. Streaming media	29. Logic chip	39. Barcode scanner
10. Robot	20. Wireless devices	30. Sensor technology	40. Accelerometer

Table A.3. Disruptive innovation dictionary

128bit, 32bit, 3d, 64bit, accelerometer, access_datum, access_network, actuator, adapter, ai, airbag, amplifier, amplifier_product, analog_converter, analog_semiconductor, analogtodigital, analyzer, antibody, app, application_specific_integrate_circuit, applicationspecific, , asymmetric, atm_switch, audio, automation_system, autonomous, backlight, bandwidth_allocation, barcode_scanner, base_station, battery, battery_charger, battery_system, batteryoperated, bidirectional, biometric, blade_server, broadband, broadband_access, broadband_communication, broadband_service_provider, business_user, cable_modem, camcorder, camera, cartridge, catheter, cdrom_drive, ce_device, channel, channelize, character_recognition, chip, cinema_processor_equipment, circulator, client/server_application, client/server_system, cloud, cloud_architecture, cloud_computing, cloud_technology, cloudbased, codec, coherent, coinop, collaboration_tool, color_display, color_output, color_screen, colorchange, communication_application, communication_capability, communication_experience, communication_infrastructure, communication_network, communication_platform, communication_subsystem, communication_technology, component_technology, compression_technology, computing_device, computing_platform, connectivity, connectivity_product, consumer, consumer_application, consumer_device, consumer_electronic_device, consumergrade, content_creator, control_software, controller_chip, copper, copper_wire, cordless, core_functionality, coupler, coupling, crossplatform, ctp_system, customizable, customize, database_connectivity, datarich, datum_application, datum_communication_technology, datum_format, datum_network, datum_stream, dc_power, dcms, decryption, delivery_platform, demodulator, dense, desktop, desktop_application, desktop_computer, desktop_pc, detection_technology, detector, devops, digital, digitally, digitized, diode, disc_player, diskbased, display_device, display_product, display_system, display_technology, downloadable, dram, droplet, dsl_technology, dvd_recorder, dvr, dwdm, dwdm_system, dynamic, dynamic_block_architecture, edfa, electrode, electromechanical, electron_beam, electronic, encoder, endoscope, endpoint_device, endtoend, enduser, energyefficient, enterprise_application, enterprise_network, enterprise_server, enterpriseclass, enterprisegrade, entertainment_system, eq, ethanol, ethernet, execution_engine, featurerich, fiber_amplifier, fiber_laser, fiberbased, fibre_channel, file_format, filter, fingerprint, flash_memory, flash_memory_card, fluidic, form_factor, fracking, fuel_cell, fullfeatured, game, gaming_platform, geometry, gps, gps_technology, gpus, guidance_system, gyroscope, handheld, handset, hardware_device, hardwarebased, hdtv, highbandwidth, highdefinition, highdefinition_video, highdensity, highend, highfrequency, highperformance, highperformance_application, highpower, highpowered, highresolution, highspeed, highvoltage, highvolume, home_network, home_networking, home_pc, host_computer, host_system, hybrid, hydrophone, ic, ic_solution, idsl, image_capture, image_capture_capability, image_processing, image_processing_system, imagery, imaging_capability, imaging_device, imaging_system, imaging_technology, incar_display, indash, inductive, industry_standard_technology, infrare, infrared, inlay, input/output, intelligent, interact, interaction, interactive, interface_device, interface_solution, internet_appliance, internet_device, internetenabled, interwork, invehicle, iot, ip_network, ipbased, isdn, isolator, laser, laser_printer, lcd, lcd_monitor, led, lightweight, line_card, lithium, logic_chip, logic_device, lowcost, lowmaintenance, lownoise, lte, machine_learning, mass_spectrometry, media_format, memory_card, memory_chip, memory_device, memory_solution, memory_technology, menudriven, mesh_architecture, metrology_equipment, microarray, microcontroller, microcontrollers, microphone, micropress, microprocessor, microprocessorbased, microscope, microservice, microturbine, millimeterwave, millimeterwave_application, miniaturization, miniaturize, missioncritical, mixedsignal, mixer, mobile, modular, module, monochrome, monolithic, mosfet, motherboard, motionbased, mountable, mpeg, multichannel, multicore, multifunction, multifunctional, multilayer, multilayered, multimedia, multimedia_application, multimedia_capability, multimedia_content, multimedia_pc, multimedia_server, multiplatform, multiplex, multiplexing, multiplexor, multiport, multiprotocol, multiservice, multithreaded, native, natively, navigation_system, network, network_card, network_connectivity, network_edge, network_interface, network_product, networked, networking_equipment, networking_product, nextgeneration, nonlinear, nonmagnetic, nonvolatile, notebook_computer, objectbased, oled, online, operator_interface, optical, optics, oscillator, packaging_technology, packetbased, palmsized, parallel_processing, pcbased, penbased, photodiode, photonic, planar, platform, platformindependent, portable, pot, power_amplifier, power_control, power_generation_system, power_management, power_management_device, power_source, powerful, prebuilt, preprogrammed, programmability, programmable, protocol_language, purposebuilt, pv_module, qlikview, quality_video, rackmounted, radar, radio_system, radiography, realtime_communication, rechargeable, reconfigurable, rf_module, rfic, rfid, rich, risc_processor, robot, robotic, roombased, router, rugged, ruggedize, san, satellite_communication, scanner, scanning, screwin_cartridge_valve, seamless, search_technology, security_application, semiconductor, semiconductor_device, semiconductor_laser, semiconductor_solution, sensor, sensor_technology, sensorfusion_technology, server, server_adapter, server_technology, serverbased, service_provider_network, settop_box, signal_processing_algorithm, signal_processor, silicon_timing_device, simulator, singlemode, singlepiece, smart, smartphone, socket, software_algorithm, software_application, software_download, software_platform, software_solution, softwarebased, solar, solidstate_lighting, solution, 33erilo, splitter, sql, sram, standardsbased, steerable, stent, stepper, stereo, storage_area_network, storage_array, storage_router, storage_solution, storage_system, storage_technology, streaming, streaming_audio, streaming_media, streaming_video, subsystem, surf, switch, switching, switching_platform, switching_technology, symmetric, synchronize, synchronous, synthesizer, system, system_product, systemlevel, systemslevel_solution, tape_library, tape_storage_system, technology_platform, television_application, television_display, test_instrument, tester, touch, touch_pad, touch_screen, touchpad, touchstyk, 33eril, transceiver, transcoding, transducer, transmission_media, transport_layer, tunable, twoway, uav, ubiquitous, ultrasonic, uncompressed, uncooled, unix_system, upgradeable, 33erilog, versatile, video, video_application, video_camera, video_compression, video_compression_technology, video_conferencing, video_datum, video_delivery, video_device, video_display, video_image, video_networking, video_technology, videoconferencing, videoondemand, viewfinder, virtual, virtualization, visualization, visualization_system, voice_communication, voice_transmission, voiceenabled, voip, voltage_regulator, wavelength_division_multiplexing, wdm, wearable, web, web_application, web_application_firewall, web_server, webcam, wideband, wifi, wifi_hotspot, windows_ce, wireless, wirelessly, xml, xray_technology

Table A.4. Examples of language in IPO prospectuses with high DIS

IPO firm with DIS at 90th percentile	Language in prospectus
SYNAPTICS INC	We have developed and own an extensive array of application specific integrated circuit, or ASIC, firmware, software, pattern recognition, and touch sensing technologies, which provide us with significant competitive advantages. Our intellectual property includes more than 57 patents issued and 25 patents pending. We conduct ongoing research, development, and engineering programs that concentrate on advancing our technologies and expanding them to serve new markets, enhancing the quality and performance of our product solutions, and developing new product solutions. Our technology enables us to develop innovative, intuitive, user-friendly interfaces that address the needs of our customers and improve their competitive positions.
Oplink Communications Inc	We had 38 engineers, 12 of whom hold Ph.D. degrees, and 98 technicians and operators, involved in research and development of our products. Our engineering team has extensive design, package, processing, and software experience in the fields of fiber optic components, integrated optic interfaces and systems. To date, we have been granted 10 patents by, and have 22 patent applications pending with, the U.S. Patent and Trademark Office for various technologies and products, including DWDM interleavers, DWDM modules, multi-channel optic filter arrays, high reliability fused couplers, circulators, compact optical switches and polarization beam combiners. Research and development expense was $1.4 million. We have increased, and intend to continue to increase, our research and development budget and staff to enhance our current fiber optic components and modules, and to develop new technologies and products to serve the next-generation communication markets.
Aeroflex Holding Corp	We are a leading global provider of radio frequency, or RF, and microwave integrated circuits, components and systems used in the design, development, and maintenance of technically demanding, high-performance wireless communication systems. Our solutions include highly specialized microelectronic components and test and measurement equipment used by companies in the space, avionics, defense, commercial wireless communications, medical and other markets. We have targeted customers in these end markets because we believe our solutions address their technically demanding requirements. We were founded in 1937 and have proprietary technology that is based on extensive know-how and a long history of research and development focused on specialized technologies, often in collaboration with our customers.
LightPath Technologies	LightPath has targeted specific applications in each of these areas for new product launches in the near future. For example, in glass a spheres: laser tools, gun sights, biomedical instruments and telecommunication subsystems; in specialty optics: laser line generators, industrial tools, optical cutting/welding, scientific lasers, semiconductors metrology systems and telecommunication subsystems; and in infrared optics: thermal imaging, security cameras, thermography, gas sensing and defense targeting and tracking.
Zoran CORP	We develop and market integrated circuits, or Ics, integrated circuit cores and embedded software used by original equipment manufacturers, or OEMs, in digital video and audio products for commercial and consumer markets. We also provide complete, copy-ready system reference designs based on our technology that help our customers produce commercial and consumer products more quickly and cost-effectively. Our products consist of integrated circuits and related products used in digital versatile disc players, or DVDs, movie and home theater systems, digital cameras, and video editing systems. We also provide high performance, low-power application processors, technology, and products for the multimedia mobile telephone market.

IPO firm with DIS at 90th percentile, and without filed patents before IPO	Language in prospectus
SOFTWORKS INC	The Company does not currently have any patents or pending patent applications and relies principally on a combination of: (i) trade secret, copyright and trademark laws; (ii) nondisclosure, use restriction and other contractual restrictions and agreements; and (iii) certain technical measures to protect its technology, including, without limitation, its SST technology. SOFTWORKS, Inc. develops, markets, licenses and supports a family of enterprise systems management software products for data and storage management and performance management.

	The Company's products are developed using "SST," its proprietary combination of a design strategy, a development methodology and a set of core technologies. Recently, the Company's products have been expanded to include UNIX-based storage management products, and the Company expects to introduce NT-based storage management products in the near future.
BE INC	To date, we have no patents, and existing copyright laws afford only limited protection for our software. We offer the BeOS(R) operating system, an operating system designed for digital media applications and Internet appliances. BeOS is capable of maximizing the performance of digital media applications that run on a wide range of devices including Internet appliances, desktop PCs and high- performance multiprocessor workstations. BeOS allows users to simultaneously operate multiple audio, video, image processing and Internet-based software applications while maintaining system stability, media quality and processor performance. BeOS provides professional users and enthusiasts with a high-performance environment to quickly and easily develop applications and content and is designed to facilitate the integration of new technologies.
NORTHEAST OPTIC NETWORK INC	Currently, we have not filed any patent applications. We intend to prepare applications and to seek patent protection for our systems and services to the extent possible. We began developing technical standards for delivery of DSL-based services within our target markets through a joint effort with Bell Atlantic. We provide a wide variety of value-added services to customers, including remote network management and monitoring, network security, virtual private networks, PBX emulation, Internet access, e-commerce and other data applications. Our network supports both legacy telecommunications infrastructures, including traditional voice, and newer, more efficient packet-based communications, such as Asynchronous Transfer Mode (ATM), Frame Relay and Internet Protocol (IP).
Vertex Inc	Our success depends, in part, upon our proprietary technology, processes, trade secrets, and other proprietary information and our ability to protect this information from unauthorized disclosure and use. Vertex delivers comprehensive tax solutions that enable global businesses to transact, comply and grow with confidence We have also established an innovation lab where we design, test and incubate next-generation tax solutions and adjacent market opportunities like blockchain, payment platforms and machine learning technologies. We have pioneered tax technology for over 40 years. Today, our software enables tax determination, compliance and reporting, tax data management and document management with powerful pre-built integrations to core business applications used by most companies, particularly those applications that have a significant impact on global commerce. Our software is fueled by over 300 million data-driven effective tax rules and supports indirect tax compliance in more than 19,000 jurisdictions worldwide.
ACE COMM CORP	The Company currently has no patents or patent applications pending. The Company's products perform such functions as billing data collection, network surveillance, alarm processing and network management for some of the largest carriers and enterprises in the world. The Company's network management products consist of standardized software-based systems that enable network managers to manage voice and data communications by automating service administration, tracking network connections, detecting system errors and malfunctions, controlling network inventory assignments and configuration, monitoring traffic and performing billing functions. The Company's network management products are designed to increase the efficiency of communication operations and incorporate recent developments in object-oriented development, real-time response, client server architecture and graphical user interfaces.
NETWORK ACCESS SOLUTIONS CORP	We currently have no patents or patent applications pending. We also rely on unpatented trade secrets and know-how to maintain our competitive position. Through our CuNet (pronounced "CopperNet") branded service, we offer our customers high speed connectivity in the Bell Atlantic region using digital subscriber line, or DSL, technology. As a complement to CuNet, we offer our customers a complete suite of value-added enterprise networking solutions, including network integration, network management, network security and professional services. Our network supports both legacy telecommunications infrastructures, including traditional voice, and newer, more efficient packet- based communications, such as Asynchronous Transfer Mode (ATM), Frame Relay and Internet Protocol (IP).

Table A.5. Horse race regressions controlling for Innovation

	Dependent variable:			
	First day return	Price revision	Ln(volume)	Bid-ask spread
	(1)	(2)	(3)	(4)
DIS	28.441**	14.437***	0.473**	4.506*
	(12.549)	(5.431)	(0.198)	(2.454)
Innovation	21.562***	1.839	1.066***	1.748
	(6.741)	(2.218)	(0.226)	(1.565)
Up revision	0.809*		-0.001	0.014
	(0.423)		(0.002)	(0.014)
High tech	-0.154	-0.509	0.150**	2.896***
	(2.825)	(0.655)	(0.072)	(0.593)
VC	13.334***	-0.960	0.217***	1.841***
	(4.115)	(0.709)	(0.051)	(0.412)
EPS+	1.352	0.627	-0.491***	2.477**
	(2.373)	(0.771)	(0.153)	(1.010)
Market return	1.233**	0.285***	0.017***	0.059
	(0.573)	(0.098)	(0.006)	(0.051)
Share overhang	1.005***	0.171	0.010	0.074
	(0.377)	(0.126)	(0.007)	(0.048)
Top underwriter	2.344	0.103	0.175***	-0.727**
	(1.868)	(0.650)	(0.047)	(0.360)
Ln(Age)	-4.036**	-0.426	0.007	-1.007***
	(1.920)	(0.329)	(0.014)	(0.194)
Ln(Sales)	0.890***	0.218	0.220***	-0.513***
	(0.317)	(0.135)	(0.017)	(0.091)
Year dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.211	0.028	0.416	0.365

Note: Regressions with disruptive innovation score (DIS), the first day return, price revision, natural logarithm of trading volume or bid-ask spread as the dependent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table A.6. Variable definitions

Variable	Definition	Source
Independent variable of interest		
DIS	The weighted-frequency of disruptive technology-related words in IPO prospectus, i.e., the number of technology-related words divided by the total number of words in a prospectus, adjusted to reduce the weight of more frequent words. (%)	Prospectus
Innovation	The weighted-frequency of innovation-related words in IPO prospectus, i.e., the number of innovation-related words divided by the total number of words in a prospectus, adjusted to reduce the weight of more frequent words. (%)	Prospectus
Total innovation	Sum of the disruptive innovation and innovation scores	Prospectus
Dependent variable		
First day return	The percentage change from the offering price to the closing price on the first day of trading	SDC
Price revision	The percentage change from the midpoint of the anticipated price range provided in the IPO prospectus to the actual offering price.	SDC
Ln(volume)	Natural logarithm of trading volume on the first trading day	CRSP
Bid-ask spread	The percentage difference between the ask price and bid price divided by the average of closing bid and ask prices	CRSP
Long run abnormal return	Cumulative percentage abnormal return on the IPO stock over the 12 months after the first post-IPO trading date, computed using monthly DGTW benchmark portfolio returns, as in Daniel et al. (1997)	CRSP
Volatility	Standard deviation of monthly stock returns over 12 months after IPO trading dates.	CRSP
Patents	The number of patents filed before or after IPO dates	Noah Stoffman's website
Citations	The number of forward citations per patent	Noah Stoffman's website
Patent value (Nominal)	Value of patent in millions of nominal dollars	Noah Stoffman's website
Patent value (Real)	Value of patent deflated to 1982 (million) dollars using the CPI	Noah Stoffman's website
Control variables		
Up revision	The percentage difference between offer price and mid filling price if offer price is larger than mid filling price, else zero	SDC
Age	Firm age = IPO year – Firm founding year	SDC, Jay Ritter's website
VC	One if IPO is backed by venture capital, else zero	SDC
High tech	One if IPO is high technology firm, else zero	SDC
EPS+	One if IPO has a positive earnings per share, else zero	SDC
Top underwriter	One if IPO has an underwriter ranked in the top 8 in Jay Ritter's list, else zero	SDC, Jay Ritter's website

Market return	Buy-and-hold percentage return on CRSP Nasdaq value-weighted index on the 15-trading days prior to the IPO date	CRSP
Share overhang	The number of shares retained by pre-existing owners divided by the number of shares in the initial offering (%)	SDC
BM	Book value of common equity divided by market value of common equity (%)	COMPUSTAT
Sales	Natural log of firm sales in million dollars	COMPUSTAT
R&D	R&D spending divided by total assets, xrd/at (%)	COMPUSTAT
ROA	EBITDA divided by total assets, ebitda/at (%)	COMPUSTAT
Leverage	Long-term debt plus debt in current liabilities divided by total assets, (ldtt + dlc) / at (%)	COMPUSTAT
Tobin's Q	Computed as: ((prcc_f * csho) +pstk+dltt+dlc) / at	COMPUSTAT
Total assets	Natural log of total assets	COMPUSTAT
Tangibility	Property, plant, and equipment divided by total assets, ppent/at (%)	COMPUSTAT
Growth	Sale growths in five years after IPO year	COMPUSTAT
Negative sentiment	Percentage of negative words using Loughran and McDonald (2011) dictionary	Prospectus
Positive sentiment	Percentage of positive words using Loughran and McDonald (2011) dictionary	Prospectus

Table A.7. Regressions of IPO outcomes with additional controls

Panel A. Controls for Innovation

	Dependent variable			
	First day return	Price revision	Ln(volume)	Bid-ask spread
	(1)	(2)	(3)	(4)
DIS	29.153***	16.576***	0.683***	4.597*
	(7.903)	(4.820)	(0.177)	(2.613)
R&D/AT	0.066	-0.020	-0.0001	-0.005
	(0.064)	(0.019)	(0.001)	(0.008)
Ln(patents)	-2.276**	-0.168	0.080***	-0.030
	(1.083)	(0.718)	(0.018)	(0.275)
Ln(citations)	2.295*	-0.197	0.011	0.054
	(1.388)	(0.254)	(0.010)	(0.190)
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.205	0.022	0.406	0.385

Panel B. Control for Negative sentiment

	Dependent variable:			
	First day return	Price revision	Ln(volume)	Bid-ask spread
	(1)	(2)	(3)	(4)
DIS	38.357***	13.173**	0.495***	6.552**
	(12.731)	(6.117)	(0.184)	(2.918)
Negative Sentiment	17.463**	2.155*	0.009	3.061***
	(8.527)	(1.213)	(0.113)	(0.778)
Year dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
State dummy	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.209	0.025	0.352	0.366

Panel C. Control for Positive sentiment

	Dependent variable:			
	First day return	Price revision	Ln(volume)	Bid-ask spread
	(1)	(2)	(3)	(4)
DIS	27.860***	16.676***	0.550***	4.859**
	(8.336)	(4.172)	(0.166)	(1.995)
Positive Sentiment	11.191***	0.769	0.024	1.960**
	(3.405)	(1.800)	(0.118)	(0.972)
Year dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
State dummy	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440
Adjusted R^2	0.253	0.026	0.411	0.403

Note: Regressions of the first-day return, price revision, natural logarithm of trading volume, or bid-ask spread as the dependent variable. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, calendar year, and state dummies. Standard errors in parentheses are clustered by industry, year, and state. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table A.8 Regressions of IPO outcomes excluding the technology bubble period of 1998-1999

	Dependent variable:				
	First day return	Price revision	Ln(volume)	Bid-ask spread	LR abnormal return
	(1)	(2)	(3)	(4)	(5)
DIS	23.078**	13.095***	1.053***	8.914***	44.329**
	(11.640)	(3.521)	(0.284)	(2.794)	(21.247)
Year dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	2,859	2,859	2,859	2,859	2,115
Adjusted R^2	0.205	0.024	0.411	0.318	0.017

Note: Regressions with the first-day return, price revision, natural logarithm of trading volume, bid-ask spread, and one-year post-IPO abnormal return as the dependent variables. The sample excludes IPOs during the technology bubble period of 1998-1999. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors in parentheses are clustered by industry. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table A.9. Propensity score matching

Panel A. Full sample

	Dependent variable:									
	First day return	Price revision	Ln(volume)	Bid-ask spread	Cash	Ln(Tobin's Q)	R&D/AT	Patents	Citations	Patent value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DIS (ATE)	4.987***	1.711***	0.192***	0.840**	1.321*	0.086***	4.821***	0.153***	0.190***	0.101***
	(1.736)	(0.430)	(0.033)	(0.328)	(0.720)	(0.026)	(0.663)	(0.045)	(0.031)	(0.024)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,440	3,440	3,440	3,440	3,384	3,384	3,384	3,440	3,440	3,440
Adjusted R^2	0.207	0.041	0.439	0.410	0.597	0.326	0.668	0.208	0.228	0.173

Panel B. First quartile vs Third quartile

	Dependent variable:									
	First day return	Price revision	Ln(volume)	Bid-ask spread	Cash	Ln(Tobin's Q)	R&D/AT	Patents	Citations	Patent value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DIS (ATE)	14.058***	2.363***	0.243***	2.352***	3.945***	0.109**	6.942***	0.413***	0.336***	0.123***
	(2.391)	(0.659)	(0.047)	(0.445)	(1.097)	(0.047)	(0.955)	(0.068)	(0.049)	(0.032)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,719	1,719	1,719	1,719	1,691	1,691	1,691	1,719	1,719	1,719
Adjusted R^2	0.141	0.078	0.467	0.406	0.617	0.375	0.744	0.320	0.286	0.246

Note: Regressions with disruptive innovation as the independent variable. Propensity score matching with the average treatment effect (ATE) is reported. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. Standard errors clustered by industry are shown in parentheses. Control variables are the same as in previous tables. The subsample includes only IPO firms with DIS in the first and third quartile. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

References

Azimi, M. and Agrawal, A., 2021. Is positive sentiment in corporate annual reports informative? Evidence from deep learning. *Review of Asset Pricing Studies* 11(4), 762-805.

Baker, M. and Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. *Journal of Finance* 55(5), 2219-2257.

Bellstam, G., Bhagat, S. and Cookson, J.A., 2021. A text-based analysis of corporate innovation. *Management Science* 67(7), 4004-4031.

Bloom, N., Hassan, T.A., Kalyani, A., Lerner, J. and Tahoun, A., 2021. The diffusion of disruptive technologies. National Bureau of Economic Research, WP No. w28999.

Bowen, D.E., Frésard, L. and Hoberg, G., 2023. Rapidly evolving technologies and startup exits. *Management Science*, 69(2), 940-967.

Brav, A. and Gompers, P.A., 1997. Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and nonventure capital-backed companies. *Journal of Finance* 52(5), 1791-1821.

Chen, M.A., Wu, Q. and Yang, B., 2019. How valuable is FinTech innovation? *Review of Financial Studies* 32(5), 2062-2106.

Christensen, C. and Raynor, M., 2013. *The innovator's solution: Creating and sustaining successful growth*. Harvard Business Review Press.

Christensen, C.M. 1997. *The Innovator's Dilemma*. Boston: Harvard Business School Press.

Christensen, C.M. and Bower, J.L., 1996. Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal* 17(3), 197-218.

Cohen, L., Gurun, U.G. and Kominers, S.D. 2019. Patent trolls: Evidence from targeted firms. *Management Science 65*(12), 5461-5486.

Crain, N., Parrino, R. and Srinivasan, R., 2021. Uncertainty, prospectus content, and the pricing of initial public offerings. *Journal of Empirical Finance* 64, 160-182.

Cumming, D.J., Köchling, G., Neukirchen, D. and Posch, P.N., 2022. Does Corporate Culture Influence IPO Pricing? Available at SSRN 4032543.

Daniel, K., Grinblatt, M., Titman, S. and Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance 52*(3), 1035-1058.

Downes, L. and Nunes, P., 2013. Big bang disruption. *Harvard Business Review*, March, 44-56.

Erel, I., Stern, L. H., Tan C. and Weisbach, M. S., 2021. Selecting directors using machine learning. *Review of Financial Studies* 34, 3226-3264.

Fama, E.F. and French, K.R., 1997. Industry costs of equity. *Journal of Financial Economics* 43(2), 153-193.

Gentzkow, M., Kelly, B. T. and Taddy, M., 2019. Text as data. *Journal of Economic Literature* 57(3), 535–74.

Guiso, L., Sapienza, P., and Zingales, L., 2015. The value of corporate culture, *Journal of Financial Economics* 117, 60–76.

Guo, R.J., Lev, B. and Zhou, N., 2004. Competitive costs of disclosure by biotech IPOs. *Journal of Accounting Research 42*(2), 319-355.

Hall, B.H. and Lerner, J., 2010. The financing of R&D and innovation. *Handbook of the Economics of Innovation*, V. 1. North-Holland, 609-639.

Hanley, K.W. and Hoberg, G., 2010. The information content of IPO prospectuses. *The Review of Financial Studies 23*(7), 2821-2864

Hanley, K.W. and Hoberg, G., 2019. Dynamic interpretation of emerging risks in the financial sector. *Review of Financial Studies*, *32*(12), 4543-4603.

Hirshleifer, D., Hsu, P.H. and Li, D., 2018. Innovative originality, profitability, and stock returns. *The Review of Financial Studies 31*(7), 2553-2605.

Hoang, D. and Wiegratz, K., 2023. Machine learning methods in finance: Recent applications and prospects. *European Financial Management*, forthcoming.

Hribar, P., 2004. Discussion of competitive costs of disclosure by biotech IPOs. *Journal of Accounting Research* 42(2), 357-364.

Kogan, L., Papanikolaou, D., Seru, A. and Stoffman, N., 2017. Technological innovation, resource allocation, and growth. *Quarterly Journal of Economics* 132(2), 665–712.

Koh, P.S. and Reeb, D.M., 2015. Missing R&D. *Journal of Accounting and Economics 60*(1), 73-94.

Kravet, T. and Muslu, V., 2013. Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies* 18, 1088-1122.

Li, K., Liu, X., Mai, F. and Zhang, T., 2021b. The role of corporate culture in bad times: Evidence from the COVID-19 pandemic. *Journal of Financial and Quantitative Analysis* 56(7), 2545-2583.

Li, K., Mai, F., Shen, R. and Yan, X. 2021a. Measuring corporate culture using machine learning. The *Review of Financial Studies* 34(7): 3265-3315

Liu, X. and Wu, B.,2021. Do IPO firms misclassify expenses? Implications for IPO price formation and post-IPO stock performance. *Management Science* 67(7), 4505-4531.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66, 35–65.

Loughran, T. and McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics* 109(2), 307-326.

Loughran, T. and Ritter, J., 2004. Why Has IPO Underpricing Changed Over Time? *Financial Management* 33(3), 5-37.

Lowry, M., 2003. Why does IPO volume fluctuate so much? *Journal of Financial Economics* 67(1), 3-40.

Lyandres, E. and Palazzo, B., 2016. Cash holdings, competition, and innovation. *Journal of Financial and Quantitative Analysis 51*(6), 1823-1861.

Mann, W., 2018. Creditor rights and innovation: Evidence from patent collateral. *Journal of Financial Economics 130*(1), 25-47.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems 26*.

Myers, S., 1977, Determinants of corporate borrowing. *Journal of Financial Economics* 5(2), 147-175.

Pacelli, J., Shi, T. and Zou, Y., 2022. Communicating corporate culture in labor markets: Evidence from job postings. SSRN, Working paper number 4235342.

Ritter, J.R., 1991., The long-run performance of initial public offerings. *Journal of Finance* 46(1), 3-27.

Saidi, F. and Žaldokas, A., 2021. How does firms' innovation disclosure affect their banking relationships? *Management Science* 67(2), 742-768.

Titman, S. and Wessels, R., 1988. The determinants of capital structure choice. *Journal of Finance* 4 (1), 1-19.

Zhou, K.Z., Yim, C.K. and Tse, D.K., 2005. The effects of strategic orientations on technology-and market-based breakthrough innovations. *Journal of Marketing* 69(2), 42-60.