Boston College

Wallace E. Carroll Graduate School of Management

Department of Finance

THREE ESSAYS IN FINTECH AND CORPORATE FINANCE

a dissertation

by

XIANG ZHENG

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy in Finance

May 2021

THREE ESSAYS IN FINTECH AND CORPORATE FINANCE

XIANG ZHENG

ADVISOR:

Thomas Chemmanur

COMMITTEE:

Francesco D'Acunto, Jonathan Reuter, Hassan Tehranian

ABSTRACT:

My Ph.D. dissertation consists of three essays. The first essay studies the economic consequence of the current patent screening process on firm performance using a machine-learning approach. Using USPTO patent application data, I apply a machine-learning algorithm to analyze how the current patent examination process in the U.S. can be improved in terms of granting higher quality patents. I make use of the quasi-random assignment of patent applications to examiners to show that screening decisions aided by a machine learning algorithm lead to a 15.5% gain in patent generality. To analyze the economic consequences of current patent screening on both public and private firms, I construct an *ex-ante* measure of past false acceptance rate for each examiner by exploiting the disagreement in patent screening decisions between the algorithm and current patent examiner. I first show that patents granted by examiners with higher false acceptance rates have lower announcement returns around patent grant news. Moreover, these patents are more likely to expire early. Next, I find that public firms whose patents are granted by such examiners are more likely to get sued in patent litigation cases. Consequently, these firms cut R&D investments and have worse operating performance. Lastly, I find that private firms whose patents are granted by such examiners are less likely to exit successfully by an IPO or an M&A. Overall, this study suggests that the social and economic cost of an inefficient patent screening system is large and can be mitigated with the help of a machine learning algorithm.

The second essay studies how investor attention affects various aspects of SEOs. Mod-

els of seasoned equity offerings (SEOs) such as Myers and Majluf (1984) assume that all investors in the economy pay immediate attention to SEO announcements and the pricing of SEOs. In this paper, we analyze, theoretically and empirically, the implications of only a fraction of investors in the equity market paying immediate attention to SEO announcements. We first show theoretically that, in the above setting, the announcement effect of an SEO will be positively related to the fraction of investors paying attention to the announcement and that there will be a post-announcement stock-return drift that is negatively related to investor attention. In the second part of the paper, we test the above predictions using the media coverage of firms announcing SEOs as our main proxy for investor attention, and find evidence consistent with the above predictions. In the third part of the paper, we develop and test various hypotheses relating investor attention paid to an issuing firm to various SEO characteristics. We empirically show that institutional investor participation in SEOs, the post-SEO equity market valuation of firms, SEO underpricing, and SEO valuation are all positively related to investor attention. Lastly, we also use the number of SEC EDGAR file downloads as an alternative proxy for investor attention, and our findings are robust to this alternative investor attention measure. The results of our identification tests show that the above results are causal.

The third essay studies how the location of a lead underwriter in its network of investment banks affects various aspects of seasoned equity offerings (SEOs). We hypothesize that investment banking networks perform an important economic role in the SEO underwriting process for SEOs, namely, that of information dissemination, where the lead underwriter uses its investment banking network to disseminate information about the SEO firm to institutional investors. Consistent with the above information dissemination role, we show that firms whose SEOs are underwritten by more central lead underwriters are associated with a smaller extent of information asymmetry in the equity market. We then develop testable hypotheses based on the information dissemination role of underwriter networks for the relationship between SEO underwriter centrality and various SEO characteristics, which we test in our empirical analysis. Consistent with the above hypotheses, we find that more central lead SEO underwriters are associated with less negative SEO announcement effects; smaller SEO offer price revisions; smaller SEO discounts and underpricing; higher immediate post-SEO equity valuations for issuing firms; and greater post-SEO long-run stock returns for issuing firms. We also find that SEOs with more central lead underwriters are associated with greater institutional investor participation. Our instrumental variable (IV) analysis using the industry-average bargaining power of underwriters relative to issuers as the instrument shows that the above results are causal. Consistent with greater value creation by more central lead underwriters, we find that more central lead underwriters receive greater compensation.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Thomas Chemmanur, for his invaluable guide and support throughout my study in the Ph.D. program. I also thank my committee members, Francesco D'Acunto, Jonathan Reuter, and Hassan Tehranian, for their helpful comments and suggestions. I also would like to thank other faculty members at the Finance Department for their helpful advice along the way.

This dissertation could not have been completed without the love of my wife, Chenting. I am indebted to her, whose constant support significantly alleviated the pressures that might otherwise have hampered the completion of this dissertation as well as this degree.

Finally, I thank my parents, parents-in-law, and friends for their love and support throughout the years. I could never get to this point without any of you.

THREE ESSAYS IN FINTECH AND CORPORATE FINANCE

XIANG ZHENG

Table of Contents

ESSAY 1: How can Innovation Screening be Improved? A Machine Learning Analysis with Economic Consequences for Firm Performance

Xiang Zheng*

Abstract

Using USPTO patent application data, I apply a machine-learning algorithm to analyze how the current patent examination process in the U.S. can be improved in terms of granting higher quality patents. I make use of the quasi-random assignment of patent applications to examiners to show that screening decisions aided by a machine learning algorithm lead to a 15.5% gain in patent generality. To analyze the economic consequences of current patent screening on both public and private firms, I construct an *ex-ante* measure of past false acceptance rate for each examiner by exploiting the disagreement in patent screening decisions between the algorithm and current patent examiner. I first show that patents granted by examiners with higher false acceptance rates have lower announcement returns around patent grant news. Moreover, these patents are more likely to expire early. Next, I find that public firms whose patents are granted by such examiners are more likely to get sued in patent litigation cases. Consequently, these firms cut R&D investments and have worse operating performance. Lastly, I find that private firms whose patents are granted by such examiners are less likely to exit successfully by an IPO or an M&A. Overall, this study suggests that the social and economic cost of an inefficient patent screening system is large and can be mitigated with the help of a machine learning algorithm.

Keywords: Machine Learning, Patent Screening, Economic Consequences, Firm Performance
JEL classification: C55, G32, O31

*Ph.D. Candidate, Finance Department, Fulton Hall 341, Carroll School of Management, Boston College, Chestnut Hill, MA 02467, Tel: (617) 552-2062, Email: zhengxr@bc.edu.

1 Introduction

"The strength and vitality of the U.S. economy depends directly on effective mechanisms that protect new ideas and investments in innovation and creativity."

– The U.S. Patent and Trademark Office

The patent system is viewed as one of the most important institutions that provide firms with innovation incentives by granting them temporary monopoly rights over their inventions. This, in turn, contributes to the technological growth in the economy (see, e.g., Nordhaus (1969), Arrow (1972), and Mansfield (1986)). However, there has been considerable criticism of the patent system: critics argue that inefficient screening of patent applications reduces, instead of increases, firms' incentives to innovate (see, e.g., Cornelli and Schankerman (1999), Lemley and Shapiro (2005), Jaffe and Lerner (2011), Schankerman and Schuett (2016), and Bessen and Maskin (2009)). Many factors may have contributed to this issue. First of all, patent examiners have faced increasing time constraints over time. On the one hand, the number of patent applications filed at the U.S. Patent and Trademark Office (USPTO) has skyrocketed over the last two decades. For example, Figure 1 shows that the number of patent applications filed at the USPTO from 2001 to 2018 has increased from 345,732 in 2001 to 643,303 in 2018.[1] Inevitably, examiners need more time to review a new patent application given the increasing number of the existing patents. On the other hand, the number pf patent examiners working in the USPTO has not kept the same pace as the increasing number of newly filed patent applications according to the Figure 1. Moreover, patent examiners, on average, spend only 19 hours reviewing an application but it takes around 25 months for an application to get its screening result (Frakes and Wasserman (2017)). Second, the USPTO also faces human capital constraints, where it fails to recruit and retain the best examiners due to fierce competitions from booming private sectors (Jaffe and Lerner (2011)). Last, the incentive structure in the USPTO favors acceptances over rejections (Merges (1999), Frakes and Wasserman (2015)). Consequently, the USPTO has been criticized for granting too many low-quality patents through a weak screening process

[1] Data source: U.S. Patent Statistics Chart and Patent Examination Data from the USPTO website.

(see, e.g., Heller and Eisenberg (1998), Jaffe and Lerner (2011), and Feng and Jaravel (2020)). Economically, inefficient patent screening may mislead investors, stunt innovators' incentives, and hurt firm performance.

Motivated by the above constraints and critics of the current patent screening system in the U.S., this paper explores the question of how to improve the patent screening process with the help of machine learning algorithms to analyze large data sets that are recently available. As discussed above, the constraints faced by the USPTO and the resulting weak screening process provide a natural setting where machine learning can offer help by reducing frictions between patent examiners and the patent office as well as relaxing the constraints faced by both parties. This idea is also partly supported by the patent office itself according to a recent news article published in the Wall Street Journal, in which the patent office is currently seeking help from artificial intelligence (including machine learning) to drive efficiencies in the patent examination process.[2] The director of the USPTO, Andrei Iancu, said in the news that "our need is high and technology has advanced, so this is a good time to take advantage of these new tools to help our examiners."

The key idea here is that the patent screening process can be viewed as a prediction process. To fulfill the mandate of the Patent and Copyright Clause of the Constitution, the U.S. Patent Act (35 U.S. Code §101 – §103) requires a granted patent to be "new," "useful," and "non-obvious" with its purpose of making new discoveries public knowledge in the future by rewarding inventors with a limited exclusive right on their invention.[3] Therefore, the grant of a patent based on the U.S. patent law hinges on the prediction of its future social value to the society by a patent examiner. I argue, more specifically, that the objective of the patent office is to grant higher quality patents (i.e., patents with higher social values) while rejecting lower quality patent applications. This argument is also supported by the USPTO itself in its 2018–2022 strategic plan, that is, the most important goal for the USPTO is to continue optimizing patent quality.[4] However, as discussed earlier

[2]For the full news story, please see: https://www.wsj.com/articles/patent-office-seeks-help-from-ai-11572297295.

[3]See the "Patent and Copyright Clause" of the U.S. Constitution. To quote: [Congress shall have the power] "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries."

[4]For the full USPTO 2018-2022 strategic plan, please see: https://www.uspto.gov/sites/default/files/documents/USPTO_2018-2022_Strategic_Plan.pdf.

on, the objective of patent examiners may not be closely aligned with the objective of the patent office due to either resource constraints faced by patent examiners or typical principle-agent problems: i.e., time constraints, talent constraints, career concerns, and compensation incentives. Therefore, it is possible that a machine learning algorithm can better execute the task and mitigate the misalignment problem.

Using detailed data on both granted and rejected applications that are recently available from the USPTO website, I train a supervised machine learning algorithm that maps patent application characteristics to patent quality using earlier patent applications (i.e., using standard quality measures that capture patent's social value innovativeness).[5] I then use this trained algorithm to predict the quality of more recent patent applications out-of-sample. My out-of-sample prediction results show that the current patent examination system grants many low-quality patents while rejecting many high-quality patent applications. I also show that the above machine learning algorithm performs significantly better than an OLS regression funcion out-of-sample, in terms of predicting standard quality measures of the patents such as "citation" and "generality" measures that capture the social value of patents which have been used extensively in the literature (see, e.g., Trajtenberg et al. (1997), Hall et al. (2005), and Chemmanur et al. (2014)).[6]

Next, I want to test whether the above machine learning algorithm can do a better job with the same grant rate as patent examiners. Ideally, one wants to compare the average quality of patent applications granted by an algorithm to the average quality of the actually granted patents. However, the main challenge here is the missing counterfactuals, given that only quality information of accepted applications is observable.[7] To address this selection issue, I make use of the quasi-random assignment of patent applications to examiners who

[5]The machine learning algorithm used in this paper falls into the category of supervised learning, namely, training a prediction function that maps inputs (X) to an output (y) based on training input-output pairs. The inputs (X) used in my setting include numerical statistics of claims text, the text-based numerical vector of claims that capture the text similarity across contemporaneous patent applications, backward citations from prior patents, patent applications, foreign patents, and scientific literature, the total number of novel words, filing year dummies, inventor nationality dummy, small entity dummy, NBER classes dummies, and art unit dummies. The output (y) used in my setting includes the generality index of patents, and forward citation counts of patents.

[6]Machine learning generally makes much more accurate (out-of-sample) predictions by imposing fewer restrictions on the prediction function form compared to traditional statistical tools.

[7]This challenge is not unique in this setting. It shows up in most machine learning applications trying to improve screening efficiency (e.g., recruiting decision, admission decision, and bail decision). See Kleinberg et al. (2017) for a detailed discussion.

have different levels of leniency (or, in other words, differnt grant rates). Because of the quasi-random assignment, I argue that the average quality of patent applications, which are reviewed by examiners with different levels of leniency, is similar.[8] Following the methodology first introduced by Kleinberg et al. (2017), I divide examiners into two halves based on their leniency (grant rate): i.e., more lenient examiners approve 78% of applications in my sample while less lenient examiners accepts only 50% of patent applications. We can think of two independent screening systems with different screening thresholds are at works. I focus on patents already granted by more lenient examiners: (1) I rank them based on the above machine learning algorithm; (2) I reject additional 28% of patents from the lowest predicted quality to match the grant rate of less lenient examiners. This identification strategy enables me to compare the observable quality of patents granted by more lenient examiners with the help of my machine learning algorithm to those granted by less lenient examiners. We would expect that less lenient examiners are the better examiners given that they set a higher bar to approve patent applications. However, with the help of a machine learning algorithm, more lenient examiners are able to do a significantly better job than less lenient examiners in terms of granting higher-quality patents.

To explain the above identification clearly, Figure 2 provides an illustrative example of the above exercise. More lenient examiners approve 700 applications, while less lenient examiners approve 500 applications. I rank the 700 applications granted by more lenient examiners based on their predicted quality and reject additional 200 patents starting from the lowest predicted quality. After that, I can quantify the quality gain made by an algorithm by comparing observable quality measures (such as the patent generality index or the number of patent citations) of the 500 patents granted by an algorithm and the 500 patents granted by less lenient patent examiners. I find that the improvement in quality is economically significant in the real data: an algorithm trained against patent generality results in about a 15.5% gain in patent generality and a 35.6% gain in the number of patent citations compared to decisions made by less lenient examiners. In other words, the more lenient examiners (or worse-performed examiners) can out perform less lenient

[8]Many recent studies exploit this feature to make causal inferences in their researches (see, e.g., Maestas et al. (2013), Farre-Mensa et al. (2020), and Sampat and Williams (2019)).

examiners (or better-performed examiners) with the help of a machine learning algorithm. These results also demonstrate that a machine learning algorithm not only results in significant improvements in an objective that is targeted by the algorithm (i.e., generality), but also results in significant improvements in an alternative measure of patent quality (i.e., the number of citations) that is not targeted.[9] In addition, the above analysis also provides some suggestive evidence on why patent examiners fall short in making screening decisions. For example, the algorithm suggests that important factors (in terms predicting patent generality) include a numerical vector capturing the text similarity across different patent applications, and the measure of patent application originality capturing whether a given application leverages knowledge from many different fields. The above factors require patent examiners to spend a significant amount of time or require good examiners to do the job properly. Lastly, I also find that busier examiners make more mistakes in terms of both the false rejection side and the false acceptance side. More experienced examiners and lenient examiners indeed make less false rejection mistakes, but they also make more false acceptance mistakes.

So far, I have documented the potential quality gain by using a machine learning algorithm. Does such an algorithm also improve the economic outcomes of firms? To examine the economic consequence of the current patent screening process, the second part of the paper examines its impact on firms who get the same number of patents from examiners with different screening efficiencies. To do so, I label patents that would be rejected based on the above algorithm, to be "falsely accepted," and construct an *ex-ante* screening efficiency measure for each examiner by computing his/her past false acceptance rate. I find that firms who get patents from examiners with higher false acceptance rates end up suffering themselves as well. In particular, I find that patents granted by such examiners have on average lower announcement returns around their grant news. These patents are also more likely to expire early. Further, I find that public firms whose patents granted

[9] Although this study finds that machine learning algorithms can make better screening decisions in terms of granting higher quality patents, replacing human examiners with machine learning algorithms may incur unintended consequences. Instead, such am algorithm can serve as an auditing process, in which examiners are responsible to reexamine those patent applications identified as questionable screenings. Combining the expertise of human examiners and the strength of machine learning mitigates such concerns while achieving better screening outcomes.

by such examiners are more likely to get sued in patent litigation in both the short-term and long-term future. Consequently, they cut R&D expenditures and have worse operating performance (measured by either ROA or Cash Flow). Additionally, such a negative impact is larger for firms in the high-tech or health industries. In the cases of private firms, they are less likely to exit successfully by an IPO or an M&A in the short-term and long-term future. The above effects are also economically significant. For example, the annual ROA for public firms would increase by 1.3 percentage points, and the probability of private firms going pubic or getting acquired in three years would increase by 3.6 percentage points, if the above machine learning algorithm screened all patent applications. These above results can be also viewed as causal evidence since patent applications are randomly assigned to patent examiners whose characteristics are unlikely to be correlated with firm characteristics.

The rest of the paper is organized as follows. Section 2 discusses the relation of my paper to the existing literature. Section 3 discusses the institutional background of the patent examination process. Section 4 describes the patent application data and sample statistics. Section 5 discusses the empirical design and results of my machine learning analysis. Section 6 describes the firm-level data and discusses the empirical analysis of firm performance. Section 7 concludes.

2 Relation to the existing literature

My paper is related to four different strands in the literature. The first strand is the theoretical and legal literature that explores the question of improving the patent screening process by reforming the patent system itself. For example, Dreyfuss (2008) argues that the patent system systematically creates type II errors (i.e., erroneous grants) due to resource constraints faced by patent examiners and the incentive structure at the USPTO. Dreyfuss (2008) proposes to increase the nonobviousness threshold in order to reduce the number of type II errors: see also, e.g., Duffy (2008), Eisenberg (2008), and Mandel (2008). On the other hand, Scherer (1972) and several other theoretical papers focus on reforming the optimal patent right (i.e., patent length and breadth) to improve the innovation incentive and quality (see, e.g., Gilbert and Shapiro (1990), Matutes et al. (1996)). A set of related papers also study the cost and benefit of the patent litigation system in affecting patent validity and scope (see, e.g., Meurer (1989), Choi (1998), Lanjouw and Schankerman (2001), and Bessen and Meurer (2006)). In this paper, I depart from the above literature and analyze how machine learning techniques are able to improve the effectiveness of the patent screening process without changing the current patent system itself. Additionally, I also provide evidence of how inefficient patent screening leads to important economic effects for firms owning these patents.

The second strand is the literature that applies machine learning techniques to economics and finance research. For example, Athey and Imbens (2017) argue that supervised machine learning has great potential for prediction problems but has not been widely utilized in social science research. Several studies apply machine learning to issues in finance: e.g., measuring asset risk premia (Gu et al. (2018)), predicting stock returns (Rossi (2018)), classifying fund types (Abis (2017)), and selecting the boards of directors (Erel et al. (2018)). However, there also exist challenges to apply machine learning in social science research. Kleinberg et al. (2017) use New York judges' decisions over bail cases as a setting to discuss unique potential endogeneity problems when applying machine learning to social science and provide methodologies to address these problems using econometric identifications.[10] However, mine is the first paper to evaluate innovation screening efficiency of patent examiners by making use of the quasi-random assignment of patent applications to patent examiners to address potential selection issues.

Third, my paper also contributes to the empirical literature that measures patent quality and studies its relationship to firm performance. For example, Hall et al. (2005) empirically document that a larger number of citations per patent leads to higher market values for firms holding these patents (see also, e.g., Zucker et al. (2002)). Chemmanur et al. (2017) also show that firms with a large number of patents and citations per patent as private firms have higher IPO valuations and future operating performance. However, the innovation measures used in these studies are only *ex-post* available. Alternatively, Bowen et al. (2019) measure the technological disruptive potential of startups using textual analysis of patents

[10] See also Kleinberg et al. (2015), and Mullainathan and Spiess (2017) for detailed discussions on how to use machine learning as an applied econometrics tool.

and show that those firms with higher technological disruptive potentials are more likely to go public and less likely to be sold. Kelly et al. (2018) use textual analysis of patent text data to create indicators of technological innovation for each patent based on its textual similarity to earlier and later patents. Kogan et al. (2017) measure the economic value of a patent as the stock price announcement effect of the patent grant and study its relationship with aggregate economic growth and TFP. Kline et al. (2019) follow a similar approach to estimate the *ex-ante* value of both accepted and rejected patent applications and study the relationship between patent-induced shocks and labor productivity. Unlike these measures used in the above papers, the screening efficiency measure of examiners constructed in my paper can be viewed as an *ex-ante* measure of patent application quality that is less likely to be related to firm characteristics, given that patent applications are randomly assigned to each patent examiner within each art unit.

A set of recent papers exploits the quasi-random assignment of applications to examiners with different leniency to make causal inferences on the relationship between current innovation and follow-on innovation. For example, Farre-Mensa et al. (2020) find that obtaining its first patent causally increases a startup's subsequent growth, follow-on innovation, and VC funding. On the other hand, Sampat and Williams (2019) examine whether patents on the field of human genes affect follow-on innovation and find that gene patents on average have no quantitatively important effects on follow-on innovation. Unlike these papers, my paper focuses on the economic consequences of inefficient screening by patent examiners and studies its impact on the future performance of both public and private firms. Overall, my paper complements the above literature by documenting causal evidence of the importance of corporate innovation on subsequent performance and investment of both public and private firms.

Finally, my paper is also related to the strand of literature that analyzes the value of innovations by examining stock market reactions to innovation-related announcements. For example, Eberhart et al. (2004) examine the market valuation of firms' innovation inputs (R&D expenditures) and show that the market consistently underreacts to firms' unexpected increases in R&D expenditures. Cohen et al. (2013) also show that the stock market does not take firms' past successes in innovation into considerations when valuing their future innovation. On the other hand, Hirshleifer et al. (2013) explore the market valuation of firms' output and show that firms' innovation efficiency (measured as patents scaled by R&D expenditures) can predict firms' future stock returns. Shu et al. (2019) test whether the workload of each patent examiner can predict firms' future stock market returns and show that investors underreact to the negative effect of examiner's workload on patent quality. My paper complements the above literature by providing additional evidence that the stock market incorporates (at least partially) the quality of firms' new patents from the past performance of the patent examiners examining these patent applications prior to patent grants.

3 Patent examination process and patentability

3.1 Patent examination process

The patent examination process starts with the filing of a patent application to the USPTO, where the USPTO will forward this newly filed application to a relevant art unit for examination.[11] That patent application will be assigned to a patent examiner, who is a specialized technology employee with training and experience that are relevant to the invention, for examination. Though there are no explicit policies regarding how patent applications are assigned to examiners within each art unit, many recent studies show that patent applications are randomly assigned to examiners within each art unit: an application that has filed in the earliest date is assigned to the first available examiner (see, e.g., Maestas et al. (2013), Farre-Mensa et al. (2020), and Sampat and Williams (2019)).

After receiving a patent application, examiners first compare the claimed invention to the existing state of knowledge in the "prior art," consisting of patent documents as well as the scientific and commercial literature to determine whether the invention satisfies legal requirements for patentability. If an invention fails the patentable requirement, the examiner will issue an office action rejecting that application as not patentable and explain reasons for the rejection. Following such a rejection, the inventor may revise the application and submit

[11]There are nine patent examining group centers where each of them consists of several art units examining patents in the relevant field.

it again or withdraw the application. My paper only focuses on the earliest application of all regular non-provisional utility applications in order to mitigate the concern that these subsequent applications may not be randomly assigned (Righi and Simcoe (2018)).

3.2 The legal requirements for patentability

Patent and Copyright Clause of the Constitution (Article I, Section 8, Clause 8, of the Constitution) grants Congress the power "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." To fulfill its mandate, the U.S. Patent Act (35 U.S. Code §101) sets the requirements for patent protection as follows:

"Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvements thereof, may obtain a patent, subject to the conditions and requirements of this title."

Under the U.S. Patent Act, an invention is patentable after satisfying the following three criteria: new, useful, and non-obvious. Specifically, the novelty requirement (35 U.S. Code §102) states that an invention cannot be patented if the invention has been publicly disclosed before the applicant filed for patent protection and the usefulness requirement states that the subject matter must be useful. Usually, a patent application can easily pass both the novelty and usefulness requirements. However, the non-obvious requirement (35 U.S. Code §103), which requires the invention to be a non-obvious improvement over the prior art, is an ambiguous threshold that attracts many criticisms from the law literature for approving many low-quality patents (see, e.g., Duffy (2008), Dreyfuss (2008), Eisenberg (2008), and Mandel (2008)).

Since the goal of U.S. Patent Act is to reward patent applicants with a limited exclusive right on their invention for providing new discoveries to the public, I argue that the main objective for patent examiners is to identify and grant patents of higher quality (or higher social value) while rejecting those of lower quality.

3.3 Measuring patent quality

Recent papers start to use excess stock market returns to measure firms' private value of a patent (Kogan et al. (2017)). However, the private value of a patent is unlikely to capture the objective of patent examiners for the following reasons. First, the private value of a given patent depends not only on its own quality but also on whom the patent belongs to: i.e., a patent may have different private values to different owners, while examiners make grant decisions based on the characteristics of a patent application itself. Second, the private value of a patent can be measured only if it is filed by a public firm, while examiners need to also evaluate patent applications filed not only by public firms but also by private firms, governments, universities, and individual inventors to make grant decisions. On the other hand, citation-based measures, which have been used extensively in existing literature (see, e.g., Trajtenberg (1990), Trajtenberg et al. (1997), and Hall et al. (2005)), not only are available for any patent granted by the USPTO regardless of whom filed the patent application but also, more importantly, capture the social value (or social spillovers) of a patent (Bloom et al. (2013)).

In this paper, I use patent generality as my primary measure of patent quality: the generality index of a patent captures the industry dispersion of citing patents in the following four years after being granted.[12] Explicitly, I compute the generality index following the seexisting literature (see, e.g., Trajtenberg et al. (1997), Hall et al. (2005)): $G_i = 1 - \sum_j^{n_i} s_{ij}^2$, where s_{ij} denotes the fraction of forward citations received by patent i in patent class j from the total number of patent classes n_i and $\sum_j^{n_i} s_{ij}^2$ is the Herfindahl-Hirschman index (Hirschman (1980)). By definition, if a patent is cited by later patents that belong to more fields, the generality of this patent will be higher. For example, if a patent in the field of biology is cited by subsequent patents in social science, medical science, and engineering, we would expect this patent to have a higher degree of generality than a similar patent that received the same number of citations but all from patents in the same field. With regard to patent classes, the USPTO has developed its own U.S. Patent Classification (USPC) system

[12] I have also used citation counts (the number of citing patents in the following four years after a patent gets granted) as an alternative measure of patent quality. The results using citation counts are reported in Section A.2.2 in the Internet Appendix and are robust to the findings in the main paper.

that consists of more than 450 unique classes and 150,000 subclasses. However, USPC classes provide no straightforward link to the established product and industry classifications (Marco et al. (2015a)). Hall et al. (2001) developed a hierarchical classification (NBER classification) by aggregating USPC classes into 37 (two-digit) sub-categories.[13] Therefore, I construct two generality measures based on either the USPC classification or the NBER classification. All results presented in empirical sections are using the generality measure based on the NBER classification.[14]

4 Patent application data and sample selection

4.1 Patent application data

I collect data on patent applications from the USPTO website that provides various research datasets.[15] In particular, I collect patent application examination data from Patent Examination Research Dataset (Graham et al. (2018) and Marco et al. (2017)), patent application claims data from Patent Claims Research Dataset (Marco et al. (2019)), patent application citation data from Office Action Research Dataset for Patents (Lu et al. (2017)) and PatentsView, and patent assignment data from Patent Assignment Dataset (Marco et al. (2015b)).[16]

4.1.1 Turning patent claims text into numerical variables

The claim section in each patent application defines the extent of the protection sought in a patent application. A typical patent contains several claims, where each claim represents an original contribution and thereby being viewed as a good measure of the real invention in a patent (Tong and Frame (1994)). If claims in a patent application are very similar or closed

[13]The NBER classification comes from the NBER Patent Data Project: https://sites.google.com/site/patentdataproject.

[14]The results are quantitatively similar using the generality measure based on the USPC classification and are reported in Section A.2.1 in the Internet Appendix.

[15]For a complete list of research datasets provided by the USPTO please see: https://www.uspto.gov/ip-policy/economic-research/research-datasets.

[16]Public PAIR data have been recently available from the USPTO website. Though not all patent applications received by the USPTO are included in Public PAIR, more than 83% of all patent applications are available after the implementation of The American Inventors Protection Act (AIPA) in late 2000. For regular utility patent applications that this paper focuses on, inclusion in Public PAIR increases to 95% since 2001 as a consequence of AIPA according to Graham et al. (2018).

to claims in other patent applications, we would expect that the quality (innovativeness) of this patent application to be low. To capture the similarity of each patent application filed in a given year compared to all patent applications filed in that year, I take all claims text in each patent application to produce a vector of 50 dimensions from claims text using the *Word2vec* algorithm.[17] I use this vector of 50 numerical variables as well as numerical statistics of claims and other patent application characteristics discussed in the later section as input variables in my machine learning prediction, where I find the prediction accuracy of my machine learning algorithm is improved with this set of text-based variables.

4.2 Summary statistics

Table 1 reports summary statistics of numerical variables for all patent applications used in my machine learning prediction. Out of 637,305 applications, 434,496 (68.2%) are approved, 236,643 of which have non-zero 4-year forward citations: the average 4-year forward citations and the generality index per patent among patents with non-zero citations are 3.886, and 0.072, respectively. In terms of numerical statistics of claims, each patent application on average has 2.791 independent claims and 15.528 dependent claims, where the average length of an independent claim (around 138 words) tends to be longer than that of a dependent claim (around 42 words). The average number of novel words per patent is 0.309.[18] I also compute the originality index for each patent application, which is defined similarly as generality except that it based on backward citations each application has made. The average backward patent citations and the originality index are 8.511, and 0.166, respectively. In addition to citing prior patents, a patent application may also cite prior applications, scientific literature, and foreign patents. The average backward citations from patent application, scientific literature, and foreign patent citations are 2.755, 3.837, and 2.905, respectively.[19] Besides patent application characteristics, 26.9% of patent

[17]The *Word2vec* algorithm learns vector representations of words from the input text corpus and places words that share similar context in the corpus in close proximity to one another in the vector space, where the vector space is set to 50 dimensions (see, e.g., Mikolov et al. (2013a), Mikolov et al. (2013b), and Mikolov et al. (2013c) for details).

[18]The number of novel words for each patent is produced by Balsmeier et al. (2018), which I used as an input variable when I train my algorithm. My results remain quantitatively similar without including the number of novel words.

[19]I exclude citations made by examiners when counting backward citations for each patent application.

applications are submitted by small entities and 43.9% of primary inventors are from the U.S.

5 Machine learning prediction design and results

The empirical design to analyze the efficiency of the patent screening process follows three steps (Kleinberg et al. (2017)). First, I partition my sample into a training set and a test set, as described in Subsection 5.1. Second, I train an algorithm using the training set by mapping the characteristics a patent application to its quality and present results in Subsection 5.2. Third, I evaluate the predicting accuracy of my algorithm using patent applications in the out-of-sample test set and present results in Subsection 5.3. Last, I test whether my prediction function can improve screening decisions of actual patent examiners by comparing the decision of my algorithm to that of patent examiners and present relevant results in Subsection 5.4.

5.1 Sample partition

I use the unique application number to merge across different data sets and obtain an initial sample of 3,473,251 patent applications with screening outcomes available (i.e., either granted or rejected) filed at the USPTO from 2001 to 2014.[20] When we train a machine learning algorithm to compare its prediction with human decisions, we have to make sure the data used to train the algorithm is *ex-ante* available for actual examiners in the test set in order to make fair comparisons. In my setting, I use not only patent application characteristics but also patent outcomes of earlier applications in the training set. Since my outcome variable used to train the algorithm, the generality index of patent applications, is constructed based on 4-year forward citations and is only available four years after each application being granted, I set a 4-year gap between the training sample and the test sample. In particular, I use applications filed from 2001 to 2005, which have their screening status available before 2006 for the training sample to train my machine learning algorithm, and use applications filed from 2010 to 2013 with their status available before 2014 for the

[20] The patent application claims data from the "Patent and Patent Application Claims Research Dataset for Academia and Researchers" section is available until the end of 2014.

test sample to evaluate my algorithm.[21]

Partitioning my sample in this way, both my trained machine learning algorithm and the quality measure of patent applications in the training sample is available at the beginning of 2010. In other words, whatever information needed to train my algorithm is also available for patent examiners in the test sample. Such a sample partition allows me to make a fair comparison between my algorithm and actual examiners in terms of screening any patent application in the test set. Figure 3 presents the sample partition along the timeline. The final sample used in my machine learning prediction consists of 280,243 patent applications in the training set and 357,101 patent applications in the test set.

5.2 Training a machine learning algorithm

To train a supervised machine learning algorithm, I need both input variables of patent application characteristics and an output variable of patent application quality from applications in the training data: the output variable y is the generality index of each patent as described in Subsection 3.3; and input variables, X, include numerical statistics of claims text as described in Subsection 4.2, the text-based numerical vector of claims, backward citations from prior patents, patent applications, foreign patents, and scientific literature, the total number of novel words, filing year dummies, inventor nationality dummy, small entity dummy, NBER classes dummies, and art unit dummies. As I mentioned earlier in Subsection 5.1, my training set consists of 280,243 patent applications, which includes 81,352 rejected applications, 83,558 accepted applications with zero 4-year forward citations, and 115,333 accepted applications with the number of 4-year forward citations larger than zero. Since the number of the 4-year forward citation to construct the generality index of a patent (an accepted application) needs to be larger than zero, 115,333 accepted applications with their generality index available are used for training the machine learning algorithm.

The prediction function I train is called "Extreme Gradient Boosting," which is an

[21] I have partitioned my sample using alternative ways: partition the whole sample randomly to a training sample and test sample; partition the whole sample along the time but without a 4-year gap. Though these alternative ways of sample partition are subject to concerns raised in this section, results using these alternative ways of sample partition are similar to the main findings in this paper.

ensemble method of decision trees based on tree boosting.[22] A decision tree is a tree-like prediction function that can be trained by splitting the data set into subsets based on particular values of input variables, where the process is repeated until splitting no longer adds value to predictions (see, e.g., Rokach and Maimon (2008)). Since a single decision tree may produce a weak learning function subject to noise, gradient boosting algorithms optimize a cost function by iteratively choosing a weak learning function that follows the negative gradient direction to produce a strong learning function (see, e.g., Friedman (2001) and Chen and Guestrin (2016)). The strength of an Extreme Gradient Boosting algorithm is finding the best feature across different subsamples. In addition, I implement 5-fold cross-validation when training the algorithm in order to alleviate the in-sample over-fitting problem.

Figure 6 shows 10 important features identified by the machine learning algorithm in terms of predicting patent generality. The most important feature is the numerical vector that captures the text similarity across different patent applications filed in the same year. The 50 variables in this feature collectively explain 43.8% of the total predictive power in the trained algorithm. The second most important feature is the originality measure of patent applications, which capture the dispersed knowledge cited by each patent application. These two features together explain 70.9% of the total predictive power in the trained algorithm, suggesting that patent applications with original ideas are more likely to be high-quality patents. Other important features include number of cited scientific literature, cited patents, claims, and words in claims. Interestingly, the inventor's nationality also explains 1.5% of the total predictive power in the trained algorithm.

5.3 Evaluating the out-of-sample predicting performance of my machine learning algorithm and OLS

In this subsection, I compare the out-of-sample predicting performance between my machine learning algorithm and an OLS function. In an OLS regression, I regress patent generality on all input variables used in the machine learning prediction with patent applications in

[22] Section A.1 of the Internet Appendix provides for a detailed discussion about the supervised machine learning problem and the Extreme Gradient Boosting algorithm.

the training sample. I then use the fitted model to predict generality of patent applications in the test set. Figure 4 presents the correlation between predicted generality and actual generality using patent applications in the out-of-sample test set. The left panel of Figure 4 plots the predicted generality based on my machine learning algorithm against the actual generality of granted patents, where I find that most of the data is centered around the 45-degree line, suggesting that the accuracy of the out-of-sample prediction is high. Yet the right panel of Figure 4 plots the predicted generality based on an OLS regression against the actual generality of granted patents, where the out-of-sample fitting is much less close to the 45-degree line.[23]

Next, I test whether my algorithm is able to identify patent applications with the highest quality (i.e., so-called "tail innovation"). In particular, I compare the predicted generality distribution from my algorithm to that from an OLS function and presents the results in Table 2. The second column of Table 2 shows that only 20.5% of patent applications identified as the top 1 % highest quality in the predicted generality by my an ML algorithm are also identified as the top 1 % highest quality by an OLS function. The actual generality of patent applications identified as the top 1% generality by my machine learning algorithm, as reported in the third column of Table 2, is 0.171, which is significantly higher than that of patent applications identified as the top 1% generality by OLS reported in the fourth column of Table 2: 0.136. The difference between my machine learning algorithm and OLS is persistent and significant when we compare the results of my machine learning algorithm and OLS in terms of top 5%, 10%, and 25% of the quality distribution as reported in the second, third, and fourth rows of Table 2.

5.4 Improve screening decisions with my machine learning algorithm

5.4.1 Do examiners reject high-quality patents?

To answer this question, I examine the grant rate of actual examiners across patent applications with different predicted generality. To visualize the results, I divide patent applications

[23] Formally, the out-of-sample mean square error (MSE) of my algorithm is 0.032. I also separately regress the actual generality on predicted generality by my machine learning algorithm and the OLS function. I find that the coefficient of predicted generality by my machine learning algorithm is 0.838, while that from OLS is 0.374.

in the test set equally into 1,000 bins based on their predicted generality and compute the grant rate of patent applications made by actual examiners in each of these 1,000 bins. Figure 5 plots the correlation between the grant rate of actual examiners and the average predicted generality of patent applications in each bin. I find that the grant rate of examiners indeed increases with the predicted generality of patent applications. However, I also notice that there is a significant portion of patent applications with very high predicted quality (i.e., patent applications in the rightmost bins) being rejected by actual examiners.[24]

5.4.2 Using variation in the leniency of examiners to quantify the improvement of screening decisions by my algorithm

One way to quantify the potential quality gain achieved by my algorithm is to rank all patent applications based on my predicted generality, and then set the grant rate of my algorithm to be the same as that of examiners. I can then compare the average generality of all patent applications granted by my algorithm to the average generality of the actually granted patents. However, measuring the improvement in this way may be misleading since I do not have information on the actual generality of those patent applications rejected by examiners but approved by my algorithm. To address this issue, I make use of the fact that patent applications are randomly assigned to examiners who have different grant rates: more lenient examiners (i.e., with an above-median grant rate) accept around 77.6% of patent applications and less lenient examiners accept 49.5% of patent applications. Given all patents granted by more lenient examiners, I can reject additional applications based on predicted generality to match the grant rate of less lenient examiners (i.e., examiners with a below-median grant rate). Now, I can compare the average actual generality of applications granted by my algorithm to that of applications granted by less lenient examiners.

More importantly, comparing across examiners with different leniency allows me to track the quality (generality) of marginal applications that get rejected. Figure 7 shows the results of such comparisons. I sort patent applications by predicted generality and divided them equally into 20 bins. The black bar at the bottom of a given bin shows the fraction of

[24] Figures IA.1 and A.3 in the Internet Appendix show similar results using the generality measure based on the USPC classification and the number of citations.

patent applications being rejected by more lenient examiners. The red bar on the top of the black bar in a given bin shows the fraction of additional applications being rejected by less lenient examiners, while the blue bar on the top of the black bar in a given bin shows the share of additional applications would be by my algorithm. The top panel of Figure 7 shows that less lenient examiners would reject additional applications from patent applications in both the low- and high-quality bins. However, the bottom panel of Figure 7 shows that my machine learning algorithm would reject additional applications starting from the lowest quality of predicted generality, suggesting that examiners do not screen out the low-quality applications identified by my algorithm.

Next, I quantify the quality gain for the above exercise by comparing the actual outcome resulting from examiners to that from my algorithm. I find that the magnitude of improvement in generality (by comparing the actual generality of patents granted by my algorithm to that granted by less lenient examiners) is 15.5%. Moreover, I find that training my algorithm using generality also leads to a significant improvement in citations of granted patents. In particular, I find that the magnitude of improvement in citations (by comparing the actual 4-year forward citations of patents granted by my algorithm to that granted by less lenient examiners) is 35.6%.[25]

5.4.3 Why do examiners underperform?

To answer this question, I link examiners' characteristics with their screening performance. I measure examiners' screening performance based on the disagreement between my machine learning predictions and actual screening decisions of patent examiners, I first compute the number of applications granted by actual examiners within each art unit in any given year. Then I rank all patent applications filed within each art unit in a given year based on their predicted generality by my algorithm and hypothetically grant the same number of patent applications as examiners within each art unit in a given year. So far, each patent application has an actual grant decision made by examiners and a hypothetical grant decision made by my machine learning algorithm. Finally, I label a patent to be "falsely accepted" if it is

[25] I have trained a similar algorithm using the number of 4-year forward citations to proxy patent quality, where I find that the magnitude of improvement in the number of citations reaches 28.7%. All results based on the number of citations are reported in Figures A.3 and A.4 in the internet appendix.

accepted by an actual examiner but rejected by my algorithm and label a patent application to be " falsely rejected" if it is rejected by an actual examiner but accepted by my algorithm. I then construct the following four measures of examiners' screening performance in a given year: the number of falsely rejected cases, the false rejection rate, the number of falsely accepted cases, and the false acceptance rate. I also construct the following three measures of examiner's characteristics: the number of years worked in the patent office, the number of patent applications reviewed in a given year, and examiner leniency. In particular, I test the relationship between examiners' characteristics with their screening performance with the following regression.

$$\begin{aligned} y_{i,t} &= \alpha + \beta_1 \#YearsEmployed_{i,t} + \beta_2 Ln(\#Applications)_{i,t} \\ &\quad + \beta_3 ExaminerLeniency_{i,t-1} + Art\ Unit_a + Status\ Year_t + \epsilon_{i,t} \end{aligned} \tag{1}$$

where i indexes patent examiner; a indexes art unit; and t indexes the issue year of a patent. y includes # False Rejection, False Rejection Rate, # False Acceptance, and False Acceptance Rate. $\#YearsEmployed_{i,t}$ measures the number of years worked in the patent office. $\#Applications_{i,t}$ measures the number of patent applications reviewed in a given year. $ExaminerLeniency_{i,t}$: measures the grant rate in the past year. $Art\ Unit_a$ and $Status\ Year_t$ indicate the art unit fixed effect and the status year fixed effect.

Table 3 presents the results of the above regression. The negative coefficients of *#YearsEmployed* and *ExaminerLeniency* in columns (1) and (2) of Table 3 suggest that more experienced patent examiners and more lenient patent examiners tend to make less false rejection mistakes. The positive coefficient of *#YearsEmployed*, *Ln(#Applications)*, and *ExaminerLeniency* in columns (3) and (4) of Table 3 suggest that more experienced patent examiners, busier examiners, and more lenient patent examiners all tend to make more false acceptance mistakes. In addition, the positive coefficient of *Ln(#Applications)* in column (1) of Table 3 suggests that busier examiners also tend to make more false rejection mistakes.

5.4.4 Robustness test #1: learn and predict examiner behavior directly

I also build another algorithm to learn from examiners' behaviors (or their revealed preference/objective function) directly as a robustness test. In particular, I train another algorithm to learn examiners' behavior based on examiners' grant decisions in earlier patents and then predict their grant decisions in more recent patents. This exercise allows me to quantify the quality loss from inconsistent decisions made by patent examiners over time.[26]

Figure 8 shows the results of comparisons similar in Figure 7. The black bar at the bottom of a given bin shows the fraction of patent applications being rejected by more lenient examiners. The pink bar on the top of the black bar in a given bin shows the fraction of additional applications being rejected by less lenient examiners, while the dark green bar on the top of the black bar in a given bin shows the share of additional applications would be by predicted examiners. The top panel of Figure 8 shows that less lenient examiners would reject additional applications from patent applications in both the low and high predicted grant probability bins, suggesting that examiners make inconsistent decisions over time. More importantly, such inconsistent decisions made by actual examiners result in worse outcomes in terms of patent quality. The quality gain made by predicted examiners (who make consistent decisions over time) is also significant: the magnitude of improvement in generality and citations is 3.7% and 15.3%, respectively. This finding also suggests that patent examiners face a harder job of reviewing patent applications over time.

5.4.5 Robustness test #2: disagreement between humans and machine algorithms, and early patent expiration

In this subsection, I test whether these actually granted patents in the out-of-sample test set, which would be rejected by my algorithm, should or should not be granted in the first place as another robustness test. To measure the disagreement between my machine learning predictions and actual screening decisions of patent examiners, I first compute the number of applications granted by actual examiners within each art unit in any given year. Then I rank all patent applications filed within each art unit in a given year based on their

[26]One thing needs to be point out: this test cannot quantify the quality loss from consistent bias made by actual examiners.

predicted generality by my algorithm and hypothetically grant the same number of patent applications as examiners within each art unit in a given year. So far, each patent application has an actual grant decision made by examiners and a hypothetical grant decision made by my machine learning algorithm. Finally, I label a patent to be "falsely accepted" if it is accepted by an actual examiner but rejected by my algorithm.

Section 154 of the U.S. Patent Law (35 U.S. Code §154 (a)) sets forth the term of a utility patent filed on or after June 8, 1995, in the U.S. to be 20 years from the earliest filing date of the application on which the patent was granted. Section 41 of the U.S. Patent Law (35 U.S. Code §41 (b) & (c)) states that maintenance fees are required to be paid in every certain period in order to maintain utility patents in force.[27] If these "falsely accepted" patents should not be granted in the first place, we would expect that these patents are more likely to get expired early as a result of delaying and defaulting in payment of maintenance fees. In particular, I test whether these "falsely accepted" patents are properly maintained with the following regression.

$$y_i = \alpha + \beta FalseAccept_i + ArtUnit_a + IssueYear_t + Small\&MicroEntity_s + USPC_j + \epsilon_i, \quad (2)$$

where i indexes patent; a indexes art unit; t indexes the issue year of a patent; s indexes the size of a patentee; and j indexes the USPC class. y represents the patent-maintenance related dummies indicating the following four aspects: payment of maintenance fee in the 4th year, payment of maintenance fee in the 8th year, maintenance fee reminder mailed, patents expired for failure to pay maintenance fees. $FalseAccept_i$ is a dummy variable, equaling to one if a patent is accepted by actual examiners but would be rejected by my algorithm. $ArtUnit_a$, $IssueYear_t$, $SmallEntity_s$, and $USPC_j$ represent art unit fixed effects, issue year fixed effects, small entity dummies, and USPC class fixed effects.[28]

Table 4 presents the results of regressing Equation (2). The negative coefficients of *FalseAccept* in columns (1) and (2) of Table 4 suggest that "falsely accepted" patents are less likely to be maintained four years and eight years after being granted. The positive

[27] A patentee needs to pay maintenance fees before the 4th year, the 8th year, and the 12 years to keep its patent in force.

[28] A patentee only needs to pay 1/2 or 1/4 of maintenance fees paid by a large entity if it is a small entity or a micro entity.

coefficient of *FalseAccept* in column (3) of Table 4 suggests that patentees who own these "falsely accepted" patents are more likely to receive maintenance fee reminders. Further, the positive coefficient of *FalseAccept* in column (4) of Table 4 indicates that these "falsely accepted" patents are more likely to expire for the failure of paying maintenance fees. These results collectively show that these "falsely accepted" patents turn out to be not very useful to their holders.

6 Innovation screening and firm performance

In this section, I extend my empirical analysis to further study the economic consequences of inefficient innovation screenings on firm performance in an econometric context that we are familiar with. First, I describe firm data as well as an *ex-ante* measure of innovation screening efficiency of patent examiners in Subsection 6.1. Second, I present empirical findings on the relationship between innovation screening efficiency and stock market returns of public firms in Subsection 6.2. Third, I discuss empirical results on the effect of innovation screening on the subsequent operating performance of public firms in Subsection 6.3. Lastly, I also examine the effect of innovation screening on subsequent exits of private firms in Subsection 6.4.

6.1 Firm data and sample selection

I use all patent applications that have filed since 2010 with their screening results available by 2018 in my analysis. In addition to the data on patent applications and patent examiners that I have used in the previous section, I have also collected data on patent assignees from the USPTO website, accounting and financial data for public firms from Compustat and CRSP, firm characteristics and VC financing for private firms from VentureXpert. I match each dataset with firm names standardized by the NBER Patent Data Name Standardization Routine.[29] By construction, both public and private firms analyzed in this section should have at least one patent application filed since 2010.

[29] The name standardization routine comes from the NBER Patent Data Project: https://sites.google.com/site/patentdataproject.

6.1.1 Measure innovation screening efficiency

I construct an *ex-ante* measure of innovation screening efficiency based on the disagreement between my machine learning predictions and actual screening decisions of patent examiners. Based on the "falsely accepted" label I have assigned to each patent as described in Section 5.4.5, I compute the false acceptance rate of each examiner using all patent applications he/she has examined prior to any newly filed patent application. Specifically, I calculate the false acceptance rate of examiner e in art unit a who reviews patent application p at date t as follows:

$$ExaminerFalseAcceptRate_{p,e,t,a} = \frac{\#FalseAccept_{e,t,a}}{\#Reviewed_{e,t,a}}, \tag{3}$$

where $\#Reviewed_{e,t,a}$ and $\#FalseAccept_{e,t,a}$ are the numbers of patents reviewed and falsely accepted by examiner j prior to date t, respectively.[30] A simple plot in Figure 9 shows that the false acceptance rate of patent examiners has been increasing since 2010, which is consistent with my findings in the previous section that patent examiners are less able to screen in high-quality patents over time.

To match the time horizon of financial and accounting data on firm performance, I further measure the patent screening of examiners associated with each firm in each quarter by averaging false acceptance rates of examiners who have examined that firm's patent applications in the past three years (i.e., a three-year rolling window).[31] For example, the false acceptance rate of firm i in quarter q is calculated as follows:

$$AvgExaminerFalseAcceptRate_{i,q} = \frac{1}{N}\sum_{a=1}^{N}\left(\sum_{t=q-13}^{q-1} ExaminerFalseAcceptRate_{p,e,t,a}\right), \tag{4}$$

where $ExaminerFalseAcceptRate_{p,e,t,a}$ is the false acceptance rate of examiner e who reviews firm i's patent application p in the past three years, and N is the total number of patent applications filed by firm i with screening results available in the past three years.

[30] I exclude the patent application p in both the numerator and the denominator. I also exclude firms whose patent application is assigned to a patent examiner who has reviewed less than 10 patent applications prior to the patent application p. All results in this section are robust to removing the above exclusions.

[31] I have used different time windows to measure firm-level innovation screening efficiency (i.e., a 1-quarter, 1-year, and 2-year window), and all my empirical results in this section remain qualitatively similar.

By construction, the false acceptance rate of an individual examiner is *ex-ante* available for any newly filed patent application in my sample. More importantly, it is also unlikely to be correlated with firm characteristics due to the quasi-random assignment of patent applications to patent examiners within each art unit (see, e.g., Maestas et al. (2013), Farre-Mensa et al. (2020), and Sampat and Williams (2019)).

6.1.2 Summary statistics

Table 5 reports summary statistics for my measures of innovation screening efficiency as well as firm characteristics. Panel A of Table 5 presents summary statistics of stock returns for the sample of public firms at the firm-event level. The average and median false acceptance rate of an examiner who grants a firm's patents are 16.6% and 16.0%. I estimate abnormal returns using the market model with CRSP value-weighted index return as the market return, where market model variables (alphas and betas) are estimated over 150 days ending 50 days before the screening decision date of each patent application.[32] The average cumulative abnormal returns over a 3-trading-day window around patent grant news are 3.2 basis points (bps), while the average 1-quarter, 2-quarter, 3-quarter, and 4-quarter buy-and-hold abnormal returns are 0%, -0.6%, -1.8%, and -3.8%, respectively.[33]

Panel B of Table 5 presents the summary statistics of firm performance and firm characteristics for public firms at the firm-quarter level. The average false acceptance rate for public firms is 16.7%; the median number of patent applications being reviewed and granted for public firms in a given three-year window are 15 and 12; the median quarterly *ROA* and *Cash Flow*, which are defined as net income and cash flow divided by total assets, are 0.6% and 1.6%. Public firms, on average, have the logarithm of book assets of 7.2, a leverage ratio of 0.2, the logarithm of the market to book ratio of 1.1, and R&D expenditures of 3.5%.[34] Most of the public firms in my sample are not involved in any patent litigation as

[32] I also estimate abnormal returns using alternative models such as Fama-French three-factor model, and Carhart four-factor model(see, e.g., Fama and French (1993), and Carhart (1997)). My results remain qualitatively similar using these alternative estimation models.

[33] The negative long-run stock return after patent being granted is somewhat surprising. However, my results are consistent with that in Cao et al. (2013) and they show that firms with patent filings (regardless of application outcomes) have negative profitability on average in the five years after IPOs.

[34] All accounting variables (i.e., *ROA*, *Cash Flow*, *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%. All regression results are qualitatively similar before winsorizing and are robust to different winsorizing thresholds.

defendants after their patents granted: the average quarterly number of patent litigation for public firms is 0.1.

Panel C of Table 5 presents the summary statistics of firm performance and firm characteristics for private firms at the firm-quarter level. The average false acceptance rate for private firms is 16.6%; the median number of patent applications being reviewed and granted for private firms in a given three-year window are 4 and 3, which is much lower compared to those for public firms. In terms of firm characteristics, the average age of private firms is 10.6; private firms on average have the logarithm of quarterly VC financing amount of 0.2, and the quarter number of VC funds of 0.3. The average rate of successful exits through IPOs or M&As is 21.2%.

6.2 Innovation screening and stock market returns of public firms

In this subsection, I test whether my measure of innovation screening efficiency can explain stock market reactions to patent grant news and predict post-granting long-run stock returns. I measure stock market reactions to patent grant news using the cumulative abnormal return on a firm's equity over a 3-trading-day window (from day -1 to day 1) around the patent grant date (*CAR [-1:1]*) and long-run stock returns using the 1-quarter, 2-quarter, 3-quarter, and 4-quarter buy-and-hold abnormal returns on a firm's equity after the patent grant date (*BHAR [1:63]*, *BHAR [1:125]*, *BHAR [1:188]*, and *BHAR [1:250]*).[35]

I separately regress each of the stock return measures on the average false acceptance rates of examiners who examine firms' patent applications and report regression results in Table 6. Panel A of Table 6 reports regression results testing the effect of innovation screening efficiency on stock market reactions to patent grant news. The coefficient of the constant term in column (1) shows that the announcement returns (*CAR [-1:1]*) on average are positive and significant, which is consistent with the findings in Kogan et al. (2017). However, the coefficient of *ExaminerFalseAcceptRate* is negative and statistically significant in column (2), suggesting that the false acceptance rates of examiners are able to explain some variation in stock market reactions to those patents they have granted.

[35] I restrict my sample in this subsection to those patents with applications publicly available before they are granted.

Economically, a one-standard-deviation increase in *ExaminerFalseAcceptRate* decreases the 3-day announcement return by 2 bps. In other words, if all patent applications were screened by my machine learning algorithm (i.e., *ExaminerFalseAcceptRate* decreases from 0.166 to 0), the 3-day announcement return would increase by 4 bps.

Panel B of Table 6 reports regression results testing the effect of innovation screening efficiency on post-granting long-run stock market returns. The coefficients of *ExaminerFalseAcceptRate* are negative and statistically significant in all four regressions, suggesting that my *ex-ante* measure of innovation screening efficiency negatively predicts firms' long-run stock market returns out of sample, thus can be viewed as *ex-ante* measures of patent quality. Economically, a one-standard-deviation increase in *ExaminerFalseAcceptRate* decreases the following 1-quarter, 2-quarter, 3-quarter, and 4-quarter buy-and-hold abnormal return by 9 bps, 20 pbs, 37 pbs, and 64 bps, respectively. In other words, if all patent applications were screened by my machine learning algorithm (i.e., *ExaminerFalseAcceptRate* decreases from 0.166 to 0), the 1-year buy-and-hold abnormal return would increase by 1.2%.

6.3 Innovation screening and subsequent operating performance of public firms

As I have shown, the average quality of patents would be higher if my algorithm granted them. If this is indeed the case, we would expect that firms should have worse performance if their patents were granted by examiners with higher false acceptance rates. In this subsection, I empirically test the effect of innovation screening efficiency on the subsequent operating performance of public firms with my baseline regression as follows:

$$y_{i,q+n} = \alpha + \beta AvgExaminerFalseAcceptRate_{i,q} + \gamma X_{i,q} + Industry_j + Quarter_q + \epsilon_{i,q}, \quad (5)$$

where i indexes firm; j indexes industry; q indexes quarter; and n equals 1, 4, 8, or 12. y is the operating performance of each public firm, which is measured using either *ROA* or *Cash Flow*. For example, $ROA_{i,q+4}$ measures the subsequent 4-quarter (or 1-year) operating performance of each public firm. $AvgExaminerFalseAcceptRate_{i,q}$ is my screening efficiency measure of examiners who have examined firm i's patent applications in the past three years (or twelve quarters) as described in 6.1.1. X is a vector of control variables including the number of patents reviewed and granted in the past three years, firm size in quarter t, leverage in quarter t, market to book ratio in quarter t, and R&D expenditures in quarter t as described in 6.1.2. $Industry_j$ and $Quarter_q$ represent two-digit SIC industry fixed effects and quarter fixed effects. All standard errors in my baseline regressions are double clustered at the firm and quarter level.

The baseline results using *ROA* as the dependent variable are reported in Table 7. Table 7 shows that coefficient of *AvgExaminerFalseAcceptRate* is negative and statistically significant in all regressions, suggesting that public firms whose patent applications are granted by examiners with higher past false acceptance rates perform worse in both the short- and long-term. These results are also economically significant: a one-standard-deviation increase in *AvgExaminerFalseAcceptRate* decreases the following 1-quarter, 1-year, 2-year, and 3-year *ROA* by 32 bps, 83 bps, 157 bps, and 156 bps, respectively. In other words, if all patent applications were screened by my machine learning algorithm (i.e., *AvgExaminerFalseAcceptRate* decreases from 0.167 to 0), ROA would increase by 0.8% and 3.9% over the following 1-quarter and 3-year periods.[36] More importantly, since patent applications are randomly assigned to patent examiners, the effect of inefficient patent screening on firm performance is likely to be causal due to the quasi-random assignment of patent applications to patent examiners.

6.3.1 Potential channels: innovation screening, subsequent R&D expenditures, and subsequent patent litigation

In this subsection, I test two potential channels behind the effect of innovation screening on firm performance. Specifically, I study the impact of innovation screening on subsequent R&D expenditures and the subsequent number of patent litigation using the same baseline specification as described in Equation (5).

Table 8 presents regression results with the subsequent R&D expenditures as the de-

[36]Due to space limitation, the baseline results using *Cash Flow* as the dependent variable are reported in Table A.1 in the Internet Appendix. The results on firm cash flows are both qualitatively and quantitatively similar to the results reported in Table 7.

pendent variable and shows that coefficient of *AvgExaminerFalseAcceptRate* is negative and statistically significant in all regressions, suggesting that firms lower their R&D expenditures after their patents reviewed by examiners with lower screening efficiency. These results are also economically significant. For example, a one-standard-deviation increase in *AvgExaminerFalseAcceptRate* decreases the following 1-quarter and 3-year R&D expenditures by 4 pbs and 45 pbs (i.e., a 1.1%, and 1.1% decrease compared to the median 1-quarter and 3-year R&D expenditures). All these results suggest that innovation screening has a causal and real effect on the innovation input of public firms that might hurt their short-term and long-term performance.

Table 9 presents regression results with the number of subsequent patent litigation as dependent variables and shows that coefficients of *AvgExaminerFalseAcceptRate* are positive and statistically significant in all regressions. Economically, a one-standard-deviation increase in *AvgExaminerFalseAcceptRate* increases the number of patent litigation in the next one quarter and three years by 0.012 and 0.134 (i.e., a 9.3% and 8.6% increase compared to the average number of patent litigation over the same period). These results suggest that firms whose patents granted by examiners with higher past false acceptance rates are more likely to be involved in subsequent patent litigation, which in turn might harm their short- and long-term performance.[37]

6.3.2 A cross-industry analysis: innovation screening and subsequent operating performance

In this subsection, I empirically test whether the effect of innovation screening on firm performance is larger in innovation-intensive industries with the following specification:

$$\begin{aligned} y_{i,q+n} \quad = \quad & \alpha + \beta_1 \mathit{AvgExaminerFalseAcceptRate}_{i,q} + \beta_2 \mathit{HiTechAndHealth} \\ & + \beta_3 \mathit{AvgExaminerFalseAcceptRate}_{i,q} \times \mathit{HiTechAndHealth} \\ & + \gamma X_{i,q} + \mathit{Industry}_j + \mathit{Quarter}_q + \epsilon_{i,q}, \end{aligned} \tag{6}$$

[37] I have also run the same set of regressions with firm fixed effects and reported the results in Table A.2 in the Internet Appendix. Most of the results remain statistically significant, suggesting that the effect of innovation screening on firms' outcome exists within each firm and persistent across different time horizons.

where *HiTechAndHealth* is a dummy variable, which equals to one if a firm belongs to the High-Tech or Health industry and zero otherwise. The High-Tech and Health industry definition is based on the Fama and French 5 industry groups.[38] I add an industry dummy (*HiTechAndHealth*) and its interaction with $AvgExaminerFalseAcceptRate_{i,q}$ to Equation (5) as described in Equation (6).

Table 10 presents the regression results of the cross-industry analysis. Panels A and B of Table 10 show that β_3 is negative and statistically significant in all regressions. These results suggest that firms in industries that rely more heavily on technological innovation experience significantly larger impact from inefficient innovation screenings. However, β_3 is not statistically significant in Panel C of Table 10, but is significantly negative in Panel D of Table 10, suggesting that the larger impact experienced by firms in the High-Tech and Health industry is related to subsequent litigation costs.

6.4 Innovation screening and subsequent exits of private firms

In this subsection, I study the relationship between innovation screening and subsequent exits of private firms with the following specification:

$$y_{i,q+n} = \alpha + \beta AvgExaminerFalseAcceptRate_{i,q} + \gamma Z_{i,q} + State_s + Industry_j + Quarter_q + \epsilon_{i,q}, \quad (7)$$

where y is *SuccessfulExit*, which measures the successful exit of each private firm. For example, $SuccessfulExit_{i,q+4}$ is a dummy variable that equals one if a firm successfully exits either by an IPO or M&A in the following 1-year (4-quarter) period, and zero otherwise; $SuccessfulExit_{i,q+12}$ is a dummy variable that equals one if a firm successfully exits either by an IPO or M&A in the following 3-year (12-quarter) period, and zero otherwise. $AvgExaminerFalseAcceptRate_{i,q}$ is my screening efficiency measure of examiners who have examined firm i's patent applications in the past three years (twelve quarters) as described in 6.1.1. Z is a vector of control variables including the number of patents reviewed and granted in the past three years, firm age in quarter t, total funding received ending in quarter $t-1$, VC funding received in quarter t, and the number of funds invested in quarter t

[38] For a complete list of four-digt SIC code in each industry provided by Kenneth R. French's data library please see: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_5_ind_port.html.

as described in 6.1.2. $State_s$, $Industry_j$, and $Quarter_q$ represent the state of incorporation fixed effects, two-digit SIC industry fixed effects, and quarter fixed effects. Standard errors are clustered at the state level.

The regression results reported in Table 11 show that the coefficient of *AvgExaminerFalseAcceptRate* is negative and statistically significant in all regressions, suggesting that private firms whose patent applications are granted by examiners with higher past false acceptance rates are less likely to exit successfully either by an IPO or by an M&A in both the short- and long-term. These results are also economically significant: a one-standard-deviation increase in *AvgExaminerFalseAcceptRate* decreases the following 1-quarter, 1-year, 2-year, and 3-year probabilities of exiting successfully by an IPO or an M&A by 15 bps, 72 bps, 139 bps, and 165 bps, respectively. In other words, if all patent applications were screened by my machine learning algorithm (i.e., *AvgExaminerFalseAcceptRate* decreases from 0.166 to 0), the probability of exiting successfully by an IPO or M&A increases by 3.6% over the following three-year period. More importantly, these results suggest that inefficient innovation screenings causally reduces the probability of subsequent exits by IPOs or M&As for private firms due to the quasi-random assignment of patent applications to patent examiners.[39]

7 Conclusion

In this paper, I examine whether the patent screening process can be improved under the current patent system in terms of granting better quality patents. I argue that examiners may not process relevant information efficiently to screen out low-quality applications due to their increasing time constraints and their incentive structure, while advanced computation algorithms, such as machine learning, have much larger capacities to process information efficiently and potentially reduce human biases. Using all utility patent applications filed at the USPTO from 2001, I train a machine learning algorithm using earlier patent applica-

[39]To make sure my empirical results are not primarily driven by the art-unit level of screening efficiency, I have also constructed a measure of art-unit adjusted innovation screening efficiency and rerun all the regressions in Sections 6.3 and 6.4 as a robustness test. Due to space limitation, the results using this alternative measure are reported in Tables A.3 and A.4 in the Internet Appendix and are consistent with my findings reported in Sections 6.3 and 6.4.

tions and predict the quality of more recent patent applications out of sample. I show that the current patent screening process screens in many low-quality patents and can be significantly improved in terms of granting higher quality patents. To address potential selection issues when evaluating the performance of my algorithm, I make use of the quasi-random assignment of patent applications to examiners who have different leniency. I find that the improvement in quality is substantial and significant: training an algorithm targeting the generality of patents results in a 15.5% gain of patent generality and a 35.6% gain of the number of patent citations. Further, regression analyses show that these patents, which would be rejected by my algorithm, are more likely to expire early, suggesting that these "falsely accepted" patents indeed turn out to be useless to their holders.

To examine the economic consequences of current patent screening, I study the impact of innovation screening efficiency on the future performance of firms who have at least one patent application filed at the USPTO since 2010. To do so, I construct an *ex-ante* efficiency measure of innovation screening by computing the false acceptance rate of examiners who examine firms' patent applications. I find that my measure of innovation screening efficiency is able to predict both the announcement return around patent grant news and the subsequent long-run stock return, and thereby can be viewed as an *ex-ante* measure of patent quality. Next, I find that public firms whose patent applications are accepted by examiners with higher false acceptance rates are likely to have lower operating performance (measured by ROA and Cash Flow), and lower their R&D expenditures; also more likely to be involved in more patent litigation in both the short-term and long-term future. Such a negative impact is larger for firms in the High-Tech and Health industry. Lastly, I find that private firms whose patent applications are accepted by examiners with higher false acceptance rates are less likely to exit successfully by an IPO or an M&A in the short-term and long-term future. The above results are also economically significant. For example, the 3-year ROA for public firms increases by 3.9 percentage points, and the 3-year probability of exiting successfully by an IPO or an M&A for private firms increases by 3.6 percentage points, if my machine learning algorithm screened all patent applications. More importantly, these findings can be interpreted as causal evidence for the economic consequences of current patent screening since patent applications are randomly assigned to patent examiners that are unlikely to be correlated with firm characteristics.

Overall, this study shows how new technologies such as machine learning algorithms can help improve human decisions and generates policy implications for policymakers at the USPTO. Machine learning algorithms can serve as a supporting tool in assisting human examiners to make better decisions. For example, human examiners may use a machine learning algorithm as their reference to double-check their screen decisions over patent applications in which machine learning algorithms have a different opinion. While human examiners may or may not change their decisions after reexaminations of those patent applications, such a reexamination process may potentially reduce human bias from their behavioral issues or the fact of the increasing time constraint faced by these examiners. Although this study finds that machine learning algorithms can make better screening decisions in terms of granting higher quality patents, replacing human examiners with machine learning algorithms may incur unintended consequences: i.e., inventors may strategically file patent applications to respond such replacements. Therefore, this study proposes, instead, that such a machine learning algorithm can serve as an auditing process at a relatively low cost to relax the constraints faced by the US patent office and patent examiners. Combining the expertise of human examiners and the strength of machine learning mitigates such concerns while achieving better screening outcomes.

References

Abis, Simona (2017), "Man vs. machine: Quantitative and discretionary equity management." *Unpublished working paper. Columbia Business School.*

Arrow, Kenneth Joseph (1972), "Economic welfare and the allocation of resources for invention." In *Readings in Industrial Economics*, 219–236, Springer.

Athey, Susan and Guido W Imbens (2017), "The state of applied econometrics: Causality and policy evaluation." *Journal of Economic Perspectives*, 31, 3–32.

Balsmeier, Benjamin, Mohamad Assaf, Tyler Chesebro, Gabe Fierro, Kevin Johnson, Scott Johnson, Guan-Cheng Li, Sonja Lück, Doug O'Reagan, Bill Yeh, et al. (2018), "Machine learning and natural language processing on the patent corpus: Data, tools, and new measures." *Journal of Economics & Management Strategy*, 27, 535–553.

Bessen, James and Eric Maskin (2009), "Sequential innovation, patents, and imitation." *The RAND Journal of Economics*, 40, 611–635.

Bessen, James E and Michael J Meurer (2006), "Patent litigation with endogenous disputes." *American Economic Review*, 96, 77–81.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013), "Identifying technology spillovers and product market rivalry." *Econometrica*, 81, 1347–1393.

Bowen, Donald E, Laurent Frésard, and Gerard Hoberg (2019), "Technological Disruptive Potential and the Evolution of IPOs and Sell-Outs." *Finance Down Under 2019 Building on the Best from the Cellars of Finance*, 19–22.

Cao, Jerry, Fuwei Jiang, and Jay R Ritter (2013), "Patent and Innovation-Driven Performance in Venture Capital-Backed IPOs." *Available at SSRN*, 2364668.

Carhart, Mark M (1997), "On persistence in mutual fund performance." *The Journal of Finance*, 52, 57–82.

Chemmanur, Thomas J, Manish Gupta, and Karen Simonyan (2017), "Management quality and innovation in private firms and the IPO market rewards to innovative activity."

Chemmanur, Thomas J, Elena Loutskina, and Xuan Tian (2014), "Corporate venture capital, value creation, and innovation." *The Review of Financial Studies*, 27, 2434–2473.

Chen, Tianqi and Carlos Guestrin (2016), "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794, ACM.

Choi, Jay Pil (1998), "Patent litigation as an information-transmission mechanism." *American Economic Review*, 1249–1263.

Cohen, Lauren, Karl Diether, and Christopher Malloy (2013), "Misvaluing innovation." *The Review of Financial Studies*, 26, 635–666.

Cornelli, Francesca and Mark Schankerman (1999), "Patent renewals and R&D incentives." *The RAND Journal of Economics*, 197–213.

Dreyfuss, Rochelle Cooper (2008), “Nonobviousness: a comment on three learned papers.” *Lewis & Clark Law Review*, 12, 431.

Duffy, John F (2008), “A Timing Approach to Patentability.” *Lewis & Clark Law Review*, 12, 343.

Eberhart, Allan C, William F Maxwell, and Akhtar R Siddique (2004), “An examination of long-term abnormal stock returns and operating performance following R&D increases.” *The Journal of Finance*, 59, 623–650.

Eisenberg, Rebecca S (2008), “Pharma’s nonobvious problem.” *Lewis & Clark Law Review*, 12, 375.

Erel, Isil, Léa H Stern, Chenhao Tan, and Michael S Weisbach (2018), “Selecting Directors Using Machine Learning.” Technical report, National Bureau of Economic Research.

Fama, Eugene F and Kenneth R French (1993), “Common risk factors in the returns on stocks and bonds.” *Journal of Financial Economics*, 33, 3–56.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist (2020), “What is a patent worth? Evidence from the US patent “lottery”.” *The Journal of Finance*, 75, 639–682.

Feng, Josh and Xavier Jaravel (2020), “Crafting intellectual property rights: Implications for patent assertion entities, litigation, and innovation.” *American Economic Journal: Applied Economics*, 12, 140–81.

Frakes, Michael D and Melissa F Wasserman (2015), “Does the us patent and trademark office grant too many bad patents: evidence from a quasi-experiment.” *Stan. L. Rev.*, 67, 613.

Frakes, Michael D and Melissa F Wasserman (2017), “Is the time allocated to review patent applications inducing examiners to grant invalid patents? Evidence from microlevel application data.” *Review of Economics and Statistics*, 99, 550–563.

Friedman, Jerome H (2001), “Greedy function approximation: a gradient boosting machine.” *Annals of statistics*, 1189–1232.

Gilbert, Richard and Carl Shapiro (1990), “Optimal patent length and breadth.” *The RAND Journal of Economics*, 106–112.

Graham, Stuart JH, Alan C Marco, and Richard Miller (2018), “The USPTO patent examination research dataset: A window on patent processing.” *Journal of Economics & Management Strategy*, 27, 554–578.

Gu, Shihao, Bryan T Kelly, and Dacheng Xiu (2018), “Empirical Asset Pricing Via Machine Learning.”

Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg (2005), “Market value and patent citations.” *The RAND Journal of Economics*, 16–38.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg (2001), “The NBER patent citation data file: Lessons, insights and methodological tools.” Technical report, National Bureau of Economic Research.

Heller, Michael A and Rebecca S Eisenberg (1998), "Can patents deter innovation? The anticommons in biomedical research." *Science*, 280, 698–701.

Hirschman, Albert O (1980), *National power and the structure of foreign trade*, volume 105. University of California Press.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li (2013), "Innovative efficiency and stock returns." *Journal of Financial Economics*, 107, 632–654.

Jaffe, Adam B and Josh Lerner (2011), *Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it.* Princeton University Press.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy (2018), "Measuring technological innovation over the long run." Technical report, National Bureau of Economic Research.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2017), "Human decisions and machine predictions." *The Quarterly Journal of Economics*, 133, 237–293.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer (2015), "Prediction policy problems." *American Economic Review*, 105, 491–95.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar (2019), "Who profits from patents? rent-sharing at innovative firms." *The Quarterly Journal of Economics*, 134, 1343–1404.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017), "Technological innovation, resource allocation, and growth." *The Quarterly Journal of Economics*, 132, 665–712.

Lanjouw, Jean O and Mark Schankerman (2001), "Characteristics of patent litigation: a window on competition." *RAND journal of economics*, 129–151.

Lemley, Mark A and Carl Shapiro (2005), "Probabilistic patents." *Journal of Economic Perspectives*, 19, 75–98.

Lu, Qiang, Amanda Myers, and Scott Beliveau (2017), "USPTO patent prosecution research data: Unlocking office action traits."

Maestas, Nicole, Kathleen J Mullen, and Alexander Strand (2013), "Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt." *American Economic Review*, 103, 1797–1829.

Mandel, Gregory N (2008), "Another missed opportunity: the Supreme Court's failure to define nonobviousness or combat hindsight bias in KSR v. Teleflex." *Lewis & Clark Law Review*, 12, 323.

Mansfield, Edwin (1986), "Patents and innovation: an empirical study." *Management science*, 32, 173–181.

Marco, AC, M Carley, S Jackson, and AF Myers (2015a), "The USPTO historical patent data files: two centuries of invention." *Unpublished, USPTO Economic Working Papter.*

Marco, Alan C, Amanda Myers, Stuart JH Graham, Paul D'Agostino, and Kirsten Apple (2015b), "The USPTO patent assignment dataset: Descriptions and analysis."

Marco, Alan C, Joshua D Sarnoff, and AW Charles (2019), "Patent claims and patent scope." *Research Policy*, 48, 103790.

Marco, Alan C, Andrew A Toole, Richard Miller, and Jesse Frumkin (2017), "USPTO Patent Prosecution and Examiner Performance Appraisal."

Matutes, Carmen, Pierre Regibeau, and Katharine Rockett (1996), "Optimal patent design and the diffusion of innovations." *The RAND Journal of Economics*, 60–83.

Merges, Robert P (1999), "As many as six impossible patent before breakfast: Property rights for business concepts and patent system reform." *Berkeley Technology Law Journal*, 14, 577.

Meurer, Michael J (1989), "The settlement of patent litigation." *The RAND Journal of Economics*, 77–91.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013a), "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781*.

Mikolov, Tomas, Quoc V Le, and Ilya Sutskever (2013b), "Exploiting similarities among languages for machine translation." *arXiv preprint arXiv:1309.4168*.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (2013c), "Linguistic regularities in continuous space word representations." In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751.

Mullainathan, Sendhil and Jann Spiess (2017), "Machine learning: an applied econometric approach." *Journal of Economic Perspectives*, 31, 87–106.

Nordhaus, William D (1969), "An economic theory of technological change." *American Economic Review*, 59, 18–28.

Righi, Cesare and Timothy Simcoe (2018), "Patent examiner specialization." *Research Policy*.

Rokach, Lior and Oded Z Maimon (2008), *Data mining with decision trees: theory and applications*, volume 69. World scientific.

Rossi, Alberto (2018), "Predicting Stock Market Returns with Machine Learning." *Unpublished working paper. The University of Maryland*.

Sampat, Bhaven and Heidi L Williams (2019), "How do patents affect follow-on innovation? Evidence from the human genome." *American Economic Review*, 109, 203–36.

Schankerman, Mark A and Florian Schuett (2016), "Screening for patent quality: Examination, fees, and the courts."

Scherer, Frederic M (1972), "Nordhaus' theory of optimal patent life: A geometric reinterpretation." *The American Economic Review*, 62, 422–427.

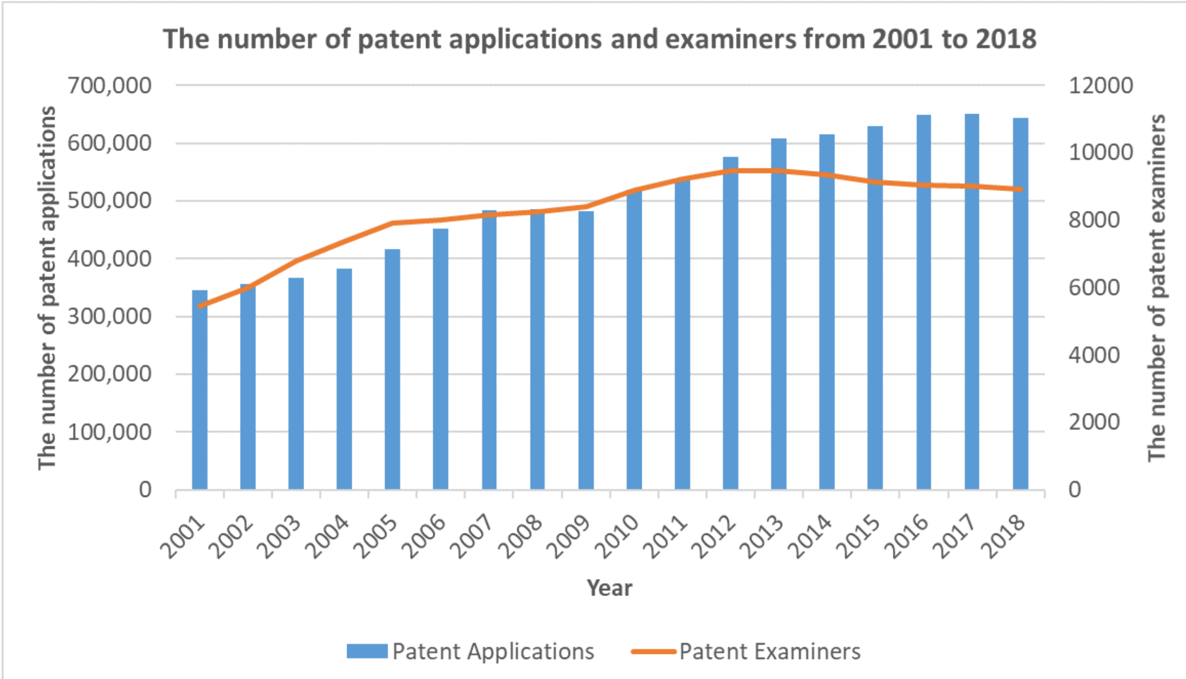
Shu, Tao, Xuan Tian, and Xintong Zhan (2019), "Patent Quality and Firm Value: Evidence from Patent Examiners Busyness." *Available at SSRN 3137894*.

Tong, Xuesong and J Davidson Frame (1994), "Measuring national technological performance with patent claims data." *Research Policy*, 23, 133–141.

Trajtenberg, Manuel (1990), "A penny for your quotes: patent citations and the value of innovations." *The RAND Journal of Economics*, 172–187.

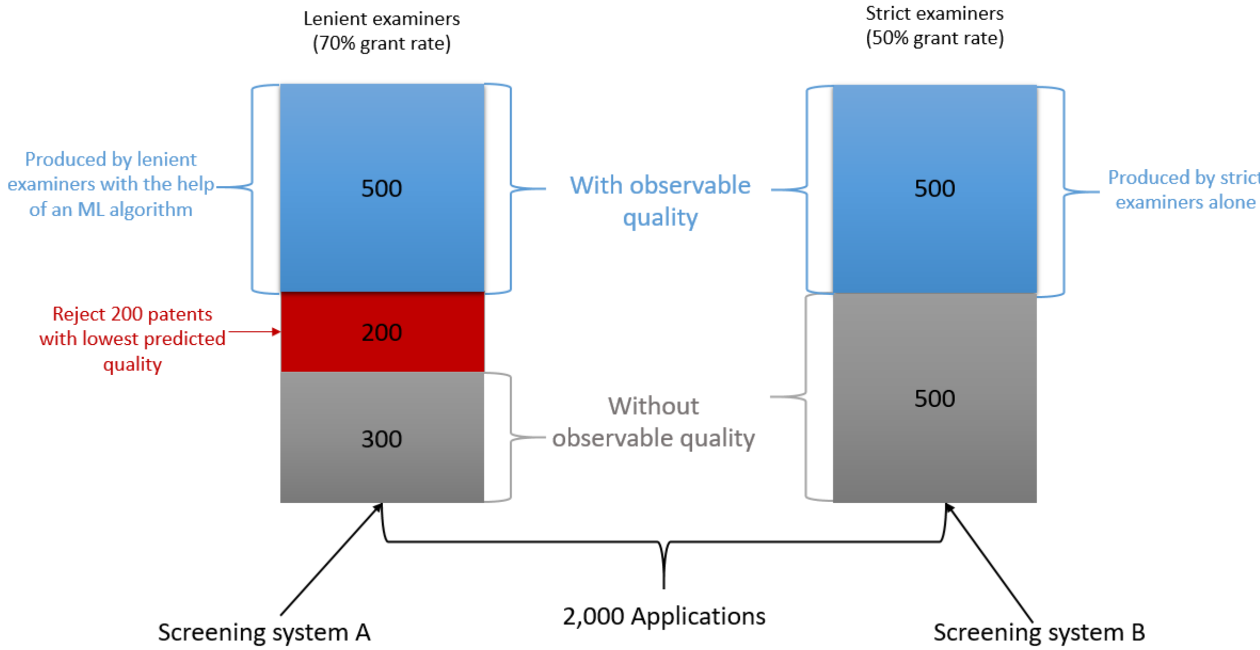
Trajtenberg, Manuel, Rebecca Henderson, and Adam Jaffe (1997), "University versus corporate patents: A window on the basicness of invention." *Economics of Innovation and new technology*, 5, 19–50.

Zucker, Lynne G, Michael R Darby, and Jeff S Armstrong (2002), "Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology." *Management science*, 48, 138–153.



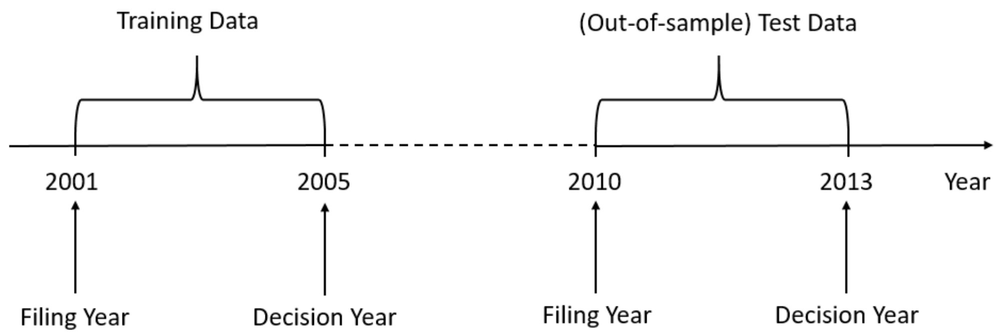
This figure shows the number of patent applications and patent examiners at the USPTO from 2001 to 2018. Each blue bin represents the number of patent applications and the yellow line represents the number of patent examiners. Data source: Patent Statistics Chart and Patent Examination Data from the USPTO website.

Figure 1: The number of patent applications and patent examiners at the USPTO from 2001 to 2018



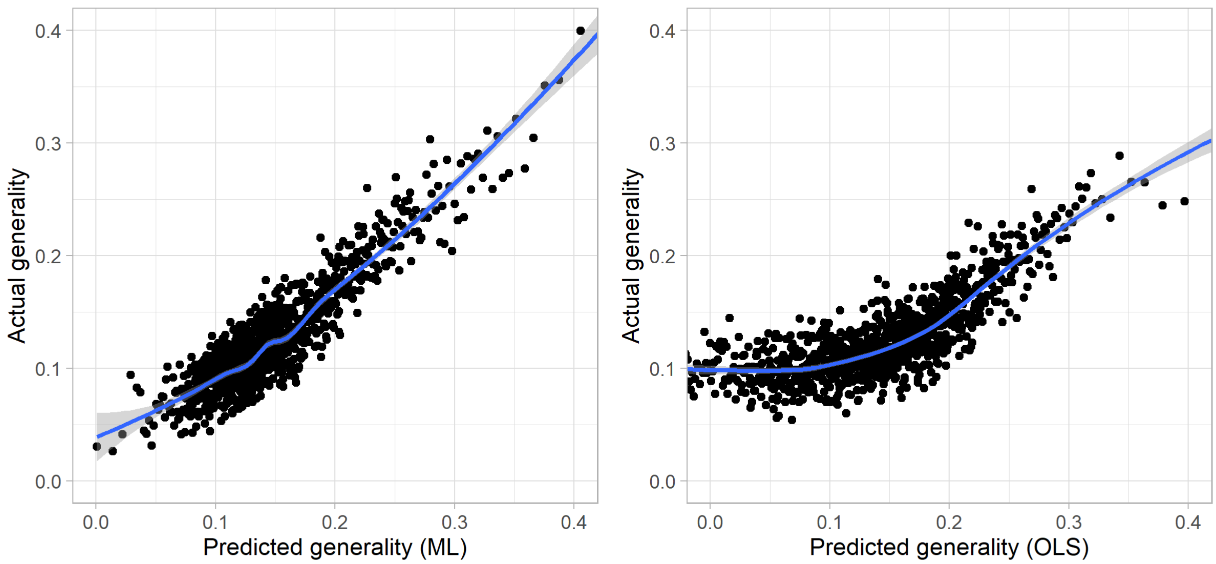
This figure provides an illustrative example of using examiner leniency to compare the performance of actual examiners and a machine learning algorithm.

Figure 2: An illustrative example of using examiner leniency to evaluate the screening performance of a machine learning algorithm



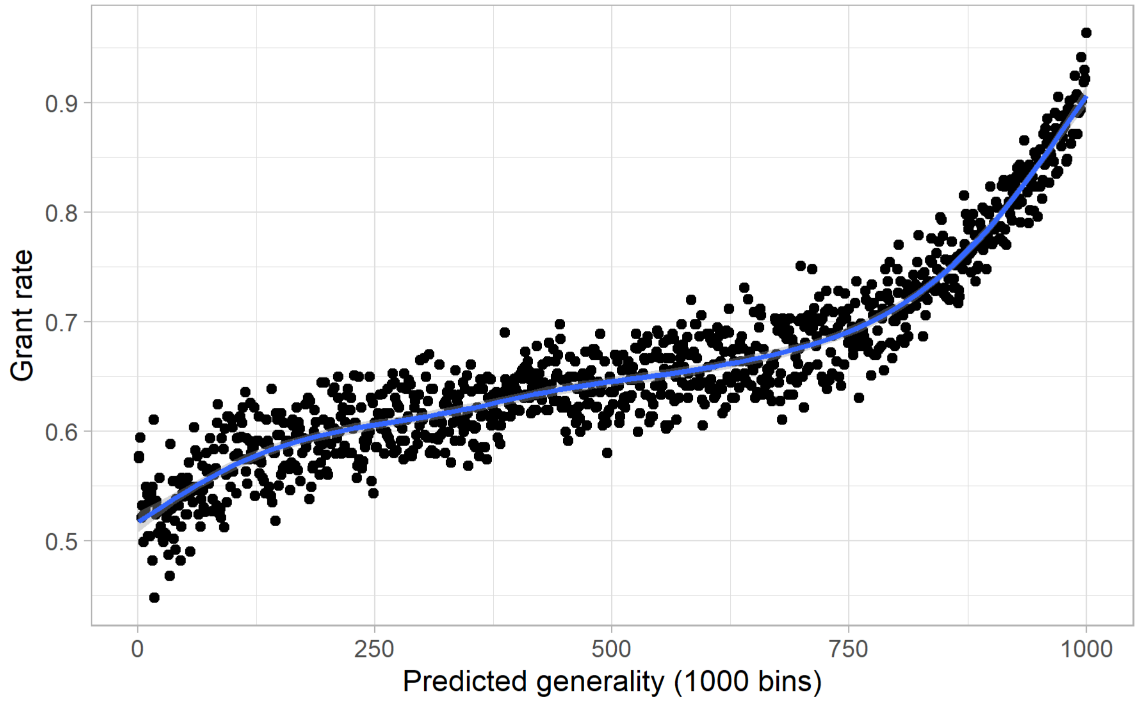
The figure shows the partition for the training and test data used for my machine learning prediction. I select applications filed from 2001 to 2005 with screening status available before the beginning of 2006 into the training set, and applications filed from 2010 to 2013 with screening status available before the beginning of 2014 into the test set. The training set is used to form the algorithm for my prediction and the test set is used to evaluate all of my results. The final sample used in my machine learning prediction consists of 280,243 patent applications in the training set and 357,101 patent applications in the test set.

Figure 3: Training and testing data used for my machine learning prediction



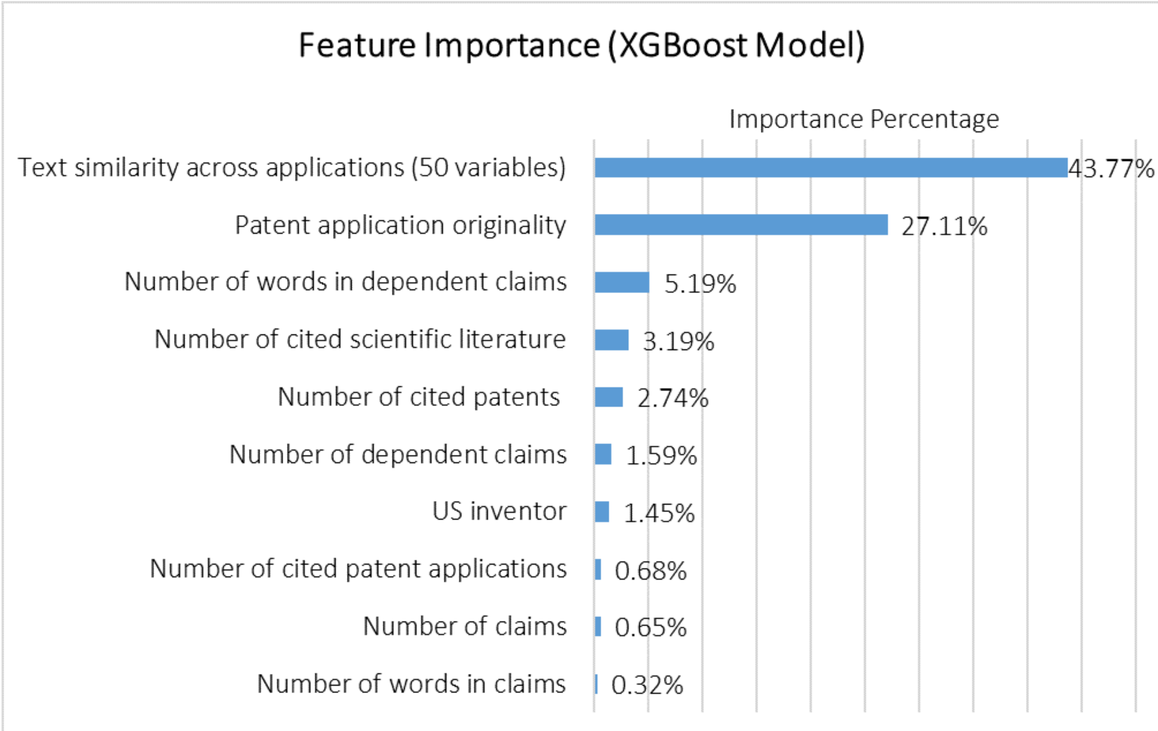
The figure shows the results of the machine learning algorithm (in the left panel) and OLS regressions (in the right panel) built using applications in the training set, applied to applications in the out-of-sample test set. The average predicted generality of patent application in each bin based on the machine learning algorithm and the OLS regression are on the x-axis of the left panel and the right panel. The actual generality is on the y-axis of both panels.

Figure 4: The relation between predicted generality and actual generality in the test set



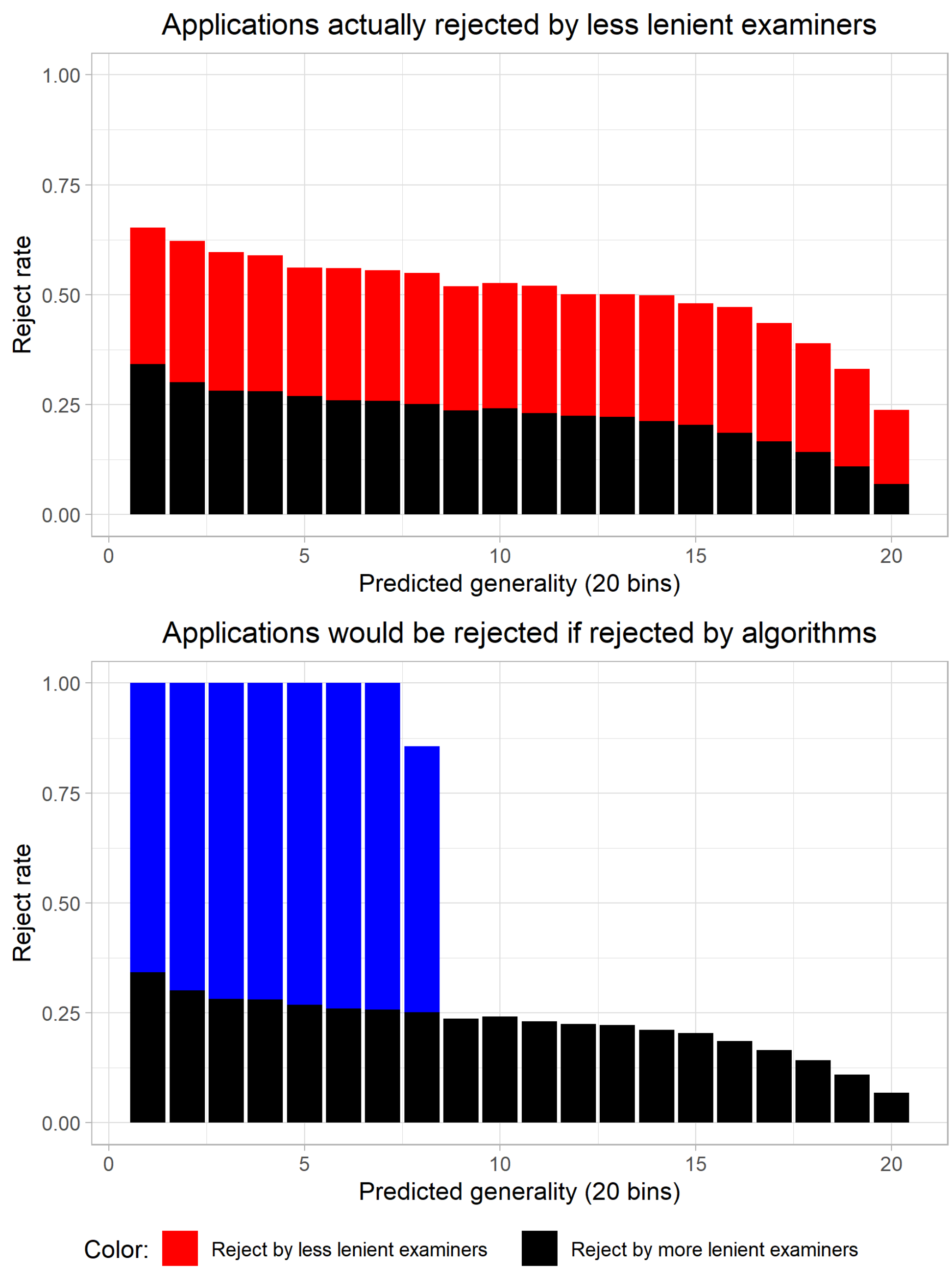
The figure shows the relation between predicted generality by the machine learning algorithm and actual examiner grant decisions. The rank of average predicted generality of all patent application in each pin is on the x-axis. The grant rate is on the y-axis.

Figure 5: The relation between predicted generality by the machine learning algorithm and actual examiner grant decisions



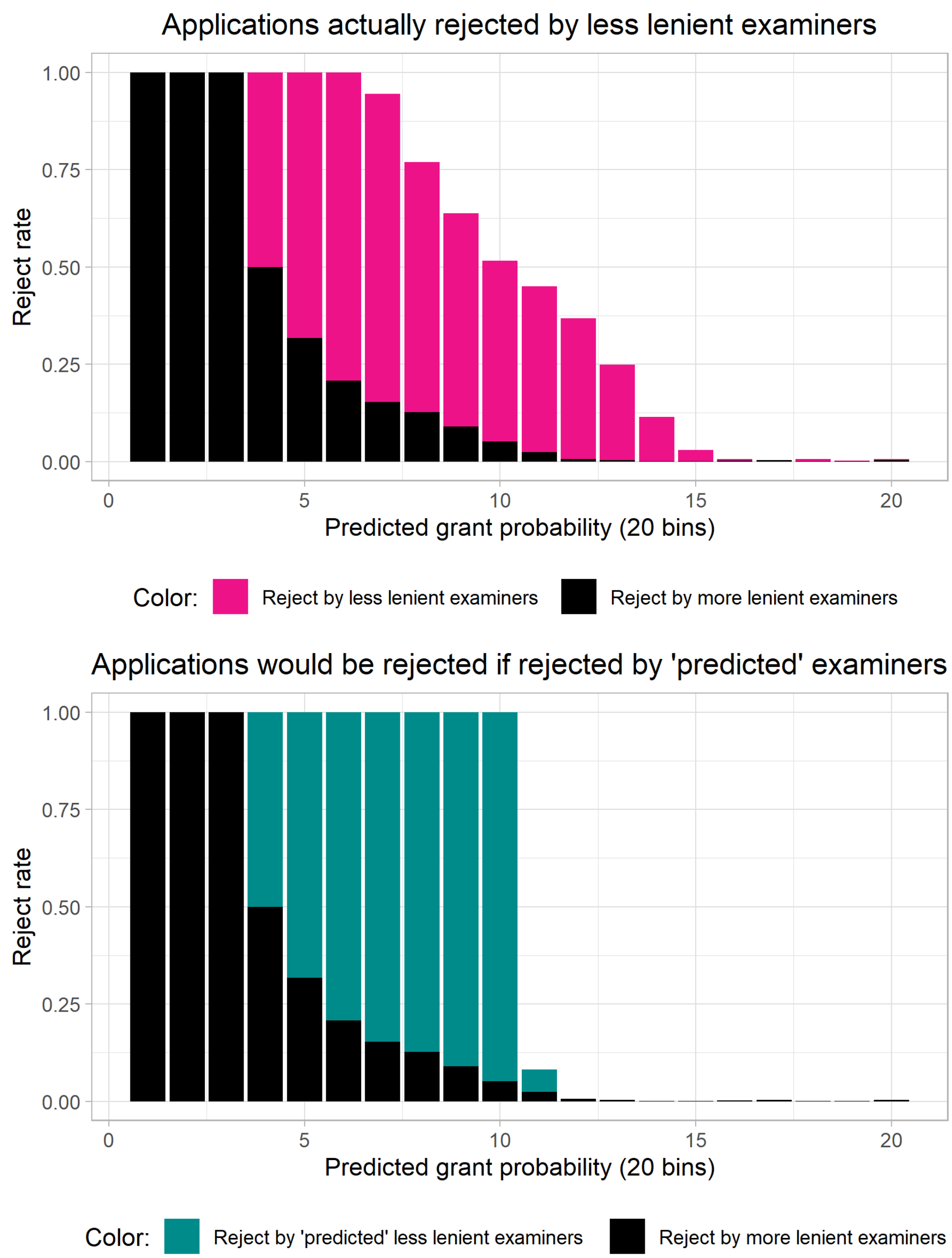
The figure shows ten important features identified by the machine learning algorithm. The predictive power by each feature measured as the percentage of total predictive power is on the x-axis. The name of each of the ten features is on the y-axis.

Figure 6: Ten important features identified by the machine learning algorithm



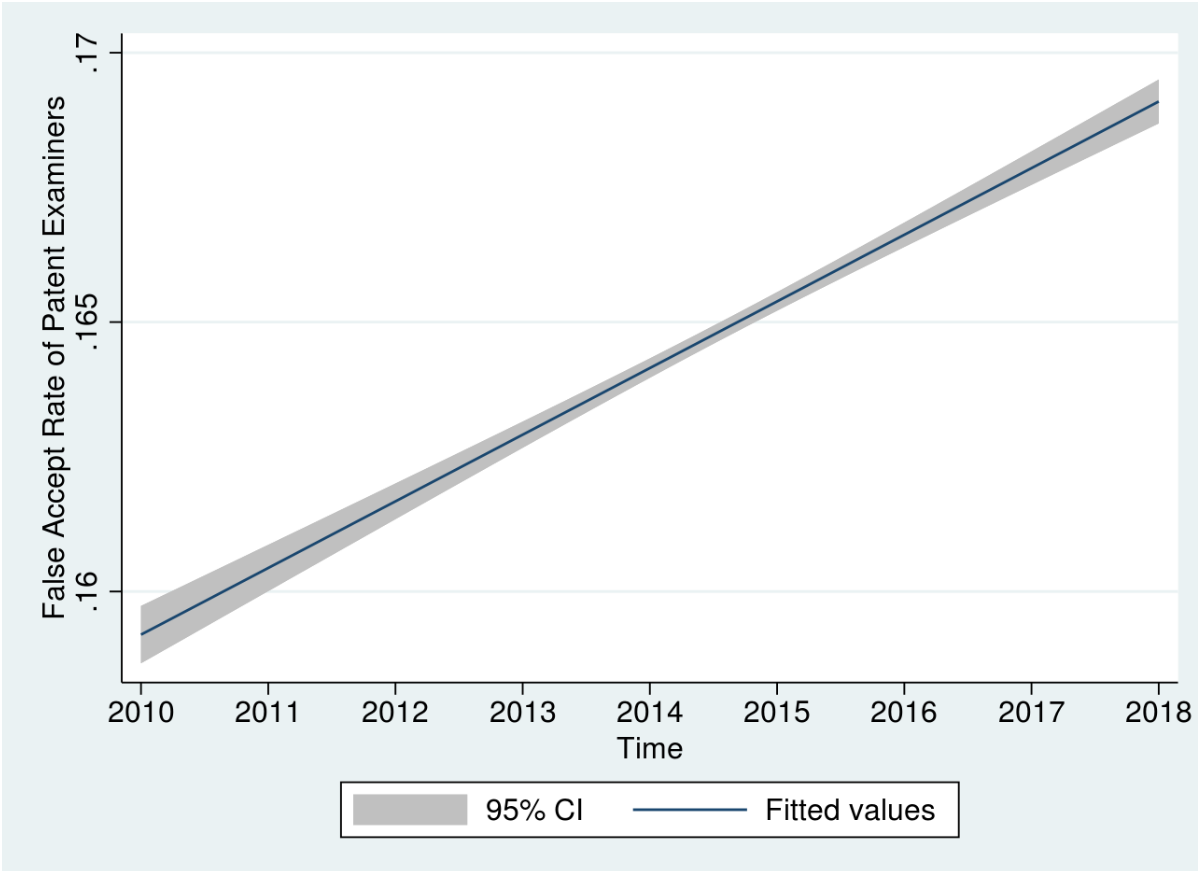
This figure shows comparison between applications rejected by stricter examiners and applications rejected by the algorithm. I divide patent applications in the test set into 20 bins by predicted generality (x-axis). In both panels, the black bar at the bottom of a given bin shows the fraction of patent applications being rejected by more lenient examiners. The red bar in the top panel shows which applications less lenient examiners actually reject. The blue bar in the below panel shows which applications my algorithm would reject to match the grant rate of less lenient examiners.

Figure 7: Comparison between applications rejected by stricter examiners and applications rejected by the algorithm



This figure shows comparison between applications rejected by actual examiners and predicted examiners. I divide patent applications in the test set into 20 bins by predicted grant probability (x-axis). In both panels, the black bar at the bottom of a given bin shows the fraction of patent applications being rejected by more lenient examiners. The pink bar in the top panel shows which applications less lenient examiners actually reject. The dark green bar in the below panel shows which applications predicted examiners would reject to match the grant rate of actual examiners.

Figure 8: Comparison between applications rejected by actual examiners and predicted examiners



The figure shows the fitted false acceptance rate of patent examiners since 2010. The time variable is on the x-axis. The solid line is the fitted value from regressing the false acceptance rate of individual patent examiners on the time variable; the gray shade represents the 95% confidence interval of the fitted value.

Figure 9: The screening efficiency of patent examiners since 2010

Table 1: Summary statistics (patent applications)

This table shows descriptive statistics for the sample of patent applications from 2001 to 2013 used in my machine learning analysis. *ForwardCitations* counts the number of future citation that each patent has received over a 4-year period after it being granted. *Generality* captures the industry dispersion of 4-year forward citing patents, which equals to one minus Herfindahl-Hirschman index of industries that citing patents belong to. *NumberIndepClaims* and *NumberDepClaims* count the number of independent claims and dependent claims for each patent application. *NumberWordsIndepClaims* and *NumberWordsDepClaims* count the total number of words in independent claims and dependent claims for each patent application. *MinNumberWordsIndepClaims* and *MinNumberWordsDepClaims* count the minimum number of words in independent claims and dependent claims for each patent application. *AvgNumberWordsIndepClaims* and *AvgNumberWordsDepClaims* count the average number of words per independent claim and per dependent claim for each patent application. *NumberCitedForeignPatents* counts the number of foreign patents that each patent application has cited. *NumberCitedForeignPatents* counts the number of novel words that each patent application has. *NumberCitedLiterature* counts the number of scientific literature that each patent application has cited. *NumberCitedApplications* counts the number of patent applications that each patent application has cited. *OriginalityApplication* captures the industry dispersion of backward cited patent applications that each patent application has made, which equals to one minus Herfindahl-Hirschman index of industries that cited patent applications belong to. *NumberCitedPatents* counts the number of patents that each patent application has cited. *OriginalityPatent* captures the industry dispersion of backward cited patents that each patent application has made, which equals to one minus Herfindahl-Hirschman index of industries that cited patents belong to. *USInventorDummy* is a dummy variable indicating whether an investor is from U.S. or not. *SmallEntityDummy* is a dummy variable indicating whether a patent application is from a small entity or not.

Panel A: Patent application quality variables

	N	Mean	Median	p10	p90	S.D.
ForwardCitations	236,643	3.464	2	1	7	6.709
Generality	236,643	0.133	0	0	0.500	0.220

Panel B: Patent application characteristics

	N	Mean	Median	p10	p90	S.D.
NumberIndepClaims	637,344	2.791	2	1	5	2.545
NumberDepClaims	637,344	15.528	14	4	27	13.438
NumberWordsIndepClaims	637,344	361.497	258	85	695	499.292
NumberWordsDepClaims	637,344	601.290	475	135	1,134	879.788
MinNumberWordsIndepClaims	637,344	115.735	92	32	210	130.584
MinNumberWordsDepClaims	637,344	21.837	17	11	30	64.500
AvgNumberWordsIndepClaims	637,344	138.185	114	51.500	235.333	136.399
AvgNumberWordsDepClaims	637,344	42.348	34.125	20.875	64.500	69.755
NumberCitedForeignPatents	637,344	2.905	0	0	7	10.680
NumberNovelWords	637,344	0.309	0	0	1	5.103
NumberCitedLiterature	637,344	3.837	0	0	6	22.254
NumberCitedApplications	637,344	2.755	0	0	5	15.446
OriginalityApplication	637,344	0.155	0	0	0.776	0.309
NumberCitedPatents	637,344	8.511	0	0	18	35.589
OriginalityPatent	637,344	0.166	0	0	0.618	0.259
USInventorDummy	637,344	0.439	0	0	1	0.496
SmallEntityDummy	637,344	0.269	0	0	1	0.444

Table 2: Comparing OLS to machine learning prediction of high-quality patents

This table compare the performance of a machine learning algorithm and an OLS function in terms of identifying high-quality patents in the test set. The first column indicates the top 1%, 5%, 10%, and 25% of the predicted generality distribution and the second column shows the percentage of applications that identified by both ML and OLS as the top 1%, 5%, 10%, and 25% of predicted generality distribution. The third and fourth columns report the actual generality among the applications within each of the predicted generality distribution that are identified either by ML only, or by OLS only. The last column shows the statistical difference between results in the third and fourth columns.

Predicted generality	ML&OLS overlap	Average actual generality for applications identified as high predicted generality by:		
		ML Only	OLS Only	Difference (t-statistic)
Top 1%	20.5%	0.281	0.218	0.063*** (6.82)
Top 5%	33.3%	0.227	0.162	0.065*** (13.33)
Top 10%	38.4%	0.198	0.136	0.062*** (16.47)
Top 25%	42.7%	0.151	0.113	0.038*** (14.95)

Table 3: Relationship between patent examiner characteristics and screening performance

The sample consists of patent examiners in the out-of-sample test set. Art Unit fixed effects and issue year fixed effects are included in all regressions. Standard error are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable Dependent Variable	# False Rejection	False Rejection Rate	# False Acceptance	False Acceptance Rate
	(1)	(2)	(3)	(4)
# Years Employed	-0.055***	-0.002***	0.066***	0.002***
	(0.002)	(0.000)	(0.003)	(0.000)
Ln(# Applications Reviewed)	1.359***	-0.052***	2.632***	0.018***
	(0.024)	(0.001)	(0.034)	(0.001)
Examiner Leniency	-4.227***	-0.317***	6.976***	0.192***
	(0.068)	(0.004)	(0.097)	(0.003)
Art Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.486	0.265	0.465	0.167
Observations	83103	83103	83103	83103

Table 4: Relationship between inefficient patent screening and subsequent patent maintenance

The sample consists of granted patents in the out-of-sample test set. *FalseAccept* equals to one if a patent is accepted by an actual examiner but rejected by my algorithm and zero otherwise as described in Section 5.4.5. Small & Micro Entity Dummies, Art Unit fixed effects, issue year fixed effects, and patent USPC class fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Payment of Maintenance Fee in the 4th Year	Payment of Maintenance Fee in the 8th Year	Maintenance Fee Reminder Mailed	Patent Expired for Failure to Pay Maintenance Fees
	(1)	(2)	(3)	(4)
FalseAccept	-0.032***	-0.011***	0.026***	0.032***
	(-19.26)	(-7.69)	(13.22)	(18.14)
Small & Micro Entity Dummies	Yes	Yes	Yes	Yes
Art Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Patent USPC Class FE	Yes	Yes	Yes	Yes
R^2	0.092	0.458	0.081	0.056
Observations	235552	235552	235552	235552

Table 5: Summary statistics (firms)

This table shows descriptive statistics for the sample of both public and private firms that have at least one patent application filed since 2010 and with status available before (and including) 2018. Panels A and B show summary statistics for the sample of public firms; Panel C shows summary statistics for the sample of private firms. *ExaminerFalseAcceptRate* is the false acceptance rate of an examiner associated with each patent application, which is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *CAR [-1:1]* is the cumulative abnormal return on a firm's equity over a 3-trading-day window (from day -1 to day 1) around each patent grant date. *BHAR [1:63]*, *BHAR [1:125]*, *BHAR [1:188]*, and *BHAR [1:250]* are the buy-and-hold abnormal returns on a firm's equity over a 63-trading-day, 125-trading-day, 188-trading-day, and 250-trading-day window after each patent grant date. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *ROA* is the ratio of quarterly net income over book assets. *Cash Flow* is the quarterly cash flow over book assets. *R&D Expenditures* are the quarterly R&D expenditures over book assets. *#PatentLitigation* counts the quarterly number of patent litigation that firms act as defendants. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. *SuccessfulExit* is a dummy, which equals one if a given private firm has exited through an IPO or an M&A by the end of my sample period and zero otherwise. *LnVCFinancingAmount* is the natural logarithm of the quarterly investment amount for each firm. *LnNumberFundInvested* is the natural logarithm of the quarterly number of invested funds for each firm. *TotalFundingToDate* is the natural logarithm of total funding each firm has received prior to a given quarter. *LnFirmAge* is the natural logarithm of firm age, which equals the current year minus the firm founding year plus one. All accounting variables (i.e., *ROA*, *Cash Flow*, *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%.

Panel A: public firm sample – stock returns (firm-event level)

	N	Mean	Median	p10	p90	S.D.
ExaminerFalseAcceptRate	115,673	0.166	0.160	0.062	0.273	0.091
CAR [-1:1]	115,664	0.032	0.016	-2.540	2.568	2.592
BHAR [1:63]	115,669	-0.004	-0.275	-16.943	17.371	15.408
BHAR [1:125]	115,353	-0.571	-0.301	-28.009	27.763	26.214
BHAR [1:188]	114,183	-1.834	-0.257	-40.514	36.723	38.290
BHAR [1:250]	112,607	-3.862	0.063	-55.678	45.625	52.162

Panel B: public firm sample – operating performance (firm-quarter level)

	N	Mean	Median	p10	p90	S.D.
AvgExaminerFalseAcceptRate	13,416	0.167	0.164	0.105	0.227	0.066
#ApplicatiosReviewe	13,416	160.624	15	2	200	839.933
#PatentsGranted	13,416	127.365	12	1	162	686.466
ROA	13,130	-0.022	0.006	-0.118	0.034	0.102
Cash Flow	12,768	-0.012	0.016	-0.111	0.043	0.101
R&D Expenditures	13,141	0.035	0.020	0	0.087	0.050
#PatentLitigation	13,416	0.131	0	0	0	0.622
FirmSize	13,354	7.196	6.985	4.214	10.560	2.429
Leverage	12,800	0.199	0.165	0	0.459	0.220
Ln(M/B)	12,718	1.130	1.051	0.075	2.255	0.909

Panel C: private firm sample (firm-quarter level)						
	N	Mean	Median	p10	p90	S.D.
SuccessExit	13,494	0.212	0	0	1	0.409
AvgExaminerFalseAcceptRate	13,494	0.166	0.164	0.080	0.248	0.077
#ApplicationsReviewed	13,494	9.773	4	1	20	23.829
#PatentsGranted	13,494	7.834	3	0	16	20.930
LnVCFinancingAmount	13,494	0.209	0	0	0	0.774
LnNumberFundInvested	13,494	0.319	0	0	1	1.203
TotalFundingToDate	13,494	0.461	0	0	2.398	1.382
FirmAge	13,494	10.620	10	6	17	4.678

Table 6: Relationship between screening efficiency of patent examiners and stock market returns of public firms

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *CAR [-1:1]* is the cumulative abnormal return on a firm's equity over a 3-trading-day window (from day -1 to day 1) around each patent grant date. *BHAR [1:63]*, *BHAR [1:125]*, *BHAR [1:188]*, and *BHAR [1:250]* are the buy-and-hold abnormal returns on a firm's equity over a 63-trading-day, 125-trading-day, 188-trading-day, and 250-trading-day window after each patent grant date. *ExaminerFalseAcceptRate* is the false acceptance rate of an examiner associated with each patent application, which is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: Relationship between screening efficiency of patent examiners and stock market reactions around each patent grant date

Dependent Variable	CAR [-1:1]	CAR [-1:1]
	(1)	(2)
ExaminerFalseAcceptRate		-0.233***
		(-2.78)
Constant	0.032***	0.071***
	(4.19)	(4.45)
Observations	115664	115664

Panel B: Relationship between screening efficiency of patent examiners and long-run stock market return

Dependent Variable	[BHAR [1:63]	BHAR [1:125]	BHAR [1:188]	BHAR [1:250]
	(1)	(2)	(3)	(4)
ExaminerFalseAcceptRates	-0.992**	-2.195***	-4.060***	-7.031***
	(-1.99)	(-2.59)	(-3.27)	(-4.14)
Constant	0.161*	-0.205	-1.158***	-2.691***
	(1.71)	(-1.28)	(-4.91)	(-8.34)
Observations	115669	115353	114184	112609

Table 7: Relationship between screening efficiency of patent examiners and subsequent operating performance of public firms

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *ROA* is the ratio of quarterly net income over book assets. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. *R&D Expenditures* are the quarterly R&D expenditures over book assets. All accounting variables (i.e., *ROA*, *Cash Flow*, *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%. Quarter fixed effects and industry (two-digit SIC code) fixed effects are included in all regressions. All regressions are OLS regressions with standard errors double clustered at the firm and quarter level. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Subsequent ROA			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	-0.048***	-0.125***	-0.238***	-0.236***
	(-4.83)	(-3.93)	(-4.35)	(-2.96)
#ApplicationsReviewed	-0.018***	-0.070***	-0.142***	-0.196***
	(-5.01)	(-5.05)	(-4.80)	(-4.08)
#PatentsGranted	0.018***	0.066***	0.136***	0.181***
	(4.87)	(4.94)	(4.72)	(3.91)
FirmSize	0.011***	0.047***	0.090***	0.130***
	(14.26)	(18.30)	(18.14)	(15.41)
Leverage	-0.084***	-0.296***	-0.481***	-0.636***
	(-13.18)	(-14.22)	(-11.66)	(-8.87)
Ln(M/B)	0.013***	0.045***	0.084***	0.128***
	(11.22)	(12.35)	(11.27)	(9.78)
R&D Expenditures	-0.936***	-3.466***	-6.545***	-9.661***
	(-18.58)	(-23.38)	(-20.92)	(-18.97)
Constant	-0.143***	-0.748***	-1.589***	-2.577***
	(-5.14)	(-6.95)	(-8.55)	(-6.74)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.409	0.523	0.543	0.537
Observations	11954	10536	8170	5995

Table 8: Relationship between screening efficiency of patent examiners and subsequent R&D expenditures of public firms

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *R&D Expenditures* are the quarterly R&D expenditures over book assets. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. *R&D Expenditures* is the R&D expenditure over the book value total assets. All accounting variables (i.e., *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%. Quarter fixed effects and industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. All regressions are OLS regressions with standard errors double clustered at the firm and quarter level. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Subsequent R&D Expenditures			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	-0.006*	-0.021**	-0.036*	-0.069**
	(-1.85)	(-2.00)	(-1.77)	(-2.15)
#ApplicationsReviewed	0.000	0.001	0.004	0.001
	(0.38)	(0.26)	(0.49)	(0.06)
#PatentsGranted	0.001	0.004	0.008	0.022*
	(0.56)	(1.11)	(1.05)	(1.76)
FirmSize	-0.002***	-0.011***	-0.025***	-0.042***
	(-7.50)	(-10.82)	(-11.43)	(-10.38)
Leverage	-0.001	-0.013**	-0.062***	-0.128***
	(-0.74)	(-2.10)	(-4.07)	(-5.02)
Ln(M/B)	0.002***	0.006***	0.012***	0.017***
	(3.62)	(4.04)	(3.55)	(2.87)
R&D Expenditures	0.808***	2.961***	5.465***	7.679***
	(25.61)	(27.49)	(23.01)	(17.34)
Constant	0.032***	0.197***	0.482***	0.938***
	(4.85)	(5.98)	(9.10)	(8.09)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.760	0.796	0.784	0.767
Observations	11965	10572	8215	6039

Table 9: Relationship between screening efficiency of patent examiners and the subsequent number of patent litigation of public firms

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *#PatentLitigation* counts the quarterly number of patent litigation that firms act as defendants. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. *R&D Expenditures* are the quarterly R&D expenditures over book assets. All accounting variables (i.e., *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%. Quarter fixed effects and industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. All regressions are OLS regressions with standard errors double clustered at the firm and quarter level. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Subsequent #PatentLitigation			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	0.184*	0.749***	1.331**	2.043**
	(1.70)	(2.63)	(2.49)	(2.21)
#ApplicationsReviewed	0.029**	0.119**	0.260***	0.506***
	(2.02)	(2.57)	(2.73)	(3.25)
#PatentsGranted	0.004	0.017	0.031	-0.030
	(0.34)	(0.42)	(0.36)	(-0.21)
FirmSize	0.069***	0.270***	0.571***	0.924***
	(11.13)	(11.89)	(11.51)	(10.77)
Leverage	-0.284***	-1.229***	-2.699***	-4.457***
	(-5.63)	(-6.35)	(-6.24)	(-5.81)
Ln(M/B)	0.018**	0.066**	0.155**	0.347***
	(2.45)	(2.41)	(2.52)	(3.03)
R&D Expenditures	0.780***	2.868***	5.351***	7.308***
	(8.33)	(9.26)	(7.98)	(6.02)
Constant	-0.244	-0.787	-2.523**	-4.667**
	(-1.03)	(-1.28)	(-2.02)	(-2.53)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.121	0.159	0.179	0.197
Observations	12204	11197	9228	7182

Table 10: Relationship between screening efficiency of patent examiners and subsequent operating performance of public firms (A cross-industry analysis)

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *ROA* is the ratio of quarterly net income over book assets. *R&D Expenditure* is the R&D expenditure over the book value total assets. *#PatentLitigation* counts the quarterly number of patent litigation that firms act as defendants. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *HiTechAndHealth* is a dummy, which equals one if a firm belongs to the High-Tech industry or the Health industry based on Fama and French 5 industry groups. Control variables are defined as in Table 7. Quarter fixed effects and industry (two-digit SIC code) fixed effects are included in all regressions. All regressions are OLS regressions with standard errors double clustered at the firm and quarter level. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: Relationship between screening efficiency of patent examiners and subsequent ROA

Dependent Variable	Subsequent ROA			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
A: AvgExaminerFalseAcceptRate	-0.015	-0.036	-0.148**	-0.209**
	(-1.53)	(-1.15)	(-2.52)	(-2.41)
B: HiTechAndHealth	0.012***	0.030**	0.027	0.017
	(3.26)	(2.50)	(1.26)	(0.48)
A × B	-0.058***	-0.155**	-0.161	-0.043
	(-3.10)	(-2.55)	(-1.52)	(-0.27)
Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.409	0.523	0.543	0.537
Observations	11954	10536	8170	5995

Panel B: Relationship between screening efficiency of patent examiners and subsequent R&D expenditures

Dependent Variable	Subsequent R&D Expenditures			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
A: AvgExaminerFalseAcceptRate	-0.003	-0.017**	-0.035**	-0.073**
	(-1.18)	(-2.27)	(-2.11)	(-2.45)
B: HiTechAndHealth	0.006***	0.028***	0.064***	0.101***
	(4.77)	(6.72)	(7.49)	(6.58)
A × B	-0.001	0.016	0.054	0.098*
	(-0.20)	(0.90)	(1.51)	(1.66)
Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.758	0.791	0.775	0.753
Observations	11965	10572	8215	6039

Panel C: Relationship between screening efficiency of patent examiners and subsequent patent litigation

Dependent Variable	Subsequent #PatentLitigation			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
A: AvgExaminerFalseAcceptRate	-0.052	0.210	0.247	-0.076
	(-0.47)	(0.76)	(0.52)	(-0.10)
B: HiTechAndHealth	-0.015	0.060	0.092	-0.002
	(-0.39)	(0.55)	(0.44)	(-0.01)
A × B	0.467**	1.151**	2.265**	4.255**
	(2.35)	(2.23)	(2.35)	(2.55)
Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.106	0.139	0.154	0.167
Observations	12204	11197	9228	7182

Table 11: Relationship between screening efficiency of patent examiners and subsequent exits of private firms

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *SuccessfulExit* is a dummy, which equals one if a given private firm has exited through an IPO or an M&A by the end of my sample period and zero otherwise. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted applications for each firm in the past three years as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *LnVCFinancingAmount* is the natural logarithm of the quarterly investment amount for each firm. *LnNumberFundInvested* is the natural logarithm of the quarterly number of invested funds for each firm. *TotalFundingToDate* is the natural logarithm of total funding each firm has received prior to a given quarter. *LnFirmAge* is the natural logarithm of firm age, which equals the current year minus the firm founding year plus one. Year fixed effects, industry (two-digit SIC code) fixed effects, and state fixed effects are included in all regressions. *t*-statistics are in parentheses. All regressions are OLS regressions with standard errors clustered at the state level. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Subsequent SuccessfulExit			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	-0.019***	-0.094***	-0.181***	-0.214***
	(-4.03)	(-4.73)	(-3.89)	(-4.73)
#PatentsGranted	0.006***	0.021***	0.038***	0.059***
	(3.23)	(4.55)	(3.45)	(3.41)
#ApplicationsReviewed	-0.003	-0.008	-0.013	-0.032
	(-1.28)	(-1.37)	(-1.04)	(-1.68)
InvestmentAmount	-0.005**	-0.005**	0.007*	0.004
	(-2.58)	(-2.28)	(1.83)	(0.56)
NumberFundInvested	0.005***	0.010***	0.010**	0.017***
	(4.01)	(3.77)	(2.49)	(3.42)
TotalFundingToDate	0.001	0.004***	0.008***	0.009***
	(1.27)	(4.28)	(3.09)	(3.48)
LnFirmAge	0.010***	0.020***	0.022	0.019
	(3.34)	(2.70)	(1.55)	(0.83)
Constant	-0.038***	-0.045	0.076	0.135
	(-6.94)	(-1.21)	(0.74)	(0.98)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
R^2	0.013	0.026	0.041	0.051
Observations	13476	12543	10020	7412

Appendix to "How can Innovation Screening be Improved? A Machine Learning Analysis with Economic Consequences for Firm Performance"

A.1 The supervised machine learning problem and the algorithm used in this paper

A.1.1 The supervised machine learning problem

Supervised learning is a machine learning problem of learning a function that maps input variables to an output variable using the training data with both input and output variables available. The goal of supervised learning is to predict well with a new out-of-sample dataset (which we usually called it the test data).

In the context of this paper, I use the training data to construct $\hat{f}(X) = \hat{y}$ from input variables X about patent applications to predict an outcome variable y about the performance of patent applications such that $\hat{f}(X)$ predicts well out of sample. Specifically, I use the training data to train $f(X)$ as follows:

$$\hat{y} = \hat{f}(X) \quad = \quad \arg\min_{f \in \mathcal{F}} L(f(X), y) + R(f(X)), \tag{A.1}$$

where $L(f(X), y)$ is the training loss function, $\mathcal{F}$ is the set of all possible functions f, and $R(f(X))$ is the regularization term.

The goal of minimizing the training loss function is to increase the in-sample prediction accuracy as much as possible, while adding the regularization term is to avoid in-sample over fitting by penalizing the algorithm for choosing more expressive functions.

A.1.2 The "Extreme Gradient Boosting" algorithm

The "Extreme Gradient Boosting" algorithm (XGBoost) is an implementation of gradient boosting machines, which is used for the supervised machine learning prediction described above (see, e.g., Chen and Guestrin (2016) and Friedman (2001)). XGBoost is a decision tree ensemble based on tree boosting. A decision tree ensemble consists of a set of decision trees, where each tree i itself is a prediction function $f_i(X)$. Tree boosting is to train the each prediction function $f_i(X)$ using an additive strategy: add one new tree at a time from what we have learned. Specifically, we have

$$\hat{y}_0 = \hat{f}_0(X) \quad = \quad 0 \tag{A.2}$$
$$\hat{y}_1 = \hat{f}_1(X) \quad = \quad \hat{f}_0(X) + f_1(X) = f_1(X) \tag{A.3}$$
$$\hat{y}_2 = \hat{f}_2(X) \quad = \quad \hat{f}_1(X) + f_2(X) = f_1(X) + f_2(X) \tag{A.4}$$
$$\cdots$$
$$\hat{y}_t = \hat{f}_t(X) \quad = \quad \hat{f}_{t-1}(X) + f_t(X) = \sum_{i=1}^{t} f_i(X), \tag{A.5}$$

and the goal at step t is to find $f_t(X)$ that solves the following minimization problem:

$$\hat{y}_t = \hat{f}_t(X) \quad = \quad \arg\min_{f \in \mathcal{F}} L(f_t(X) + \hat{y}_{t-1}, y) + R(f_t(X)). \tag{A.6}$$

Here, each prediction function $f_i(X)$ and the corresponding regularization term $R(f_i(X))$ are defined as

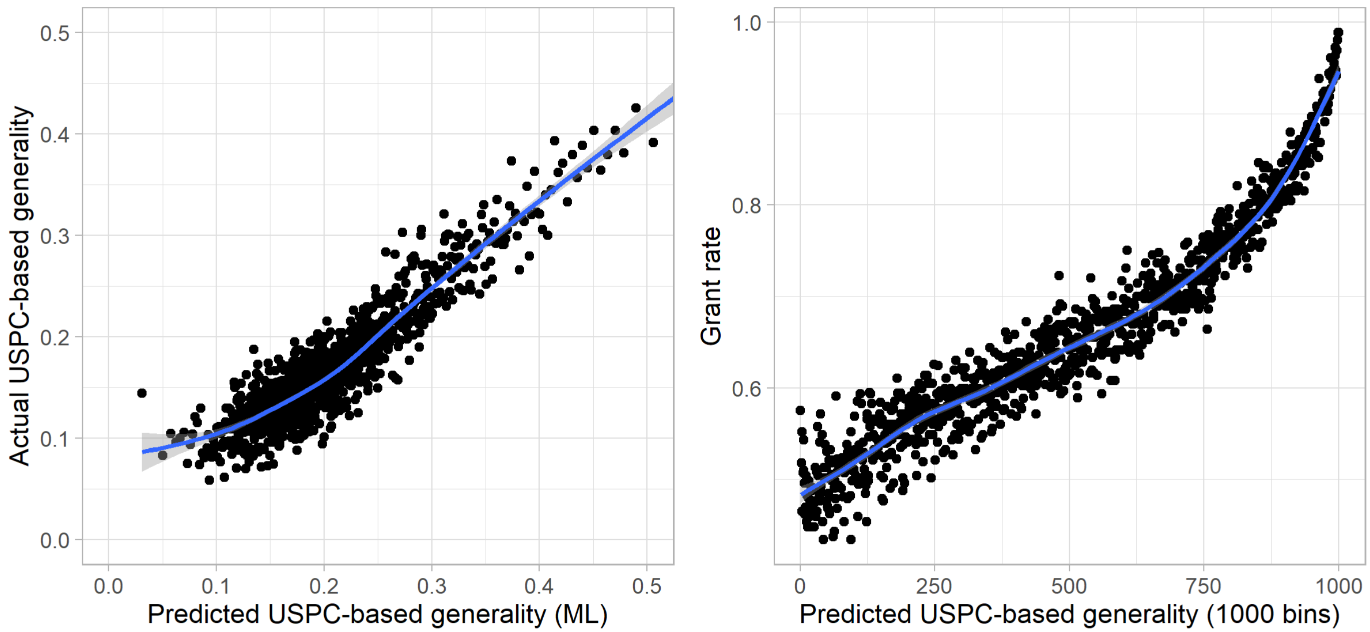
$$f_i(X) = \omega_{q(X)}, q : \mathbb{R}^m \to T, \omega \in \mathbb{R}^T, \tag{A.7}$$

$$R(f_i(X)) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{A.8}$$

where ω are the leaf weights, q is a function mapping each data point to the corresponding leaf index, T is the total number of leaves in the tree, both γ and λ are parameters to weight each of these two complexity measures in order to avoid over-fitting (see Chen and Guestrin (2016) for a detailed discussion).

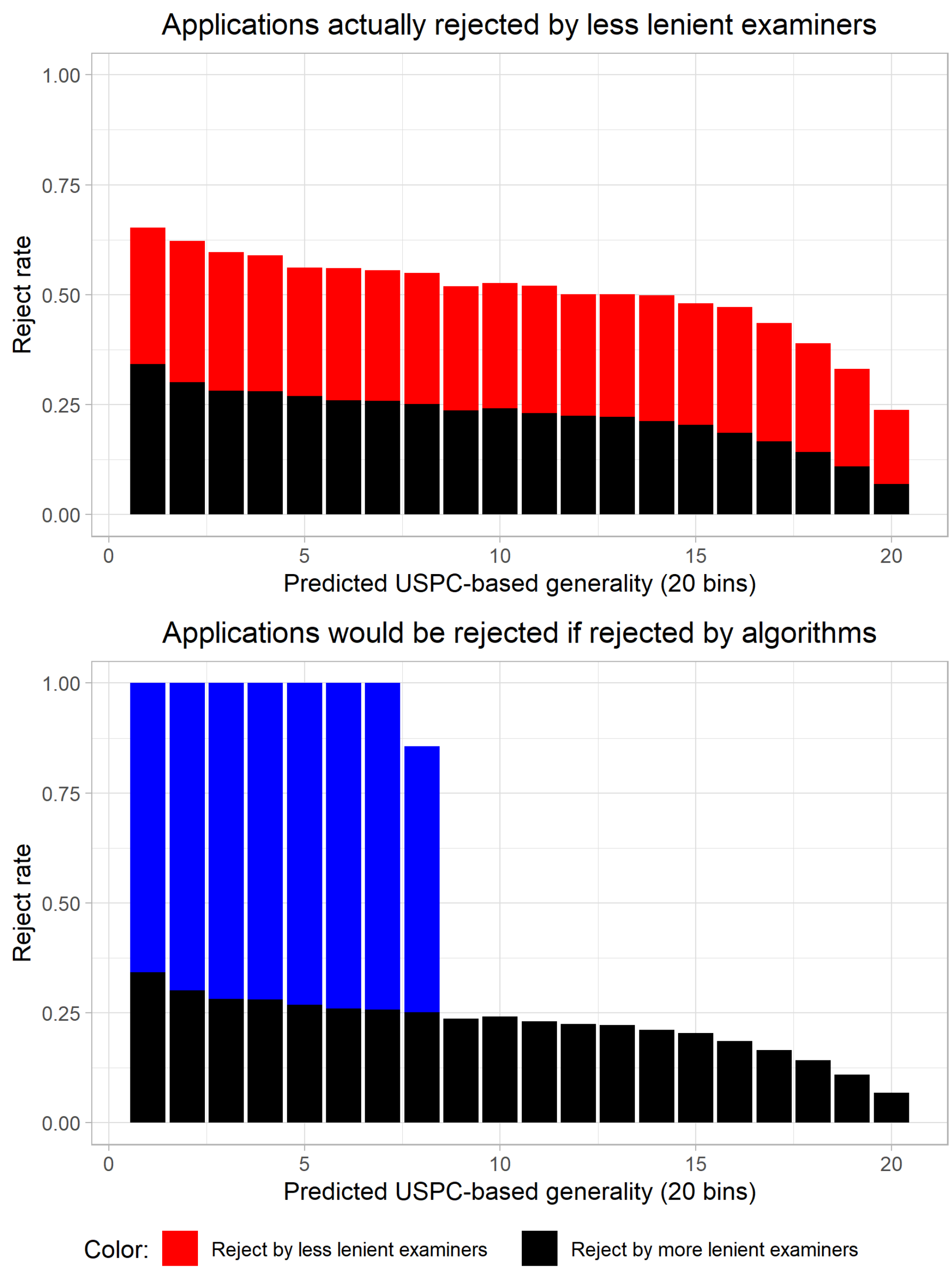
A.2 Additional Figures

A.2.1 Results using the generality measure based on the USPC classification



The figure shows the relation between predicted USPC-based generality, actual USPC-based generality, and grant rate of actual examiners in the test set. In the left panel, the average predicted USPC-based generality of patent applications in each bin based on my machine learning algorithm is on the x-axis and the actual USPC-based generality is on the y-axis. In the right panel, the rank of the average predicted USPC-based generality of patent applications in each bin based on my machine learning algorithm is on the x-axis and the grant rate is on the y-axis.

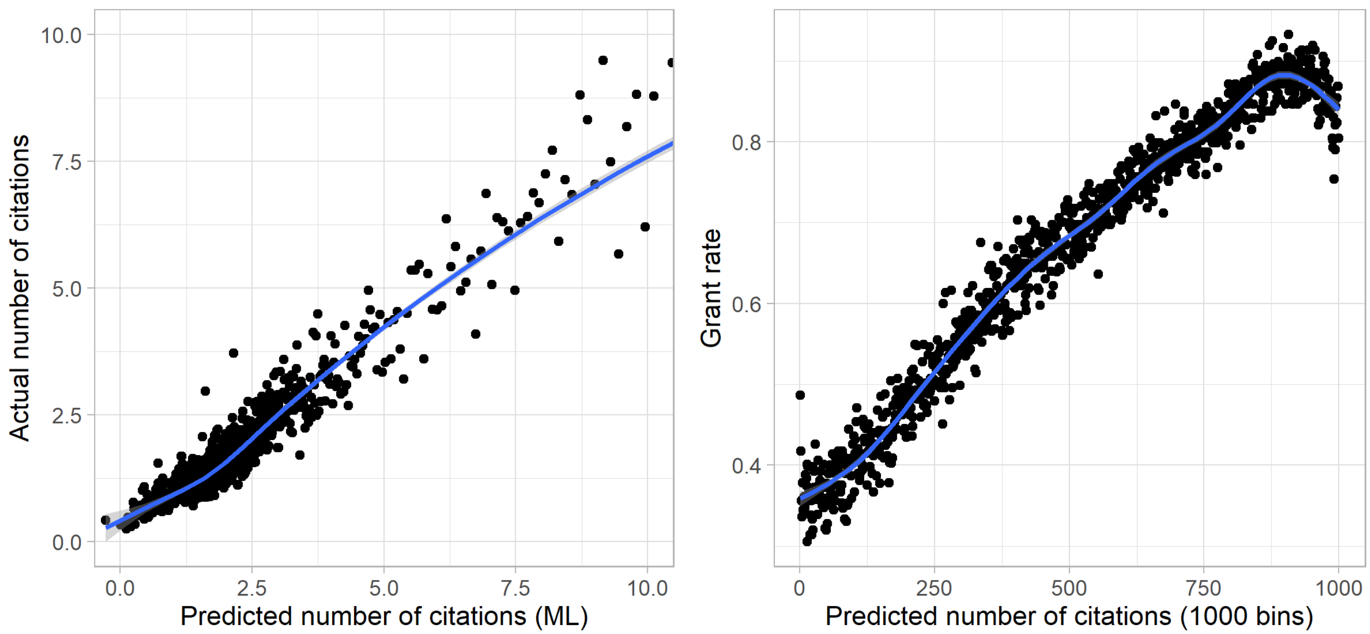
Figure IA.1: The relation between predicted USPC-based generality, actual USPC-based generality, and grant rate of actual examiners in the test set



This figure shows comparison between applications rejected by stricter examiners and applications rejected by the algorithm. I divide the sample up equally into 20 bins by predicted USPC-based generality (x-axis). In both panels, the black bar at the bottom of each bin shows the fraction of patent applications rejected by more lenient examiners. The red bar in the top panel shows which applications less lenient examiners actually reject. The blue bar in the below panel shows which applications my algorithm would reject to match the grant rate of less lenient examiners.

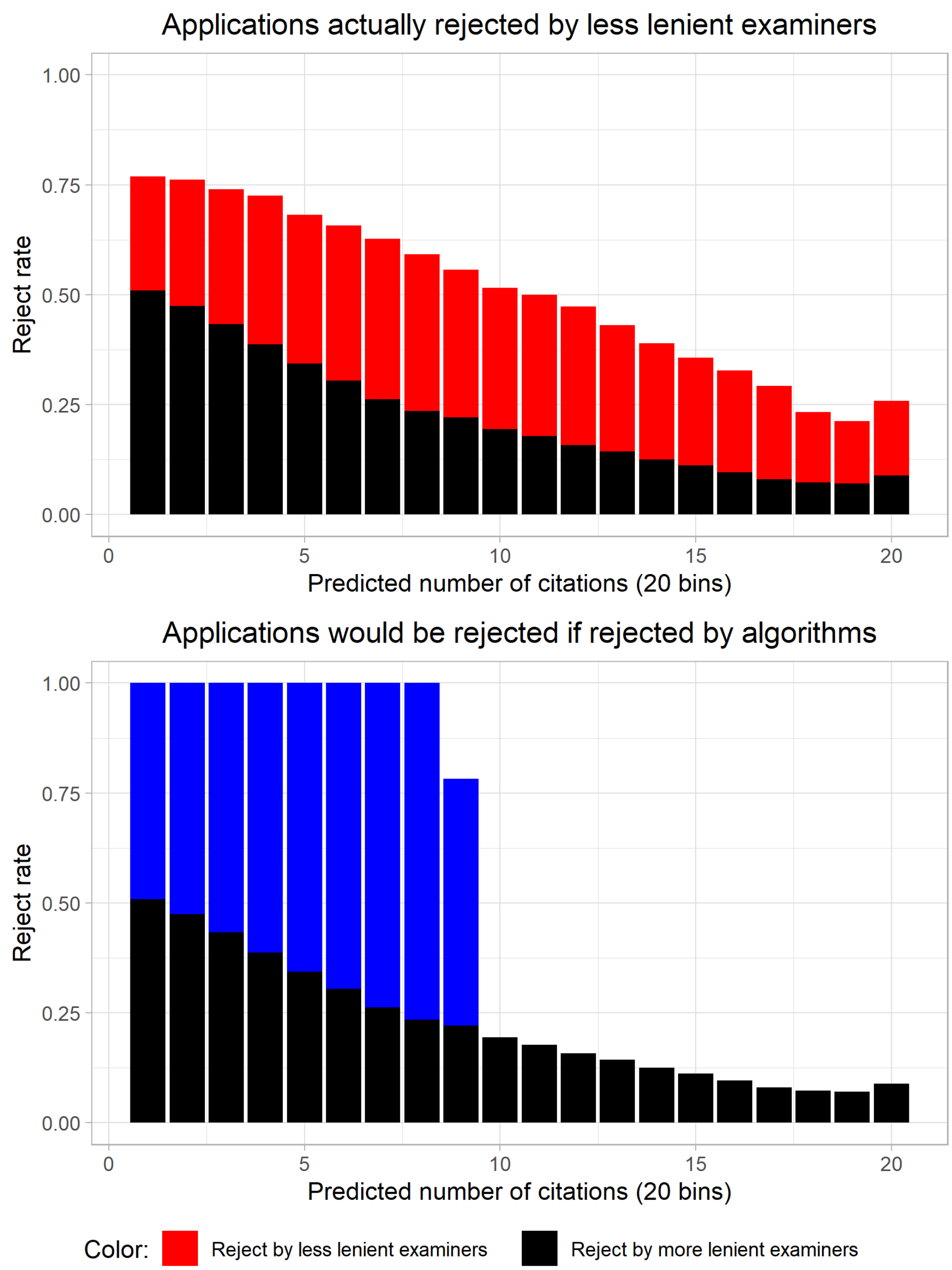
Figure A.2: Comparison between applications rejected by stricter examiners and applications rejected by the algorithm

A.2.2 Results using the number of citations



The figure shows the relation between predicted citations, actual citations, and grant rate of actual examiners in the test set. In the left panel, the average predicted number of citations of patent application in each bin based on my machine learning algorithm is on the x-axi and the actual citation is on the y-axis. In the right panel, the rank of the average predicted number of citations of patent application in each bin based on my algorithm is on the x-axis and the grant rate is on the y-axis.

Figure A.3: The relation between predicted citations, actual citations, and grant rate of actual examiners in the test set



This figure shows comparison between applications rejected by stricter examiners and applications rejected by the algorithm. I divide patent applications in the test set into 20 bins by the predicted number of citations (x-axis). In both panels, the black bar at the bottom of each bin shows the fraction of patent applications rejected by more lenient examiners. The red bar in the top panel shows which applications less lenient examiners actually reject. The blue bar in the below panel shows which applications my algorithm would reject to match the grant rate of less lenient examiners.

Figure A.4: Comparison between applications rejected by stricter examiners and applications rejected by the algorithm

A.3 Additional Tables

Table A.1: Relationship between screening inefficiency of patent examiners and subsequent operating performance of public firms (Cash Flow)

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *Cash Flow* is the quarterly cash flow over book assets. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. *R&D Expenditures* are the quarterly R&D expenditures over book assets. All accounting variables (i.e., *Cash Flow*, *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%. Quarter fixed effects and industry (two-digit SIC code) fixed effects are included in all regressions. All regressions are OLS regressions with standard errors double clustered at the firm and quarter level. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Subsequent Cash Flow			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	-0.045***	-0.119***	-0.241***	-0.266***
	(-5.10)	(-3.97)	(-4.35)	(-3.28)
#ApplicationsReviewed	-0.020***	-0.076***	-0.160***	-0.237***
	(-5.44)	(-5.38)	(-5.24)	(-4.71)
#PatentsGranted	0.019***	0.073***	0.155***	0.224***
	(5.33)	(5.36)	(5.24)	(4.63)
FirmSize	0.011***	0.046***	0.088***	0.124***
	(13.89)	(17.39)	(16.96)	(14.07)
Leverage	-0.074***	-0.254***	-0.390***	-0.470***
	(-11.65)	(-11.99)	(-8.95)	(-6.13)
Ln(M/B)	0.011***	0.037***	0.069***	0.101***
	(9.84)	(10.17)	(8.90)	(7.29)
R&D Expenditures	-0.924***	-3.381***	-6.361***	-9.494***
	(-18.61)	(-22.71)	(-20.03)	(-18.08)
Constant	-0.130***	-0.694***	-1.464***	-2.340***
	(-4.77)	(-6.58)	(-8.00)	(-6.28)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.432	0.536	0.546	0.542
Observations	11622	10125	7744	5591

Table A.2: Relationship between screening inefficiency of patent examiners and subsequent outcomes of public firms (A within-firm analysis)

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *ROA* is the ratio of quarterly net income over book assets. *R&D Expenditure* is the R&D expenditure over the book value total assets. *#PatentLitigation* counts the quarterly number of patent litigation that firms act as defendants. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#Patents-Granted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. Quarter fixed effects and firm fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: Relationship between screening inefficiency of patent examiners and subsequent ROA

Dependent Variable	Subsequent ROA			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	-0.013	-0.045*	-0.091***	-0.147***
	(-1.23)	(-1.90)	(-2.62)	(-3.50)
#ApplicationsReviewed	-0.002	0.018**	0.036***	0.033**
	(-0.59)	(2.38)	(2.93)	(2.05)
#PatentsGranted	0.002	-0.014*	-0.017	-0.008
	(0.49)	(-1.93)	(-1.49)	(-0.55)
FirmSize	0.014***	0.025***	-0.007	-0.012
	(6.78)	(4.91)	(-0.85)	(-1.01)
Leverage	-0.087***	-0.218***	-0.174***	-0.012
	(-12.34)	(-12.30)	(-5.69)	(-0.28)
Ln(M/B)	0.023***	0.068***	0.079***	0.075***
	(18.13)	(22.16)	(15.36)	(10.40)
R&D Expenditures	-0.365***	-1.081***	-1.090***	-0.741***
	(-13.53)	(-16.98)	(-10.60)	(-5.03)
Constant	-0.117**	-0.361***	-0.087	-0.377*
	(-2.06)	(-2.85)	(-0.51)	(-1.95)
Firm FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.692	0.875	0.936	0.964
Observations	11954	10536	8170	5995

Panel B: Relationship between screening inefficiency of patent examiners and subsequent R&D expenditures

Dependent Variable	Subsequent R&D Expenditures			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	-0.007*	-0.019*	-0.036**	-0.034*
	(-1.96)	(-1.76)	(-2.16)	(-1.70)
#ApplicationsReviewed	0.000	-0.010***	-0.022***	-0.031***
	(0.25)	(-2.97)	(-3.81)	(-3.92)
#PatentsGranted	-0.000	0.008**	0.017***	0.029***
	(-0.27)	(2.48)	(3.07)	(3.85)
FirmSize	-0.007***	-0.017***	-0.013***	-0.021***
	(-9.24)	(-7.14)	(-3.12)	(-3.46)
Leverage	0.008***	0.018**	0.024	0.001
	(3.05)	(2.17)	(1.64)	(0.05)
Ln(M/B)	-0.001	-0.006***	-0.007***	-0.010***
	(-1.36)	(-4.20)	(-2.95)	(-2.95)
R&D Expenditures	0.351***	0.838***	0.611***	0.016
	(35.03)	(28.58)	(12.29)	(0.23)
Constant	0.056***	0.182***	0.219***	0.394***
	(2.65)	(3.12)	(2.62)	(4.21)
Firm FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.844	0.917	0.956	0.975
Observations	11965	10572	8215	6039

Panel C: Relationship between screening inefficiency of patent examiners and subsequent patent litigation

Dependent Variable	Subsequent #PatentLitigation			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
AvgExaminerFalseAcceptRate	0.140	0.511**	0.594	1.008*
	(1.49)	(2.01)	(1.37)	(1.82)
#ApplicationsReviewed	-0.122***	-0.474***	-0.949***	-1.248***
	(-4.21)	(-5.89)	(-6.26)	(-5.84)
#PatentsGranted	0.006	0.002	-0.006	-0.110
	(0.22)	(0.02)	(-0.04)	(-0.54)
FirmSize	-0.008	-0.049	-0.138	-0.201
	(-0.42)	(-0.92)	(-1.33)	(-1.30)
Leverage	-0.187***	-0.816***	-1.236***	-0.989*
	(-2.92)	(-4.38)	(-3.38)	(-1.84)
Ln(M/B)	-0.011	-0.066**	-0.169***	-0.204**
	(-0.98)	(-2.04)	(-2.68)	(-2.17)
R&D Expenditures	0.142	0.520	0.586	0.610
	(0.58)	(0.76)	(0.46)	(0.32)
Constant	0.024	0.261	2.837*	4.158
	(0.05)	(0.19)	(1.65)	(1.52)
Firm FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.499	0.691	0.788	0.871
Observations	12204	11197	9228	7182

Table A.3: Relationship between screening inefficiency of patent examiners and subsequent outcomes of public firms (Robustness tests)

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *ROA* is the ratio of quarterly net income over book assets. *R&D Expenditure* is the R&D expenditure over the book value total assets. *#PatentLitigation* counts the quarterly number of patent litigation that firms act as defendants. $AvgExaminerFalseAcceptRate_{Adj}$ is defined as the average (art-unit adjusted) false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past three-year rolling window as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln(M/B)* is the natural logarithm of the market to book ratio. All accounting variables (i.e., *ROA*, *R&D Expenditures*, *Leverage*, *Ln(M/B)*) are winsorized at 0.1% and 99.9%. Quarter fixed effects and industry (two-digit SIC code) fixed effects are included in all regressions. All regressions are OLS regressions with standard errors double clustered at the firm and quarter level. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: Relationship between screening inefficiency of patent examiners and subsequent ROA

Dependent Variable	Subsequent ROA			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
$\text{AvgExaminerFalseAcceptRate}_{\text{Adj}}$	-0.041***	-0.093***	-0.200***	-0.231***
	(-3.84)	(-2.70)	(-3.33)	(-2.68)
#ApplicationsReviewed	-0.018***	-0.068***	-0.139***	-0.193***
	(-4.85)	(-4.92)	(-4.70)	(-4.02)
#PatentsGranted	0.017***	0.065***	0.133***	0.178***
	(4.72)	(4.82)	(4.63)	(3.86)
Size	0.011***	0.047***	0.090***	0.130***
	(14.24)	(18.27)	(18.12)	(15.41)
Leverage	-0.085***	-0.297***	-0.483***	-0.638***
	(-13.22)	(-14.29)	(-11.73)	(-8.93)
Market to Book	0.013***	0.045***	0.085***	0.129***
	(11.29)	(12.41)	(11.32)	(9.81)
R&D Expenditures	-0.935***	-3.463***	-6.539***	-9.653***
	(-18.56)	(-23.36)	(-20.91)	(-18.97)
Constant	-0.149***	-0.764***	-1.618***	-2.605***
	(-5.36)	(-7.07)	(-8.66)	(-6.81)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.409	0.523	0.543	0.537
Observations	11954	10536	8170	5995

Panel B: Relationship between screening inefficiency of patent examiners and subsequent R&D expenditures

Dependent Variable	Subsequent R&D Expenditures			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
$\text{AvgExaminerFalseAcceptRate}_{\text{Adj}}$	-0.006*	-0.020	-0.042*	-0.079**
	(-1.76)	(-1.62)	(-1.86)	(-2.24)
#ApplicationsReviewed	0.000	0.001	0.004	0.001
	(0.42)	(0.33)	(0.52)	(0.11)
#PatentsGranted	0.001	0.004	0.008	0.022*
	(0.54)	(1.06)	(1.03)	(1.72)
Size	-0.002***	-0.011***	-0.025***	-0.042***
	(-7.49)	(-10.81)	(-11.43)	(-10.38)
Leverage	-0.001	-0.014**	-0.062***	-0.129***
	(-0.76)	(-2.12)	(-4.08)	(-5.05)
Market to Book	0.002***	0.006***	0.012***	0.017***
	(3.64)	(4.06)	(3.57)	(2.89)
R&D Expenditures	0.808***	2.961***	5.466***	7.680***
	(25.63)	(27.51)	(23.02)	(17.36)
Constant	0.031***	0.195***	0.479***	0.931***
	(4.83)	(5.95)	(9.12)	(8.08)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.760	0.796	0.784	0.767
Observations	11965	10572	8215	6039

Panel C: Relationship between screening inefficiency of patent examiners and subsequent patent litigation

Dependent Variable	Subsequent #PatentLitigation			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
$\text{AvgExaminerFalseAcceptRate}_{\text{Adj}}$	0.229*	0.890**	1.659**	2.642**
	(1.70)	(2.57)	(2.55)	(2.35)
#ApplicationsReviewed	0.029**	0.116**	0.253***	0.497***
	(2.00)	(2.51)	(2.69)	(3.24)
#PatentsGranted	0.005	0.019	0.034	-0.025
	(0.36)	(0.48)	(0.40)	(-0.18)
Size	0.069***	0.270***	0.572***	0.925***
	(11.14)	(11.91)	(11.53)	(10.78)
Leverage	-0.284***	-1.227***	-2.695***	-4.449***
	(-5.63)	(-6.35)	(-6.24)	(-5.81)
Market to Book	0.018**	0.065**	0.153**	0.344***
	(2.42)	(2.37)	(2.49)	(3.01)
R&D Expenditures	0.777***	2.854***	5.322***	7.244***
	(8.31)	(9.22)	(7.94)	(5.99)
Constant	-0.226	-0.713	-2.397*	-4.487**
	(-0.96)	(-1.16)	(-1.94)	(-2.46)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
R^2	0.121	0.159	0.179	0.197
Observations	12204	11197	9228	7182

Table A.4: Relationship between screening inefficiency of patent examiners and subsequent exits of private firms (Robustness tests)

The sample consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *SuccessfulExit* is a dummy, which equals one if a given private firm has exited through an IPO or an M&A by the end of my sample period and zero otherwise. $AvgExaminerFalseAcceptRate_{Adj}$ is defined as the average (art-unit adjusted) false acceptance rates of examiners that are related to all granted applications for each firm in the past three years as described in Section 6.1.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by my machine learning algorithm. *#ApplicationsReviewed* and *#PatentsGranted* count the number of patent applications being reviewed and accepted for each firm in a given past three-year rolling window. *LnVCFinancingAmount* is the natural logarithm of the quarterly investment amount for each firm. *LnNumberFundInvested* is the natural logarithm of the quarterly number of invested funds for each firm. *TotalFundingToDate* is the natural logarithm of total funding each firm has received prior to a given quarter. *LnFirmAge* is the natural logarithm of firm age, which equals the current year minus the firm founding year plus one. Year fixed effects, industry (two-digit SIC code) fixed effects, and state fixed effects are included in all regressions. *t*-statistics are in parentheses. All regressions are OLS regressions with standard errors clustered at the state level. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	Subsequent SuccessfulExit			
	1 Quarter	1 Year	2 Years	3 Years
	(1)	(2)	(3)	(4)
$\text{AvgExaminerFalseAcceptRate}_{\text{Adj}}$	-0.012	-0.064**	-0.130**	-0.151***
	(-1.23)	(-2.36)	(-2.28)	(-2.79)
#PatentsGranted	0.006***	0.019***	0.035***	0.056***
	(2.85)	(4.07)	(3.22)	(3.19)
#ApplicationsReviewed	-0.002	-0.006	-0.010	-0.028
	(-1.04)	(-1.03)	(-0.79)	(-1.46)
InvestmentAmount	-0.005**	-0.005**	0.007*	0.004
	(-2.61)	(-2.34)	(1.80)	(0.53)
NumberFundInvested	0.005***	0.010***	0.010**	0.017***
	(4.03)	(3.79)	(2.55)	(3.51)
TotalFundingToDate	0.001	0.004***	0.008***	0.010***
	(1.29)	(4.33)	(3.12)	(3.53)
LnFirmAge	0.010***	0.020**	0.022	0.019
	(3.34)	(2.69)	(1.54)	(0.81)
Constant	-0.038***	-0.044	0.078	0.136
	(-7.09)	(-1.16)	(0.75)	(0.97)
Industry FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
R^2	0.012	0.026	0.040	0.050
Observations	13476	12543	10020	7412

ESSAY 2: The Role of Investor Attention in Seasoned Equity Offerings: Theory and Evidence

Thomas J. Chemmanur* Karen Simonyan† Yu Wang‡ Xiang Zheng§

Abstract

Models of seasoned equity offerings (SEOs) such as Myers and Majluf (1984) assume that all investors in the economy pay immediate attention to SEO announcements and the pricing of SEOs. In this paper, we analyze, theoretically and empirically, the implications of only a fraction of investors in the equity market paying immediate attention to SEO announcements. We first show theoretically that, in the above setting, the announcement effect of an SEO will be positively related to the fraction of investors paying attention to the announcement and that there will be a post-announcement stock-return drift that is negatively related to investor attention. In the second part of the paper, we test the above predictions using the media coverage of firms announcing SEOs as our main proxy for investor attention, and find evidence consistent with the above predictions. In the third part of the paper, we develop and test various hypotheses relating investor attention paid to an issuing firm to various SEO characteristics. We empirically show that institutional investor participation in SEOs, the post-SEO equity market valuation of firms, SEO underpricing, and SEO valuation are all positively related to investor attention. Lastly, we also use the number of SEC EDGAR file downloads as an alternative proxy for investor attention, and our findings are robust to this alternative investor attention measure. The results of our identification tests show that the above results are causal.

Keywords: Seasoned Equity Offerings; Limited Attention; Announcement Effect; Post-announcement Drift.

JEL classification: G23; G24; G32

*Professor of Finance and Hillenbrand Distinguished Fellow, Finance Department, Fulton Hall 336, Carroll School of Management, Boston College, Chestnut Hill, MA 02467, Tel: (617) 552-3980, Fax: (617) 552-0431, Email: chemmanu@bc.edu.

†Associate Professor of Finance, Sawyer Business School, Suffolk University, 8 Ashburton Place, Boston, MA 02108, Email: ksimonya@suffolk.edu.

‡Assistant Professor of Finance, Walton College of Business, University of Arkansas, AR 72701, Tel: (479) 575-7149, Email: ywang@walton.uark.edu.

§Ph.D. Candidate, Finance Department, Fulton 341, Carroll School of Management, Boston College, Chestnut Hill, MA 02467, Tel: (617) 552-2062, Email: zhengxr@bc.edu.

1 Introduction

Equity issues are an important source of external financing for corporations. Corresponding to their importance, there is a large theoretical and empirical literature in corporate finance that has studied various phenomena around seasoned equity offerings (SEOs). In particular, there is an important theoretical literature (see, e.g., Myers and Majluf (1984) or Giammarino and Lewis (1988)) that has attempted to explain the widely documented negative SEO announcement effect (see, e.g., Asquith and Mullins (1986) or Masulis and Korwar (1986)). The theoretical literature has focused on the asymmetric information facing the firm in the equity market as the main driving force to explain the negative announcement effect of an equity issue. Further, in models such as Myers and Majluf (1984), a crucial assumption is that all investors pay immediate attention to the equity issue announcement. The objective of this paper is to relax the above assumption, assuming instead that only a fraction of investors in the equity market pay attention to the SEO announcement, while the remaining fraction update their beliefs in a delayed manner after the announcement. We then analyze the consequences of such partial investor attention paid to SEOs theoretically and empirically in this paper.

In the first part of the paper, our focus is on theoretically analyzing a setting where an SEO conveys a negative signal to the equity market, but where, unlike in Myers and Majluf (1984), a fraction of investors do not pay immediate attention to the equity issue and update their beliefs about the firm announcing the equity issue only in a delayed manner. We show that, in the above setting with limited investor attention, the equity market underreacts to the SEO announcement (compared to the full attention setting). Further, we show that the announcement effect of an equity issue is increasing in investor attention (the fraction of investors paying attention to the SEO announcement). We then show that there will be a post-announcement stock return drift (driven by inattentive investors engaging in delayed updating of their beliefs after the SEO announcement). Additionally, this post-announcement stock return drift will be negatively related to the extent of investor attention paid to the SEO announcement. Finally, our model implies that both the abnormal stock return upon an SEO announcement and the post-announcement stock return drift will have predictive power for the subsequent operating performance of the firm.[1]

In the second part of the paper, we empirically test the implications of the above theory for the SEO announcement effect and the SEO post-announcement drift. We conduct the above empirical analyses using data on SEOs from 2000 to 2018. Following several papers in the IPO literature, we make use of the media coverage of an SEO firm over certain period of time (one to eight weeks) prior to its SEO announcement as a proxy for investor attention

[1] We would like to emphasize here that the goal of our model is not to argue for the negativity of the SEO issuance as a signal for the firm value (i.e., the negativity of the announcement effect upon SEO *per se*, as implied already by the model of Myers and Majluf (1984)), but rather, to study the split of the overall stock market reaction to SEO issuance between the immediate announcement effect and the subsequent post-SEO announcement stock return drift depending on the level of investor attention.

(see, e.g., Liu, Sherman, and Zhang (2014) or Bajo, Chemmanur, Simonyan, and Tehranian (2016)). Our baseline results from the above empirical analyses are as follows. First, the announcement effect of an equity issue is positively related to the investor attention paid to the SEO announcement: i.e., while the announcement effect is negative, it is larger in magnitude for SEOs with greater investor attention paid to the announcement. Second, the post-announcement stock return drift is decreasing in the investor attention paid to the SEO announcement: i.e., the post-announcement drift, while it is also negative, is decreasing in magnitude with greater investor attention.[2] Third, both the above variables (i.e., the SEO announcement effect and the post-announcement stock return drift) have predictive power for the future operating performance of an issuing firm (as confirmed by running multivariate regressions of post-SEO operating performance on the SEO announcement effect and on the post-announcement stock return drift).

We conduct two different identification tests to establish the causality of our baseline results. First, it may be argued that SEO firms with certain firm characteristics (omitted in our baseline regressions) may be more likely to attract investor attention, so that the baseline results we document above may be driven by such omitted variables rather than the investor attention received by the firm's SEO announcement. To rule out the above omitted variable problem, our first identification test analyzes the relationship between the "abnormal" media coverage received by the SEO firm prior to its SEO announcement (where abnormal media coverage is defined as the media coverage immediately prior to the SEO announcement minus the media coverage one year previously) and its SEO announcement effect and post-announcement stock return drift. Second, it may be argued that there may be some informational or other confounding events occurring before the SEO announcement that affects both the media coverage received by the firm prior to its SEO announcement and the relevant SEO characteristics (namely, the SEO announcement effect and the post-announcement stock return drift). To control for this type of endogeneity, we instrument for the investor attention received by the SEO firm immediately before the SEO announcement using the media coverage received by the firm one year before the SEO announcement. Using the above instrument, we conduct an instrumental variable (IV) analysis of the relationship between investor attention and the SEO announcement effect and also the relationship

[2] A recent SEO conducted by Moderna, Inc. demonstrates how the extent of investor attention paid to firms conducting SEOs splits the overall equity market reaction to the SEO announcement between the announcement day stock returns (SEO announcement effect) and post-announcement day stock returns (post-SEO stock return drift). Moderna announced a seasoned offering of 17.6 million shares at \$76 each on May 18, 2020 after reporting earlier that day eye-catching results on its early-stage human trials for a Covid-19 vaccine which attracted the largest news search since 2008 according to Google Trends. With such a high level of investor attention Moderna realized a negative 10.4% return on the day of its SEO announcement. Still, Moderna realized a significantly negative post-announcement stock return drift in the following ten days after the SEO announcement. Yet Moderna was able to make use of this high level of investor attention to boost its SEO valuation: according to NASDAQ news the company is trading at about 218 times of its estimated future sales.

between investor attention and the post-announcement stock return drift.[3] The results of the above two identification tests are also consistent with the predictions of our theory, thus confirming that the relationships we documented earlier in our baseline analyses are causal.

In the third part of the paper, we extend our analysis to study the relationship between investor attention and the pricing and characteristics of the SEO itself (in the U.S., the actual SEO occurs four to six weeks after the SEO announcement). We first develop testable hypotheses regarding the relationship between the investor attention received by a firm immediately before the actual SEO issue and the pricing of the SEO and other SEO characteristics. In order to develop these testable hypotheses, we start by assuming that, for institutional investors to participate in a firm's SEO, they not only need to receive information about various aspects of the firm from the SEO underwriter, but also to pay attention to or "recognize" this information. This last assumption is in the spirit of Merton's (1987) investor recognition or attention model, which assumes that an investor will incorporate a security into his portfolio only if he pays attention to (or acquires information about) that security by incurring a cost. While Merton (1987) posits several possible sources of this "attention" or "recognition" cost, he views this cost mainly as arising from the cost of investors becoming aware of (or familiar with) a firm: in his setting, investors consider investing only in the stock of firms with which they have a certain level of familiarity. In a similar vein, we can assume that institutional and other investors consider investing in only the stocks of those SEO firms which they have become familiar with by incurring an "attention cost". Then, if a larger number of institutions have paid attention to a firm's SEO, we would expect to find, *ceteris paribus*, a larger number of institutional investors investing in the equity of the SEO firm. Further, if the demand for the SEO firm's equity from institutional investors is higher for SEOs receiving greater investor attention, we expect the market clearing price of the equity of such firms to be higher (for a given supply of shares offered in the SEO). We therefore expect to find a positive relationship between investor attention and SEO firm market valuations immediately post-SEO. As we discuss in more detail in Subsection 7.1, if SEO underpricing is unrelated to investor attention (e.g., driven only by considerations of information extraction as argued by Benveniste and Spindt (1989)), then we expect to find a positive relationship between investor attention and firm valuation at the SEO offer price as well. If, however, SEO underpricing is itself positively related to investor attention (as implied by the theoretical SEO model of Chemmanur and Jiao (2011) or by the IPO model of Liu, Lu, Sherman, and Zhang (2019)), then the relationship between investor attention and firm valuation at the SEO offer price will turn ambiguous.

We test the above hypotheses using the media coverage received by the firm prior to the actual equity issue (i.e., after the SEO announcement but before the pricing of the SEO)

[3]It should be noted that our IV analysis using the media coverage received by the firm one year before the SEO announcement as the instrument rules out the possibility that our results are driven by asymmetric information rather than investor attention, since it is unlikely that any private information held by firm insiders is so long-lived (i.e., having a one year horizon).

as a proxy for investor attention. First, we find that the institutional investor participation in an SEO is increasing in the investor attention received by the SEO firm. This result also holds after we control for SEO underpricing. Second, we find that the post-SEO secondary market valuation of the SEO firm is increasing in investor attention. This result holds regardless of whether the market valuation proxy is constructed using the SEO issue day closing price, or using the stock price one quarter after the completion of the SEO. Third, we find that the underpricing of an SEO (as measured by the stock return from the SEO offer price to the closing price on the SEO issue day) is positively related to the investor attention received by the SEO firm. Fourth, we find that firm valuation at the SEO offer price is also positively related to the investor attention received by the SEO firm. We conduct two identification tests: first, we analyze the relationship between the "abnormal" media coverage received by the SEO firm and various SEO characteristics; and second, we conduct an IV analysis using the media coverage received by the SEO firm one year prior to the SEO announcement as an instrument for the media coverage received by the firm immediately before the SEO. The above two identification tests establish that the baseline results we discussed earlier are causal.

We also rerun all of our baseline regressions using the pre-SEO EDGAR file downloads of SEO firms as an alternative proxy for investor attention as a robustness check. We construct this alternative measure of investor attention for each SEO firm by counting the number of downloads for 10-K, 10-Q, and 8-K filings of the SEO firm over a certain period of time (namely, 1 week, 2 weeks, 1 month, and 2 months) prior to the SEO announcement date or prior to the SEO issue date. The results using this alternative measure of investor attention are very similar to those using media coverage, suggesting that our findings in this paper are robust to alternative measures of investor attention, and thereby providing strong support for our testable hypotheses.

This paper has several important takeaways. It shows that investor attention is an important factor affecting equity market's reaction to the announcements of SEOs. Because not all investors pay immediate attention to SEO announcements, not all information about SEOs gets incorporated immediately upon SEO announcements into issuing firms' stock prices, but it takes a longer time as more and more investors pay delayed attention to the SEOs. Further, the more attention investors pay to issuing firms the greater the SEO announcement effect and the smaller the post-SEO announcement drift. This has important implications for issuing firms' shareholders who should expect to realize negative returns not only upon their firms' SEO announcements but also for a significant period of time after that depending on the extent of investor attention paid to their firms. Our results also show that the extent of investor attention paid to SEO firms increases the SEO valuation received by such firms. This demonstrates that firms considering SEOs may try to draw investors' attention towards their firms prior to announcing their plans to the equity market or they may time their SEOs to be conducted during periods when their firms enjoy higher levels of investor attention. Finally, investor attention affects not only SEO valuation, but also post-SEO secondary market valuation of issuing firms' by increasing the participation of institutional investors in the SEOs of such firms. This has important implications for issuing firms as well as equity market investors.

The rest of this paper is organized as follows. Section 2 discusses how our paper is related to the existing literature and describes its contribution relative to this literature. Section 3 presents the set-up of our theoretical analysis of the relationship between investor attention, the announcement effect of an SEO, and the post-announcement stock return drift, and Section 4 develops results and describes the testable implications of our theoretical model. Section 5 describes our data and discusses our proxies for investor attention. Section 6 describes our empirical tests and results on the relationship between the investor attention received by a firm prior to an SEO and the SEO announcement effect and the post-announcement stock return drift as well as our empirical analysis of the predictive power of the SEO announcement effect and post-SEO stock return drift for post-SEO operating performance. Section 7 develops testable hypotheses for the relationship between the investor attention received by an SEO firm and various SEO characteristics and presents our empirical tests (and results) of the above hypotheses. Section 9 concludes. Online Appendix A.1 gives a list of constants used in various propositions and proofs. The proofs of all propositions are confined to Online Appendix A.2. Online Appendix A.3 presents some additional empirical tests not included in the main text due to space limitations.

2 Relationship to the Existing Literature and Contribution

Our paper is related to several strands in the literature. The first strand is the theoretical and empirical literature on the stock market reaction to SEO announcements: see, e.g., Myers and Majluf (1984) or Giammarino and Lewis (1988). The theoretical model of Myers and Majluf (1984) suggests that equity issues will have a negative announcement effect in a setting of asymmetric information, since they convey that insiders of a firm announcing an equity issue have less favorable private information about their firm's future prospects. Since, in Myers and Majluf (1984), all investors pay immediate attention to the equity issue announcement, there will be no post-announcement drift in their setting. Thus, our model can be viewed as building on the Myers and Majluf (1984) setting where the announcement of an equity issue conveys a negative signal to the equity market, but where, unlike in Myers and Majluf (1984), a fraction of investors do not pay immediate attention to the equity issue and update their beliefs about the firm announcing an equity issue only in a delayed manner, thereby giving rise to a significant post-announcement stock return drift.[4]

[4]There is also a large empirical literature documenting the negative stock market reaction to the announcement of equity issues: see, e.g., Asquith and Mullins (1986) or Masulis and Korwar (1986). Asquith and Mullins (1986) document a significant negative SEO announcement effect and find that the extent of price reduction is negatively related to the size of the equity issue.

The second strand is the theoretical and empirical literature on the pricing of SEOs as well as the discounting and underpricing of SEOs. Two theoretical models of the pricing of SEOs are those of Chemmanur and Jiao (2011) and Gerard and Nanda (1993). These papers develop theoretical rationales for the pricing of SEOs, and, in particular, for SEO discounts and underpricing based on asymmetric information (albeit driven by different motivations). Unlike the above papers, our focus in the third part of this paper is on empirically analyzing the implications of investor attention for SEO underpricing, post-SEO firm valuation, and institutional investor participation in SEOs.

There is also a large empirical literature on the underpricing of SEOs. Since Smith (1977), who first empirically documented a significant SEO underpricing, the academic literature has offered various explanations for this phenomenon. Loderer, Sheehan, and Kadlec (1991) document a more significant SEO underpricing for stocks listed on the Nasdaq than stocks listed on other exchanges such as NYSE and Amex. Corwin (2003) studies the determinants of SEO underpricing such as offer size, uncertainty of firm value, the magnitude of preoffer returns, price rounding, and the pricing relative to the bid quote. Altınkılıç and Hansen (2003) decompose SEO discounting into a predictable component and a surprise component, and argue that the surprise component is used by underwriters as a channel to release additional information to investors. Gao and Ritter (2010) study the effect of various offer methods on SEO characteristics such as discount and underpricing. Gibson, Safieddine, and Sonti (2004) show that SEO firms with the greatest increase in institutional investment around the issue date significantly outperform those with the greatest decrease in institutional investment. Chemmanur, He, and Hu (2009) analyze the relationship between institutional trading around SEOs and various SEO characteristics, and conclude that their findings are consistent with institutions being able to produce information about the firm making the SEO. Huang and Zhang (2011) document a negative relationship between the number of managing underwriters and SEO discount. Gustafson (2018) documents a higher offer price and lower post-issue return for over-night SEO offerings than non-overnight offerings. Unlike the above empirical papers, our focus in the second part of this paper is on the relationship between investor attention paid to a firm making an SEO and SEO underpricing, immediate post-SEO firm valuation, and the participation of institutional investors in the SEO which has not been analyzed before in the literature.[5]

The third strand is the theoretical literature on limited attention. Hirshleifer and Teoh (2003) use a static limited-attention model where only a fraction of investors pay attention to public information immediately and correctly to study the effects of firms' different presentations of financial disclosure and reporting on market prices. Hirshleifer, Lim, and

[5] Pinto-Gutiérrez (2018) empirically analyzes the relationship between the media coverage received by an SEO firm prior to its offering and the SEO discount, and also the relationship between the above media coverage and the abnormal stock return during the three-day window around the day of the equity issue (not the SEO announcement day). However, the above paper does not analyze any of the relationships that we study in this paper.

Teoh (2011) use a related model to analyze the interpretation of different earnings components and investors' underreaction to earnings announcements and overreaction to accruals. Our model builds on the above two static models by introducing random supply shocks on trading dates, so that we are able to explicitly characterize the post-announcement drift following SEO announcements.[6] The broader "investor recognition" or "investor attention" literature that builds on Merton (1987)'s model is also related to our paper: we build on this literature to develop testable hypotheses for our empirical analysis of the relationship between investor attention and various SEO characteristics such as SEO underpricing and post-SEO equity valuation. One paper from this literature that is related to ours is Liu, Lu, Sherman, and Zhang (2019) who develop a model, in the context of IPO, in which underwriters attract potential investors to an IPO by offering underpriced shares. In their setting, IPO underpricing is a way of compensating investors for their cost of paying attention to the firm going public.

The fourth and final strand in the literature our paper is related to is the empirical literature on investor attention in the context of IPOs. Bajo, Chemmanur, Simonyan, and Tehranian (2016) study two functions of IPO underwriters, information dissemination and information extraction, within IPO underwriter networks and find that higher investor attention leads to more favorable IPO characteristics, using pre-IPO media coverage as a proxy for investor attention. Chemmanur, Krishnan, and Yu (2018) analyze the role of backing by venture capitalists (VCs) in enhancing the investor attention paid to an IPO firm and the role of this enhanced investor attention in offering favorable IPO characteristics. Da, Engelberg, and Gao (2011) use Google Search Volume Index data to proxy for retail investor attention and document a positive relationship between the retail investor attention and the initial return of IPOs and a negative relationship between the retail investor attention and the long-run stock return performance after IPO. Liu, Sherman, and Zhang (2014) use media coverage as a proxy for investor attention to document a positive relationship between investor attention and post-IPO long-run stock return performance. Cook, Kieschnick, and Van Ness (2006), document the positive effects of underwriters' promotional efforts in IPOs on IPO offer price revision, and IPO initial return.[7] Unlike the above literature, that has focused on the effect of investor attention in IPOs, the focus of our paper is the role of investor attention in SEOs. In particular, we are the first to analyze the effect of pre-SEO investor attention on the stock market reaction to SEO announcements, and also the first to analyze the effect of investor attention on various SEO characteristics.

[6]Peng (2005) applies the setting of limited attention to regimes such as the learning process of investors; Peng and Xiong (2006) applies such a setting to investors' category learning and consequent return comovement when investors also suffer from overconfidence.

[7]The broader literature on the role of media coverage in the financial market is also distantly related. Engelberg and Parsons (2011) establish the causal effect of media coverage on investor trading by studying the trading in local markets following local papers reporting the earnings announcements of S&P 500 firms. Fang and Peress (2009) document a negative relationship between media coverage and stock returns, consistent with the explanation that media coverage diminishes information asymmetry and thus decreases the expected return of stocks in equilibrium.

3 Model Setup

We develop a discrete-time dynamic model to study the relationship between investor attention paid to SEO announcements and the SEO announcement effects and SEO post-announcement drifts. The model builds upon the SEO model of Myers and Majluf (1984) and the static limited attention model of Hirshleifer and Teoh (2003). By introducing a supply shock into the model, we are able to explicitly represent the drift and study the corresponding comparative statics on both the announcement effect and the post-announcement drift.

3.1 Timeline

There are four dates in the model (Figure 1): $t = 0, 1, 2, 3$.

t=0	t=1	t=2	t=3
Investors form their initial portfolios based on their prior beliefs of the firm.	An upcoming SEO is announced. Attentive investors pay attention to the announcement, but inattentive investors do not. Investors trade.	Inattentive investors notice that they missed the SEO announcement at t=1 and correct their beliefs in a delayed manner at t=2. Investors trade again.	All payoffs are realized.

Figure 1: **Timeline of Model**

At $t = 0$, investors are endowed with homogeneous wealth (or equal shares of the asset). There is no trading on this day. All investors hold homogeneous prior beliefs on the payoff of the asset. At $t = 1$, an upcoming SEO is announced. Attentive investors update their beliefs conditional on the announcement; inattentive investors do not update their beliefs (still hold the prior beliefs). Investors trade to rebalance their portfolios. At $t = 2$, inattentive investors realize that they missed the SEO announcement (at $t = 1$) and correct their beliefs in a delayed manner. There is no change in attentive investors' beliefs at $t = 2$. Investors then trade again. At $t = 3$, asset payoff is realized and there is no further trading.

3.2 Assets and the SEO Announcement

There are two assets in the market: a risky asset issued by the SEO firm and a riskfree asset.

Riskfree asset. The riskfree asset offers a net return of r, which is normalized to 0.[8] The riskfree asset has unlimited supply.

Risky asset. The SEO firm issues a risky asset, which can be naturally interpreted as a stock of the firm or, equivalently, as the liquidation value of the firm in the end. The

[8] The results of the model are qualitatively the same if we allow r to be a nonzero constant. So without loss of generality, we set r to be equal to zero to keep the model simple in exhibition.

terminal payoff of the risky asset is represented by a random variable f:

$$f = \mu + z, \text{ where } \mu = E(f) \text{ and } z \sim N(0, \sigma_0^2). \tag{1}$$

The unconditional expected supply of the risky asset is $\bar{x}$ and there is an independent supply shock $x_t \sim N(0, \sigma_x^2)$ at each period of $t = 1$ and $t = 2$, i.e., the aggregate supply of the risky asset at t is $\bar{x} + \sum_{s=1}^{t} x_s$ where $x_s \sim N(0, \sigma_x^2)$, for $t = 1, 2$.[9]

The SEO announcement. On date $t = 1$, a public signal $e_1 = z + \epsilon_1$ is revealed by the SEO announcement, where $\epsilon_1 \sim N(0, \sigma_e^2)$.[10] The error term ϵ_1 is independent of all other shocks in the model. We can interpret the public signal e_1 as an exogenous signal conveyed by the SEO (equity issue) announcement to the stock market about the firm's future cash flows.[11]

3.3 Market Participants

The continuum of investors consists of two types of investors: attentive investors ("type-a") and inattentive investors ("type-u"). The total mass of investors is 1; a fraction of f^a are attentive, and the rest, $f^u = 1 - f^a$, are inattentive. We use i as the generic index for "type", i.e., $i = a$ for attentive investors and $i = u$ for inattentive investors.

Attentive investors (indexed by type a). An attentive investor updates his/her beliefs immediately on the SEO announcement at date $t = 1$. Since no investor in the market observes any private signal, the equilibrium prices do not contain additional information about the terminal payoff of the risky asset. However, the equilibrium price does reflect information about the current supply shock. Therefore, attentive investors always keep track of contemporaneous supply shocks as they pay attention to all public signals immediately.

Inattentive investors (indexed by type u). Because of limited attention, inattentive investors do not pay immediate attention to the public signal e_1 made available to them by the SEO announcement at $t = 1$ and delay their belief updating on e_1 till $t = 2$. Also because of their limited attention, they are unaware of their delay even though they may notice the change in equilibrium prices from S_0 to S_1, hence they are not able to figure

[9] The supply shock is not observable directly. However, since there is no private signal in the model, investors may be able to figure out the total supply shock from the equilibrium price if they do know (pay attention to) all public signals available contemporaneously (e.g., attentive investors at $t = 1$). More will be discussed in the next subsection.

[10] Our objective in this paper is not to endogenously show that the expected announcement effect of an SEO is negative, which has already been shown in the theoretical literature by Myers and Majluf (1984) or Giammarino and Lewis (1988). Given this, we wish to take the signal conveyed by an SEO announcement, e_1, as exogenous, and theoretically analyze, for the first time in the literature, how the equity market reaction to this signal is modified if a fraction of the investors do not pay immediate attention to the signal conveyed by the SEO announcement.

[11] We focus primarily on the case where $e_1 < 0$ in our analysis of the SEO announcement effect and post-SEO announcement drift. As documented extensively by empirical literature, the average announcement effect of an equity issue is negative: see, e.g., Asquith and Mullins (1986). Theoretical models of equity issues such as Myers and Majluf (1984) predict a negative announcement effect for an equity issue as well.

out the supply shock x_1 right away. Instead, on a later date, $t = 2$, they notice that they missed the SEO announcement (or, equivalently, they are finally able to evaluate the effect of SEO on the firm value) and update their beliefs based on e_1 in a delayed manner and rebalance their portfolios. This assumption is similar in spirit to the assumptions made by Hirshleifer and Teoh (2003) and Hirshleifer, Lim, and Teoh (2011).[12,13]

Utility. All investors hold the constant-absolute-risk-aversion (CARA) utility with a common risk aversion parameter ρ. On each trading date ($t = 0, 1, 2$), they all optimally choose their demands $\{D_t^i\}_{i\in\{a,u\}}$ of the risky asset to maximize their personal expected utilities on terminal wealth,

$$\max_{D_t^i} E_t^i(-\exp[-\rho W_3^i]), \text{ for } i \in \{a, u\} \text{ and } t = 0, 1, 2, \tag{2}$$

subject to the following budget constraints

$$W_{t+1}^i = W_t^i + D_t^i(S_{t+1} - S_t), \text{ for } t = 0, 1, \tag{3}$$

$$W_3^i = W_2^i + D_2^i(f - S_2). \tag{4}$$

4 Equilibrium and Results

We calculate the updating of beliefs by moving forward as more information arrives on each subsequent date. In contrast, we solve the equilibrium prices and demands by moving backwards, since investors' demand depends on their expectation of the capital gain in each subsequent period.

4.1 Bayesian Updating of Beliefs

The information set for an investor of type i at time t is denoted by $\mathcal{F}_t^i$.

At $t = 0$, all investors hold the prior belief: $f = \mu + z$, where μ is the unconditional expectation of f and $z \sim N(0, \sigma_0^2)$. Since μ is a constant, the updating of beliefs occurs only on the random component z in later periods.

At $t = 1$, an attentive investor, type a, pays attention to the SEO announcement e_1, and has an information set $\mathcal{F}_1^a = \{e_1\}$. The posterior belief is

$$z|_{\mathcal{F}_1^a} \sim N(\hat{z}_1^a, (\sigma_1^a)^2), \text{ where } \hat{z}_1^a = (\sigma_1^a)^2\sigma_e^{-2}e_1 \text{ and } (\sigma_1^a)^{-2} = \sigma_0^{-2} + \sigma_e^{-2}. \tag{5}$$

[12] We can also interpret the inattention to the SEO announcement as the inability to evaluate the effect of the announcement immediately. Since the SEO announcement may occur significantly ahead of the actual offering, investors may wait for more updates about the firm performance (and thus to evaluate the firm stock) before the actual offering to make their trading decisions.

[13] Once inattentive investors pay attention to e_1 and understand the components in the equilibrium price S_1, they are able to figure out the supply shock x_1 retroactively at $t = 2$ and thus they learn about x_2 by observing the equilibrium price S_2.

An inattentive investor, type u, does not pay attention immediately to the SEO announcement e_1, and hence still holds the prior belief, i.e.,

$$z|_{\mathcal{F}_1^u} \sim N(\hat{z}_1^u, (\sigma_1^u)^2), \text{ where } \hat{z}_1^u = 0 \text{ and } \sigma_1^u = \sigma_0. \tag{6}$$

At $t = 2$, an attentive investor, type a, has no change in his/her information set, $\mathcal{F}_2^a = \mathcal{F}_1^a = \{e_1\}$ and therefore has no change in belief, i.e.,

$$z|_{\mathcal{F}_2^a} \sim N(\hat{z}_2^a, (\sigma_2^a)^2), \text{ where } \hat{z}_2^a = (\sigma_2^a)^2\sigma_e^{-2}e_1 \text{ and } (\sigma_2^a)^{-2} = \sigma_0^{-2} + \sigma_e^{-2}. \tag{7}$$

An inattentive investor, type u, now notices the upcoming SEO, i.e., the public signal e_1, (or is finally able to interpret the effect of the SEO announcement on the firm's fundamental value f), so the information set is now $\mathcal{F}_2^u = \{e_1\}$. The posterior belief of a type-u investor is

$$z|_{\mathcal{F}_2^u} \sim N(\hat{z}_2^u, (\sigma_2^u)^2), \text{ where } \hat{z}_2^u = (\sigma_2^u)^2\sigma_e^{-2}e_1 \text{ and } (\sigma_2^u)^{-2} = \sigma_0^{-2} + \sigma_e^{-2}. \tag{8}$$

Therefore, investors of both types have same posterior beliefs, i.e., $\hat{z}_2^i$ and σ_2^i are both independent of $i = a$ or u, and hence can be denoted by $\hat{z}_2$ and σ_2 respectively for conciseness and without ambiguity.[14]

4.2 Equilibrium Prices and Demands

On each trading date ($t = 0, 1, 2$), given their updated beliefs of z, investors decide their optimal demands $\{D_t^i\}_{i\in\{a,u\}}$ for the risky asset to maximize their expected CARA utilities of terminal wealth $E_t^i(-\exp[-\rho W_3^i])$. At each t, the equilibrium price S_t clears the market, i.e.,[15]

$$\int D_t^i di = f^a D_t^a + f^u D_t^u = \bar{x} + \sum_{s=1}^{t} x_s, \text{ for } t = 0, 1, 2. \tag{9}$$

Proposition 1 (The Equilibrium Prices and Investors' Optimal Demands)

(i) For $t = 0, 1, 2$, the equilibrium price S_t has the following expressions respectively:

$$\begin{aligned} S_2 &= \mu + \hat{z}_2 - \rho\sigma_2^2(\bar{x} + x_1 + x_2), && (10)\\ S_1 &= \mu + \frac{A_a}{A_a + A_u}\hat{z}_1^a - \rho(B_0\bar{x} + B_1 x_1), && (11)\\ S_0 &= \mu - \rho\frac{Q_a + Q_u + 1}{P_a + P_u}\bar{x}, && (12) \end{aligned}$$

where the definitions of all constants A_a, A_u, B_0, B_1, P_a, P_u, Q_a, and Q_u are listed in Appendix A.1.

[14]Notice that although all investors have the same posterior belief at $t = 2$, their conditional expectations of $\hat{z}_2$ at $t = 1$, i.e., $E_1^i[\hat{z}_2]$ for $i \in \{a, u\}$, are different, because the SEO announcement e_1 is in $\mathcal{F}_1^a$ and hence deterministic for attentive investors at $t = 1$ but not in $\mathcal{F}_1^u$ and hence still random for inattentive investors at $t = 1$.

[15]Here we apply the convention that $\sum_{s=M}^{N} x_s = 0$ for any integers $N < M$.

(ii) For $t = 0, 1, 2$, the optimal demands of the risky asset by investors of type $i \in \{a, u\}$ are respectively

$$\begin{aligned}
D_2^i &= \rho^{-1}\sigma_2^{-2}(\mu + \hat{z}_2 - S_2) \text{ for } i \in \{a, u\}, && (13)\\
D_1^a &= \rho^{-1}\frac{A_a}{f^a}(\mu + \hat{z}_1^a - S_1) - [\frac{A_a}{f^a}(\sigma_1^a)^2 - 1](\bar{x} + x_1), && (14)\\
D_1^u &= \rho^{-1}\frac{A_u}{f^u}(\mu - S_1) - [\frac{A_u}{f^u}\sigma_0^2 - 1]\bar{x}, && (15)\\
D_0^a &= \rho^{-1}\frac{P_a}{f^a}(\mu - S_0) - \frac{Q_a}{f^a}\bar{x}, && (16)\\
D_0^u &= \rho^{-1}\frac{P_u}{f^u}(\mu - S_0) - \frac{Q_u}{f^u}\bar{x}, && (17)
\end{aligned}$$

where the definitions of all constants are listed in Appendix A.1.

The equilibrium prices on all trading dates are in the form of "μ+(investors' belief on z)-(a term of $\bar{x}$ and x_s for $x \leq t$)". Generally speaking, if investors interpret the public signal from the announcement at $t = 1$ as good news on the terminal firm value, i.e., $e_1 > 0$, then investors modify their beliefs on z upward and thus the equilibrium prices increase; if, however, the announcement is interpreted as bad news on the terminal firm value, i.e., $e_1 < 0$, then investors modify their beliefs on z downward and thus the equilibrium prices decrease. The term containing $\bar{x}$ and $x_s(x \leq t)$ represents a compensation (risk premium) for holding the risky asset by investors.

On each date, the optimal demand of risky asset by an investor increases with the investor's conditional expectation of z. Observe that investors' demands at $t = 2$ are homogeneous regardless of their attention type. This is because at $t = 2$ both attentive and inattentive investors have their beliefs updated correctly on the SEO announcement e_1, thus they all have homogeneous beliefs and hence homogeneous demands. In contrast, the demands at $t = 1$ and $t = 0$ depend on the attention type since only attentive investors pay attention to the SEO announcement e_1 immediately at $t = 1$ and therefore hold different beliefs from inattentive investors.

4.3 SEO Announcement Effect and Post-SEO Announcement Drift

In this subsection, we study the abnormal stock returns (announcement effects) at $t = 1$ and the corresponding post-announcement stock return drifts from $t = 1$ to $t = 2$. This is done by looking at the differences in the equilibrium prices of the risky asset across time. Because the supply shocks have a mean of zero and the analysis of announcement effects and post-announcement drifts is unrelated to risk premium, without loss of generality, we follow Hirshleifer and Teoh (2003) and set $\bar{x} = x_t = 0$ (for $t = 1, 2$) in this subsection for our analysis on the announcement effect and post-announcement drift around SEOs.

By taking the difference between (11) and (12), we rewrite the price change of the risky asset from $t = 0$ to $t = 1$ as follows

$$S_1 - S_0 = \frac{A_a}{A_a + A_u} \frac{\sigma_0^{-2} + \sigma_e^{-2}}{\sigma_e^{-2}} e_1 - \rho[(B_0 - \frac{Q_a + Q_u + 1}{P_a + P_u})\bar{x} - B_1 x_1] \tag{18}$$

The first term represents the average change in investors' beliefs (from 0 to $\hat{z}_1^a$ by attentive investors, diluted by the zero change in inattentive investors' beliefs) and the second term represents the change in risk premium because of both uncertainty resolution and supply shock. Since the supply shock x_1 is on average zero and the change in risk premium is not the focus of our study, we silence the terms containing $\bar{x}$ and x_1 by setting both $\bar{x}$ and x_1 equal to zero, and focus on the first component to analyze the effect of investor attention on SEO announcement effect.

Proposition 2 (The Announcement Effect of an SEO)

(i) Let the public signal $e_1 < 0$. Then, the abnormal stock return upon the announcement of an SEO will be negative with its magnitude increasing in the realization e_1 of the announcement, given by:

$$\frac{A_a}{A_a + A_u} \frac{\sigma_0^{-2} + \sigma_e^{-2}}{\sigma_e^{-2}} e_1 < 0, \tag{19}$$

where the constants A_a and A_u are both positive and increasing functions of f^a and f^u, respectively (defined in Appendix A.1).

(ii) For any given public signal e_1 from the SEO announcement, the magnitude of the abnormal stock return upon announcement will be increasing in the proportion of investors who are attentive to the announcement.

Intuitively, as more investors pay immediate attention to the public signal revealed by the SEO announcement, i.e., the higher the fraction f^a of attentive investors in the equity market, the greater the immediate updating of beliefs reflecting the information contained in the signal e_1. This means that the equilibrium price S_1 reflects a larger proportion of the information contained in e_1, thus creating an announcement effect of a larger magnitude.

We now turn to calculating the post-SEO announcement stock return drift as a function of investor attention. When fewer investors delay their belief updating till $t = 2$ (i.e., the larger the fraction f^a of attentive investors in the market), the smaller the proportion of information reflected in the post-SEO announcement drift. We can calculate the price change given by $S_2 - S_1$, by taking the difference between (10) and (11):

$$S_2 - S_1 = \frac{A_u}{A_a + A_u} \frac{\sigma_0^{-2} + \sigma_e^{-2}}{\sigma_e^{-2}} e_1 - \rho[(\sigma_2^2 - B_0)\bar{x} + (\sigma_2^2 - B_1)x_1 + \sigma_2^2 x_2]. \tag{20}$$

The price change $S_2 - S_1$ consists of two parts: the first part is the delayed belief update by inattentive investors with respect to the public signal e_1 at SEO; the second part is the change in risk premium as a combination of uncertainty resolution over time and the additional supply shock realized contemporaneously. Since the supply shocks x_1 and x_2 are on average zero and the change in risk premium is not the focus of our study, we follow Hirshleifer and Teoh (2003) and silence the terms on $\bar{x}$ and x_t by setting $\bar{x} = x_t = 0$ (for $t = 1, 2$), and focus on the first component of the price change $S_2 - S_1$ to analyze the effect of investor attention on the post-SEO announcement stock return drift.

Proposition 3 (Post-SEO Announcement Stock Return Drift)

(i) Let the public signal $e_1 < 0$. Then, there will be a negative post-SEO announcement stock return drift given by

$$\frac{A_u}{A_a + A_u} \frac{\sigma_0^{-2} + \sigma_e^{-2}}{\sigma_e^{-2}} e_1 < 0, \tag{21}$$

where the constants A_a and A_u are both positive and increasing functions of f^a and f^u respectively (defined in Appendix A.1).

(ii) For any given public signal e_1 at the SEO announcement, the magnitude of the post-SEO announcement stock return drift decreases as the proportion of attentive investors f^a increases.

4.4 Implications and Testable Hypotheses

Our model generates several testable implications and we develop corresponding testable hypotheses for our empirical analysis.

1. *Relationship between a proxy for investor attention and the abnormal stock return upon SEO announcements:* Proposition 2 of our model predicts a positive relationship between the extent of investor attention paid to a given SEO announcement and the magnitude of the abnormal stock return upon that announcement. Since the abnormal stock return is on average negative upon SEO announcements, in the spirit of Myers and Majluf (1984) and also as shown in the next section of this paper, our model predicts a more negative abnormal stock return when more investor attention is paid to the SEO announcement. This is the first hypothesis that we test here (**H1**). We use a proxy for investor attention (namely, media coverage) to test this hypothesis.

2. *Relationship between a proxy for investor attention and the post-announcement drift following SEO announcements:* Proposition 3 of our model predicts a negative relationship between the extent of investor attention paid to a given SEO announcement and the magnitude of the post-announcement stock return drift. Since our model predicts that the post-announcement drift overall will be negative, we expect a less negative drift when more investor attention is paid to the SEO announcement. This is the second hypothesis that we test here (**H2**). We use a proxy for investor attention (namely, media coverage) to test the above hypothesis.

3. *The predictive power of SEO announcement effect and SEO post-announcement drift for long-term firm performance:* as shown in Propositions 2 and 3, both the abnormal

stock return upon the SEO announcement and the subsequent post-announcement drift are positively correlated with the information released at the SEO announcement, about the firm's future cash flows. In other words, not all information conveyed by the SEO announcement about the issuing firm is incorporated into the firm's price immediately upon the announcement but is split between the SEO announcement effect and the post-announcement drift. Therefore, we expect that both the SEO announcement effect and the post-announcement drift to have a predictive power regarding the subsequent operating performance, that is, both the abnormal stock return upon the SEO announcement and the post-announcement drift are positively correlated with the long-term firm cash flow, and, more broadly, with the post-SEO operating performance of the firm. This is the third hypothesis that we test here (**H3**). We use multiple proxies for the firm operating performance (e.g., ROA, and cash flow) to test this hypothesis.

5 Data and Sample Selection

We collect data on SEOs from the Securities Data Company (SDC) Global New Issues database. We first obtain the list of all SEOs conducted in the U.S. from 2000 to 2018 and then select only offerings of common shares (thus excluding all other types of offerings such as real estate investment trusts, units, rights, spin-offs, American Depository Receipts, etc.) from this list. We collect data on SEO firms' media coverage from RavenPack News Analytics (Dow Jones Edition). RavenPack covers news items from Dow Jones Newswires, regional editions of Wall Street Journal, Barron's, and MarketWatch starting from January 1, 2000 (thus the starting date of our sample period is determined by the availability of media coverage data collected from RavenPack). We obtain accounting data from Compustat; stock return data from the Center for Research in Security Prices (CRSP); analyst forecast data from the Institutional Brokers' Estimation System (IBES) database; and institutional holdings data from Thomson Reuters' institutional holdings (13F) database.

5.1 Measures of Investor Attention and Summary Statistics

We use the pre-SEO media coverage of firms conducting SEOs as our proxy for the amount of attention paid by market investors to SEO firms. We construct our measures of investor attention for each SEO firm by counting the number of news items mentioning the firm over a certain period of time (namely, 1 week, 2 weeks, 1 month, and 2 months) prior to the SEO announcement date (*NumNewsFile*) or prior to the SEO issue date (*NumNewsIss*). For example, *NumNewsFile [-60:-1]* and *NumNewsIss [-60:-1]* are the numbers of news items covering an SEO firm over a two-month period (60-day period from day -60 to day -1) prior to the SEO announcement date and prior to the SEO issue date, respectively. We also construct abnormal investor attention measures (*AbnNumNewsFile* and *AbnNumNewsIss*) as the difference between the media coverage of an SEO firm immediately prior to its

SEO as described above and the media coverage of the same firm exactly one year before its SEO announcement date. In other words, e.g., *AbnNumNewsFile [-60:-1]* is equal to *NumNewsFile [-60:-1]* minus *PriorYrNumNewsFile [-60:-1]*, where *PriorYrNumNewsFile [-60:-1]* is the number of news items covering an SEO firm over a two-month period ending one year prior to its SEO announcement date.

5.2 Summary Statistics of SEO Characteristics, Investor Attention, and Other Control Variables

Panel A of Table 1 reports the summary statistics of our investor attention measures for SEO firms in our sample. The average numbers of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO announcement dates are 2.15, 4.04, 8.32, and 15.62, respectively; while the average numbers of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO issue dates are 3.02, 5.10, 10.13, and 19.60, respectively. Further, the mean abnormal media coverage proxies measuring abnormal investor attention both prior to the SEO announcement date and prior to the SEO issue date are positive, suggesting that SEO firms receive somewhat more investor attention prior to their SEOs. For example, the mean abnormal numbers of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to the SEO announcement date are 0.49, 0.87, 1.82, and 3.21, respectively; while the mean abnormal numbers of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to the SEO issue date are 1.62, 2.28, 3.50, and 5.62, respectively.

Panel B of Table 1 reports the summary statistics of various SEO firm characteristics as well as certain SEO characteristics.[16] For example, the average book value of SEO firms' assets at the end of the fiscal year prior to the SEO announcement is $505 million, the mean return on assets (ROA measured at the end of the first post-announcement fiscal quarter) is -3.74%, the mean industry-adjusted Q ratio (measured using the issue day closing price) is -0.041, the mean SEO underpricing (the percentage difference between the issue day closing price and the SEO offer price) is 3.6%, the mean midpoint of initial filing price range is $24.1, and the mean number of institutional investors holding SEO firm shares at the end of the first post-issue fiscal quarter is 132.

6 Investor Attention and the Market Reaction to SEO Announcements

In this section, we present our empirical findings on how the extent of investor attention paid to firms prior to their SEOs affects the market reaction to the announcements of these SEOs.

[16]We winsorize all firm and SEO characteristics variables at the 0.5% and 99.5% levels to reduce potential biases in our analysis caused by outliers. Our results without winsorization are qualitatively similar to those reported in this paper.

We first present the summary statistics of SEO announcement effect and the results of baseline regressions on the relationship between investor attention and SEO announcement effect in Section 6.1. Next, we present the summary statistics of SEO post-announcement drift and the results of baseline regressions on the relationship between investor attention and SEO post-announcement drift in Subsection 6.2. Further, in Subsection 6.3 we examine the relationship between the market reaction to SEO announcements (namely, the announcement effect and the post-announcement drift) and the post-announcement operating performance of SEO firms. Finally, we address potential endogeneity concerns by presenting a set of robustness tests and instrumental variable analyses in Subsection 6.4.

6.1 Investor Attention and SEO Announcement Effects

In this subsection, we first present the summary statistics of SEO announcement effects. We estimate SEO announcement effect as the cumulative abnormal return (CAR) over a certain window around the SEO announcement date. We estimate abnormal returns using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO announcement date.[17] Panel A of Table 2 reports the summary statistics of SEO announcement effects measured using various event windows and their statistical significance. The mean cumulative abnormal return on the SEO announcement day, *CAR [0:0]*, is -0.76%, which is statistically significantly different from zero at the 1% level. We will use *CAR [0:0]* as our main measure of SEO announcement effect in our subsequent tests. Further, the mean cumulative abnormal returns upon SEO announcements over the 3-day (*CAR [-1:1]*), 5-day (*CAR [-2:2]*), and 7-day (*CAR [-3:3]*) windows are -2.30%, -2.13%, and -2.07%, respectively. These announcement effects are statistically significantly different from zero at the 1% level as well. Our findings in Panel A of Table 2 are consistent with the existing literature which has documented negative announcement effects for SEOs.

Next, we test our first hypothesis **H1** which predicts that the more investors pay attention to the SEO firm the more negative the announcement effect of the SEO will be. We test this hypothesis in a multivariate regression setting by regressing the announcement-day abnormal return (*CAR [0:0]*) on our investor attention proxies and other controls. The announcement-day abnormal return is estimated using the market model as described above in this subsection. The independent variables of interest in our regressions are our four investor attention measures (*NumNewsFile [-7:-1]*, *NumNewsFile [-14:-1]*, *NumNewsFile [-30:-1]*, *NumNewsFile [-60:-1]*) as described in Subsection 5.1. We also add several control variables to rule out potentially confounding effects. First, we control for lead SEO underwriter reputation. Following Bajo, Chemmanur, Simonyan, and Tehranian (2016), we

[17] We also estimate abnormal returns using alternative models such as Fama-French three-factor model, and Carhart four-factor model (Fama and French (1993) and Carhart (1997)). Our results remain qualitatively similar using these alternative estimation models.

construct a measure of lead SEO underwriter reputation, *UndwrtReputation*, as the lead SEO underwriter's share of total proceeds raised in the SEO market over previous five years. In our regressions we also control for SEO firm size (*FirmSize*), which is the natural logarithm of the book value of the SEO firm's total assets at the end of the fiscal quarter prior to the SEO announcement date; the midpoint of initial filing range (*MidFilePrice*); the level of information asymmetry about the SEO firm using the earnings surprise for the fiscal quarter prior to the SEO announcement date (*PriorQtrEarnSurpFile*), where earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price; and the return on the CRSP value-weighted index over a one-month (21-trading-day) period prior to the SEO announcement date (*PriorMktRetFile*). Finally, we also include announcement year $\times$ two-digit SIC industry code fixed effects to control for time-varying unobservables across different industries.

Table 3 presents the results of our regressions of the SEO announcement effect on various investor attention proxies. The coefficient estimates of all four investor attention measures in our regressions are negative and statistically significant at the 1% level. Given that the mean abnormal returns upon SEO announcements are negative as shown in Table 2, this finding suggests that the announcements of SEOs conducted by firms which receive more attention from market investors are associated with more negative announcement-period abnormal returns. The results in Table 3 are also economically significant. For example, a one-standard-deviation increase in the number of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to the SEO announcement date (which correspond to an increase in the number of news items by approximately 5, 9, 18, and 32, respectively) decreases the announcement-day abnormal return (*CAR [0:0]*) by 0.27, 0.32, 0.33, and 0.29 percentage points, respectively (i.e., augments the negative SEO announcement effect by 35.2%, 41.4%, 43.9%, and 38.0%, respectively). These findings suggest that, indeed, the greater the extent of investor attention paid to the SEO firm the more negative the SEO announcement effect, and provide support for our hypothesis **H1**.

6.2 Investor Attention and SEO Post-Announcement Stock Return Drift

In this subsection, we first present the summary statistics of SEO post-announcement drift. We estimate the post-announcement drift as the cumulative abnormal return (CAR) over a certain window after the SEO announcement date. Abnormal returns are estimated using the market model as described in Subsection 6.1. Panel B of Table 2 reports the summary statistics of two measures for the SEO post-announcement drift and their statistical significance. The mean SEO post-announcement cumulative abnormal return over a one-month (21-trading day) period (*CAR [1:21]*) and a two-month (42-trading day) period (*CAR [1:42]*) are −3.53% and −5.63%, respectively. These SEO post-announcement drift measures are statistically significantly different from zero at the 1% level. Overall, the summary statistics in Table 2 indicate that not all information about SEOs (or the firms conducting SEOs) is incorporated in SEO firms' stock prices upon the announcements of SEOs, but that information continues to be incorporated in the stock price over a longer period of time in the form of SEO post-announcement drift.

Next, we test our second hypothesis **H2** which predicts that the more investors pay attention to the SEO firm the less negative the SEO post-announcement drift will be. We test this hypothesis in a multivariate regression setting by regressing the SEO post-announcement cumulative abnormal return over a one-month (21-trading-day) period (*CAR [1:21]*) on our investor attention proxies and the same set of control variables and fixed effects as described in Subsection 6.1. The results of our regressions are reported in Table 4. The coefficient estimates of all four investor attention measures in our regressions are positive and statistically significant at the 1% level. This suggests that the SEO post-announcement drift is less negative for those firms which receive more investor attention upon their SEO announcements. These findings are also economically significant. For example, a one-standard-deviation increase in the number of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to the SEO announcement date (which corresponds to an increase in the number of news items by approximately 5, 9, 18, and 32, respectively) increases the post-announcement one-month cumulative abnormal return (*CAR [1:21]*) by 1.24, 1.35, 1.05, and 0.74 percentage points, respectively (i.e., shrinks the negative post-announcement drift by 35.1%, 38.3%, 29.9%, and 20.9%, respectively). These results suggest that, indeed, the greater the extent of investor attention paid to SEO firms the less negative the SEO post-announcement drift, and provide support for our hypothesis **H2**.

6.3 Relationship between SEO Announcement Effect, Post-anno- uncement Stock Return Drift, and Subsequent Operating Performance

In this subsection, we test our hypothesis **H3**, which predicts that both SEO announcement effect and SEO post-announcement drift will be positively correlated with the SEO firm's post-SEO operating performance. In other words, we examine whether better market reaction upon SEO announcement (less negative announcement effect and less negative post-announcement drift) leads to better post-SEO operating performance. We measure the post-announcement operating performance of the firm conducting SEO using two proxies measured in four windows: return on assets (*ROA*) and cash flow (*Cash Flow*), each measured over one, two, three, and four fiscal quarters after the SEO announcement. *ROA* is defined as the ratio of net income to the book value of total assets, and *Cash Flow* is defined as the ratio of income before extraordinary items plus depreciation to the book value of total assets. To directly compare the predictive ability of the SEO announcement effect to that of the post-announcement drift, we use *Standardized CAR [0:0]* to proxy for the SEO announcement effect and *Standardized CAR [1:21]* to proxy for the post-announcement drift, where both variables are standardized variables with a mean of zero and a standard deviation of one. We regress these measures of post-SEO operating performance on the proxy for announcement effect (*Standardized CAR [0:0]*) and the proxy for SEO post-announcement drift (*Standardized CAR [1:21]*) while controlling for the same set of control variables and fixed effects as described in Subsection 6.1.

The results of our regressions are presented in Table 5: Panel A presents the results of our regressions using *ROA* as the dependent variable and Panel B presents the results of our regressions using *Cash Flow* as the dependent variable. Table 5 shows that the coefficient estimates of both *Standardized CAR [0:0]* and *Standardized CAR [1:21]* are positive in all four regressions in both Panels A and B and they are statistically significant for both operating performance proxies measured over two, three, and four fiscal quarters after the SEO announcement. These findings suggest that firms with better market reaction upon their SEO announcements realize better post-SEO operating performance starting two fiscal quarters after their SEO announcements. Further, our finding of *Standardized CAR [0:0]* and *Standardized CAR [1:21]* both having significantly positive coefficient estimates also suggests that the information released at the SEO announcement regarding the firm's future expected (operating) performance is incorporated into the firm's stock price not only upon the announcement of the SEO (announcement effect) but also over a longer period of time after the announcement (post-announcement drift).

Since both the announcement return and post-announcement drift return are standardized variables, the magnitudes of these coefficients directly indicate their economic significance.[18] For example, a one-standard-deviation increase in the standardized announcement-day abnormal return and a one-standard-deviation increase in the standardized one-month post-announcement cumulative abnormal return lead to an increase in *ROA* computed over three fiscal quarters after SEO by 0.79 and 0.87 percentage points, respectively. Similarly, a one-standard-deviation increase in the announcement-day abnormal return and a one-standard-deviation increase in the one-month post-announcement cumulative abnormal return lead to increases of 0.72 and 0.89 percentage points, respectively, in *Cash Flow* measured over three fiscal quarters after SEO. More importantly, the coefficients of *Standardized CAR [1:21]* are larger than the coefficients of *Standardized CAR [0:0]* in terms of both the magnitude and the statistical significance in all regressions, suggesting that the post-announcement drift has a larger predictive power in terms of predicting future operating performance than the SEO announcement effect. In other words, a larger portion of information regarding the firm's future expected performance is only incorporated in the post-announcement drift. These findings collectively provide strong support for our hypothesis **H3**.

[18] Table A.1 in the Internet Appendix also reports regression results using the original SEO announcement return and post-announcement drift. The coefficient of CAR [0:0] does not change much when we compare its coefficient in regressions with and without CAR [1:21] as reported in Table A.1. In addition, the R^2 of regressions with CAR [1:21] is larger than those without CAR [1:21]. These results collectively show that the post-announcement drift incorporates a significant portion of information regarding the firm's future expected performance that is not incorporated in the SEO announcement return.

6.4 Identification

While our baseline results are consistent with our hypotheses (**H1** through **H3**) derived from our theoretical model, our baseline empirical design may suffer from potential endogeneity problems. The first problem is due to potential omitted variables. One could argue that certain (long-term) firm characteristics omitted from our baseline analysis may affect both the extent of attention paid by investors to a firm conducting an SEO as well as the market reaction upon its SEO announcement, so that the baseline results we reported above can potentially be driven by such omitted variables rather than investor attention. In order to address this potential omitted variable problem, we regress the measures of SEO announcement effect and SEO post-announcement drift on measures of abnormal media coverage as described in Subsection 5.1, where abnormal media coverage for a given firm is computed as the media coverage received by that firm immediately prior to its SEO announcement minus the media coverage of the same firm one year before the SEO announcement.

The results of our regressions are presented in Panels A (for SEO announcement effect) and B (for SEO post-announcement drift) of Table 6. In Panel A of Table 6, all four measures of abnormal investor attention have significantly negative coefficient estimates, consistent with our baseline results in Table 3. This finding indicates that the negative relationship between investor attention and SEO announcement effect we documented in our baseline regressions was not driven by omitted variables. In Panel B of Table 6, three out of four measures for abnormal investor attention have positive coefficient estimates and one of them (*AbnNumNewsFile [-14:-1]*) is statistically significant. These results are also broadly consistent with our baseline findings in Table 4 and provide a weak indication that our baseline findings on the positive relationship between investor attention and SEO post-announcement drift is unlikely to be caused by omitted variables.

The second potential problem that our baseline analysis may suffer from is that there could be some informational or other confounding events happening prior to a firm's SEO announcement which could potentially affect both the extent of attention paid by investors to the firm as well as the market reaction upon its SEO announcement that we study here. We address this potential endogeneity concern by making use of an instrumental variable analysis. We instrument for the extent of investor attention received by the firm immediately before its SEO announcement using the media coverage received by the firm one year before the SEO announcement.[19] For example, we use *PriorYrNumNewsFile [-60:-1]*, which is the number of news items covering an SEO firm over the two-month period ending one year prior to its SEO announcement date, as our instrumental variable for *NumNewsFile [-60:-1]*. We expect the media coverage received by an SEO firm one year before its SEO announcement to be positively correlated with the media coverage received by the firm immediately before its SEO announcement; however, we do not expect the SEO characteristics we study here

[19] Liu and McConnell (2013) use a similar instrument in their instrumental variable analysis to study the role of media coverage in corporate governance.

(SEO announcement effect and SEO post-announcement drift) to be correlated with the media coverage received by the SEO firm one year before its SEO announcement.

The results of our instrumental variable analysis are presented in Panels A (for SEO announcement effect) and B (for SEO post-announcement drift) of Table 7. In our first-stage regressions we regress the SEO firm's media coverage prior to its SEO announcement on the media coverage for the same firm one year before the SEO announcement (i.e., our instrumental variable) and the same set of control variables and industry × year fixed effects as described in Subsection 6.1. Both Panels A and B of Table 7 show, consistent with our expectation discussed above, that in first-stage regressions our instrumental variables are significantly positively correlated with our investor attention measures. We also report the F-statistics of the weak instruments test (or the test of excluded instruments) for each first-stage regression in Table 7. This test is used to determine whether instrumental variables used in first-stage regressions are strong. In their survey of the literature on weak instruments, Stock, Wright, and Yogo (2002) develop benchmarks for the necessary magnitude of the F-statistic. They point out that if the number of instruments is equal to one, then the critical value of the F-statistic is 8.96. Given that the F-statistics reported for our first-stage regressions in Table 7 are all well above the critical value of 8.96, the null hypothesis that our instruments are weak is strongly rejected.

Our second-stage regressions in both Panels A and B of Table 7 show that the coefficient estimates of predicted values of investor attention measures from all first-stage regressions have the same signs as reported in baseline regressions in Tables 3 and 4, and three out of four coefficient estimates in Panel A and all four coefficient estimates in Panel B are statistically significant. These results suggest that, even after controlling for the potential endogeneity of investor attention paid to SEO firms immediately prior to their SEO announcements, firms which receive a higher level of investor attention prior to their SEO announcement dates are associated with larger (more negative) SEO announcement effects and smaller (less negative) SEO post-announcement drifts. Overall, our analysis in this subsection, which deals with the potential endogeneity of investor attention, demonstrates the robustness of our baseline findings in previous subsections on the relationship between investor attention and the market reaction upon SEO announcements.

7 Relationship between Investor Attention and SEO Characteristics

In this section, we study the relationship between the extent of investor attention paid to firms conducting SEOs and certain SEO-related offering and firm characteristics such as SEO initial returns (underpricing), SEO firm market valuation both at the SEO offer price as well as in the immediate post-SEO secondary market, and the extent of post-SEO institutional investor interest in the shares of SEO firms. We first develop testable hypotheses regarding these relations in Subsection 7.1. In subsequent subsections (7.2 to 7.5), we present our empirical findings on the relationship between investor attention and institutional investor holdings of SEO firms' equity, SEO firm market valuation in the immediate post-SEO secondary market, SEO underpricing, and SEO firm market valuation at SEO offer price. Finally, we discuss the results of our two identification tests that establish causality in Subsection 7.6.

7.1 Theory and Hypothesis Development

We first develop testable hypotheses regarding the relationship between the investor attention received by a firm immediately before its actual SEO and the pricing of the SEO and other SEO characteristics. In order to develop these testable hypotheses, we start by assuming that, for institutional investors to participate in a firm's SEO, they not only need to receive information about various aspects of the firm from SEO underwriter, but also need to pay attention to or "recognize" this information. This assumption is in the spirit of Merton's (1987) investor recognition or attention model, which assumes that an investor will incorporate a security into his portfolio only if he pays attention to (or acquires information about) that security by incurring a cost. While Merton (1987) posits several possible sources of this "attention" or "recognition" cost, he views this cost mainly as arising from the cost of investors becoming aware of (or familiar with) a firm. In his setting, investors consider investing only in the stock of firms with which they have a certain level of familiarity. Similarly, in our setting, we can assume that institutional and other investors consider investing only in the stock of those SEO firms that they have become familiar with by incurring an attention cost. Then we would expect the extent of institutional investor participation in the SEOs of firms that received greater investor attention to be greater. This is the first hypothesis that we test here (**H4**).

The above setup has implications for the valuation of equity both in the immediate aftermarket (pricing in the equity market after the SEO) and for firm valuation at the SEO offer price as well. We first discuss the relationship between investor attention and post-SEO secondary market valuation. Since the demand from investors for the equity of firms whose SEOs receive greater investor attention will be greater (for a given supply of shares offered in the SEO), the market clearing price for the equity of these firms will be higher as well. Assuming that the immediate aftermarket share price of the SEO firm is the market clearing price, this implies that there will be a positive relationship between investor attention and the immediate post-SEO market valuation of firms (**H5**).

We now turn to the relationship between investor attention and SEO initial returns as well as the relationship between investor attention and SEO firm valuation at the offer price. These relations depend on the process of price setting in SEOs. While there is no consensus in the theoretical or empirical literature on how the SEO offer price is set, there is some agreement that the office price is set at a discount to the expected market clearing price (which can be viewed as the same as the expected aftermarket price) giving rise to positive SEO initial returns (SEO underpricing). There are a number of alternative theories about the drivers of SEO underpricing. One theory, obtained from the IPO literature, is advanced by Benveniste and Spindt (1989), who argue that the equity of the firm making IPO is priced at a discount to the market clearing (immediate secondary market) price in order to ensure that institutions have an incentive to reveal their true demand for the firm's equity (i.e., it ensures that their incentive compatibility or truth-telling conditions hold). If the discount applied to the market clearing price to arrive at the SEO offer price is driven by considerations similar to those advanced by Benveniste and Spindt (1989) in the context of IPOs (i.e., unrelated to investor attention), then we would expect pre-SEO investor attention to be unrelated to SEO initial returns (**H6A**).

However, there are also some theories suggesting that there may be a positive relationship between investor attention and SEO initial returns. For example, Chemmanur and Jiao (2011) show in their theoretical analysis that SEO initial returns (underpricing) may be positively related to pre-SEO institutional demand for SEO firm equity.[20] Given that SEOs characterized by greater investor attention are likely to have greater pre-SEO institutional investor demand as well, this implies a positive relationship between investor attention and SEO initial returns (**H6B**).[21]

Consider now the relationship between investor attention and firm valuation at the SEO offer price. If SEO underpricing is unrelated to investor attention (e.g., driven only by considerations of information extraction, as posited by Benveniste and Spindt (1989)), then we would expect an unambiguously positive relationship between investor attention and firm valuation at the SEO offer price (**H7A**). On the other hand, if SEO underpricing is positively related to investor attention (e.g., following the arguments made by Chemmanur and Jiao (2011) discussed above), then the predicted relationship between investor attention and firm valuation at the SEO offer price becomes ambiguous (**H7B**). This is because the greater secondary market price associated with greater investor attention may potentially be offset by even greater SEO underpricing associated with greater investor attention, so that the relationship between investor attention and firm valuation at the SEO offer price may turn negative.

[20] See Proposition 8 of Chemmanur and Jiao (2011).

[21] An alternative theory that suggests a positive relationship between investor attention and SEO underpricing is provided by Liu, Sherman, and Zhang (2014) and Liu, Lu, Sherman, and Zhang (2019). They argue, in the context of IPOs, that IPO underpricing is a way of compensating investors for their cost of paying attention to the IPO firm. In a similar vein, it may be argued that SEO underpricing (initial returns) is a way of enhancing the investor attention paid to an SEO by implicitly compensating investors for their cost of paying attention to the firm making SEO. Given this alternative theory, we will show some specifications in our empirical analysis of SEO valuation, post-SEO secondary market valuation, and institutional investor participation in SEOs where we control for the extent of SEO initial returns (underpricing).

7.2 Investor Attention and Post-SEO Participation of Institutional Investors

In this subsection, we test our hypothesis **H4** which predicts that a greater extent of investor attention received by a firm prior to its SEO will be associated with greater institutional investor ownership of the firm's equity after the SEO. We measure the extent of institutional investors' ownership of issuing firm's equity after its SEO by the number of institutional investors holding firm's shares at the end of the first quarter after the SEO (*InstN*). We regress *InstN* on our four investor attention measures (*NumNewsIss*) as described in Subsection 5.1 and other control variables including underwriter reputation, firm size, the midpoint of initial filing range, the level of information asymmetry about the SEO firm measured by the earnings surprise one quarter prior to the SEO issue date, one-month stock market return prior to the SEO issue date, and issue year $\times$ two-digit SIC industry code fixed effects. In our regressions we include only those SEOs for which the number of days between the SEO announcement date and the SEO issue date is greater than the number of days that we use to measure investor attention. For example, if in a regression we use the investor attention measured over the 7-day window prior to the SEO issue date, then this regression is estimated using only a sub-sample of SEOs with at least a 7-day gap between the SEO announcement date and the SEO issue date.

We report the results of our regressions in Table 8. In regression specifications (2), (4), (6), and (8) we include SEO underpricing as an additional control variable in order to control for the potential effect of SEO underpricing on the post-SEO institutional investor ownership of the issuing firm's equity. Table 8 demonstrates that all four investor attention measures have significantly positive coefficient estimates in all regressions (with and without controlling for SEO underpricing), suggesting that a firm which receives more investor attention prior to its SEO is likely to have a greater number of institutional investors as shareholders after the SEO. The positive coefficient estimates of SEO underpricing (statistically significant in regression specifications (2) and (4)) provide further support for our theoretical prediction that firms conducting SEOs may leave more money on the table to attract more institutional investors to invest in their firms' equity. These results are also economically significant. For example, a one-standard-deviation increase in the number of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO issue dates (which correspond to increases in the number of news items of approximately 6, 11, 23, and 46, respectively) increases the number of institutional investors holding the SEO firms' equity by 29, 37, 40, and 39, respectively (i.e., 22.0%, 28.0%, 30.0%, and 29.8% increases in the number of institutional investors, respectively). These results indicate that a greater extent of investor attention paid to issuing firms immediately prior to their SEOs is associated with a greater number of institutional investors holding the issuing firms' equity post-SEO, and support our hypothesis **H4**.

7.3 Investor Attention and the Post-SEO Market Valuation of Issuing Firms

In this subsection, we test our hypothesis **H5** which predicts a positive relationship between investor attention received by firms immediately prior to their SEOs and their post-SEO market valuation. We measure post-issue market valuation of SEO firms using industry-adjusted Q ratios computed using either the SEO issue day closing stock price (*QFTDAdj*) or the stock price at the end of the first post-issue fiscal quarter (*QFQAdj*). We define Q ratio as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and either the SEO issue day closing price (*QFTD*) or the price at the end of the first post-issue fiscal quarter (*QFQ*). We further adjust these ratios for median industry valuation by subtracting contemporaneous 2-digit SIC code industry median Q ratios from the above Q ratios of SEO firms. We regress these two measures of post-SEO market valuation (*QFTDAdj* and *QFQAdj*) on our four investor attention measures (*NumNewsIss*) while controlling for the same set of control variables and fixed effects as described in Subsection 7.2. As discussed in Subsection 7.2, in our regressions we include only those SEOs for which the number of days between the SEO announcement date and the SEO issue date is greater than the number of days that we use to measure investor attention.

Due to space limitations, we present the results of our regressions in Panels A (using *QFTDAdj* as the dependent variable) and B (using *QFQAdj* as the dependent variable) of Table A.2 in the Internet Appendix. In regression specifications (2), (4), (6), and (8) of each panel we include SEO underpricing as an additional control variable in order to control for the potential effect of SEO underpricing on the immediate post-SEO valuation of issuing firms. Both Panels A and B of Table A.2 show that all four pre-SEO investor attention measures have significantly positive coefficient estimates in all regressions (both with and without SEO underpricing as a control variable). This suggests that firms receiving more investor attention immediately prior to their SEOs are likely to have higher post-SEO market valuations. These results are also economically significant. For example, a one-standard-deviation increase in the number of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO issue dates (which correspond to increases in the number of news items of approximately 6, 11, 23, and 46, respectively) increases the magnitude of *QFQAdj* by 0.126, 0.166, 0.184, and 0.182, respectively, which is a sizable increase compared to the mean *QFQAdj* of -0.037 in our sample. These results imply that a greater extent of investor attention paid to issuing firms immediately prior to their SEOs leads to higher post-SEO market valuations, and support our hypothesis **H4**.

7.4 Investor Attention and SEO Underpricing

In this subsection, we study the relationship between investor attention and SEO underpricing by regressing SEO underpricing on our investor attention measures and other controls. We compute SEO underpricing as the percentage difference between SEO issue day closing price and SEO offer price (*Underpricing*). We test the above hypothesis by regressing SEO underpricing on our pre-SEO investor attention measures (*NumNewsIss*) while controlling for the same set of control variables and fixed effects as described in Subsection 7.2. As discussed in Subsection 7.2, in our regressions we include only those SEOs for which the number of days between the SEO announcement date and the SEO issue date is greater than the number of days that we use to measure investor attention.

In Table 9, we report the results of our regressions using SEO underpricing (*Underpricing*) as the dependent variable. All four investor attention measures have positive and statistically significant coefficient estimates, suggesting that firms receiving more investor attention prior to their SEOs are associated with greater SEO underpricing. These results are also economically significant. For example, a one-standard-deviation increase in the number of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO issue dates (which correspond to increases in the number of news items of approximately 6, 11, 23, and 46, respectively) increases SEO underpricing by 0.265, 0.554, 0.760, and 0.683 percentage points, respectively (i.e., increases the magnitude of underpricing by 7.4%, 15.6%, 21.4%, and 19.2%, respectively). These findings suggest that, indeed, a greater extent of investor attention paid to issuing firms immediately prior to their SEOs is associated with a greater degree of SEO underpricing, and support our hypothesis **H6B**.

7.5 Investor Attention and SEO Valuation of Issuing Firms

In this subsection, we study the effect of investor attention received by firms immediately prior to their SEOs on their firm valuation at the SEO offer price. We measure SEO valuation of issuing firms using industry-adjusted Q ratios computed using SEO offer price (*QOPAdj*). We define Q ratio as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and SEO offer price (*QOP*). We further adjust these ratios for median industry valuation by subtracting contemporaneous 2-digit SIC code industry median Q ratios from the above Q ratios of SEO firms. We regress SEO valuations (*QOPAdj*) on our four investor attention measures (*NumNewsIss*) while controlling for the same set of control variables and fixed effects as described in Subsection 7.2. As discussed in Subsection 7.2, in our regressions we include only those SEOs for which the number of days between the announcement date and the issue date is greater than the number of days that we use to measure investor attention.

The results of our regressions are reported in Table 10. In regression specifications (2), (4), (6), and (8) we include SEO underpricing as an additional control variable in order to control for SEO underpricing being potentially used as a compensation for investor attention. Table 10 shows that the coefficient estimates of all four pre-SEO investor attention measures are significantly positive in all regressions and remain unchanged with and without SEO underpricing as a control variable. This suggests that firms receiving more investor attention immediately prior to their SEOs are likely to have higher SEO valuation. These results are also economically significant. For example, a one-standard-deviation increase in the number of news items covering SEO firms over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO issue dates (which correspond to increases in the number of news items of approximately 6, 11, 23, and 46, respectively) increases the magnitude of *QOPAdj* by 0.202, 0.266, 0.322, and 0.228, respectively, which is a sizable increase compared to the mean *QOPAdj* of -0.036 in our sample. These results imply that a greater extent of investor attention paid to issuing firms immediately prior to their SEOs leads to higher SEO valuations. Combined with our findings in Subsection 7.4, the results here support our hypothesis **H7B**.

7.6 Identification

In order to address the potential endogeneity problems in our analysis of the effect of investor attention on various SEO-related offering and firm characteristics (discussed previously in Subsection 6.4), we perform a similar set of robustness tests and instrumental variable analyses as in Subsection 6.4. First, we regress SEO underpricing, SEO valulation and post-SEO secondary market valuation of issuing firms, and post-SEO institutional investors' participation in issuing firm's equity ownership on our four abnormal investor attention measures while controlling for the same set of control variables and fixed effects as described in Subsection 7.2. For brevity, the results of these regressions are presented in the Internet Appendix of this paper: Tables A.3, A.4, A.5, and A.6. The coefficient estimates of all four abnormal investor attention measures in these regressions have the same signs as those reported in our baseline results and are statistically significant in all of the *QOPAdj*, *QFTDAdj* and *InstN* regressions, and in two of the *Underpricing* and *QFQAdj* regressions. Overall, these regression results on the relationship between abnormal investor attention and various SEO-related offering and firm characteristics are consistent with the results of our baseline regressions.

Next, we also implement instrumental variable analyses of the effect of investor attention on various SEO-related offering and firm characteristics making use of the same instrumental variables as described in Subsection 6.4. The results of our instrumental variable analyses for post-SEO institutional investor participation, SEO underpricing, and SEO valuation at offer price are reported in Tables 11, 12, and 13, respectively.[22] Our first-stage regressions in all

[22]Due to space limitation, the results of our instrumental variable analyses for post-SEO secondary market valuation is reported in Table A.7 in the Internet Appendix.

these tables show that our instrumental variables are significantly and positively correlated with our investor attention variables and the F-statistics of the weak instruments test are well above the critical value of 8.96. Our second-stage regressions in Tables 11, 12, and 13 show that the coefficient estimates of predicted values of investor attention measures from first-stage regressions have the same signs as those reported in the baseline results in Tables 8, 9, and 10, and are statistically significant. These results suggest that, even after controlling for the potential endogeneity of investor attention paid to SEO firms immediately prior to their SEOs, firms which receive a higher level of investor attention prior to their SEOs are associated with greater participation of institutional investors in their post-SEO equity ownership, higher post-SEO secondary market valuations, larger SEO underpricing, and higher firm valuation at the offer price. Overall, our instrumental variable analyses demonstrate the robustness of our baseline findings in previous subsections.

8 Robustness Check: Using the Number of SEC EDGAR File Downloads as an Alternative Measure of Investor Attention

8.1 Measures of Investor Attention Using the Number of SEC EDGAR File Downloads

We collect data on SEC EDGAR File Downloads for SEO firms from the EDGAR Log File Data Set in the SEO website.[23] The EDGAR Log File Data Set records user access statistics on its EDGAR system from January 2003 to June 2017. We follow Lee, Ma, and Wang (2015) to clean this dataset with the following two steps. First, we classify those daily IP addresses downloading more than 50 unique firms filings as robots and remove those observations from our analysis. Second, we restrict our analysis to downloads for 10-K,10-Q, and 8-K filings, which contain fundamental information about each firm. After processing the dataset with the above two steps, we match it with our sample of SEO firms. We use the pre-SEO EDGAR file downloads fpr SEO firms as our alternative proxy for the amount of attention paid by market investors to SEO firms. We construct our measures of investor attention for each SEO firm by counting the number of downloads for 10-K,10-Q, and 8-K filings of the SEO firm over a certain period of time (namely, 1 week, 2 weeks, 1 month, and 2 months) prior to the SEO announcement date (*NumEdgarFile*) or prior to the SEO issue date (*NumEdgarIss*). For example, *NumEdgarFile [-60:-1]* and *NumEdgarIss [-60:-1]* are the numbers of unique IP addresses downloaded 10-K, 10-Q, and 8-K filings of the SEO firm in the SEC EDGAR system over a two-month period (60-day period from day -60 to day -1) prior to the SEO announcement date and prior to the SEO issue date,

[23] For the complete EDGAR Log File Data Set, please see: https://www.sec.gov/dera/data/edgar-log-file-data-set.html.

respectively.

Panel A of Table 1 reports the summary statistics of these alternative investor attention measures. The average numbers of unique IP addresses downloaded 10-K, 10-Q, and 8-K filings of the SEO firm in the SEC EDGAR system over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO announcement dates are 24.7, 43.4, 81.2, and 144.2, respectively; while the average numbers of unique IP addresses downloaded 10-K, 10-Q, and 8-K filings of the SEO firm in the SEC EDGAR system over the 1-week, 2-week, 1-month, and 2-month periods prior to their SEO issue dates are 28.9, 47.9, 87.7, and 157.8, respectively.

8.2 Empirical Results Using the Number of SEC EDGAR File Downloads as an Alternative Measure of Investor Attention

Panel A of Table 14 presents the results of replicating our regressions of the SEO announcement effect in Table 3 using this alternative measure of investor attention. The coefficient estimates of these alternative investor attention measures in all four regressions are negative and statistically significant, consistent with what we find in Table 3. The results of replicating our regressions in Table 4 using this alternative measure of investor attention are reported in Panel B of Table 14. The coefficient estimates of all four investor attention measures are also positive and statistically significant at the 1% level. These findings suggest that our results are robust to alternative investor attention measures, and thereby providing strong support for our hypotheses **H1** and **H2**.

We also rerun our regressions on various SEO characteristics using the number of SEC EDGAR file downloads as an alternative measure of investor attention. Table 15 reports these regression results. Consistent with our findings in 7, firms that receive a higher level of investor attention prior to their SEOs, measured as a larger number of SEC EDGAR file downloads, are associated with a larger SEO underpricing, a greater participation of institutional investors in their post-SEO equity ownership, and a higher firm valuation at the offer price.[24]

9 Conclusion

Models of seasoned equity offerings such as Myers and Majluf (1984) assume that all investors in the economy pay immediate attention to SEO announcements and the pricing of SEOs. In this paper, we relax the above assumption and analyze, theoretically and empirically, the implications of a fraction of investors in the equity market paying only delayed attention to SEO announcements. We first show theoretically that, in the above setting,

[24]Tables A.8 and A.9 in the Internet Appendix also report regression results using the abnormal EDGAR filing downloads as a measure of investor attention, where all the results are quantitatively similar to those reported in Tables 14 and 15. The set of abnormal EDGAR filing downloads measures is defined in the same way as the set of abnormal media coverage measures described in Section 5.1.

the announcement effect of an SEO will be positively related to the fraction of investors paying attention to the announcement and that there will be a post-announcement stock-return drift that is negatively related to investor attention. In the first part of the paper, we test the above predictions using the media coverage of firms announcing SEOs as a proxy for investor attention, and find evidence consistent with the above predictions. In the second part of the paper, we develop and test various hypotheses relating investor attention paid to the SEO firm (between the announcement date and the issue date) to various SEO characteristics. We empirically show that institutional investor participation in SEOs, the post-SEO secondary equity market valuation of issuing firms, SEO underpricing, and SEO valuation are all positively related to investor attention. The results of our identification tests show that the above results are causal.

References

Altınkılıç, Oya, and Robert S Hansen, 2003, Discounting and underpricing in seasoned equity offers, *Journal of Financial Economics* 69, 285–323.

Asquith, Paul, and David W. Mullins, 1986, Equity issues and offering dilution, *Journal of Financial Economics* 15, 61–89.

Bajo, Emanuele, Thomas J. Chemmanur, Karen Simonyan, and Hassan Tehranian, 2016, Underwriter networks, investor attention, and initial public offerings, *Journal of Financial Economics* 122, 376–408.

Benveniste, Lawrence M., and Paul A. Spindt, 1989, How investment bankers determine the offer price and allocation of new issues, *Journal of Financial Economics* 24, 343–361.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chemmanur, Thomas J., Shan He, and Gang Hu, 2009, The role of institutional investors in seasoned equity offerings, *Journal of Financial Economics* 94, 384–411.

Chemmanur, Thomas J., and Yawen Jiao, 2011, Institutional Trading, Information Production, and the SEO Discount: A Model of Seasoned Equity Offerings, *Journal of Economics & Management Strategy* 20, 299–338.

Chemmanur, Thomas J., Karthik Krishnan, and Qianqian Yu, 2018, Venture capital backing, investor attention, and initial public offerings, *Unpublished working paper,* Boston College.

Cook, Douglas O., Robert Kieschnick, and Robert A. Van Ness, 2006, On the marketing of IPOs, *Journal of Financial Economics* 82, 35–61.

Corwin, Shane A., 2003, The determinants of underpricing for seasoned equity offers, *Journal of Finance* 58, 2249–2279.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *Journal of Finance* 66, 1461–1499.

Engelberg, Joseph E., and Christopher A. Parsons, 2011, The causal impact of media in financial markets, *Journal of Finance* 66, 67–97.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 2023–2052.

Gao, Xiaohui, and Jay R. Ritter, 2010, The marketing of seasoned equity offerings, *Journal of Financial Economics* 97, 33–52.

Gerard, Bruno, and Vikram Nanda, 1993, Trading and manipulation around seasoned equity offerings, *Journal of Finance* 48, 213–245.

Giammarino, Ronald M., and Tracy Lewis, 1988, A theory of negotiated equity financing, *The Review of Financial Studies* 1, 265–288.

Gibson, Scott, Assem Safieddine, and Ramana Sonti, 2004, Smart investments by smart money: Evidence from seasoned equity offerings, *Journal of Financial Economics* 72, 581–604.

Gustafson, Matthew T., 2018, Price pressure and overnight seasoned equity offerings, *Journal of Financial and Quantitative Analysis* 53, 837–866.

Hirshleifer, David, Sonya S. Lim, and Siew Hong Teoh, 2011, Limited investor attention and stock market misreactions to accounting information, *The Review of Asset Pricing Studies* 1, 35–73.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Huang, Rongbing, and Donghang Zhang, 2011, Managing underwriters and the marketing of seasoned equity offerings, *Journal of Financial and Quantitative Analysis* 46, 141–170.

Lee, Charles MC, Paul Ma, and Charles CY Wang, 2015, Search-based peer firms: Aggregating investor perceptions through internet co-searches, *Journal of Financial Economics* 116, 410–431.

Liu, Baixiao, and John J. McConnell, 2013, The role of the media in corporate governance: Do the media influence managers' capital allocation decisions?, *Journal of Financial Economics* 110, 1–17.

Liu, Laura Xiaolei, Ruichang Lu, Ann E. Sherman, and Yong Zhang, 2019, Attracting attention in an IPO, *Unpublished working paper,* Peking University.

Liu, Laura Xiaolei, Ann E. Sherman, and Yong Zhang, 2014, The long-run role of the media: Evidence from initial public offerings, *Management Science* 60, 1945–1964.

Loderer, Claudio F., Dennis P. Sheehan, and Gregory B. Kadlec, 1991, The pricing of equity offerings, *Journal of Financial Economics* 29, 35–57.

Masulis, Ronald W., and Ashok N. Korwar, 1986, Seasoned equity offerings: An empirical investigation, *Journal of Financial Economics* 15, 91–118.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.

Myers, Stewart C., and Nicholas S. Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, *Journal of Financial Economics* 13, 187–221.

Peng, Lin, 2005, Learning with information capacity constraints, *Journal of Financial and Quantitative Analysis* 40, 307–329.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Pinto-Gutiérrez, Cristian A., 2018, The effect of media coverage on the pricing of seasoned equity offerings, *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration* 36, 432–449.

Smith, Clifford W., 1977, Alternative methods for raising capital: Rights versus underwritten offerings, *Journal of Financial Economics* 5, 273–307.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo, 2002, A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business & Economic Statistics* 20, 518–529.

Table 1: **Summary statistics**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *NumNewsFile [-60:-1]* and *NumNewsIss [-60:-1]* are the numbers of news items covering SEO firms over a two-month period (from day -60 to day -1) prior to their SEO announcement dates and prior to their SEO issue dates, respectively. *PriorYrNumNewsFile [-60:-1]* is the number of news items covering SEO firms over a two-month period ending one year prior to their SEO announcement dates. *AbnNumNewsFile [-60:-1]* is the abnormal media coverage over a two-month period (from day -60 to day -1) prior to the SEO announcement date, which is defined as the difference between *NumNewsFile [-60:-1]* and *PriorYrNumNewsFile [-60:-1]*. *AbnNumNewsIss [-60:-1]* is the abnormal media coverage over a two-month period (from day -60 to day -1) prior to the SEO issue date, which is defined as the difference between *NumNewsIss [-60:-1]* and *PriorYrNumNewsFile [-60:-1]*. *NumEdgarFile [-60:-1]* and *NumEdgarIss [-60:-1]* are the numbers of unique IP addresses downloaded 10-K, 10-Q, and 8-K filings of the SEO firm in the SEC EDGAR systemthe numbers of news items covering SEO firms over a two-month period (from day -60 to day -1) prior to their SEO announcement dates and prior to their SEO issue dates, respectively. *AbnNumEdgarFile [-60:-1]* and *AbnNumEdgarIss [-60:-1]* are the abnormal EDGAR filing downloads over a two-month period (from day -60 to day -1) prior to the SEO announcement dates and prior to the SEO issue dates, which is defined in the same way as *AbnNumNewsFile [-60:-1]* and *AbnNumNewsIss [-60:-1]*. Other media coverage and EDGAR filing downloads measures are defined in a similar fashion and their precise definitions can be found in Sections 5.1 and 8. *ROA* is the ratio of net income over the book value of total assets at the end of the first post-announcement fiscal quarter. *Cash Flow* is the ratio of income before extraordinary items plus depreciation to the book value of total assets at the end of the first post-announcement fiscal quarter. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *QOPAdj* is the industry-adjusted Q ratio calculated using the SEO offer price. *QFTDAdj* and *QFQAdj* are the industry-adjusted Q ratios calculated using the SEO issue day closing price and the price at the end of the first post-issue fiscal quarter, respectively. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and either the SEO issue day closing price (*QFTDAdj*) or the price at the end of the first post-issue fiscal quarter (*QFQAdj*). Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *InstN* is the number of institutional investors holding SEO firms' shares at the end of the first post-issue fiscal quarter. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. *PriorQtrEarnSurpFile* and *PriorQtrEarnSurpIss* are the earnings surprises one quarter prior to the SEO announcement date and prior to the SEO issue date, respectively. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* and *PriorMktRetIss* are the returns on the CRSP value-weighted index over one-month (21-trading-day) periods prior to the SEO announcement date and prior to the SEO issue date, respectively.

Panel A: Summary statistics of investor attention measures

	N	Mean	Median	Min.	Max.	S.D.
NumNewsFile [-7:-1]	6,309	2.148	1	0	173	5.249
NumNewsFile [-14:-1]	6,309	4.044	2	0	287	9.013
NumNewsFile [-30:-1]	6,309	8.329	5	0	18	17.601
NumNewsFile [-60:-1]	6,309	15.620	10	0	31	32.142
NumNewsIss [-7:-1]	4,862	3.016	1	0	176	6.305
NumNewsIss [-14:-1]	4,510	5.100	3	0	355	11.080
NumNewsIss [-30:-1]	3,577	10.135	7	0	20	23.038
NumNewsIss [-60:-1]	2,713	19.602	13	3	35	45.532
PriorYrNumNewsFile [-7:-1]	6,309	1.663	0	0	146	4.506
PriorYrNumNewsFile [-14:-1]	6,309	3.178	1	0	239	7.597
PriorYrNumNewsFile [-30:-1]	6,309	6.524	3	0	15	15.201
PriorYrNumNewsFile [-60:-1]	6,309	12.417	8	0	26	29.007
AbnNumNewsFile [-7:-1]	6,309	0.485	0	-49	126	3.929
AbnNumNewsFile [-14:-1]	6,309	0.866	0	-56	212	5.846
AbnNumNewsFile [-30:-1]	6,309	1.805	0	-5	10	9.835
AbnNumNewsFile [-60:-1]	6,309	3.203	1	-7	15	16.024
AbnNumNewsIss [-7:-1]	4,862	1.616	0	-56	129	5.220

AbnNumNewsIss [-14:-1]	4,510	2.280	1	-119	280	8.422
AbnNumNewsIss [-30:-1]	3,577	3.492	2	-5	12	14.612
AbnNumNewsIss [-60:-1]	2,713	5.608	3	-8	19	26.033
NumEdgarFile [-7:-1]	4,738	24.691	14	0	55	52.495
NumEdgarFile [-14:-1]	4,738	43.404	26	0	96	93.733
NumEdgarFile [-30:-1]	4,738	81.236	50	0	173	178.085
NumEdgarFile [-60:-1]	4,738	144.232	90	0	309	310.195
NumEdgarIss [-7:-1]	3,762	28.946	17	0	69	43.394
NumEdgarIss [-14:-1]	3,466	47.946	31	0	110	72.547
NumEdgarIss [-30:-1]	2,837	87.745	59	0	195	136.070
NumEdgarIss [-60:-1]	2,249	157.800	113	0	338	231.681
AbnNumEdgarIss [-7:-1]	4,359	20.907	11	-4	59	45.900
AbnNumEdgarIss [-14:-1]	4,359	31.969	17	-7	88	73.830
AbnNumEdgarIss [-30:-1]	4,359	50.796	27	-16	140	124.766
AbnNumEdgarIss [-60:-1]	4,359	79.451	41	-28	218	213.726
AbnNumEdgarFile [-7:-1]	4,359	11.277	4	-7	36	39.397
AbnNumEdgarFile [-14:-1]	4,359	18.502	7	-12	59	60.513
AbnNumEdgarFile [-30:-1]	4,359	32.087	13	-22	100	110.125
AbnNumEdgarFile [-60:-1]	4,359	53.741	20	-39	171	195.291

Panel B: Summary statistics of SEO and firm characteristics

	N	Mean	Median	Min.	Max.	S.D.
ROA	6,194	-3.741	0.200	-75.153	11.279	10.572
Cash Flow	5,345	-3.390	0.515	-80.311	12.947	11.251
Underpricing	6,006	3.556	2.227	-20.661	38.321	6.583
QOPAdj	6,189	-0.036	-0.009	-6.873	13.298	2.048
QFTDAdj	5,902	-0.041	-0.005	-7.234	13.749	2.144
QFQAdj	6,182	-0.037	-0.009	-7.902	12.086	1.939
InstN	6,079	131.664	102	1	907	128.516
UndwrtReputation	6,309	0.036	0.007	0	0.193	0.050
FirmSize	6,174	6.225	6.150	1.515	12.506	2.122
MidFilePrice	6,009	24.073	18.700	0.350	158.550	23.017
PriorQtrEarnSurpIss	5,469	-0.077	0.001	-13.393	2.400	1.040
PriorQtrEarnSurpFile	5,382	0.028	0	-4.030	8.889	0.768
PriorMktRetIss	6,300	0.014	0.017	-0.127	0.140	0.039
PriorMktRetFile	6,309	0.012	0.016	-0.164	0.151	0.044

Table 2: **Summary statistics of SEO announcement effects and SEO post-announcement drift**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *CAR [0: 0]* is the abnormal return on SEO firm's equity on the SEO announcement day. CAR [-1:1] is the cumulative abnormal return on SEO firm's equity over a 3-day window (from day -1 to day +1) around the SEO announcement date. *CAR [-2:2]* is the cumulative abnormal return on SEO firm's equity over a 5-day window (from day -2 to day +2) around the SEO announcement date. *CAR [-3:3]* is the cumulative abnormal return on SEO firm's equity over a 7-day window (from day -3 to day +3) around the SEO announcement date. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *CAR [1:42]* is the cumulative abnormal return on SEO firm's equity over a 42-day window (from day 1 to day 42) after the SEO announcement date. Abnormal returns are estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO announcement date. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Summary statistics of SEO announcement effects

	N	Mean	Median	Min.	Max.	S.D.	t-Statistic (Means = 0)	z-Statistic (Medians = 0)
CAR [0:0]	5,821	-0.761	-0.445	-19.573	17.677	4.302	-13.492***	-15.307***
CAR [-1:1]	5,818	-2.298	-1.698	-34.642	33.101	7.821	-22.411***	-27.192***
CAR [-2:2]	5,815	-2.131	-1.794	-38.164	49.046	9.901	-16.415***	-22.709***
CAR [-3:3]	5,815	-2.068	-1.890	-41.185	62.389	11.425	-13.804***	-20.123***

Panel B: Summary statistics of SEO post-announcement drift

	N	Mean	Median	Min.	Max.	S.D.	t-Statistic (Means = 0)	z-Statistic (Medians = 0)
CAR [1:21]	5,828	-3.530	-2.778	-65.918	68.419	17.546	-15.358***	-18.543***
CAR [1:42]	5,829	-5.625	-4.198	-98.238	101.048	26.018	-16.506***	-18.918***

Table 3: **Relationship between investor attention and SEO announcement effects**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *CAR [0:0]* is the abnormal return on SEO firm's equity on the SEO announcement day. *NumNewsFile [-7:-1]*, *NumNewsFile [-14:-1]*, *NumNewsFile [-30:-1]*, and *NumNewsFile [-60:-1]* are measures of investor attention prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	CAR [0:0]			
	(1)	(2)	(3)	(4)
NumNewsFile [-7:-1]	-0.051***			
	(-3.15)			
NumNewsFile [-14:-1]		-0.035***		
		(-3.85)		
NumNewsFile [-30:-1]			-0.019***	
			(-4.15)	
NumNewsFile [-60:-1]				-0.009***
				(-3.66)
UndwrtReputation	-1.625	-1.549	-1.568	-1.590
	(-1.11)	(-1.06)	(-1.07)	(-1.09)
FirmSize	0.157***	0.170***	0.176***	0.172***
	(3.00)	(3.23)	(3.35)	(3.26)
PriorQtrEarnSurpFile	0.260***	0.259***	0.257***	0.258***
	(3.23)	(3.21)	(3.18)	(3.20)
PriorMktRetFile	0.662	0.507	0.434	0.512
	(0.43)	(0.33)	(0.28)	(0.33)
MidFilePrice	0.009***	0.009***	0.009***	0.009***
	(2.65)	(2.76)	(2.73)	(2.67)
Constant	-7.308	-7.410	-7.516	-7.395
	(-1.20)	(-1.22)	(-1.24)	(-1.22)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.170	0.171	0.171	0.171
Observations	4735	4735	4735	4735

Table 4: **Relationship between investor attention and SEO post-announcement drift**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *NumNewsFile [-7:-1]*, *NumNewsFile [-14:-1]*, *NumNewsFile [-30:-1]*, and *NumNewsFile [-60:-1]* are measures of investor attention prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	CAR [1:21]			
	(1)	(2)	(3)	(4)
NumNewsFile [-7:-1]	0.236***			
	(3.60)			
NumNewsFile [-14:-1]		0.150***		
		(4.09)		
NumNewsFile [-30:-1]			0.060***	
			(3.25)	
NumNewsFile [-60:-1]				0.023**
				(2.28)
UndwrtReputation	-8.611	-8.918	-8.718	-8.592
	(-1.45)	(-1.51)	(-1.47)	(-1.45)
FirmSize	0.649***	0.609***	0.649***	0.697***
	(3.07)	(2.87)	(3.04)	(3.25)
PriorQtrEarnSurpFile	0.706**	0.712**	0.721**	0.718**
	(2.16)	(2.18)	(2.21)	(2.20)
PriorMktRetFile	16.078**	16.806***	17.058***	16.824***
	(2.58)	(2.70)	(2.73)	(2.70)
MidFilePrice	-0.031**	-0.033**	-0.032**	-0.031**
	(-2.37)	(-2.47)	(-2.40)	(-2.33)
Constant	45.202*	45.516*	45.333*	44.661*
	(1.84)	(1.85)	(1.84)	(1.81)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.155	0.156	0.155	0.154
Observations	4742	4742	4742	4742

Table 5: **Relationship between the SEO announcement effect, post-announcement drift, and subsequent operating performance**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000 - 2018. $ROA_{1(2,3,4)}$ is the ratio of net income over the book value of total assets measured over one (two, three, four) quarters after the SEO announcement. $Cash\ Flow_{1(2,3,4)}$ is the ratio of income before extraordinary items plus depreciation to the book value of total assets measured over one (two, three, four) quarters after the SEO announcement. *Standardized CAR [0:0]* is the standardized abnormal return on SEO firm's equity on the SEO announcement day with its mean of zero and standard deviation of one. *Standardized CAR [1:21]* is the standardized cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date with its mean of zero and standard deviation of one. *NumNewsFile [-7:-1]*, *NumNewsFile [-14:-1]*, *NumNewsFile [-30:-1]*, and *NumNewsFile [-60:-1]* are measures of investor attention prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between SEO announcement effect, post-announcement drift, and subsequent ROA

Dependent Variable	ROA_1	ROA_2	ROA_3	ROA_4
	(1)	(2)	(3)	(4)
Standardized CAR [0:0]	0.182	0.459**	0.787**	0.716*
	(1.56)	(2.16)	(2.54)	(1.76)
Standardized CAR [1:21]	0.286**	0.509**	0.868***	1.236***
	(2.42)	(2.39)	(2.79)	(3.05)
UndwrtReputation	3.269	7.207	10.963*	14.924*
	(1.30)	(1.58)	(1.66)	(1.73)
FirmSize	1.262***	2.498***	3.706***	4.862***
	(14.44)	(15.79)	(16.09)	(16.17)
PriorQtrEarnSurpFile	-0.740***	-1.075***	-2.145***	-2.573***
	(-5.33)	(-4.24)	(-5.77)	(-5.29)
PriorMktRetFile	0.497	-0.838	-0.769	-1.638
	(0.19)	(-0.18)	(-0.11)	(-0.18)
MidFilePrice	0.032***	0.064***	0.092***	0.126***
	(5.72)	(6.37)	(6.31)	(6.53)
Constant	-2.402	-7.847	51.650*	-79.531***
	(-0.23)	(-0.42)	(1.92)	(-2.81)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.462	0.514	0.531	0.539
Observations	4724	4688	4561	4438

Panel B: Relationship between SEO announcement effect, post-announcement drift, and subsequent Cash Flow

Dependent Variable	Cash $Flow_1$	Cash $Flow_2$	Cash $Flow_3$	Cash $Flow_4$
	(1)	(2)	(3)	(4)
Standardized CAR [0:0]	0.184	0.438*	0.722**	0.464
	(1.38)	(1.80)	(2.03)	(0.98)
Standardized CAR [1:21]	0.240*	0.519**	0.903**	1.274***
	(1.78)	(2.12)	(2.52)	(2.72)
UndwrtReputation	2.991	7.494	12.083	17.188
	(0.99)	(1.36)	(1.50)	(1.63)
FirmSize	1.493***	2.985***	4.424***	5.795***
	(14.50)	(15.89)	(16.04)	(16.06)
PriorQtrEarnSurpFile	-0.768***	-1.110***	-2.161***	-2.711***
	(-4.99)	(-3.93)	(-5.17)	(-4.97)
PriorMktRetFile	0.149	-2.389	-3.995	-4.646
	(0.05)	(-0.40)	(-0.46)	(-0.41)
MidFilePrice	0.035***	0.070***	0.102***	0.140***
	(5.06)	(5.59)	(5.59)	(5.83)
Constant	-3.819	-9.295	52.822*	-80.270**
	(-0.34)	(-0.45)	(1.79)	(-2.57)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.472	0.525	0.543	0.553
Observations	4076	4010	3858	3728

Table 6: **Relationship between abnormal investor attention and market reaction upon SEO announcement**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000 - 2018. *CAR [0:0]* is the abnormal return on SEO firm's equity on the SEO announcement day. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *AbnNumNewsFile [-7:-1]*, *AbnNumNewsFile [-14:-1]*, *AbnNumNewsFile [-30:-1]*, and *AbnNumNewsFile [-60:-1]* are measures of abnormal investor attention prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between abnormal investor attention and SEO announcement effect

Dependent Variable	CAR [0:0]			
	(1)	(2)	(3)	(4)
AbnNumNewsFile [-7:-1]	-0.041**			
	(-2.42)			
AbnNumNewsFile [-14:-1]		-0.041***		
		(-3.35)		
AbnNumNewsFile [-30:-1]			-0.019***	
			(-2.73)	
AbnNumNewsFile [-60:-1]				-0.011**
				(-2.44)
UndwrtReputation	-1.672	-1.508	-1.492	-1.471
	(-1.14)	(-1.03)	(-1.02)	(-1.00)
FirmSize	0.120**	0.122**	0.121**	0.121**
	(2.36)	(2.41)	(2.39)	(2.39)
PriorQtrEarnSurpFile	0.260***	0.262***	0.259***	0.258***
	(3.23)	(3.24)	(3.21)	(3.20)
PriorMktRetFile	0.585	0.582	0.491	0.499
	(0.38)	(0.38)	(0.32)	(0.32)
MidFilePrice	0.008***	0.009***	0.009***	0.009***
	(2.60)	(2.68)	(2.69)	(2.64)
Constant	-7.013	-7.014	-7.060	-6.933
	(-1.15)	(-1.15)	(-1.16)	(-1.14)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.169	0.170	0.169	0.169
Observations	4735	4735	4735	4735

Panel B: Relationship between abnormal investor attention and post-announcement drift				
Dependent Variable	CAR [1:21]			
	(1)	(2)	(3)	(4)
AbnNumNewsFile [-7:-1]	0.071			
	(1.06)			
AbnNumNewsFile [-14:-1]		0.079*		
		(1.65)		
AbnNumNewsFile [-30:-1]			0.019	
			(0.70)	
AbnNumNewsFile [-60:-1]				-0.008
				(-0.46)
UndwrtReputation	-7.751	-8.075	-7.936	-7.581
	(-1.34)	(-1.40)	(-1.37)	(-1.31)
FirmSize	0.802***	0.799***	0.804***	0.810***
	(3.98)	(3.96)	(3.98)	(4.01)
PriorQtrEarnSurpFile	2.122	2.123	2.150	2.160
	(1.56)	(1.56)	(1.58)	(1.59)
PriorMktRetFile	18.330***	18.311***	18.473***	18.505***
	(2.92)	(2.92)	(2.95)	(2.95)
MidFilePrice	-0.031**	-0.031**	-0.031**	-0.030**
	(-2.21)	(-2.26)	(-2.22)	(-2.17)
Constant	44.363*	44.370*	44.362*	44.271*
	(1.86)	(1.86)	(1.86)	(1.85)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.155	0.155	0.155	0.155
Observations	4742	4742	4742	4742

Table 7: **Instrumental variable analysis of the relationship between investor attention and market reaction upon SEO announcement**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000 - 2018. *CAR [0: 0]* is the abnormal return on SEO firm's equity on the SEO announcement day. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *NumNewsFileHat [-7:-1]*, *NumNewsFileHat [-14:-1]*, *NumNewsFileHat [-30:-1]*, and *NumNewsFileHat [-60:-1]* are predicted values of investor attention variables as described in Table 1 (*NumNewsFile [-7:-1]*, *NumNewsFile [-14:-1]*, *NumNewsFile [-30:-1]*, and *NumNewsFile [-60:-1]*) from first-stage regressions. *PriorYrNumNewsFile [-7:-1]*, *PriorYrNumNewsFile [-14:-1]*, *PriorYrNumNewsFile [-30:-1]*, and *PriorYrNumNewsFile [-60:-1]* are instrumental variables which measure investor attention one year prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and SEO announcement effect

Dependent Variable	1st-stage	CAR [0:0]	1st-stage	CAR [0:0]	1st-stage	CAR [0:0]	1st-stage	CAR [0:0]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.570***							
	(37.31)							
NumNewsFileHat [-7:-1]		-0.031						
		(-1.06)						
PriorYrNumNewsFile [-14:-1]			0.832***					
			(58.28)					
NumNewsFileHat [-14:-1]				-0.022*				
				(-1.79)				
PriorYrNumNewsFile [-30:-1]					0.899***			
					(71.49)			
NumNewsFileHat [-30:-1]						-0.017***		
						(-3.03)		
PriorYrNumNewsFile [-60:-1]							1.003***	
							(87.87)	
NumNewsFileHat [-60:-1]								-0.009***
								(-3.00)
UndwrtReputation	0.562	-1.646	4.127**	-1.597	9.694***	-1.580	19.804***	-1.597

	(0.46)	(-1.22)	(2.20)	(-1.19)	(2.86)	(-1.18)	(3.68)	(-1.19)
FirmSize	0.388***	0.141***	0.366***	0.150***	0.535***	0.170***	0.429**	0.168***
	(8.79)	(2.71)	(5.39)	(2.99)	(4.35)	(3.42)	(2.18)	(3.39)
PriorQtrEarnSurpFile	0.070	0.259***	0.098	0.258***	0.064	0.257***	0.078	0.258***
	(1.02)	(3.50)	(0.95)	(3.49)	(0.34)	(3.47)	(0.26)	(3.48)
PriorMktRetFile	2.777**	0.594	1.981	0.500	-0.204	0.440	0.926	0.510
	(2.14)	(0.42)	(1.00)	(0.35)	(-0.06)	(0.31)	(0.16)	(0.36)
MidFilePrice	0.004	0.009***	0.011***	0.009***	0.024***	0.009***	0.026**	0.009***
	(1.51)	(2.84)	(2.71)	(2.92)	(3.16)	(2.96)	(2.17)	(2.90)
Constant	-2.644	-2.250	-8.475	-2.303	-5.405	-2.229	-10.969	-2.288
	(-0.63)	(-0.49)	(-1.33)	(-0.51)	(-0.47)	(-0.49)	(-0.60)	(-0.50)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.170		0.170		0.171		0.171
Observations	4735	4735	4735	4735	4735	4735	4735	4735
F Statistics	1391.92		3396.03		5110.44		7720.30	

Panel B: Relationship between investor attention and post-announcement drift

Dependent Variable	1st-stage	CAR [1:21]	1st-stage	CAR [1:21]	1st-stage	CAR [1:21]	1st-stage	CAR [1:21]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.571***							
	(37.33)							
NumNewsFileHat [-7:-1]		0.389***						
		(3.29)						
PriorYrNumNewsFile [-14:-1]			0.832***					
			(58.30)					
NumNewsFileHat [-14:-1]				0.194***				
				(3.91)				
PriorYrNumNewsFile [-30:-1]					0.899***			
					(71.40)			
NumNewsFileHat [-30:-1]						0.084***		
						(3.75)		
PriorYrNumNewsFile [-60:-1]							1.003***	
							(87.72)	
NumNewsFileHat [-60:-1]								0.041***
								(3.54)

UndwrtReputation	0.503	-8.763	4.020**	-9.077*	9.534***	-8.860	19.483***	-8.754
	(0.41)	(-1.61)	(2.14)	(-1.67)	(2.82)	(-1.63)	(3.62)	(-1.61)
FirmSize	0.391***	0.527**	0.369***	0.541***	0.535***	0.570***	0.437**	0.590***
	(8.85)	(2.50)	(5.44)	(2.67)	(4.35)	(2.83)	(2.22)	(2.94)
PriorQtrEarnSurpFile	0.070	0.697**	0.098	0.710**	0.064	0.722**	0.079	0.718**
	(1.03)	(2.32)	(0.95)	(2.37)	(0.34)	(2.41)	(0.27)	(2.39)
PriorMktRetFile	2.761**	15.559***	1.944	16.785***	-0.299	17.134***	0.744	16.786***
	(2.13)	(2.71)	(0.99)	(2.93)	(-0.08)	(2.99)	(0.13)	(2.93)
MidFilePrice	0.004	-0.032***	0.011***	-0.033***	0.024***	-0.032***	0.026**	-0.031***
	(1.51)	(-2.65)	(2.72)	(-2.75)	(3.17)	(-2.67)	(2.16)	(-2.59)
Constant	-2.645	-0.627	-8.476	-0.796	-5.398	-2.016	-10.968	-1.767
	(-0.63)	(-0.03)	(-1.33)	(-0.04)	(-0.47)	(-0.11)	(-0.60)	(-0.10)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.154		0.156		0.155		0.153
Observations	4742	4742	4742	4742	4742	4742	4742	4742
F Statistics	1393.29		3398.43		5097.53		7693.97	

Table 8: **Relationship between investor attention and post-SEO participation of institutional investors**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *InstN* is the number of institutional investors holding SEO firms' shares at the end of the first post-issue fiscal quarter. *NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]* are measures of investor attention prior to the SEO issue date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year $\times$ industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	InstN							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NumNewsIss [-7:-1]	4.593***	4.591***						
	(17.58)	(17.54)						
NumNewsIss [-14:-1]			3.330***	3.326***				
			(18.70)	(18.62)				
NumNewsIss [-30:-1]					1.723***	1.718***		
					(14.74)	(14.64)		
NumNewsIss [-60:-1]							0.866***	0.863***
							(13.16)	(13.06)
Underpricing		0.484**		0.436*		0.355		0.311
		(2.16)		(1.91)		(1.31)		(0.95)
UndwrtReputation	19.050	20.258	22.962	23.691	71.773**	71.129**	79.433**	77.654*
	(0.69)	(0.73)	(0.79)	(0.82)	(2.06)	(2.04)	(1.97)	(1.92)
FirmSize	43.146***	43.206***	42.261***	42.290***	43.815***	43.834***	45.555***	45.535***
	(41.47)	(41.25)	(38.61)	(38.36)	(33.09)	(32.75)	(29.20)	(28.89)
PriorQtrEarnSurpIss	-1.822	-1.602	-1.649	-1.453	-1.459	-1.313	-1.869	-1.731
	(-1.59)	(-1.39)	(-1.45)	(-1.27)	(-1.10)	(-0.99)	(-1.28)	(-1.18)
PriorMktRetIss	-7.902	-11.573	-0.806	-4.363	31.472	28.633	24.887	22.354
	(-0.24)	(-0.36)	(-0.02)	(-0.13)	(0.80)	(0.72)	(0.52)	(0.47)
MidFilePrice	1.032***	1.042***	1.067***	1.076***	1.238***	1.246***	1.336***	1.345***
	(14.31)	(14.42)	(14.02)	(14.12)	(13.26)	(13.32)	(11.84)	(11.88)
Constant	-262.282**	-268.652***	-271.063***	-276.603***	-304.249***	-308.994***	-344.022*	-432.622***
	(-2.53)	(-2.59)	(-2.64)	(-2.70)	(-2.76)	(-2.80)	(-1.77)	(-2.62)

Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.729	0.728	0.744	0.743	0.766	0.765	0.781	0.780
Observations	3883	3854	3569	3541	2796	2770	2156	2132

Table 9: **Relationship between investor attention and SEO underpricing**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]* are measures of investor attention prior to the SEO issue date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	Underpricing			
	(1)	(2)	(3)	(4)
NumNewsIss [-7:-1]	0.042**			
	(2.00)			
NumNewsIss [-14:-1]		0.050***		
		(3.44)		
NumNewsIss [-30:-1]			0.033***	
			(3.52)	
NumNewsIss [-60:-1]				0.015***
				(2.93)
UndwrtReputation	-4.869**	-4.688**	-1.553	1.696
	(-2.22)	(-1.97)	(-0.56)	(0.55)
FirmSize	-0.363***	-0.400***	-0.500***	-0.423***
	(-4.42)	(-4.50)	(-4.78)	(-3.57)
PriorQtrEarnSurpIss	-0.285***	-0.279***	-0.237**	-0.321***
	(-3.44)	(-3.28)	(-2.43)	(-3.04)
PriorMktRetIss	4.582*	4.735*	3.681	1.736
	(1.77)	(1.72)	(1.17)	(0.47)
MidFilePrice	-0.014**	-0.014**	-0.009	-0.011
	(-2.50)	(-2.20)	(-1.23)	(-1.35)
Constant	14.443*	14.525*	15.429*	10.844
	(1.74)	(1.72)	(1.75)	(0.85)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.228	0.243	0.288	0.326
Observations	3920	3601	2817	2166

Table 10: **Relationship between investor attention and SEO valuation of issuing firms**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *QOPAdj* is the industry-adjusted Q ratio calculated using the SEO offer price. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and the SEO offer price. Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]* are measures of investor attention prior to the SEO issue date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	QOPAdj							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NumNewsIss [-7:-1]	0.032***	0.032***						
	(4.00)	(4.00)						
NumNewsIss [-14:-1]			0.024***	0.024***				
			(4.29)	(4.35)				
NumNewsIss [-30:-1]					0.014***	0.014***		
					(4.23)	(4.25)		
NumNewsIss [-60:-1]							0.005***	0.005***
							(3.04)	(3.13)
Underpricing		-0.010		-0.010		-0.009		-0.015*
		(-1.55)		(-1.43)		(-1.19)		(-1.95)
UndwrtReputation	4.103***	4.019***	3.746***	3.660***	2.737***	2.670***	2.321**	2.278**
	(4.88)	(4.77)	(4.17)	(4.06)	(2.81)	(2.73)	(2.34)	(2.29)
FirmSize	-0.494***	-0.498***	-0.487***	-0.491***	-0.442***	-0.446***	-0.409***	-0.415***
	(-15.78)	(-15.78)	(-14.53)	(-14.53)	(-12.07)	(-12.04)	(-10.78)	(-10.82)
PriorQtrEarnSurpIss	0.022	0.019	0.021	0.018	-0.007	-0.009	-0.019	-0.024
	(0.69)	(0.60)	(0.64)	(0.55)	(-0.20)	(-0.27)	(-0.57)	(-0.71)
PriorMktRetIss	-0.127	-0.071	-0.305	-0.249	-0.209	-0.215	0.176	0.150
	(-0.13)	(-0.07)	(-0.29)	(-0.24)	(-0.19)	(-0.19)	(0.15)	(0.13)
MidFilePrice	0.030***	0.030***	0.030***	0.030***	0.025***	0.025***	0.021***	0.021***
	(13.89)	(13.79)	(12.65)	(12.56)	(9.59)	(9.54)	(7.80)	(7.72)

Constant	-1.815	-1.666	-1.903	-1.759	-1.644	-1.510	10.848**	15.464***
	(-0.57)	(-0.52)	(-0.59)	(-0.55)	(-0.53)	(-0.49)	(2.25)	(3.78)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.120	0.119	0.119	0.118	0.126	0.125	0.173	0.173
Observations	3940	3911	3621	3593	2835	2809	2185	2161

Table 11: **Instrumental variable analysis of the relationship between investor attention and post-SEO participation of institutional investors**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *InstN* is the number of institutional investors holding SEO firms' shares at the end of the first post-issue fiscal quarter. *NumNewsIssHat [-7:-1]*, *NumNewsIssHat [-14:-1]*, *NumNewsIssHat [-30:-1]*, and *NumNewsIssHat [-60:-1]* are predicted values of investor attention variables as described in Table 1 (*NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]*) from first-stage regressions. *PriorYrNumNewsFile [-7:-1]*, *PriorYrNumNewsFile [-14:-1]*, *PriorYrNumNewsFile [-30:-1]*, and *PriorYrNumNewsFile [-60:-1]* are instrumental variables which measure investor attention one year prior to the SEO announcement date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year $\times$ industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	1st-stage	InstN	1st-stage	InstN	1st-stage	InstN	1st-stage	InstN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.566***							
	(22.51)							
NumNewsIssHat [-7:-1]		9.437***						
		(13.98)						
PriorYrNumNewsFile [-14:-1]			0.687***					
			(32.49)					
NumNewsIssHat [-14:-1]				5.386***				
				(16.87)				
PriorYrNumNewsFile [-30:-1]					0.667***			
					(38.12)			
NumNewsIssHat [-30:-1]						2.285***		
						(13.73)		
PriorYrNumNewsFile [-60:-1]							0.840***	
							(42.00)	
NumNewsIssHat [-60:-1]								1.105***
								(13.72)
Underpricing	0.021	0.329	0.053***	0.266	0.113***	0.254	0.165*	0.223
	(1.52)	(1.52)	(2.62)	(1.26)	(2.96)	(1.04)	(1.96)	(0.78)

UndwrtReputation	3.753**	5.929	8.929***	9.600	9.892**	67.705**	21.689**	75.417**
	(2.16)	(0.22)	(3.46)	(0.36)	(2.01)	(2.17)	(2.07)	(2.13)
FirmSize	0.825***	37.479***	1.102***	38.176***	1.898***	41.647***	2.578***	43.646***
	(12.71)	(30.09)	(11.37)	(33.03)	(10.26)	(32.12)	(6.34)	(30.06)
PriorQtrEarnSurpIss	-0.003	-1.492	-0.043	-1.308	-0.091	-1.225	-0.079	-1.643
	(-0.04)	(-1.36)	(-0.42)	(-1.24)	(-0.48)	(-1.03)	(-0.21)	(-1.28)
PriorMktRetIss	3.196	-32.790	2.184	-15.350	2.455	25.656	10.008	18.819
	(1.57)	(-1.05)	(0.73)	(-0.50)	(0.44)	(0.72)	(0.81)	(0.45)
MidFilePrice	-0.000	1.054***	0.005	1.074***	0.006	1.249***	-0.018	1.356***
	(-0.07)	(15.25)	(0.79)	(15.24)	(0.43)	(14.92)	(-0.60)	(13.63)
Constant	-8.597	-217.755**	-7.670	-247.806***	-5.408	-296.564***	-13.311	-416.064***
	(-1.32)	(-2.19)	(-0.84)	(-2.61)	(-0.35)	(-3.01)	(-0.31)	(-2.86)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.699		0.732		0.763		0.779
Observations	3854	3854	3541	3541	2770	2770	2132	2132
F Statistics	506.58		1055.81		1452.81		1763.62	

Table 12: **Instrumental variable analysis of the relationship between investor attention and SEO underpricing**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *NumNewsIssHat [-7:-1]*, *NumNewsIssHat [-14:-1]*, *NumNewsIssHat [-30:-1]*, and *NumNewsIssHat [-60:-1]* are predicted values of investor attention variables as described in Table 1 (*NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]*) from first-stage regressions. *PriorYrNumNewsFile [-7:-1]*, *PriorYrNumNewsFile [-14:-1]*, *PriorYrNumNewsFile [-30:-1]*, and *PriorYrNumNewsFile [-60:-1]* are instrumental variables which measure investor attention one year prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	1st-stage	Underpricing	1st-stage	Underpricing	1st-stage	Underpricing	1st-stage	Underpricing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.569***							
	(22.83)							
NumNewsIssHat [-7:-1]		0.090*						
		(1.78)						
PriorYrNumNewsFile [-14:-1]			0.683***					
			(32.60)					
NumNewsIssHat [-14:-1]				0.064**				
				(2.49)				
PriorYrNumNewsFile [-30:-1]					0.666***			
					(38.28)			
NumNewsIssHat [-30:-1]						0.031**		
						(2.32)		
PriorYrNumNewsFile [-60:-1]							0.840***	
							(42.41)	
NumNewsIssHat [-60:-1]								0.016**
								(2.53)
UndwrtReputation	3.468**	-5.000**	8.479***	-4.778**	10.361**	-1.540	22.102**	1.689
	(2.02)	(-2.49)	(3.31)	(-2.22)	(2.12)	(-0.62)	(2.14)	(0.62)
FirmSize	0.817***	-0.420***	1.069***	-0.427***	1.809***	-0.492***	2.496***	-0.429***
	(12.94)	(-4.53)	(11.30)	(-4.68)	(10.02)	(-4.87)	(6.32)	(-3.93)

PriorQtrEarnSurpIss	0.000	-0.284***	-0.016	-0.278***	-0.101	-0.237***	-0.159	-0.321***
	(0.01)	(-3.76)	(-0.17)	(-3.61)	(-0.59)	(-2.74)	(-0.45)	(-3.46)
PriorMktRetIss	3.240	4.366*	2.695	4.653*	2.606	3.693	10.054	1.726
	(1.60)	(1.85)	(0.91)	(1.87)	(0.47)	(1.31)	(0.82)	(0.54)
MidFilePrice	-0.001	-0.014***	0.004	-0.014**	0.005	-0.009	-0.021	-0.011
	(-0.15)	(-2.72)	(0.54)	(-2.43)	(0.42)	(-1.38)	(-0.73)	(-1.54)
Constant	-8.291	14.931**	-6.765	14.699*	-3.482	15.388**	-11.290	10.892
	(-1.28)	(1.98)	(-0.74)	(1.92)	(-0.22)	(1.96)	(-0.26)	(0.97)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.227		0.243		0.288		0.326
Observations	3920	3920	3601	3601	2817	2817	2166	2166
F Statistics	521.35		1062.51		1465.18		1798.93	

Table 13: **Instrumental variable analysis of the relationship between investor attention and SEO valuation of issuing firms**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *QOPAdj* is the industry-adjusted Q ratio calculated using the SEO offer price. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and the SEO offer price. Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *NumNewsIssHat [-7:-1]*, *NumNewsIssHat [-14:-1]*, *NumNewsIssHat [-30:-1]*, and *NumNewsIssHat [-60:-1]* are predicted values of investor attention variables as described in Table 1 (*NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]*) from first-stage regressions. *PriorYrNumNewsFile [-7:-1]*, *PriorYrNumNewsFile [-14:-1]*, *PriorYrNumNewsFile [-30:-1]*, and *PriorYrNumNewsFile [-60:-1]* are instrumental variables which measure investor attention one year prior to the SEO announcement date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	1st-stage	QOPAdj	1st-stage	QOPAdj	1st-stage	QOPAdj	1st-stage	QOPAdj
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.568***							
	(22.77)							
NumNewsIss [-7:-1]		0.037*						
		(1.89)						
PriorYrNumNewsFile [-14:-1]			0.681***					
			(32.46)					
NumNewsIss [-14:-1]				0.016*				
				(1.67)				
PriorYrNumNewsFile [-30:-1]					0.664***			
					(38.13)			
NumNewsIss [-30:-1]						0.007		
						(1.47)		
PriorYrNumNewsFile [-60:-1]							0.838***	
							(42.24)	
NumNewsIss [-60:-1]								0.004*
								(1.89)
Underpricing	0.020	-0.011*	0.052***	-0.009	0.106***	-0.008	0.157*	-0.015**

	(1.48)	(-1.72)	(2.63)	(-1.48)	(2.86)	(-1.15)	(1.91)	(-2.15)
UndwrtReputation	3.550**	4.005***	8.723***	3.714***	10.428**	2.715***	21.760**	2.291***
	(2.06)	(5.21)	(3.40)	(4.54)	(2.13)	(3.11)	(2.10)	(2.63)
FirmSize	0.827***	-0.503***	1.089***	-0.476***	1.857***	-0.418***	2.549***	-0.404***
	(13.03)	(-14.09)	(11.48)	(-13.69)	(10.24)	(-11.68)	(6.43)	(-11.47)
PriorQtrEarnSurpIss	0.006	0.019	-0.001	0.017	-0.076	-0.010	-0.106	-0.025
	(0.09)	(0.66)	(-0.01)	(0.60)	(-0.44)	(-0.33)	(-0.30)	(-0.83)
PriorMktRetIss	3.040	-0.091	2.305	-0.207	2.282	-0.179	9.814	0.169
	(1.50)	(-0.10)	(0.78)	(-0.22)	(0.41)	(-0.18)	(0.80)	(0.16)
MidFilePrice	-0.000	0.030***	0.004	0.030***	0.006	0.025***	-0.019	0.021***
	(-0.07)	(15.17)	(0.65)	(13.89)	(0.49)	(10.68)	(-0.66)	(8.78)
Constant	-8.597	-1.616	-7.520	-1.866	-5.114	-1.664	-12.899	15.372***
	(-1.32)	(-0.56)	(-0.82)	(-0.64)	(-0.33)	(-0.60)	(-0.30)	(4.28)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.119		0.117		0.123		0.172
Observations	3911	3911	3593	3593	2809	2809	2161	2161
F Statistics	518.49		1053.76		1453.65		1783.83	

Table 14: **Robustness Check: Relationship between investor attention and market reaction upon SEO announcement**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000 - 2018. *CAR [0:0]* is the abnormal return on SEO firm's equity on the SEO announcement day. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *NumEdgarFile [-7:-1]*, *NumEdgarFile [-14:-1]*, *NumEdgarFile [-30:-1]*, and *NumEdgarFile [-60:-1]* are defined as the number of unique IP addresses downloaded 10-K, 10-Q, and 8-K filings of the SEO firm in the SEC EDGAR system prior to the SEO announcement date. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year $\times$ industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and SEO announcement effect

Dependent Variable	CAR [0:0]			
	(1)	(2)	(3)	(4)
NumEdgarFile [-7:-1]	-0.003**			
	(-2.08)			
NumEdgarFile [-14:-1]		-0.001*		
		(-1.76)		
NumEdgarFile [-30:-1]			-0.001**	
			(-2.07)	
NumEdgarFile [-60:-1]				-0.001**
				(-2.28)
UndwrtReputation	-1.539	-1.526	-1.526	-1.538
	(-1.05)	(-1.04)	(-1.04)	(-1.05)
FirmSize	0.102**	0.101*	0.104**	0.107**
	(1.97)	(1.95)	(2.01)	(2.07)
PriorQtrEarnSurpFile	0.177**	0.176**	0.175**	0.176**
	(2.23)	(2.22)	(2.21)	(2.22)
PriorMktRetFile	1.398	1.386	1.400	1.421
	(0.88)	(0.87)	(0.88)	(0.90)
MidFilePrice	0.009**	0.008**	0.009**	0.008**
	(2.46)	(2.43)	(2.45)	(2.45)
Constant	-3.118	-3.273	-3.195	-3.200
	(-0.49)	(-0.52)	(-0.50)	(-0.50)
Industry $\times$ Year FE	Yes	Yes	Yes	Yes
R^2	0.187	0.187	0.187	0.188
Observations	3631	3631	3631	3631

Panel B: Relationship between investor attention and post-announcement drift				
Dependent Variable	CAR [1:21]			
	(1)	(2)	(3)	(4)
NumEdgarFile [-7:-1]	0.015**			
	(2.57)			
NumEdgarFile [-14:-1]		0.009***		
		(2.68)		
NumEdgarFile [-30:-1]			0.004**	
			(2.31)	
NumEdgarFile [-60:-1]				0.002**
				(2.05)
UndwrtReputation	-8.581	-8.689	-8.633	-8.551
	(-1.36)	(-1.38)	(-1.37)	(-1.36)
FirmSize	0.750***	0.743***	0.748***	0.749***
	(3.39)	(3.35)	(3.37)	(3.37)
PriorQtrEarnSurpFile	0.236	0.240	0.245	0.243
	(0.69)	(0.70)	(0.72)	(0.71)
PriorMktRetFile	7.450	7.527	7.440	7.357
	(1.09)	(1.11)	(1.09)	(1.08)
MidFilePrice	-0.025*	-0.025*	-0.025*	-0.024
	(-1.70)	(-1.71)	(-1.65)	(-1.61)
Constant	32.499	32.904	33.132	33.549
	(1.19)	(1.21)	(1.22)	(1.23)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.165	0.165	0.164	0.164
Observations	3636	3636	3636	3636

Table 15: **Robustness Check: Relationship between investor attention and SEO Characteristics**

The sample consists of seasoned equity offerings (SEOs) conducted in 2003-2017. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *InstN* is the number of institutional investors holding SEO firms' shares at the end of the first post-issue fiscal quarter. *QOPAdj* is the industry-adjusted Q ratio calculated using the SEO offer price. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and the SEO offer price. Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *NumEdgarIss [-7:-1]*, *NumEdgarIss [-14:-1]*, *NumEdgarIss [-30:-1]*, and *NumEdgarIss [-60:-1]* are defined as the number of unique IP addresses downloaded 10-K, 10-Q, and 8-K filings of the SEO firm in the SEC EDGAR system prior to the SEO issue date. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and SEO underpricing

Dependent Variable	Underpricing			
	(1)	(2)	(3)	(4)
NumEdgarIss [-7:-1]	0.012***			
	(3.76)			
NumEdgarIss [-14:-1]		0.008***		
		(3.98)		
NumEdgarIss [-30:-1]			0.004***	
			(3.00)	
NumEdgarIss [-60:-1]				0.003***
				(3.23)
UndwrtReputation	-4.176*	-3.070	0.399	2.569
	(-1.68)	(-1.13)	(0.12)	(0.73)
FirmSize	-0.358***	-0.373***	-0.435***	-0.401***
	(-3.90)	(-3.72)	(-3.66)	(-3.01)
PriorQtrEarnSurpIss	-0.054	-0.053	-0.019	-0.092
	(-0.92)	(-0.87)	(-0.26)	(-1.21)
PriorMktRetIss	4.626	4.892	3.419	0.414
	(1.43)	(1.41)	(0.87)	(0.09)
MidFilePrice	-0.010	-0.010	-0.004	-0.007
	(-1.37)	(-1.24)	(-0.46)	(-0.70)
Constant	24.447***	24.301***	22.680***	3.820
	(3.61)	(3.50)	(3.00)	(0.43)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.243	0.258	0.283	0.323
Observations	3098	2827	2287	1841

Panel B: Relationship between investor attention and post-SEO participation of institutional investors

Dependent Variable	InstN			
	(1)	(2)	(3)	(4)
NumEdgarIss [-7:-1]	0.414***			
	(10.43)			
NumEdgarIss [-14:-1]		0.303***		
		(11.81)		
NumEdgarIss [-30:-1]			0.171***	
			(9.45)	
NumEdgarIss [-60:-1]				0.127***
				(9.86)
Underpricing	0.430*	0.459*	0.358	0.349
	(1.65)	(1.71)	(1.19)	(1.00)
UndwrtReputation	-14.336	7.273	34.537	39.472
	(-0.45)	(0.21)	(0.86)	(0.87)
FirmSize	44.148***	43.728***	45.225***	45.933***
	(36.58)	(34.07)	(29.96)	(26.22)
PriorQtrEarnSurpIss	-0.921	-1.041	-0.688	-0.623
	(-1.05)	(-1.18)	(-0.68)	(-0.59)
PriorMktRetIss	71.393*	77.791*	93.431*	94.574*
	(1.71)	(1.78)	(1.90)	(1.68)
MidFilePrice	1.331***	1.432***	1.442***	1.536***
	(14.49)	(14.68)	(13.23)	(12.17)
Constant	-163.906*	-169.931*	-222.445**	-356.006***
	(-1.89)	(-1.96)	(-2.37)	(-3.10)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.724	0.739	0.760	0.773
Observations	3038	2773	2245	1809

Panel C: Relationship between investor attention and SEO valuation of issuing firms				
Dependent Variable	QOPAdj			
	(1)	(2)	(3)	(4)
NumEdgarIss [-7:-1]	0.004***			
	(4.21)			
NumEdgarIss [-14:-1]		0.002***		
		(3.39)		
NumEdgarIss [-30:-1]			0.002***	
			(4.07)	
NumEdgarIss [-60:-1]				0.001***
				(3.65)
Underpricing	-0.004	-0.002	0.001	-0.002
	(-0.60)	(-0.31)	(0.15)	(-0.32)
UndwrtReputation	3.239***	2.779***	2.765***	2.123**
	(4.33)	(3.46)	(2.98)	(2.05)
FirmSize	-0.440***	-0.410***	-0.411***	-0.419***
	(-15.87)	(-13.78)	(-11.86)	(-10.66)
PriorQtrEarnSurpIss	0.011	0.010	-0.011	-0.018
	(0.60)	(0.57)	(-0.51)	(-0.82)
PriorMktRetIss	0.297	0.368	-0.310	-0.533
	(0.30)	(0.36)	(-0.27)	(-0.41)
MidFilePrice	0.025***	0.022***	0.022***	0.023***
	(11.83)	(9.79)	(8.79)	(8.03)
Constant	0.731	0.540	0.217	1.382
	(0.36)	(0.26)	(0.10)	(0.52)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.131	0.120	0.132	0.146
Observations	3096	2825	2285	1840

Appendix to "The Role of Investor Attention in Seasoned Equity Offerings: Theory and Evidence"

A.1 List of Constants in Propositions and Proofs

$$
\begin{aligned}
A_a &\equiv f^a(\sigma_1^a)^{-2}[1+\rho^{-2}(\sigma_1^a)^{-2}\sigma_x^{-2}] > 0, && \text{(A.1)}\\
A_u &\equiv f^u\sigma_0^{-2}[1+\frac{\frac{1}{2}\sigma_0^{-2}}{\rho^2\sigma_x^2+\frac{1}{2}\sigma_e^{-2}}] > 0, && \text{(A.2)}\\
B_0 &\equiv \frac{A_a(\sigma_1^a)^2+A_u\sigma_0^2}{A_a+A_u} > 0, && \text{(A.3)}\\
B_1 &\equiv \frac{A_a(\sigma_1^a)^2+f^u}{A_a+A_u} > 0, && \text{(A.4)}\\
E &\equiv \frac{A_a}{f^a}(\frac{A_u}{A_a+A_u})^2+(\sigma_1^a)^{-2}\sigma_e^2\sigma_0^{-2} > 0, && \text{(A.5)}\\
F_0 &= \frac{A_u}{A_a+A_u}[\frac{A_a}{f^a}B_0-\frac{A_a}{f^a}(\sigma_1^a)^2+1], && \text{(A.6)}\\
F_1 &= \frac{A_u}{A_a+A_u}[\frac{A_a}{f^a}B_1-\frac{A_a}{f^a}(\sigma_1^a)^2+1], && \text{(A.7)}\\
G &\equiv \frac{A_a}{f^a}[B_1-(\sigma_1^a)^2]^2+2B_1-(\sigma_1^a)^2-\frac{F_1^2}{E}+\rho^{-2}\sigma_x^{-2}, && \text{(A.8)}\\
H_0 &\equiv B_1+\frac{A_a}{A_a+A_u}\frac{F_1}{E}, && \text{(A.9)}\\
H_1 &\equiv -\frac{A_a}{f^a}B_0B_1+\rho^{-2}(\sigma_1^a)^{-2}\sigma_x^{-2}(B_0+B_1)-\rho^{-2}\sigma_x^{-2}+\frac{F_0F_1}{E}, && \text{(A.10)}\\
J &\equiv \frac{A_u}{f^u}(\frac{A_a}{A_a+A_u})^2+(\sigma_1^a)^{-2}\sigma_e^2\sigma_0^{-2} > 0, && \text{(A.11)}\\
K &\equiv \frac{A_u}{f^u}\frac{B_1^2}{J}(\sigma_1^a)^{-2}\sigma_e^2\sigma_0^{-2}+\rho^{-2}\sigma_x^{-2} > 0, && \text{(A.12)}\\
L_0 &\equiv \frac{B_1}{J}(\sigma_1^a)^{-2}\sigma_e^2\sigma_0^{-2}, && \text{(A.13)}\\
P_a &\equiv f^a[\frac{1}{E}(\frac{A_a}{A_a+A_u})^2+\frac{H_0^2}{G}]^{-1}, && \text{(A.14)}\\
Q_a &\equiv f^a[\frac{1}{E}(\frac{A_a}{A_a+A_u})^2+\frac{H_0^2}{G}]^{-1}(B_0+\frac{A_a}{A_a+A_u}\frac{F_0}{E}+\frac{H_0H_1}{G}), && \text{(A.15)}\\
P_u &\equiv f^u[\frac{1}{J}(\frac{A_a}{A_a+A_u})^2+\frac{L_0^2}{K}]^{-1}, && \text{(A.16)}\\
Q_u &\equiv f^u\{B_0[\frac{1}{J}(\frac{A_a}{A_a+A_u})^2+\frac{L_0^2}{K}]^{-1}-\frac{A_u}{f^u}(B_0-\sigma_0^2)1\}. && \text{(A.17)}
\end{aligned}
$$

Both A_a and A_u are positive because they both consist of sums and products of variances terms (σ's) and positive parameters (ρ, f^a, and f^u). This further confirms the positivity of B_0, B_1, E, J, and K.

A.2 Proof of Propositions

Proof of Proposition 1. We solve the investors' utility maximization problems (UMP) backwards.

- At $t = 2$, an investor of type i (for both $i = a$ and $i = u$) solves the following UMP

$$\max_{D_2^i} E_2^i[-\exp(-\rho W_3^i)], \text{ where } W_3^i = W_2^i + D_2^i(f - S_2) \tag{A.18}$$

The only random component here is $f = \mu + z$, which follows normal distribution, hence the above UMP is equivalent to

$$\max_{D_2^i} D_2^i(\mu + E_2^i[z] - S_2) - \frac{\rho}{2} D_2^i V_2^i[z] D_2^i = D_2^i(\mu + \hat{z}_2 - S_2) - \frac{\rho}{2}(D_2^i)^2 \sigma_2^2 \tag{A.19}$$

By the standard optimization procedure, the optimal demand is therefore

$$D_2^i = \frac{\mu + \hat{z}_2 - S_2}{\rho\sigma_2^2}, \text{ for } i = a, u. \tag{A.20}$$

To clear the markets, $\sum_{i=a,u} D_2^i = \bar{x} + x_1 + x_2$, hence

$$\bar{x} + x_1 + x_2 = \frac{\mu + \hat{z}_2 - S_2}{\rho\sigma_2^2}, \tag{A.21}$$

and the equilibrium price at $t = 2$ is therefore

$$S_2 = \mu + \hat{z}_2 - \rho\sigma_2^2(\bar{x} + x_1 + x_2). \tag{A.22}$$

The consequent value function (optimized utility), after substituting in (A.20) and (A.22), is

$$E_2^i[-\exp(-\rho W_3^i)] \quad = \quad -\exp\{-\rho W_2^i - \frac{1}{2}\rho^2\sigma_2^2(\bar{x} + x_1 + x_2)^2\} \tag{A.23}$$

- At $t = 1$, an investor of type i maximizes the following expected utility

$$E_1^i[-\exp(-\rho W_3^i)] = E_1^i[-\exp\{-\rho W_2^i - \frac{1}{2}\rho^2\sigma_2^2(\bar{x} + x_1 + x_2)^2\}]. \tag{A.24}$$

Since the information set of an investor (and thus the corresponding posterior belief on z) depends on the type of the investor, the calculation for (A.24) is carried out separately for type $i = a$ and type $i = u$.

Type-a investors. As to be confirmed, the equilibrium price follows a linear structure that combines the public signal e_1 and the supply shock x_1. Once an attentive investor correctly observes the public signal e_1, he/she can back out the contemporaneous supply shock x_1 from the equilibrium price. Hence, the supply shock x_1 is essentially "known" to a type-a investor and not a random variable in his/her UMP at $t = 1$, and the only relevant random variable here is $x_2 \sim (0, \sigma_x^2)$. Therefore, continuing from (A.24), we get

$$E_1^a[-\exp(-\rho W_3^a)] \quad \propto \quad -\exp\Big(-\rho\{W_1^a + D_1^a[\mu + \hat{z}_1^a - \rho(\sigma_1^a)^2(\bar{x} + x_1) - S_1] + \frac{\rho}{2}(\sigma_1^a)^2(\bar{x} + x_1)^2\} + \frac{1}{2}\rho^2(\sigma_1^a)^2[1 + \rho^{-2}(\sigma_1^a)^{-2}\sigma_x^{-2}]^{-1}[D_1^a - (\bar{x} + x_1)]^2\Big) \tag{A.25}$$

The standard optimization procedure derives the optimal demand by an attentive investor as

$$D_1^a \quad = \quad \rho^{-1}\frac{A_a}{f^a}(\mu + \hat{z}_1^a - S_1) - [\frac{A_a}{f^a}(\sigma_1^a)^2 - 1](\bar{x} + x_1), \tag{A.26}$$

where we applied the constants A_a and A_u as defined in Appendix A.1.

Type-u investors. Inattentive investors are not aware of the SEO announcement immediately at $t = 1$, thus they are unable to back out the exact number of x_1 from the equilibrium price contemporaneously either.[25] Therefore, the calculation of (A.24) for $i = u$ involves taking two expectations: one with respect to the random variable $\hat{z}_2 = \sigma_2^2\sigma_e^{-2}e_1 \sim N(0, \sigma_2^4\sigma_e^{-4}(\sigma_0^2 + \sigma_e^2))$, the other with respect to the random variable $x_1 + x_2 \sim N(0, 2\sigma_x^2)$. Indeed,

$$E_1^u[-\exp(-\rho W_3^u)] \quad \propto \quad -\exp\Big(-\rho\{W_1^u + D_1^u[\mu - \rho\sigma_2^2\bar{x} - S_1] + \frac{\rho}{2}\sigma_2^2\bar{x}^2\} + \frac{\rho^2}{2}(D_1^u)^2\sigma_2^2\sigma_e^{-2}\sigma_0^2 + \frac{1}{2}[\rho^2\sigma_2^2 + \frac{1}{2}\sigma_x^{-2}]^{-1}\rho^4\sigma_2^4(D_1^u - \bar{x})^2\Big) \tag{A.27}$$

The standard optimization procedure implies the optimal demand by an inattentive investor as

$$D_1^u \quad = \quad \rho^{-1}\frac{A_u}{f^u}(\mu - S_1) - [\frac{A_u}{f^u}\sigma_0^2 - 1]\bar{x}. \tag{A.28}$$

The equilibrium price of the risky asset at $t = 1$ is thus

$$S_1 \quad = \quad \mu + \frac{A_a}{A_a + A_u}\hat{z}_1^a - \rho(B_0\bar{x} + B_1 x_1), \tag{A.29}$$

assuming the market clearing condition $\bar{x} + x_1 = f^a D_1^a + f^u D_1^u$ holds.

- At $t = 0$, all investors maximize their expected utility based on their prior belief on the fundamental value of the firm's stock. The calculation is in principle similar to the one for $t = 1$.

[25] At $t = 2$, however, as inattentive investors realized that they missed the SEO announcement at $t = 1$, they could retroactively find the value of x_1 when they looked back at S_1, and thus when they make their portfolio rebalance decision at $t = 2$, x_2 (rather than $x_1 + x_2$ as a whole) is the only random component they do not know directly (but then can be learned from the equilibrium price S_2, same as for type-a investors).

Type-a investors. The calculation of $E_0^a[-\exp(-\rho W_3^a)]$ consists of two expectations of $E_1^a[-\exp(-\rho W_3^a)]$: one with respect to $\hat{z}_1^a \sim N(0, \sigma_2^4\sigma_e^{-4}(\sigma_0^2+\sigma_e^2))$, the other with respect to $x_1 \sim N(0,\sigma_x)$. In fact,

$$
\begin{aligned}
& E_0^a[-\exp(-\rho W_3^a)] \\
\propto & -\exp\Big(-\rho D_0^a(\mu - \rho B_0\bar{x} - S_0) + \frac{1}{2E}[\rho^2(D_0^a)^2(\frac{A_a}{A_a+A_u})^2 + 2\rho^2 D_0^a \frac{A_a}{A_a+A_u} F_0\bar{x}] \\
& + \frac{\rho^2}{2G}(H_0 D_0^a + H_1\bar{x})^2\Big) \qquad \text{(A.30)}
\end{aligned}
$$

By maximizing (A.30), we obtain the optimal demand of a type-a investor at $t=0$ as

$$
D_0^a \quad = \quad \rho^{-1}\frac{P_a}{f^a}(\mu - S_0) - \frac{Q_a}{f^a}\bar{x}. \qquad \text{(A.31)}
$$

Type-u investors. The calculation for $E_0^u[-\exp(-\rho W_3^u)]$ is in essence similar to that for $E_0^a[-\exp(-\rho W_3^a)]$, and we eventually obtain the UMP as

$$
E_0^u[-\exp(-\rho W_3^u)] = -\exp\{-\rho D_0^u(\mu - \rho B_0\bar{x} - S_0) + \frac{\rho^2}{2K}(L_0 D_0^u + L_1\bar{x})^2\}, \qquad \text{(A.32)}
$$

and the optimal demand of a type-u investor at $t=0$ as

$$
D_0^u \quad = \quad \rho^{-1}\frac{P_u}{f^u}(\mu - S_0) - \frac{Q_u}{f^u}\bar{x}, \qquad \text{(A.33)}
$$

The market clearing condition $f^a D_0^a + f^u D_0^u = \bar{x}$ implies that the equilibrium price at $t=0$ is

$$
S_0 \quad = \quad \mu - \rho\frac{Q_a+Q_u+1}{P_a+P_u}\bar{x}. \qquad \text{(A.34)}
$$

This completes the proof for Proposition 1.

Proof of Proposition 2.

(i) The calculation of (19) is straightforward by taking the difference between (11) and (12) and then setting both of $\bar{x}$ and x_1 to zero, i.e.,

$$
(S_1 - S_0)|_{\bar{x}=x_1=0} = \frac{A_a}{A_a+A_u}\frac{\sigma_0^{-2}+\sigma_e^{-2}}{\sigma_e^{-2}}e_1. \qquad \text{(A.35)}
$$

From the discussion in Appendix A.1, both A_a and A_u are positive, and thus the coefficient of e_1 is positive. Since $e_1 < 0$, the right hand side of (A.35) is negative.

(ii) For any given e_1, the magnitude of the abnormal stock return (A.35) depends on the coefficient of e_1, and it suffices to show that this coefficient is an increasing function of f^a. In fact,

$$\frac{\partial}{\partial f^a}\Big(\frac{A_a}{A_a+A_u}\frac{\sigma_0^{-2}+\sigma_e^{-2}}{\sigma_e^{-2}}\Big)=\frac{A_aA_u}{f^af^u(A_a+A_u)^2}\frac{\sigma_0^{-2}+\sigma_e^{-2}}{\sigma_e^{-2}}>0, \tag{A.36}$$

where we apply the fact that $f^u = 1 - f^a$ and the positivity of constants A_a and A_u (as discussed in Appendix A.1).

This completes the proof of Proposition 2.

Proof of Proposition 3.

(i) The calculation of (21) is by taking the difference between (10) and (11) and then setting all of $\bar{x}$, x_1, and x_2 to zero, i.e.,

$$(S_2-S_1)|_{\bar{x}=x_1=x_2=0}=\frac{A_u}{A_a+A_u}\frac{\sigma_0^{-2}+\sigma_e^{-2}}{\sigma_e^{-2}}e_1. \tag{A.37}$$

From the discussion in Appendix A.1, both A_a and A_u are positive, and thus the coefficient of e_1 is positive. Since $e_1 < 0$, the right hand side of (A.37) is negative.

(ii) For any given e_1, the magnitude of the post-announcement drift (A.37) depends on the coefficient of e_1, and it suffices to show that this coefficient is a decreasing function of f^a. In fact,

$$\frac{\partial}{\partial f^a}\Big(\frac{A_u}{A_a+A_u}\frac{\sigma_0^{-2}+\sigma_e^{-2}}{\sigma_e^{-2}}\Big)=-\frac{A_aA_u}{f^af^u(A_a+A_u)^2}\frac{\sigma_0^{-2}+\sigma_e^{-2}}{\sigma_e^{-2}}<0, \tag{A.38}$$

where we apply the fact that $f^u = 1 - f^a$ and the positivity of constants A_a and A_u (as discussed in Appendix A.1).

This completes the proof of Proposition 3.

A.3 Additional Empirical Results

Table A.1: Relationship between the SEO announcement effect, post-announcement drift, and subsequent operating performance

The sample consists of seasoned equity offerings (SEOs) conducted in 2000 - 2018. $ROA_{1(2,3,4)}$ is the ratio of net income over the book value of total assets measured over one (two, three, four) quarters after the SEO announcement. $Cash\ Flow_{1(2,3,4)}$ is the ratio of income before extraordinary items plus depreciation to the book value of total assets measured over one (two, three, four) quarters after the SEO announcement. *CAR [0:0]* is the abnormal return on SEO firm's equity on the SEO announcement day. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *NumNewsFile [-7:-1]*, *NumNewsFile [-14:-1]*, *NumNewsFile [-30:-1]*, and *NumNewsFile [-60:-1]* are measures of investor attention prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year $\times$ industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between SEO announcement effect, post-announcement drift, and subsequent ROA

Dependent Variable	ROA_1		ROA_2		ROA_3		ROA_4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CAR [0:0]	0.043	0.042	0.108**	0.107**	0.186***	0.183**	0.171*	0.166*
	(1.59)	(1.56)	(2.20)	(2.16)	(2.59)	(2.54)	(1.80)	(1.76)
CAR [1:21]		0.016**		0.029**		0.049***		0.070***
		(2.42)		(2.39)		(2.79)		(3.05)
UndwrtReputation	3.127	3.269	6.956	7.207	10.513	10.963*	14.254*	14.924*
	(1.24)	(1.30)	(1.53)	(1.58)	(1.59)	(1.66)	(1.66)	(1.73)
FirmSize	1.276***	1.262***	2.522***	2.498***	3.748***	3.706***	4.921***	4.862***
	(14.62)	(14.44)	(15.96)	(15.79)	(16.29)	(16.09)	(16.37)	(16.17)
PriorQtrEarnSurpFile	-0.729***	-0.740***	-1.054***	-1.075***	-2.119***	-2.145***	-2.539***	-2.573***
	(-5.25)	(-5.33)	(-4.16)	(-4.24)	(-5.69)	(-5.77)	(-5.22)	(-5.29)
PriorMktRetFile	0.765	0.497	-0.355	-0.838	0.114	-0.769	-0.410	-1.638
	(0.29)	(0.19)	(-0.07)	(-0.18)	(0.02)	(-0.11)	(-0.05)	(-0.18)
MidFilePrice	0.032***	0.032***	0.064***	0.064***	0.091***	0.092***	0.123***	0.126***
	(5.63)	(5.72)	(6.28)	(6.37)	(6.21)	(6.31)	(6.41)	(6.53)
Constant	-1.595	-2.313	-6.381	-7.664	54.152**	51.963*	-79.342***	-79.156***
	(-0.15)	(-0.22)	(-0.34)	(-0.41)	(2.01)	(1.93)	(-2.80)	(-2.79)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.461	0.462	0.513	0.514	0.530	0.531	0.538	0.539
Observations	4724	4724	4688	4688	4561	4561	4438	4438

Panel B: Relationship between SEO announcement effect, post-announcement drift, and subsequent Cash Flow

Dependent Variable	Cash $Flow_1$		Cash $Flow_2$		Cash $Flow_3$		Cash $Flow_4$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CAR [0:0]	0.044	0.043	0.105*	0.102*	0.173**	0.168**	0.114	0.108
	(1.42)	(1.38)	(1.85)	(1.80)	(2.09)	(2.03)	(1.03)	(0.98)
CAR [1:21]		0.014*		0.030**		0.051**		0.073***
		(1.78)		(2.12)		(2.52)		(2.72)
UndwrtReputation	2.820	2.991	7.095	7.494	11.334	12.083	16.119	17.188
	(0.93)	(0.99)	(1.29)	(1.36)	(1.40)	(1.50)	(1.53)	(1.63)
FirmSize	1.504***	1.493***	3.007***	2.985***	4.465***	4.424***	5.853***	5.795***
	(14.63)	(14.50)	(16.02)	(15.89)	(16.21)	(16.04)	(16.23)	(16.06)
PriorQtrEarnSurpFile	-0.761***	-0.768***	-1.092***	-1.110***	-2.135***	-2.161***	-2.682***	-2.711***
	(-4.94)	(-4.99)	(-3.87)	(-3.93)	(-5.11)	(-5.17)	(-4.91)	(-4.97)
PriorMktRetFile	0.530	0.149	-1.521	-2.389	-2.396	-3.995	-2.326	-4.646
	(0.16)	(0.05)	(-0.26)	(-0.40)	(-0.28)	(-0.46)	(-0.21)	(-0.41)
MidFilePrice	0.034***	0.035***	0.069***	0.070***	0.100***	0.102***	0.138***	0.140***
	(5.00)	(5.06)	(5.52)	(5.59)	(5.51)	(5.59)	(5.73)	(5.83)
Constant	-3.107	-3.738	-7.738	-9.113	55.517*	53.131*	-79.633**	-79.932**
	(-0.27)	(-0.33)	(-0.38)	(-0.44)	(1.88)	(1.80)	(-2.55)	(-2.56)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.472	0.472	0.525	0.525	0.542	0.543	0.552	0.553
Observations	4076	4076	4010	4010	3858	3858	3728	3728

Table A.2: **Relationship between investor attention and post-SEO secondary market valuation of issuing firms**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *QFTDAdj* and *QFQAdj* are the industry-adjusted Q ratios calculated using the SEO issue day closing price and the price at the end of the first post-issue fiscal quarter, respectively. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and either the SEO issue day closing price (*QFTDAdj*) or the price at the end of the first post-issue fiscal quarter (*QFQAdj*). Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]* are measures of investor attention prior to the SEO issue date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and post-SEO market valuation measured using the SEO issue day closing price

Dependent Variable	QFTDAdj							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NumNewsIss [-7:-1]	0.033***	0.033***						
	(4.03)	(3.98)						
NumNewsIss [-14:-1]			0.025***	0.024***				
			(4.31)	(4.21)				
NumNewsIss [-30:-1]					0.015***	0.014***		
					(4.26)	(4.14)		
NumNewsIss [-60:-1]							0.005***	0.005***
							(3.03)	(2.99)
Underpricing		0.013*		0.013*		0.014*		0.008
		(1.82)		(1.86)		(1.80)		(0.96)
UndwrtReputation	4.069***	4.141***	3.711***	3.784***	2.759***	2.795***	2.403**	2.412**
	(4.65)	(4.73)	(3.97)	(4.04)	(2.72)	(2.75)	(2.34)	(2.35)
FirmSize	-0.519***	-0.517***	-0.512***	-0.509***	-0.465***	-0.461***	-0.425***	-0.426***
	(-15.92)	(-15.79)	(-14.65)	(-14.51)	(-12.16)	(-12.00)	(-10.81)	(-10.78)
PriorQtrEarnSurpIss	0.016	0.020	0.015	0.019	-0.013	-0.010	-0.027	-0.025
	(0.48)	(0.59)	(0.44)	(0.55)	(-0.37)	(-0.28)	(-0.78)	(-0.70)
PriorMktRetIss	-0.146	-0.177	-0.315	-0.352	-0.391	-0.408	-0.097	-0.072

	(-0.14)	(-0.17)	(-0.29)	(-0.32)	(-0.34)	(-0.35)	(-0.08)	(-0.06)
MidFilePrice	0.031***	0.032***	0.031***	0.031***	0.025***	0.026***	0.022***	0.022***
	(13.79)	(13.88)	(12.57)	(12.66)	(9.44)	(9.51)	(7.68)	(7.74)
Constant	-1.421	-1.590	-1.516	-1.697	-1.194	-1.392	15.880***	15.817***
	(-0.43)	(-0.48)	(-0.46)	(-0.51)	(-0.37)	(-0.43)	(3.76)	(3.75)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.119	0.121	0.117	0.119	0.124	0.126	0.174	0.176
Observations	3915	3911	3597	3593	2813	2809	2163	2161

Panel B: Relationship between investor attention and post-SEO market valuation measured using the price at the end of the first post-issue fiscal quarter

Dependent Variable	QFQAdj							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NumNewsIss [-7:-1]	0.021***	0.020***						
	(2.79)	(2.74)						
NumNewsIss [-14:-1]			0.015***	0.015***				
			(2.90)	(2.86)				
NumNewsIss [-30:-1]					0.008***	0.008**		
					(2.74)	(2.57)		
NumNewsIss [-60:-1]							0.004**	0.004**
							(2.42)	(2.34)
Underpricing		0.004		0.006		0.011		0.003
		(0.68)		(0.92)		(1.55)		(0.42)
UndwrtReputation	4.149***	4.148***	3.571***	3.572***	2.516***	2.499***	2.151**	2.105**
	(5.31)	(5.30)	(4.28)	(4.28)	(2.76)	(2.74)	(2.35)	(2.30)
FirmSize	-0.470***	-0.469***	-0.457***	-0.454***	-0.426***	-0.419***	-0.408***	-0.405***
	(-16.21)	(-16.04)	(-14.71)	(-14.52)	(-12.45)	(-12.15)	(-11.67)	(-11.48)
PriorQtrEarnSurpIss	0.004	0.005	0.003	0.004	-0.014	-0.012	-0.019	-0.018
	(0.12)	(0.16)	(0.10)	(0.15)	(-0.44)	(-0.37)	(-0.60)	(-0.57)
PriorMktRetIss	-0.051	-0.092	-0.090	-0.142	-0.162	-0.311	0.526	0.363
	(-0.06)	(-0.10)	(-0.09)	(-0.15)	(-0.16)	(-0.30)	(0.48)	(0.33)
MidFilePrice	0.021***	0.021***	0.019***	0.020***	0.017***	0.017***	0.014***	0.014***
	(10.54)	(10.57)	(8.95)	(9.00)	(6.98)	(7.04)	(5.51)	(5.55)
Constant	-2.997	-3.064	-3.024	-3.114	-2.864	-3.038	10.584**	15.185***
	(-1.02)	(-1.04)	(-1.02)	(-1.05)	(-0.99)	(-1.05)	(2.38)	(4.03)

Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.133	0.130	0.133	0.130	0.146	0.142	0.218	0.212
Observations	3934	3905	3615	3587	2829	2803	2180	2156

Table A.3: **Relationship between abnormal investor attention and post-SEO participation of institutional investors in the ownership of issuing firms' equity**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *InstN* is the number of institutional investors holding SEO firms' shares at the end of the first post-issue fiscal quarter. *AbnNumNewsIss [-7:-1]*, *AbnNumNewsIss [-14:-1]*, *AbnNumNewsIss [-30:-1]*, and *AbnNumNewsIss [-60:-1]* are measures of abnormal investor attention prior to the SEO issue date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	InstN							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AbnNumNewsIss [-7:-1]	2.421***	2.440***						
	(8.70)	(8.74)						
AbnNumNewsIss [-14:-1]			1.550***	1.558***				
			(7.37)	(7.39)				
AbnNumNewsIss [-30:-1]					0.504***	0.505***		
					(3.45)	(3.44)		
AbnNumNewsIss [-60:-1]							0.361***	0.359***
							(3.70)	(3.67)
Underpricing		0.599**		0.647***		0.620**		0.580*
		(2.58)		(2.71)		(2.19)		(1.70)
UndwrtReputation	21.500	23.189	29.579	31.053	76.068**	75.535**	79.412*	76.987*
	(0.75)	(0.81)	(0.97)	(1.02)	(2.09)	(2.07)	(1.88)	(1.82)
FirmSize	47.181***	47.285***	47.740***	47.854***	49.835***	49.989***	51.592***	51.704***
	(45.33)	(45.16)	(43.51)	(43.38)	(37.98)	(37.79)	(33.20)	(33.00)
PriorQtrEarnSurpIss	-2.003*	-1.736	-1.924	-1.642	-1.796	-1.549	-2.280	-2.036
	(-1.69)	(-1.46)	(-1.61)	(-1.37)	(-1.30)	(-1.12)	(-1.50)	(-1.33)
PriorMktRetIss	7.363	2.939	17.214	12.219	41.214	37.114	35.144	31.680
	(0.22)	(0.09)	(0.49)	(0.35)	(1.00)	(0.90)	(0.70)	(0.63)
MidFilePrice	1.017***	1.028***	1.058***	1.069***	1.221***	1.231***	1.301***	1.312***
	(13.62)	(13.75)	(13.25)	(13.38)	(12.51)	(12.59)	(11.02)	(11.08)
Constant	-291.507***	-299.458***	-307.050***	-315.648***	-339.162***	-348.055***	-379.480*	-490.675***

	(-2.72)	(-2.79)	(-2.86)	(-2.94)	(-2.95)	(-3.02)	(-1.86)	(-2.84)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.709	0.709	0.718	0.718	0.744	0.744	0.760	0.759
Observations	3883	3854	3569	3541	2796	2770	2156	2132

Table A.4: **Relationship between abnormal investor attention and post-SEO secondary market valuation of issuing firms**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *QFTDAdj* and *QFQAdj* are the industry-adjusted Q ratios calculated using the SEO issue day closing price and the price at the end of the first post-issue fiscal quarter, respectively. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and either the SEO issue day closing price (*QFTDAdj*) or the price at the end of the first post-issue fiscal quarter (*QFQAdj*). Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *AbnNumNewsIss [-7:-1]*, *AbnNumNewsIss [-14:-1]*, *AbnNumNewsIss [-30:-1]*, and *AbnNumNewsIss [-60:-1]* are measures of abnormal investor attention prior to the SEO issue date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year $\times$ industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between abnormal investor attention and post-SEO market valuation measured using the SEO issue day closing price

Dependent Variable	QFTDAdj							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AbnNumNewsIss [-7:-1]	0.026***	0.025***						
	(3.00)	(2.98)						
AbnNumNewsIss [-14:-1]			0.023***	0.023***				
			(3.55)	(3.49)				
AbnNumNewsIss [-30:-1]					0.015***	0.014***		
					(3.60)	(3.53)		
AbnNumNewsIss [-60:-1]							0.006**	0.006**
							(2.37)	(2.34)
Underpricing		0.013*		0.015**		0.015**		0.009
		(1.91)		(2.01)		(1.97)		(1.09)
UndwrtReputation	4.056***	4.130***	3.659***	3.735***	2.666***	2.703***	2.319**	2.326**
	(4.63)	(4.71)	(3.90)	(3.98)	(2.62)	(2.66)	(2.26)	(2.26)
FirmSize	-0.494***	-0.492***	-0.479***	-0.477***	-0.422***	-0.419***	-0.395***	-0.396***
	(-15.68)	(-15.56)	(-14.33)	(-14.22)	(-11.67)	(-11.54)	(-10.56)	(-10.54)
PriorQtrEarnSurpIss	0.014	0.018	0.013	0.017	-0.015	-0.011	-0.029	-0.026
	(0.43)	(0.55)	(0.38)	(0.50)	(-0.43)	(-0.32)	(-0.84)	(-0.75)
PriorMktRetIss	-0.059	-0.095	-0.194	-0.239	-0.326	-0.349	-0.076	-0.054

	(-0.06)	(-0.09)	(-0.18)	(-0.22)	(-0.28)	(-0.30)	(-0.06)	(-0.04)
MidFilePrice	0.031***	0.031***	0.031***	0.031***	0.025***	0.025***	0.021***	0.022***
	(13.73)	(13.82)	(12.51)	(12.60)	(9.33)	(9.41)	(7.62)	(7.69)
Constant	-1.576	-1.751	-1.737	-1.926	-1.536	-1.745	15.561***	15.490***
	(-0.48)	(-0.53)	(-0.52)	(-0.58)	(-0.48)	(-0.54)	(3.68)	(3.67)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.117	0.119	0.115	0.117	0.122	0.125	0.173	0.175
Observations	3915	3911	3597	3593	2813	2809	2163	2161

Panel B: Relationship between abnormal investor attention and post-SEO market valuation measured using the price at the end of the first post-issue fiscal quarter

Dependent Variable	QFQAdj							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AbnNumNewsIss [-7:-1]	0.011	0.012						
	(1.49)	(1.51)						
AbnNumNewsIss [-14:-1]			0.010*	0.011*				
			(1.82)	(1.90)				
AbnNumNewsIss [-30:-1]					0.006	0.006		
					(1.56)	(1.51)		
AbnNumNewsIss [-60:-1]							0.002	0.002
							(1.01)	(1.00)
Underpricing		0.005		0.007		0.012*		0.004
		(0.75)		(1.03)		(1.69)		(0.55)
UndwrtReputation	4.157***	4.157***	3.570***	3.569***	2.496***	2.479***	2.137**	2.087**
	(5.32)	(5.31)	(4.27)	(4.26)	(2.73)	(2.71)	(2.32)	(2.27)
FirmSize	-0.453***	-0.451***	-0.435***	-0.433***	-0.399***	-0.394***	-0.384***	-0.381***
	(-16.13)	(-15.97)	(-14.64)	(-14.48)	(-12.32)	(-12.07)	(-11.53)	(-11.38)
PriorQtrEarnSurpIss	0.003	0.004	0.002	0.003	-0.015	-0.013	-0.020	-0.019
	(0.09)	(0.13)	(0.06)	(0.12)	(-0.48)	(-0.40)	(-0.65)	(-0.61)
PriorMktRetIss	0.013	-0.033	-0.013	-0.072	-0.120	-0.277	0.559	0.391
	(0.01)	(-0.04)	(-0.01)	(-0.07)	(-0.12)	(-0.27)	(0.51)	(0.36)
MidFilePrice	0.021***	0.021***	0.019***	0.020***	0.017***	0.017***	0.014***	0.014***
	(10.50)	(10.54)	(8.92)	(8.97)	(6.92)	(6.98)	(5.45)	(5.49)
Constant	-3.123	-3.192	-3.169	-3.266	-3.048	-3.226	10.440**	14.953***
	(-1.06)	(-1.08)	(-1.07)	(-1.10)	(-1.05)	(-1.12)	(2.34)	(3.97)

Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.131	0.129	0.131	0.129	0.144	0.141	0.215	0.209
Observations	3934	3905	3615	3587	2829	2803	2180	2156

Table A.5: **Relationship between abnormal investor attention and SEO underpricing**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *AbnNumNewsIss [-7:-1]*, *AbnNumNewsIss [-14:-1]*, *AbnNumNewsIss [-30:-1]*, and *AbnNumNewsIss [-60:-1]* are measures of abnormal investor attention prior to the SEO issue date as described in Subsection 5.1. *UndwrtReputation* is the reputation measure of the lead underwriter, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in the previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO issue date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean estimates of earnings and actual earnings adjusted by price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	Underpricing			
	(1)	(2)	(3)	(4)
AbnNumNewsIss [-7:-1]	0.021			
	(0.97)			
AbnNumNewsIss [-14:-1]		0.032*		
		(1.95)		
AbnNumNewsIss [-30:-1]			0.021*	
			(1.86)	
AbnNumNewsIss [-60:-1]				0.010
				(1.45)
UndwrtReputation	-4.842**	-4.661*	-1.613	1.593
	(-2.20)	(-1.95)	(-0.58)	(0.51)
FirmSize	-0.326***	-0.324***	-0.394***	-0.326***
	(-4.11)	(-3.81)	(-3.97)	(-2.88)
PriorQtrEarnSurpIss	-0.287***	-0.283***	-0.242**	-0.327***
	(-3.45)	(-3.31)	(-2.48)	(-3.09)
PriorMktRetIss	4.717*	4.997*	3.844	1.855
	(1.83)	(1.81)	(1.22)	(0.50)
MidFilePrice	-0.014**	-0.014**	-0.009	-0.012
	(-2.52)	(-2.22)	(-1.28)	(-1.42)
Constant	14.172*	14.020*	14.711*	9.903
	(1.71)	(1.65)	(1.67)	(0.78)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.227	0.241	0.285	0.324
Observations	3920	3601	2817	2166

Table A.6: **Relationship between abnormal investor attention and SEO valuation of issuing firms**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *QOPAdj* is the industry-adjusted Q ratio calculated using the SEO offer price. *AbnNumNewsIss [-7:-1]*, *AbnNumNewsIss [-14:-1]*, *AbnNumNewsIss [-30:-1]*, and *AbnNumNewsIss [-60:-1]* are measures of abnormal investor attention prior to the SEO issue date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	QOPAdj							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AbnNumNewsIss [-7:-1]	0.024***	0.024***						
	(2.97)	(2.97)						
AbnNumNewsIss [-14:-1]			0.022***	0.022***				
			(3.54)	(3.60)				
AbnNumNewsIss [-30:-1]					0.014***	0.014***		
					(3.61)	(3.58)		
AbnNumNewsIss [-60:-1]							0.006**	0.006**
							(2.39)	(2.42)
Underpricing		-0.010		-0.009		-0.008		-0.014*
		(-1.46)		(-1.29)		(-1.02)		(-1.81)
UndwrtReputation	4.090***	4.009***	3.695***	3.612***	2.645***	2.582***	2.238**	2.193**
	(4.85)	(4.75)	(4.10)	(4.00)	(2.71)	(2.64)	(2.25)	(2.20)
FirmSize	-0.471***	-0.474***	-0.456***	-0.459***	-0.402***	-0.404***	-0.380***	-0.384***
	(-15.53)	(-15.54)	(-14.21)	(-14.21)	(-11.58)	(-11.55)	(-10.52)	(-10.54)
PriorQtrEarnSurpIss	0.021	0.018	0.019	0.016	-0.009	-0.011	-0.021	-0.026
	(0.65)	(0.56)	(0.59)	(0.50)	(-0.26)	(-0.31)	(-0.62)	(-0.76)
PriorMktRetIss	-0.043	0.009	-0.186	-0.137	-0.140	-0.156	0.200	0.169
	(-0.04)	(0.01)	(-0.18)	(-0.13)	(-0.13)	(-0.14)	(0.17)	(0.14)
MidFilePrice	0.030***	0.030***	0.030***	0.029***	0.025***	0.024***	0.021***	0.021***
	(13.83)	(13.73)	(12.59)	(12.51)	(9.49)	(9.44)	(7.74)	(7.66)
Constant	-1.963	-1.824	-2.113	-1.987	-1.969	-1.858	10.666**	15.131***
	(-0.62)	(-0.57)	(-0.66)	(-0.62)	(-0.64)	(-0.60)	(2.21)	(3.69)

Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.118	0.117	0.117	0.116	0.124	0.123	0.172	0.171
Observations	3940	3911	3621	3593	2835	2809	2185	2161

Table A.7: **Instrumental variable analysis of the relationship between investor attention and post-SEO market valuation of issuing firms**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000-2018. *QFTDAdj* and *QFQAdj* are the industry-adjusted Q ratios calculated using the SEO issue day closing price and the price at the end of the first post-issue fiscal quarter, respectively. *NumNewsIssHat [-7:-1]*, *NumNewsIssHat [-14:-1]*, *NumNewsIssHat [-30:-1]*, and *NumNewsIssHat [-60:-1]* are predicted values of investor attention variables as described in Table 1 (*NumNewsIss [-7:-1]*, *NumNewsIss [-14:-1]*, *NumNewsIss [-30:-1]*, and *NumNewsIss [-60:-1]*) from first-stage regressions. *PriorYrNumNewsFile [-7:-1]*, *PriorYrNumNewsFile [-14:-1]*, *PriorYrNumNewsFile [-30:-1]*, and *PriorYrNumNewsFile [-60:-1]* are instrumental variables which measure investor attention one year prior to the SEO announcement date as described in Table 1. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year $\times$ industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and post-SEO market valuation measured using the SEO issue day closing price

Dependent Variable	1st-stage	QFTDAdj	1st-stage	QFTDAdj	1st-stage	QFTDAdj	1st-stage	QFTDAdj
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.568***							
	(22.77)							
NumNewsIssHat [-7:-1]		0.037*						
		(1.83)						
PriorYrNumNewsFile [-14:-1]			0.681***					
			(32.46)					
NumNewsIssHat [-14:-1]				0.016				
				(1.60)				
PriorYrNumNewsFile [-30:-1]					0.664***			
					(38.13)			
NumNewsIssHat [-30:-1]						0.007		
						(1.39)		
PriorYrNumNewsFile [-60:-1]							0.838***	
							(42.24)	
NumNewsIssHat [-60:-1]								0.004*
								(1.78)
Underpricing	0.020	0.013**	0.052***	0.014**	0.106***	0.015**	0.157*	0.008

	(1.48)	(1.97)	(2.63)	(2.14)	(2.86)	(2.19)	(1.91)	(1.16)
UndwrtReputation	3.550**	4.130***	8.723***	3.840***	10.428**	2.841***	21.760**	2.426***
	(2.06)	(5.18)	(3.40)	(4.52)	(2.13)	(3.14)	(2.10)	(2.70)
FirmSize	0.827***	-0.521***	1.089***	-0.493***	1.857***	-0.433***	2.549***	-0.415***
	(13.03)	(-14.07)	(11.48)	(-13.67)	(10.24)	(-11.64)	(6.43)	(-11.43)
PriorQtrEarnSurpIss	0.006	0.020	-0.001	0.018	-0.076	-0.011	-0.106	-0.025
	(0.09)	(0.65)	(-0.01)	(0.60)	(-0.44)	(-0.35)	(-0.30)	(-0.82)
PriorMktRetIss	3.040	-0.194	2.305	-0.308	2.282	-0.370	9.814	-0.053
	(1.50)	(-0.21)	(0.78)	(-0.31)	(0.41)	(-0.36)	(0.80)	(-0.05)
MidFilePrice	-0.000	0.032***	0.004	0.031***	0.006	0.026***	-0.019	0.022***
	(-0.07)	(15.26)	(0.65)	(13.99)	(0.49)	(10.65)	(-0.66)	(8.81)
Constant	-8.597	-1.547	-7.520	-1.807	-5.114	-1.552	-12.899	15.723***
	(-1.32)	(-0.52)	(-0.82)	(-0.60)	(-0.33)	(-0.54)	(-0.30)	(4.25)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.121		0.118		0.125		0.176
Observations	3911	3911	3593	3593	2809	2809	2161	2161
F Statistics	518.49		1053.76		1453.65		1783.83	

Panel B: Relationship between investor attention and post-SEO market valuation measured using the price at the end of the first post-issue fiscal quarter

Dependent Variable	1st-stage	QFQAdj	1st-stage	QFQAdj	1st-stage	QFQAdj	1st-stage	QFQAdj
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PriorYrNumNewsFile [-7:-1]	0.568***							
	(22.75)							
NumNewsIssHat [-7:-1]		0.039**						
		(2.13)						
PriorYrNumNewsFile [-14:-1]			0.681***					
			(32.42)					
NumNewsIssHat [-14:-1]				0.016*				
				(1.72)				
PriorYrNumNewsFile [-30:-1]					0.664***			
					(38.07)			
NumNewsIssHat [-30:-1]						0.007		
						(1.55)		
PriorYrNumNewsFile [-60:-1]							0.838***	

							(42.17)	
NumNewsIssHat [-60:-1]								0.004**
								(2.16)
Underpricing	0.020	0.004	0.052***	0.006	0.106***	0.011*	0.156*	0.003
	(1.47)	(0.65)	(2.62)	(1.00)	(2.84)	(1.77)	(1.90)	(0.45)
UndwrtReputation	3.434**	4.098***	8.638***	3.567***	10.292**	2.506***	21.607**	2.101***
	(1.99)	(5.75)	(3.36)	(4.71)	(2.10)	(3.08)	(2.08)	(2.62)
FirmSize	0.826***	-0.490***	1.088***	-0.456***	1.856***	-0.415***	2.550***	-0.408***
	(13.00)	(-14.80)	(11.45)	(-14.18)	(10.21)	(-12.43)	(6.42)	(-12.60)
PriorQtrEarnSurpIss	0.006	0.005	-0.001	0.004	-0.075	-0.012	-0.106	-0.018
	(0.10)	(0.19)	(-0.01)	(0.16)	(-0.44)	(-0.42)	(-0.30)	(-0.65)
PriorMktRetIss	3.002	-0.167	2.249	-0.146	2.198	-0.306	9.894	0.357
	(1.48)	(-0.20)	(0.76)	(-0.17)	(0.39)	(-0.33)	(0.80)	(0.37)
MidFilePrice	-0.000	0.022***	0.004	0.020***	0.006	0.017***	-0.019	0.014***
	(-0.06)	(11.64)	(0.65)	(9.95)	(0.49)	(7.89)	(-0.66)	(6.34)
Constant	-8.592	-2.873	-7.512	-3.103	-5.100	-3.063	-12.882	15.215***
	(-1.32)	(-1.07)	(-0.82)	(-1.16)	(-0.33)	(-1.19)	(-0.30)	(4.61)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.129		0.130		0.142		0.212
Observations	3905	3905	3587	3587	2803	2803	2156	2156
F Statistics	517.55		1051.18		1449.09		1778.37	

Table A.8: **Robustness Check: Relationship between abnormal investor attention and market reaction upon SEO announcement**

The sample consists of seasoned equity offerings (SEOs) conducted in 2000 - 2018. *CAR [0:0]* is the abnormal return on SEO firm's equity on the SEO announcement day. *CAR [1:21]* is the cumulative abnormal return on SEO firm's equity over a 21-day window (from day 1 to day 21) after the SEO announcement date. *AbnNumEdgarFile [-7:-1]*, *AbnNumEdgarFile [-14:-1]*, *AbnNumEdgarFile [-30:-1]*, and *AbnNumEdgarFile [-60:-1]* are measures of abnormal EDGAR filing downloads prior to the SEO announcement date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpFile* is the earnings surprise one quarter prior to the SEO announcement date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO announcement date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and SEO announcement effect

Dependent Variable	CAR [0:0]			
	(1)	(2)	(3)	(4)
AbnNumEdgarFile [-7:-1]	-0.004**			
	(-2.14)			
AbnNumEdgarFile [-14:-1]		-0.003**		
		(-2.14)		
AbnNumEdgarFile [-30:-1]			-0.002**	
			(-2.46)	
AbnNumEdgarFile [-60:-1]				-0.001**
				(-2.28)
UndwrtReputation	-1.549	-1.477	-1.441	-1.465
	(-0.99)	(-0.95)	(-0.92)	(-0.94)
FirmSize	0.087	0.088	0.091*	0.091*
	(1.60)	(1.64)	(1.68)	(1.69)
PriorQtrEarnSurpFile	0.198**	0.197**	0.195**	0.196**
	(2.32)	(2.31)	(2.29)	(2.30)
PriorMktRetFile	1.531	1.522	1.541	1.571
	(0.94)	(0.93)	(0.94)	(0.96)
MidFilePrice	0.008**	0.008**	0.008**	0.008**
	(2.28)	(2.30)	(2.30)	(2.25)
Constant	-0.289	-0.294	-0.237	-0.222
	(-0.07)	(-0.07)	(-0.05)	(-0.05)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.186	0.186	0.187	0.186
Observations	3342	3342	3342	3342

Panel B: Relationship between investor attention and post-announcement drift				
Dependent Variable	CAR [1:21]			
	(1)	(2)	(3)	(4)
AbnNumEdgarFile [-7:-1]	0.018**			
	(2.33)			
AbnNumEdgarFile [-14:-1]		0.015***		
		(2.94)		
AbnNumEdgarFile [-30:-1]			0.007**	
			(2.31)	
AbnNumEdgarFile [-60:-1]				0.003*
				(1.82)
UndwrtReputation	-8.258	-8.776	-8.644	-8.448
	(-1.23)	(-1.31)	(-1.29)	(-1.26)
FirmSize	0.775***	0.761***	0.759***	0.761***
	(3.34)	(3.28)	(3.27)	(3.27)
PriorQtrEarnSurpFile	0.243	0.243	0.254	0.252
	(0.66)	(0.66)	(0.69)	(0.68)
PriorMktRetFile	8.563	8.648	8.503	8.391
	(1.22)	(1.23)	(1.21)	(1.19)
MidFilePrice	-0.022	-0.024	-0.022	-0.021
	(-1.47)	(-1.54)	(-1.47)	(-1.40)
Constant	3.278	3.539	2.938	2.803
	(0.17)	(0.19)	(0.15)	(0.15)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.163	0.164	0.163	0.162
Observations	3347	3347	3347	3347

Table A.9: **Robustness Check: Relationship between abnormal investor attention and SEO Characteristics**

The sample consists of seasoned equity offerings (SEOs) conducted in 2003-2017. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *InstN* is the number of institutional investors holding SEO firms' shares at the end of the first post-issue fiscal quarter. *QOPAdj* is the industry-adjusted Q ratio calculated using the SEO offer price. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and the SEO offer price. Industry adjustment is performed by subtracting contemporaneous 2-digit SIC code industry median Q ratios from SEO firms' Q ratios. *AbnNumEdgarIss [-7:-1]*, *AbnNumEdgarIss [-14:-1]*, *AbnNumEdgarIss [-30:-1]*, and *AbnNumEdgarIss [-60:-1]* are measures of abnormal EDGAR filing downloads prior to the SEO issue date as described in Table 1. *UndwrtReputation* is the lead SEO underwriter's reputation measure, which is defined as the lead underwriter's share of total proceeds raised in the SEO market in previous five years. *FirmSize* is the natural logarithm of the book value of total assets at the end of the fiscal quarter prior to the SEO announcement date. *PriorQtrEarnSurpIss* is the earnings surprise one quarter prior to the SEO issue date. Earnings surprise is defined as the difference between the mean earnings estimate and actual earnings divided by the stock price. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *MidFilePrice* is the midpoint of initial filing range. Year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relationship between investor attention and SEO underpricing

Dependent Variable	Underpricing			
	(1)	(2)	(3)	(4)
AbnNumEdgarIss [-7:-1]	0.011***			
	(2.93)			
AbnNumEdgarIss [-14:-1]		0.007***		
		(2.96)		
AbnNumEdgarIss [-30:-1]			0.004**	
			(2.44)	
AbnNumEdgarIss [-60:-1]				0.002**
				(2.03)
UndwrtReputation	-4.515	-3.298	1.078	3.110
	(-1.61)	(-1.08)	(0.29)	(0.77)
FirmSize	-0.295***	-0.290***	-0.377***	-0.325**
	(-2.98)	(-2.69)	(-2.98)	(-2.33)
PriorQtrEarnSurpIss	-0.033	-0.032	0.022	-0.078
	(-0.49)	(-0.46)	(0.26)	(-0.84)
PriorMktRetIss	5.188	5.520	3.926	0.597
	(1.48)	(1.47)	(0.92)	(0.13)
MidFilePrice	-0.009	-0.010	-0.004	-0.007
	(-1.24)	(-1.15)	(-0.43)	(-0.71)
Constant	22.607***	23.358***	22.651**	5.921
	(3.08)	(3.02)	(2.33)	(0.56)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.236	0.252	0.279	0.315
Observations	2791	2544	2057	1677

Panel B: Relationship between investor attention and post-SEO participation of institutional investors				
Dependent Variable	InstN			
	(1)	(2)	(3)	(4)
AbnNumEdgarIss [-7:-1]	0.226***			
	(4.83)			
AbnNumEdgarIss [-14:-1]		0.189***		
		(5.89)		
AbnNumEdgarIss [-30:-1]			0.082***	
			(3.63)	
AbnNumEdgarIss [-60:-1]				0.075***
				(4.82)
Underpricing	0.534*	0.586**	0.463	0.535
	(1.93)	(2.06)	(1.45)	(1.46)
UndwrtReputation	5.496	27.378	72.467	92.097*
	(0.15)	(0.71)	(1.58)	(1.77)
FirmSize	47.046***	47.418***	49.185***	49.281***
	(35.99)	(34.04)	(30.38)	(26.83)
PriorQtrEarnSurpIss	-0.877	-1.023	-0.480	-0.445
	(-0.83)	(-0.96)	(-0.38)	(-0.34)
PriorMktRetIss	74.744*	74.466	90.874*	82.015
	(1.65)	(1.57)	(1.70)	(1.35)
MidFilePrice	1.272***	1.364***	1.358***	1.475***
	(12.88)	(12.93)	(11.56)	(10.94)
Constant	-164.476*	-151.573	-206.445*	-297.353**
	(-1.74)	(-1.56)	(-1.70)	(-2.20)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.716	0.729	0.752	0.763
Observations	2731	2490	2015	1645

Panel C: Relationship between investor attention and SEO valuation of issuing firms				
Dependent Variable	QOPAdj			
	(1)	(2)	(3)	(4)
AbnNumEdgarIss [-7:-1]	0.003***			
	(3.10)			
AbnNumEdgarIss [-14:-1]		0.002**		
		(2.16)		
AbnNumEdgarIss [-30:-1]			0.001**	
			(2.20)	
AbnNumEdgarIss [-60:-1]				0.000
				(1.31)
Underpricing	-0.001	0.000	0.004	0.001
	(-0.20)	(0.08)	(0.53)	(0.08)
UndwrtReputation	3.467***	2.886***	2.884***	2.689**
	(4.22)	(3.29)	(2.81)	(2.35)
FirmSize	-0.416***	-0.378***	-0.359***	-0.362***
	(-14.39)	(-12.24)	(-10.09)	(-9.10)
PriorQtrEarnSurpIss	0.025	0.025	0.001	-0.008
	(1.28)	(1.25)	(0.04)	(-0.31)
PriorMktRetIss	0.719	0.861	0.094	-0.217
	(0.70)	(0.80)	(0.08)	(-0.16)
MidFilePrice	0.025***	0.022***	0.021***	0.021***
	(11.33)	(9.25)	(8.04)	(7.41)
Constant	1.412	1.536	3.053	4.029
	(0.66)	(0.69)	(1.12)	(1.34)
Industry × Year FE	Yes	Yes	Yes	Yes
R^2	0.131	0.119	0.125	0.133
Observations	2790	2543	2056	1676

ESSAY 3: Underwriter Networks, Information Asymmetry, and Seasoned Equity Offerings

Thomas J. Chemmanur* Karen Simonyan† Xiang Zheng‡

Abstract

Using various "centrality" measures from Social Network Analysis (SNA), we analyze, for the first time in the literature, how the location of a lead underwriter in its network of investment banks affects various aspects of seasoned equity offerings (SEOs). We hypothesize that investment banking networks perform an important economic role in the SEO underwriting process for SEOs, namely, that of information dissemination, where the lead underwriter uses its investment banking network to disseminate information about the SEO firm to institutional investors. Consistent with the above information dissemination role, we show that firms whose SEOs are underwritten by more central lead underwriters are associated with a smaller extent of information asymmetry in the equity market. We then develop testable hypotheses based on the information dissemination role of underwriter networks for the relationship between SEO underwriter centrality and various SEO characteristics, which we test in our empirical analysis. Consistent with the above hypotheses, we find that more central lead SEO underwriters are associated with less negative SEO announcement effects; smaller SEO offer price revisions; smaller SEO discounts and underpricing; higher immediate post-SEO equity valuations for issuing firms; and greater post-SEO long-run stock returns for issuing firms. We also find that SEOs with more central lead underwriters are associated with greater institutional investor participation. Our instrumental variable (IV) analysis using the industry-average bargaining power of underwriters relative to issuers as the instrument shows that the above results are causal. Consistent with greater value creation by more central lead underwriters, we find that more central lead underwriters receive greater compensation.

Keywords: Seasoned Equity Offerings; Equity Issues; Announcement Effect; Information Asymmetry; Investment Banking Networks
JEL classification: G24; G32

*Professor of Finance and Hillenbrand Distinguished Fellow, Finance Department, Fulton Hall 336, Carroll School of Management, Boston College, Chestnut Hill, MA 02467, Tel: (617) 552-3980, Fax: (617) 552-0431, Email: chemmanu@bc.edu.

†Associate Professor of Finance, Sawyer Business School, Suffolk University, 8 Ashburton Place, Boston, MA 02108, Email: ksimonya@suffolk.edu.

‡Ph.D. Candidate, Finance Department, Fulton Hall 341, Carroll School of Management, Boston College, Chestnut Hill, MA 02467, Tel: (617) 552-2062, Email: zhengxr@bc.edu.

1 Introduction

The crucial role of information asymmetry between firm insiders and outsiders in generating the negative announcement effect of (abnormal stock return) seasoned equity offerings (SEOs) has been analyzed extensively in the literature both theoretically (see, e.g., Myers and Majluf (1984); Giammarino and Lewis (1988)) and empirically (see, e.g., Asquith and Mullins (1986)). Given that information asymmetry and the resulting negative announcement effects of equity issues have been known to create permanent wealth losses for shareholders of issuing firms, there has been considerable interest in analyzing the mechanism through which the above information asymmetry facing firms making SEOs may be reduced, potentially lowering the wealth losses of shareholders. One mechanism that has been examined theoretically in the literature is the role of the underwriters. For example, Chemmanur and Fulghieri (1994) argue theoretically that the economic role of the underwriter in an equity issue (initial public offering [IPO] or SEO) is to produce information about a firm making the equity issue and to convey this information to investors making use of their reputation.

In this paper, we empirically analyze, for the first time, a complementary mechanism through which a lead SEO underwriter may reduce the information asymmetry facing firms making SEOs, namely, by making use of the network of investment banks connected to it. We then analyze the implications of this reduction in information asymmetry on the SEO announcement effect. We also study the effect of the position of a lead SEO underwriter in its investment banking network for various important phenomena that have been documented in the context of SEOs, namely, SEO discount and SEO underpricing; market valuation of firms conducting SEOs; the participation of institutional investors in SEOs; and long-run post-SEO stock returns of issuing firms.

The theoretical literature (see, e.g., Chemmanur and Fulghieri (1994)) has argued that the role of an underwriter in an equity issue is that of an information producer, who produces noisy information about an issuing firm in the process of conducting the due diligence associated with the equity issue. The underwriter may then convey this information to investors making use of its reputation as a certifying mechanism (thus mitigating concerns about their incentives to suppress negative information about the firm and report only favorable information). Chemmanur and Fulghieri (1994) assume that underwriters are able to costlessly (and credibly) convey the above information to potential investors in an equity issue. In this paper, we explicitly consider the transmission mechanism of the above noisy information from the underwriter to institutional investors and how institutions process the above information when deciding whether or not to invest in a firm's SEO. Given that over time an SEO underwriter may repeatedly interact with other investment banks when they together underwrite equity issues as a part of underwriting syndicates (thus forming an "investment banking network"), we conjecture that the network of investment banks to which a lead SEO underwriter is connected may serve as an important mechanism to credibly transmit the above information. Each investment bank in the above network is likely to have repeated interactions with a subset of institutional investors, who may potentially invest in the SEO. A lead underwriter who is better connected to various investment banks in its network will be able to convey the information it has produced (about the true value of the SEO firm) to these investment banks more efficiently. Further, given the repeated interactions between a particular investment bank in the lead underwriter's network and a subset of institutional investors, that investment bank will, in turn, be able to more credibly (and efficiently) convey the above information to these institutional investors.

The above argument implies that the information asymmetry faced by institutional investors about the firm whose SEO is underwritten by a better-connected lead underwriter (i.e., more central to its network and with more extensive underwriter network) will be smaller, since such a lead underwriter will be able to more efficiently convey the information it has produced about the firm making the SEO to institutional investors. This lower information asymmetry, in turn, will have several implications for the equity market's reaction to the announcement of an SEO; for the participation of institutional investors in that SEO; for various SEO characteristics; and for the post-SEO stock returns of issuing firms. We will test the above implications in our empirical analysis.

We characterize the relative position of a lead SEO underwriter in its network of investment banks by making use of six different measures from the social network analysis (SNA) literature. The first measure is *Degree*, which is the number of other unique investment banks that the lead SEO underwriter had connections with (either as a lead underwriter or as a member of an IPO or an SEO syndicate) in the five-year period prior to the SEO year. As is clear from a comparison of our graphical illustration of investment banking (underwriter) networks in 1980 (in Figure 1) and in 2017 in (Figure 2), such networks have grown dramatically over the last three to four decades. In Figures 1 and 2, the size of the node associated with an investment bank represents the magnitude of its *Degree* centrality within its investment banking network. In addition to *Degree* centrality, there are five other measures from the SNA literature that capture additional information about the underwriter network (and which we use in our empirical analysis). These are: *Outdegree*, *Indegree*, *Eigenvector*, *Betweenness*, and *2-StepReach*. To give one example, *Eigenvector* centrality captures not only the number of other investment banks that an underwriter is connected to, but also captures the importance of each of the above investment banks. We define and discuss these six measures in more detail in Section 6.

We first summarize the empirical results from our baseline analysis. First, firms whose SEOs are underwritten by more central lead underwriters are characterized by a smaller extent of information asymmetry in the equity market. Thus, such firms are associated with smaller analyst forecast errors, smaller analyst forecast dispersion, and smaller bid-ask spreads. Second, we find that SEOs underwritten by more central lead SEO underwriters are associated with less negative announcement effects. Third, SEOs underwritten by more central lead underwriters are associated with smaller absolute values of SEO offer price revisions. Fourth, SEOs underwritten by more central lead SEO underwriters are associated with smaller SEO discounts and smaller SEO underpricing. Fifth, firms whose SEOs are underwritten by more central lead SEO underwriters have higher immediate post-SEO equity market valuations. Sixth, firms whose SEOs are underwritten by more central lead underwriters are associated with larger institutional investor holdings. Finally, firms whose SEOs are underwritten by more central lead underwriters are associated with better post-SEO long-run stock returns. All the above empirical results hold even after controlling for lead SEO underwriter reputation. Overall, our results are consistent with the notion that more central lead SEO underwriters are able to more efficiently disseminate information about these SEO firms to institutional investors, thereby reducing the information asymmetry faced by these SEO firms with respect to investors in the equity market.

It may be argued that our baseline empirical analysis may be driven by the endogenous matching between SEO firms and underwriters: i.e., higher quality (intrinsic value) firms may hire more central underwriters to underwrite their SEOs, so that our baseline empirical results may be driven (arguably) by the quality of the SEO firm rather than any causal effect of lead SEO underwriter centrality. To address this issue and thereby establish the causality of lead underwriter centrality, we rely on two identification tests. First, we conduct an instrumental variable (IV) analysis, using the relative bargaining power of an SEO firm's industry (*IRBP*) as our instrumental variable. *IRBP* is defined as the average number of book runners for IPOs and SEOs in each industry over the three years prior to the SEO. It is likely that firms in an industry with higher relative bargaining power (*IRBP*) are more likely to hire more central underwriters than firms in an industry with lower relative bargaining power: we establish the relevance of this IV empirically as well. The average number of book runners in each industry is unlikely to affect the SEO characteristics of a specific firm, thereby satisfying the exclusion restriction for this IV. Our IV analysis reveals that, even after controlling for the potential endogenous matching between firms making SEOs and lead underwriters, all our baseline empirical findings continue to hold, suggesting that more central lead SEO underwriters causally create more value when underwriting SEOs.

The second test we use to establish the causality of lead underwriter centrality in value creation in SEOs is by analyzing lead underwriter compensation. We rely on the fact that, if more central lead SEO underwriters are able to create more value for firms conducting SEOs, they will be compensated to a greater extent in a competitive SEO underwriting market. Consistent with this hypothesis, our analysis of lead underwriter compensation shows that more central lead underwriters are indeed compensated with larger gross spreads, larger management fees, and larger underwriting fees.

The rest of this paper is organized as follows. Section 2 discusses how our paper is related to the existing literature and describes its contribution relative to this literature. Section 3 discusses the underlying theory and develops testable hypotheses. Section 4 describes our identification strategy. Section 5 describes our sample selection procedure. Section 6 describes our measures of lead SEO underwriter centrality. Section 7 describes our empirical results. Section 8 concludes.

2 Relation to the existing literature and contribution

Our paper is related to several strands in the theoretical and empirical literature. The first strand is the theoretical literature on the announcement effect of an equity issue and the role of asymmetric information in affecting the magnitude of this announcement effect (e.g., Myers and Majluf (1984) and Giammarino and Lewis (1988)) and the large empirical literature studying the determinants of this announcement effect: see, e.g., Asquith and Mullins (1986) and Masulis and Korwar (1986). There is also a significant theoretical literature analyzing the role played by financial intermediaries in mitigating the effect of the asymmetric information facing firms at the time of an equity issue: see, e.g., Chemmanur and Fulghieri (1994), Booth and Smith (1986), or Titman and Trueman (1986). We contribute to the above literature by empirically analyzing, for the first time, the role of underwriter networks and lead underwriter centrality in reducing the information asymmetry facing firms making SEOs and the effect of the reduction in information asymmetry on the announcement effect of SEOs.

The second strand is the theoretical and empirical literature on the pricing of SEOs, the discounting and underpricing of SEOs, as well as other SEO characteristics. Two theoretical models of the pricing of SEOs are those of Chemmanur and Jiao (2011) and Gerard and Nanda (1993). These papers develop theoretical rationales for the pricing of SEOs (in particular, for SEO discounts and SEO underpricing) in an asymmetric information setting based on information production by institutional investors (in the case of Chemmanur and Jiao (2011)) and on SEO price manipulation (in the case of Gerard and Nanda (1993)). There is also a large empirical literature on the discounting and underpricing of SEOs. Loderer, Sheehan, and Kadlec (1991) document more significant SEO underpricing for stocks listed on the Nasdaq than for stocks listed on other exchanges such as the NYSE and Amex. Corwin (2003) studies the determinants of SEO underpricing such as offer size, uncertainty of firm value, the magnitude of preoffer returns, price rounding, and pricing relative to the bid quote. Altınkılıç and Hansen (2003) decompose SEO discounting into a predictable component and a surprise component, and argue that the surprise component is used by underwriters as a channel to release additional information to investors. Gao and Ritter (2010) study the effect of various offer methods on SEO characteristics such as discounting and underpricing. Gibson, Safieddine, and Sonti (2004) show that SEO firms with the greatest increase in institutional investment around the issue date significantly outperform those with the greatest decrease in institutional investment. Chemmanur, He, and Hu (2009) analyze the relation between institutional trading around SEOs and various SEO characteristics, and present findings consistent with institutions being able to produce information about the firm making the SEO. Huang and Zhang (2011) document a negative relation between the number of underwriters managing an SEO and the SEO discount. Gustafson (2018) documents a higher offer price and lower post-issue stock returns for overnight SEO offerings than for non-overnight offerings. We contribute to the above literature by analyzing, for the first time, the relation between lead underwriter centrality and various SEO characteristics such as SEO discount and underpricing; SEO offer price revision; post-SEO equity market valuation of issuing firms; institutional investor participation in SEOs; and long-run post-SEO stock returns of issuing firms.[1]

The third strand is the literature on the role of social networks in the financial market. Bajo, Chemmanur, Simonyan, and Tehranian (2016) empirically analyze the role of underwriter centrality in IPOs. They show that investment banking networks allow lead IPO underwriters to induce institutional investors to pay greater attention to the firm they take public, leading to larger absolute values of IPO offer price revisions, greater IPO and secondary market valuations of issuing firms, and higher IPO initial returns.[2] It should be noted that while IPOs and SEOs have some similarities (in the sense that both involve the issuance of equity), they represent fundamentally different economic settings in the following ways. First, since, unlike IPOs, SEOs involve equity issuance by firms whose stock is already trading in the equity market, there are important phenomena in SEOs that are not present in IPOs: two of these are the announcement effect of SEOs and SEO discount, which we focus on in the current paper. Second, since there is no agreed upon valuation for an IPO firm prior to the start of equity market trading, information extraction from institutional investors about true firm value is likely to be the dominant economic function

[1]To the extent that we also study the relation between lead SEO underwriter centrality and long-run post-SEO stock returns of issuing firms, the empirical literature on long-run post-SEO stock returns of issuing firms is also related to our paper: see, e.g., Loughran and Ritter (1995) or Ritter (2003).

[2]See also Chuluun (2015), who shows that IPO book managers with more central and cohesive networks are associated with larger IPO offer price revisions and underpricing.

of lead underwriters in the context of IPOs (see, e.g., Benveniste and Spindt (1989)). In contrast, since the firm's equity is already trading in the equity market prior to an SEO, equity value is more or less well-established in the context of an SEO, so that considerations of information dissemination (rather than information extraction) are likely to be the dominant economic role played by lead underwriters in SEOs. Consistent with the information dissemination role of the lead underwriter dominating in SEOs, we find that SEOs with more central lead underwriters are associated with smaller absolute values of SEO price revisions, and smaller SEO underpricing (as well as smaller SEO discounts). This contrasts with Bajo, Chemmanur, Simonyan, and Tehranian (2016) who find the opposite results in the IPO setting, which suggests that information extraction is the dominant economic role of lead underwriters in IPOs. Given the important economic differences between SEOs and IPOs, it is important to analyze the precise economic role of lead underwriter networks in SEOs.[3]

The contribution of this paper relative to the existing literature is threefold. First, this is the first paper to establish the role of lead SEO underwriters in disseminating information about firms conducting SEOs using the investment banking networks connected to them and thus reducing the information asymmetry facing SEO firms. Second, this is the first paper to document the relation between the lead SEO underwriter centrality and the SEO announcement effect, the SEO discount, and various other SEO characteristics such as post-SEO market valuation of issuing firms, SEO underpricing, institutional investor participation in SEOs, and long-run post-SEO stock returns of issuing firms. Third, underwriter reputation has been seen in the existing SEO literature as an important measure capturing the effectiveness of lead SEO underwriters as financial intermediaries. We extend this liter-

[3]There are also several other important papers analyzing social networks in the financial market or financial intermediary setting that are less closely related to our paper. For example, Hochberg, Ljungqvist, and Lu (2007) study how networks of venture capitalists (VCs) affect the investment performance of VC funds. They show that VC funds whose parent firms enjoy more influential network positions realize significantly better performance (measured by the proportion of portfolio investments successfully exited through an IPO or a sale to another company). Engelberg, Gao, and Parsons (2012) show that, when banks and firms are connected through interpersonal linkages, interest rates are markedly reduced. A large body of work also exists on board and CEO connectedness. For example, Larcker, So, and Wang (2013) investigate the connectedness of corporate board members across firms, and they show that firms with the best-connected boards earn on average substantially higher future excess returns compared with firms with the worst-connected boards. Similarly, El-Khatib, Fogel, and Jandik (2015) study the effects of CEO connectedness on acquisition performance.

ature by showing that various SNA measures associated with the lead underwriter can serve as important additional measures in assessing the effectiveness of lead SEO underwriters.

3 Theory and hypotheses development

In this section, we discuss the relevant theory and develop testable hypotheses for our empirical analyses. The theoretical literature (see, e.g., Chemmanur and Fulghieri (1994)) has argued that the role of an underwriter (whether in an SEO or an IPO) is that of an information producer, who produces noisy information about an issuing firm in the process of conducting due diligence. The underwriter may then convey this information to investors making use of its reputation as a certifying mechanism (thus mitigating concerns about their incentives to suppress negative information about the firm and report only favorable information). While Chemmanur and Fulghieri (1994) assume that underwriters can costlessly convey their information to potential investors in an equity issue, it is useful to explicitly consider the transmission mechanism of the above noisy information from the underwriter to institutional investors and how institutions process the above information when deciding whether or not to invest in the SEO of a firm. To accomplish this, here we introduce a network of investment banks connected (to a greater or lesser degree) to the lead SEO underwriter, with each investment bank having repeated interactions with a subset of institutional investors who may potentially invest in the SEO. A lead underwriter who is better connected to various investment banks will be able to convey the information it has produced about the true value of the SEO firm to these investment banks more efficiently. Given the repeated interactions between a particular investment bank in the lead underwriter's network and certain institutions, that investment bank, in its turn, will be able to more credibly (and efficiently) convey the information about the SEO firm's value to these institutions.

The above argument implies that the information asymmetry faced by institutions about the firm whose SEO is underwritten by a better-connected lead underwriter (more central to its network and with more extensive underwriter network) will be smaller, since such a lead underwriter will be able to more efficiently convey the information it has produced about the firm making the SEO to institutions which have repeated interactions with the investment banks in its network. This lower information asymmetry, in turn, will have several implications for the market's reaction to the announcement of an SEO; for various SEO characteristics; for post-SEO stock returns of issuing firms; and for the participation of institutions in that SEO. We will discuss in more detail below.

3.1 Underwriter centrality and information asymmetry

As we argued earlier, firms whose SEOs are underwritten by more central lead underwriters are likely to be characterized by a smaller extent of information asymmetry, since such lead underwriters are likely to be able to transmit the information they have produced about the firms conducting the SEOs to investors more efficiently through the networks of investment banks connected to them. This is the first hypothesis that we test here (**H1**).

3.2 Underwriter centrality and SEO announcement effects

The seminal paper by Myers and Majluf (1984) has shown that the announcement effect of an equity issue will be negative in the presence of information asymmetry. Further, the more severe the information asymmetry, the more negative the announcement effect. As we argued above, the extent of information asymmetry facing a firm conducting an SEO will be smaller if the SEO is underwritten by a lead underwriter who is more centrally located in its network. This, in turn, implies that the announcement effect of SEOs underwritten by more central lead underwriters will be algebraically larger (less negative). This is the next hypothesis that we test here (**H2**).

3.3 Underwriter centrality and the absolute value of SEO offer price revisions

We now discuss how the centrality of a lead SEO underwriter in its network may affect the SEO offer price revision in the SEO. First, it is useful to consider how the lead underwriter sets the initial offer price range. We assume here that the initial offer price range is chosen such that the middle point of this range is equal to the underwriter's expectation of the true value of the firm (based on its information production while conducting due diligence about the SEO firm). Second, subsequent to filing the registration statement with the initial SEO offer price range, the lead underwriter, in the process of conducting a road show and bookbuilding for the SEO, may perform two economically important roles.

The first such role the lead underwriter performs is that of disseminating information about the SEO firm to various institutions and other investors (through the network of investment banks connected to it). Because more central lead underwriters will be able to disseminate information more efficiently (and accurately) to institutions, the final offer price established as a result of such information dissemination during the book-building process will be closer to the midpoint of the initial SEO offer price range for offerings underwritten by such lead underwriters. This implies that a negative relation would be expected between lead underwriter centrality and the absolute value of the SEO offer price revision (**H3A**).

The second such role the lead underwriter performs is that of extracting information from institutions about their demand for the SEO firm's equity. The lead underwriter makes use of its investment banking network to extract information from institutions about their demand for the SEO firm's shares. The offer price is revised upward or downward from the midpoint of the initial offer price range depending on the information extracted by the lead underwriter from institutions. A more central underwriter may be in a better position to extract information useful for valuing the SEO firm's shares from institutions, making use of the investment banks in its network. If this is the case, we would expect a positive relation between lead underwriter centrality and the absolute value of the SEO offer price revision (**H3B**).

While our focus in this paper is on the relation between lead underwriter centrality and its effectiveness in information dissemination about the value of the SEO firm to investors, it is likely that both the above economic functions, namely, information dissemination and information extraction, occur during the book-building and roadshow process in an SEO. Which of the above economic roles of a lead SEO underwriter dominates is an empirical question which we will answer using our empirical analysis.

3.4 Underwriter centrality, the SEO discount, and SEO underpricing

Our earlier arguments imply that institutions' cost of evaluating SEOs underwritten by more central underwriters will be lower, given the lower information asymmetry they face with respect to the firms conducting these SEOs. In their theoretical analysis of SEOs, Chemmanur and Jiao (2011) argue that the SEO discount (the discount of the SEO offer price with respect to the previous day's closing price in the secondary market) is a compensation to institutional investors for their cost of information production about the firm making the SEO. If this is indeed the case, we would expect a negative relationship between the centrality of the SEO underwriter in its network of investment banks and the SEO discount (**H4**), given the negative relationship between underwriter centrality and the institutions' cost of producing information about the firm.

SEO underpricing is usually defined as the return from the SEO offer price and the closing price in the secondary market on the day of the SEO. SEO underpricing can be viewed as a mirror image of the SEO discount, since investors who buy at the offer price can profit from selling shares at the higher price prevailing on average at the close of that day. In other words, similar to the negative relationship between underwriter centrality and the SEO discount, we would also expect a negative relationship between underwriter centrality and SEO underpricing (given the negative relationship between underwriter centrality and investors' cost of producing information about the firm). This is the next hypothesis that we test here (**H5**).

3.5 Underwriter centrality and SEO firm market valuations

So far, we have argued that more central SEO underwriters will convey information about the SEO firm more efficiently to institutional investors, reducing the information asymmetry faced by institutions about the firm (and thereby their cost of evaluating the firm). This implies that the immediate post-SEO market valuations of firms whose SEOs are underwritten by more central lead underwriters will be greater.[4] This is the next hypothesis that we test here (**H6**).

[4]Intuitively, given the lower information asymmetry they face in evaluating these firms, institutions are likely to apply a lower cost of capital in computing their value.

3.6 Underwriter centrality and SEO firm equity ownership by institutional investors

As we argued above, institutional investors are likely to face a smaller extent of information asymmetry when considering investing in an SEO underwritten by a more central lead SEO underwriter. If we assume that institutional investors spend resources in evaluating each firm for investment, this implies that institutions' cost of producing information about SEOs underwritten by a more central lead underwriter will also be smaller. Assuming that institutions prefer to invest in SEO firms that are cheaper for them to evaluate, this implies that there will be a positive relation between underwriter centrality and the participation of institutional investors in an SEO (**H7**), since institutions are unlikely to invest in the SEOs which are harder for them to evaluate.

3.7 Underwriter centrality and SEO firm long-run stock returns

Even though, as we discussed above, more centrally located underwriters may be better at disseminating information to institutions and other investors in the SEO as well as in the secondary market, not all value-relevant information may be fully reflected in the secondary market prices immediately. In other words, some of this information may get reflected in the secondary market price of the SEO firm only gradually over time. If this is the case, then we would expect SEOs underwritten by more centrally located lead underwriters to have better long-run post-SEO stock returns (**H8**).[5]

[5] If all the information disseminated by the lead underwriter through its investment banking network is reflected in the SEO firm's immediate secondary market valuation, we would not expect any relationship between underwriter centrality and the long-run stock return of the SEO firm. However, there may be a variety of reasons why the immediate after-market stock price may not efficiently reflect all the information disseminated by the lead underwriter about the SEO firm: e.g., bounded rationality of retail investors in the immediate SEO after-market. Note that all long-run stock return studies around corporate events require the assumption of bounded rationality or limited market efficiency, similar to the one we make here. One may consider this to be a strong assumption, but, given the large empirical literature documenting post-event drift following earnings announcements and many other corporate events (see, e.g., Foster, Olsen, and Shevlin (1984) and Bernard and Thomas (1989)), one has to at least consider the possibility that the information revealed by many corporate actions taken by a firm is not always instantaneously reflected in its stock price.

4 Identification strategy

4.1 An instrumental variable analysis

We separate out two possible channels through which more central lead underwriters may obtain more favorable SEO characteristics. The first is selectivity, that is, more central lead underwriters may underwrite the SEOs of higher quality firms. The second is incremental value creation, that is, for a given firm quality, more central lead underwriters may be able to generate lower information asymmetry, less negative announcement effects, and other favorable SEO characteristics. We distinguish between the above two channels using an IV analysis. We choose the relative bargaining power of SEO firm's industry (*IRBP*) as our instrumental variable, which is defined as the average number of book runners for IPOs and SEOs in a given industry over three-year period prior to the SEO filing year. The rationale for our choice of this instrumental variable is that the relative bargaining power of firms in a certain industry (in terms of the ability of the firms in that industry to negotiate better deals with investment banks for various services that such banks provide to the firms in that industry) plays an important role when the firms in that industry decide which underwriters to hire to conduct their SEOs. The competition for underwriting business between book runners increases with the number of book runners participating in a given industry, which in turn increases the bargaining power of issuers in that industry. Therefore, we expect the firms in an industry with relatively higher bargaining power to be able to hire more central underwriters compared to the firms in an industry with relatively lower bargaining power (given that higher underwriter centrality is expected to generate certain benefits for such firms in terms of their SEO characteristics). Further, our instrumental variable is likely to satisfy the exclusion restriction (the requirement for our instrumental variable not to be correlated with the dependent variables in our regressions) as the average number of book runners in a given industry is unlikely to be correlated with the SEO announcement effects and various SEO characteristics of specific firms in that industry.[6]

[6] Huang and Zhang (2011) and Jeon, Lee, Nasser, and Via (2015) use similar IVs for the number of underwriters in their settings.

4.2 Analysis of lead SEO underwriter compensation

We conduct a test of underwriters' value creation in underwriting SEOs by analyzing lead SEO underwriter compensation. We argue that, if more central lead SEO underwriters are able to create more value through disseminating information about firms making SEOs more efficiently to institutions, they would get compensated more in the competitive market of underwriting business.

5 Data and sample selection

We collect data on SEOs from the Securities Data Company (SDC)/Platinum Global New Issues database. We first obtain the list of all SEOs conducted in the U.S. in 1980-2017 and select only offerings of common shares (thus excluding all other types of offerings such as real estate investment trusts, units, rights, spin-offs, American Depository Receipts, etc.) from this list. We then exclude financial firms (firms with standard industrial classification [SIC] codes between 6000 and 6999) and firms without underwriter information. We collect accounting and stock return data from Compustat and the Center for Research in Security Prices (CRSP), respectively; analyst coverage and analyst forecast data from the Institutional Brokers' Estimation System (IBES) database; and institutional holdings data from Thomson Reuters' institutional holdings (13F) database.

6 Measures characterizing underwriter networks

The literature on social network analysis (SNA) argues that the importance of an agent in its network depends on how central the agent is in that network (e.g., Lazarsfeld, Berelson, and Gaudet (1968)). The SNA literature borrows from the graph theory the concept of adjacency matrix, which is a way of representing agents' connections to other agents within their network, and which is used to measure the centrality of each agent in their network (e.g., Wasserman and Faust (1994)). We consider two underwriters to be connected if they were a part of the same syndicate (or multiple syndicates) which has underwritten an IPO or an SEO over the past five years. An adjacency matrix A in our setting is an $n \times n$ matrix where a_{ij} element of this matrix indicates a connection between underwriter i and underwriter j, and n is the total number of underwriters in the network.[7] In other words, a_{ij} equals to one if investment bank i has co-underwritten an IPO or an SEO with investment bank j over the past five years, and zero otherwise. We construct six different network centrality measures for lead SEO underwriters to capture various aspects of underwriter networks.

6.1 Degree, Indegree, and Outdegree

Our first three measures of lead SEO underwater centrality are *Degree*, *Indegree*, and *Outdegree*. These three measures count (in slightly different ways) the number of connections of an underwriter with other underwriters in its network. *Degree* is defined as the number of other unique underwriters with whom a given underwriter has co-underwritten an IPO or an SEO over the last five years. Formally, *Degree* for underwriter i is $d_i = \sum_{j=1}^{n} a_{ij}$, where a_{ij} is an element of the adjacency matrix. We expect a lead underwriter with higher *Degree* centrality to have a greater capacity to disseminate information about the issuing firm (whose SEO it underwrites) through the larger number of connections that this lead underwriter has in its network. One potential concern about this centrality measure is that *Degree* may be correlated with the size of the network, although this may not be a problem in our cross-sectional analysis. To rule out any potential bias from a change in the size of the network over time, we normalize *Degree* by the size of the network $(n-1)$ and use normalized *Degree* in our empirical analysis.

An adjacency matrix we use to construct *Degree* centrality measure is not directed; in other words it does not specify whether a particular investment bank has acted as a lead underwriter and invited another investment bank to join a syndicate to underwrite an IPO or an SEO (and form a connection), or whether that investment bank was invited by another lead underwriter to join an underwriting syndicate. An adjacency matrix can also be directed if it differentiates between underwriters who are invited to be syndicate members and those who lead underwriting syndicates. In a directed network, we can define two additional variations of *Degree* centrality, namely *Indegree* and *Outdegree*. *Indegree* for

[7]We construct our lead SEO underwriter centrality measures also using only SEO events. Our empirical findings using these alternative measures of lead SEO underwriter centrality are similar to those reported in this paper.

underwriter i is defined as the number of other unique underwriters who have invited underwriter i to be a syndicate member over the past five years. For example, $a_{ij} = 1$ and $a_{ji} = 0$ if underwriter j invites underwriter i into an IPO or an SEO underwriting syndicate over the past five years. *Indegree* captures the passive role that a given underwriter plays in its connection. An underwriter with a higher *Indegree* centrality may not have a large capacity to disseminate and propagate information; however it may still be a valuable member of an underwriting syndicate due to its repeated business with institutional investors. In contrast, *Outdegree* for underwriter i is defined as the number of other unique underwriters that underwriter i, acting as a lead underwriter, has invited to join an IPO or an SEO underwriting syndicate over the past five years. For example, $a_{ij} = 1$ and $a_{ji} = 0$ if underwriter i invites underwriter j to join an IPO or an SEO underwriting syndicate. *Outdegree* captures the active role that a given underwriter plays in its connection and reflects its greater capacity to disseminate and propagate information through the network. We normalize both *Indegree* and *Outdegree* by the size of the network $(n-1)$ as well.

6.2 Eigenvector

Our fourth measure of network centrality, *Eigenvector*, captures the importance of an underwriter in its network. It is a weighted sum of connections that a particular underwriter has in its network, where the weights are determined by assigning relative scores to all connections of that underwriter based on the centrality of each connected underwriter in the network. For example, if underwriter i and underwriter j have the same number of connections in their network then they will have the same *Degree* centrality. However, if most of underwriter i's connections are well-connected themselves, while most of underwriter j's connections are not well-connected, then underwriter i will have a higher *Eigenvector* centrality compared to underwriter j. Formally, *Eigenvector* for underwriter i is $e_i = \frac{1}{\lambda}\sum_{j=1}^{n} a_{ij}e_j$, where λ is the largest eigenvalue of the adjacency matrix and e is the eigenvector centrality score. In our setting, we expect an underwriter with higher *Eigenvector* centrality to be able to disseminate information within the network more efficiently given that its connections are better connected themselves with other underwriters in that network. We also normalize *Eigenvector* by the largest possible eigenvector element value in the network.

6.3 Betweenness

Our fifth measure of lead underwriter centrality, *Betweenness*, captures the extent to which an underwriter is able to act as an intermediary between two other underwriters (or groups of underwriters) who are not otherwise connected. For underwriter i, *Betweenness* is defined as the number of shortest paths between pairs of other underwriters passing through underwriter i. For example, underwriter i will have a higher *Betweenness* centrality compared to underwriter j, if the number of shortest paths between pairs of other underwriters passing through underwriter i is greater than that passing through underwriter j. Formally, *Betweenness* for underwriter i is $b_i = \sum_{i \neq j \neq k} \frac{p_{jki}}{p_{jk}}$, where p_{jk} is the number of shortest paths from underwriter j to underwriter k, and p_{ijk} is the number of shortest paths from underwriter j to underwriter k passing through underwriter i. In our setting, we expect an underwriter with higher *Betweenness* centrality to be able to disseminate information within its network more efficiently given its position within the network serving as a bridge linking other underwriters with each other.

6.4 2-StepReach

Our last measure of lead underwriter centrality, *2-StepReach*, counts the number of both direct and indirect connections which are two or less steps away from the lead underwriter. In general, *k-Step Reach* counts the number of other distinct agents within the network that a given agent can reach in k or less steps. When $k = 1$, *1-Step Reach* centrality is equivalent to *Degree* centrality. Unlike *Degree*, *2-StepReach* for underwriter i counts not only the number of direct connections but also the number of distinct indirect connections that can be reached within two steps from underwriter i. For example, if underwriter i and underwriter j have the same number of direct connections within their network then they will have the same Degree centrality. However, underwriter i will have a higher *2-StepReach* centrality compared to underwriter j, if underwriter j's direct connections themselves are connected to a larger number of other underwriters (that underwriter i is not connected to directly), while underwriter j's direct connections are connected to a smaller number of other underwriters (that underwriter j is not connected to directly). We expect a lead underwriter with a larger *2-StepReach* centrality to have a greater capacity to disseminate information within its network through both its direct as well as indirect connections. We also normalize *2-StepReach* by the size of the network $(n-1)$.

7 Empirical results

In this section, we present our methodology and empirical findings. Table 1 reports the summary statistics of all variables used in our empirical analysis.[8] Panel A of Table 1 reports the summary statistics of our six lead SEO underwriter centrality measures as well as our lead SEO underwriter reputation measure. In our data sample, in a given year lead SEO underwriters on average were directly connected to 18.7% of other underwriters in their networks, as measured by *Degree*. Further, lead SEO underwriters have formed most of these connections by inviting other underwriters into underwriting syndicates (lead SEO underwriters on average have invited 17.0% of other underwriters in their networks to join underwriting syndicates, as measured by *OutDegree*), while some other connections were formed when lead SEO underwriters were invited by other lead underwriters to join their underwriting syndicates (lead SEO underwriters on average were invited by 5.3% of other underwriters in their networks to join underwriting syndicates, as measured by *InDegree*). Next, lead SEO underwriters were on average on 5.3% of the shortest paths between pairs of underwriters within their network, as measured by *Betweenness*; and lead SEO underwriters were on average able to reach 70.0% of other underwriters within their networks counting both direct and indirect (two-step away) connections, as measured by *2-StepReach*. Finally, our measure of lead SEO underwriter reputation, *MktShare*, which is the lead SEO underwriter's share in the total proceeds raised in both IPO and SEO markets over the past five years, indicates that lead SEO underwriters on average raised 5.0% of the total proceeds both in the IPO and SEO markets over the previous five years.

Panel B of Table 1 reports the summary statistics of SEO announcement period stock returns. We estimate abnormal returns using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated

[8]We winsorize all variables except for our six measures of underwriter centrality at the 0.5% and 99.5% levels to reduce potential biases caused by outliers. Our results without winsorization are qualitatively similar to those reported in this paper.

over a 150-day period ending 50 days prior to the SEO announcement date.[9] The average cumulative abnormal returns upon SEO announcements over 3-day (-1 to +1) and 5-day (-2 to +2) windows are -2.4%, and -2.5%, respectively, which is consistent with other studies reporting negative announcement period returns for SEOs. Lastly, Panel C of Table 1 reports the summary statistics of SEO and firm characteristics. For example, in our data sample the SEO offer price is at 4.5% discount compared to the previous day's closing price; the issue day closing price is 2.7% higher than the SEO offer price; and institutional investors hold on average 55.7% of SEO firms' shares after the SEO.

We first examine the effect of lead SEO underwriter centrality on the reduction in the extent of information asymmetry faced by SEO firms in the market and present our empirical results in Subsection 7.1. Next, we examine the effect of lead SEO underwriter centrality on SEO announcement period returns in Subsection 7.2. We continue our analysis by examining the effect of lead SEO underwriter centrality on various SEO characteristics, namely, the absolute value of SEO offer price revision, SEO discount and SEO underpricing, SEO firm market valuation, SEO firm post-issue long-term stock returns, and institutional investor holdings of SEO firms' equity in Subsections 7.3, 7.4, 7.5, 7.6, and 7.7, respectively. Lastly, we discuss potential endogeneity concerns and present our instrumental variable analyses in Subsection 7.8.

7.1 Lead SEO underwriter centrality and information asymmetry facing SEO firms

In this subsection, we test our hypothesis **H1** which predicts that firms whose SEOs are underwritten by more central lead underwriters are likely to face a lower extent of information asymmetry in the financial market. We test this hypothesis by regressing several measures of information asymmetry on our six lead SEO underwriter centrality measures as described in Section 6 and other controls. Following the existing literature (see, e.g., Brennan and Subrahmanyam (1995), Krishnaswami and Subramaniam (1999), Clarke and Shastri (2000), and Chemmanur, Paeglis, and Simonyan (2009)), we make use of three different proxies for

[9] We also estimate abnormal returns using alternative models such as Fama-French three-factor model, and Carhart four-factor model (see, e.g., Fama and French (1993), and Carhart (1997)). Our results remain qualitatively similar using these alternative estimation models.

information asymmetry faced by SEO firms in the financial market: analysts' forecast error (*ForError*), the dispersion in analysts' forecasts (*Dispersion*), and the average daily bid-ask spread (*B-S Spread*). *ForError* is the absolute difference between the mean EPS forecasted by the financial analysts following the SEO firm and the actual EPS divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *Dispersion* is the standard deviation in EPS estimates forecasted by the financial analysts following the SEO firm divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *B-S Spread* is the average daily bid-ask spread over one-month (21-trading-day) period after the SEO issue date. The daily bid-ask spread of stock i on day t is $D\_Spread_{i,t} = (Ask_{i,t} - Bid_{i,t})/M_{i,t}$, where $Ask_{i,t}$ and $Bid_{i,t}$ are the ask price and bid price of stock i on day t from the CRSP daily data, and $M_{i,t}$ is the mean of $Ask_{i,t}$ and $Bid_{i,t}$. We also add several control variables to rule out potential confounding effects. Following Bajo, Chemmanur, Simonyan, and Tehranian (2016), we construct a measure of lead SEO underwriter reputation, *MktShare*, using the lead SEO underwriter's share of total proceeds raised in both IPO and SEO markets over the five-year period prior to the SEO issue year. Given that *MktShare* is highly correlated with our lead SEO underwriter centrality measures, we include the residuals from regressing *MktShare* on our six lead SEO underwriter centrality measures (*xMktShare*) as a control variable in our regressions to mitigate multicollinearity problems. In our regressions we also control for SEO offer size (*OfferSize*) which is the proceeds raised through the SEO and is measured in billions of US dollars, SEO firm size (*FirmSize*) which is the book value of SEO firm's total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars, and the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date (*PriorMktRetIss*). Finally, we also include offer year $\times$ two-digit SIC industry code fixed effects to control for time-varying unobservables across different industries.

In Panels A and B of Table 2, we report the results of our regressions on the effect of lead SEO underwriter centrality on analysts' forecast error (*ForError*) and analysts' forecast dispersion (Dispersion), respectively. We find that all six lead SEO underwriter centrality measures have negative coefficient estimates in both panels except for *Indegree* in

Panel B which has an insignificantly positive coefficient estimate.[10] These results suggest that the information asymmetry faced by SEO firms in the financial market (proxied by the analysts' forecast error and analysts' forecast dispersion) is decreasing in the centrality of lead SEO underwriters. These results are also economically significant. For example, a one-standard-deviation increase in *Degree* (in other words, if the average proportion of other underwriters in the network that lead SEO underwriter is connected to increases from 18.7% to 29.4%) decreases *ForError* by 0.232 (i.e., larger than the mean *ForError* in our sample) and it decreases *Dispersion* by 0.055 (about 87.1% decrease relative to the mean *Dispersion* in our sample). These findings provide support for our hypothesis **H1**.

Panel C of Table 2 presents our regression results using the average daily bid-ask spread as the dependent variable. We find that all six lead SEO underwriter centrality measures in our regressions have negative and significant coefficient estimates.[11] These results suggest that the information asymmetry faced by SEO firms in the financial market (proxied by the average daily bid-ask spread) is decreasing in the centrality of lead SEO underwriters. These results are also economically significant. For example, a one-standard deviation increase in *Degree* decreases *B-S Spread* by 0.193 percentage points (about 21.4% decrease relative to the mean *B-S Spread* in our sample). These findings provide further support for our hypothesis **H1**.

7.2 Lead SEO underwriter centrality and SEO announcement effects

We measure SEO announcement period returns using cumulative abnormal returns (CARs) computed over two different windows around the SEO announcement date (-1 to +1 days and -2 to +2 days) and use these CARs as dependent variables in our regressions. The independent variables of interest in our regressions are our six lead SEO underwriter centrality measures as described in Section 6. Unlike in Table 2, the time window used to measure

[10] We also measure analysts' forecast errors and analysts' forecast dispersion using forecast data at the end of the fiscal year of the SEO and find quantitatively similar results. In addition, our results are also robust to using alternative measures, i.e., number of analysts following firms at the end of the fiscal quarter (year) of their SEO, changes in forecast error and forecast dispersion from the end of the fiscal quarter prior to the SEO issue to the end of the fiscal quarter of the SEO issue.

[11] Our results are also robust to using alternative measures, i.e., the bid-ask spread of firms at the end of the month (fiscal quarter) of their SEO, and changes in average daily bid-ask spread from one month (21-trading days) before the SEO issue date to one month (21-trading days) after the SEO issue date.

our six lead SEO underwriter centrality measures in this subsection is a five-year period prior to the SEO filing year since announcement effects happen before the SEO issue date. In all of our regressions, we control for underwriter reputation, SEO offer size, firm size, and one-month stock market return prior to the SEO filing day.[12] Finally, we also include filing year $\times$ two-digit SIC industry code fixed effects.

Panel A of Table 3 presents the results of our regressions using CARs cumulated over a 3-day window around the SEO announcement date (CAR [-1:1]) as the dependent variable. All six lead SEO underwriter centrality measures in our regressions have positive coefficient estimates and five of them are statistically significant (except *Indegree*). Given that the cumulative abnormal returns upon SEO announcements are negative on average, this suggests that lead SEO underwriters who are positioned more centrally within their investment banking networks are able to reduce the degree of information asymmetry faced by SEO firms in the market to a greater extent. These results are also robust to using different event windows when calculating announcement period CARs, as shown in Panel B of Table 3. In panel B of Table 3 we use CARs cumulated over a 5-day window (CAR [-2:2]), and our findings are essentially the same as in Panel A of Table 3. The results in Table 3 are also economically significant. For example, a one-standard-deviation increase in *Degree* leads to a 0.42 percentage point increase in both 3-day and 5-day CARs around SEO announcements. These findings provide supports for our hypothesis **H2**.

7.3 Lead SEO underwriter centrality and the absolute value of SEO offer price revision

In this subsection, we test our hypotheses **H3A(B)** which predicts a negative (positive) relationship between lead SEO underwriter centrality and the absolute value of SEO offer price revision. We regress the absolute value of the percentage difference between the SEO offer price and the mid-point of the initial filing range (*AbsPriceRev*) on our six lead SEO underwriter centrality measures while controlling for underwriter reputation, SEO offer size, firm size, one-month stock market return prior to the SEO issue date, and offer year

[12] All control variables are also measured over a time window ending prior to the filing year in regressions testing SEO announcement effects.

$\times$ two-digit SIC industry code fixed effects.

Table 4 presents the results of our regressions. All six lead SEO underwriter centrality measures in our regressions have negative coefficient estimates and five of them are statistically significant (except *Indegree*). These results suggest that more central lead SEO underwriters are associated with smaller absolute value of SEO offer price revisions. The insignificant coefficient estimate of *Indegree* (compared to the statistically significant coefficient estimate of *Outdegree*) indicates that lead SEO underwriters which are more likely to be invited into underwriting syndicates (rather than invite other underwriters into underwriting syndicates) have relatively lesser capacity in terms of efficient dissemination of information within investment banking networks. The results are also economically significant. For example, a one-standard-deviation increase in *Degree* decreases the absolute value of SEO offer price revisions by 1.29 percentage points (about 12.1% of the average absolute value of SEO offer price revisions). Our findings here indicate that more central lead SEO underwriters disseminate information about the firms whose SEOs they underwrite within their investment banking networks more efficiently. These findings provide support for our hypothesis **H3A** and contradict our hypothesis **H3B**.

7.4 Lead SEO underwriter centrality, SEO discount, and SEO underpricing

In this subsection, we study the effect of lead SEO underwriter centrality on SEO discount and SEO underpricing. We define SEO discount as the percentage difference between SEO offer price and the closing price on the day prior to the SEO issue day (*Discount*), and SEO underpricing as the percentage difference between the SEO issue day closing price and SEO offer price (*Underpricing*). We regress the above two variables on our six lead SEO underwriter centrality measures while controlling for underwriter reputation, SEO offer size, firm size, one-month stock market return prior to the SEO issue date, and offer year $\times$ two-digit SIC industry code fixed effects.

Table 5 reports the results of our regressions using *Discount* as the dependent variable. All six lead SEO underwriter centrality measures in our regressions have negative coefficient estimates and five of them are statistically significant (except *Indegree*). This suggests that

SEOs underwritten by more central lead SEO underwriters are likely to have smaller SEO discounts. These results are also economically significant. For example, a one-standard-deviation increase in *Degree* reduces SEO discount by 1.07 percentage points (about 23.5% decrease relative to the mean SEO discount of 4.55%). These results provide support for our hypothesis **H4**. Similar to our findings in Table 4, the insignificant coefficient estimate of *Indegree* suggests that lead SEO underwriters which are more likely to be invited into underwriting syndicates (rather than invite other underwriters into underwriting syndicates) have relatively lesser capacity in terms of efficient dissemination of information within their investment banking networks.

In Table 6 we report the results of our regressions using *Underpricing* as the dependent variable. All six lead SEO underwriter centrality measures have negative and statistically significant coefficient estimates. These results suggest that firms whose SEOs are underwritten by more central lead underwriters leave less money on the table when they conduct their SEOs. This provides support for our hypothesis **H5**. These results are also economically significant. For example, a one-standard-deviation increase in *Degree* reduces SEO underpricing by 0.55 percentage points (about 19.7% decrease relative to the mean SEO underpricing of 2.82%).

7.5 Lead SEO underwriter centrality and SEO firm market valuation

In this subsection, we test our hypothesis **H6** which predicts a positive relationship between lead SEO underwriter centrality and SEO firm market valuation. We measure market valuation of SEO firms using industry-adjusted Tobin's Q ratio (*QAdj*). We define Tobin's Q as the ratio of the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and SEO issue day closing price. We construct industry-adjusted Tobin's Q by subtracting contemporaneous 2-digit SIC code industry median Q ratio from the above proxy. We regress *QAdj* on our six lead SEO underwriter centrality measures while controlling for underwriter reputation, SEO offer size, firm size, one-month stock market return prior to the SEO issue date, and offer year $\times$ two-digit SIC industry code fixed effects.

Table 7 reports the results of our regressions. All six lead SEO underwriter centrality measures have positive coefficient estimates and five of them are statistically significant (except for *Indegree*). This suggests that firms whose SEOs are underwritten by more central lead SEO underwriters are likely to have higher market valuations on their SEOs' issue days. These results are also economically significant. For example, a one-standard-deviation increase in *Degree* on average increases industry-adjusted Tobin's Q ratio by 0.063 which is a sizable increase compared to the mean industry-adjusted Tobin's Q ratio in our sample of -0.006. These results provide support for our hypothesis **H6**.

7.6 Lead SEO underwriter centrality and SEO firm equity ownership by institutional investors

In this subsection, we study the effect of lead SEO underwriter centrality on the SEO firm equity ownership by institutional investors. We measure SEO firm equity ownership by institutional investors using two variables: the natural logarithm of the number of institutional investors holding SEO firms' equity at the end of the first quarter after the SEO (*Ln(InstNum)*) and the proportion of SEO firms' equity held by institutional investors at the end of the first quarter after the SEO (*InstProp*). We regress the above two variables on our six lead SEO underwriter centrality measures while controlling for underwriter reputation, SEO offer size, firm size, one-month stock market return prior to the SEO issue date, and offer year $\times$ two-digit SIC industry code fixed effects.

We report the results of our regressions in Panels A (for *Ln(InstNum)*) and B (for *InstProp*) of Table 8.[13] We find that all six lead SEO underwriter centrality measures have significantly positive coefficient estimates in both panels. This indicates that institutional investors are likely to have greater equity ownership (post-SEO) in the firms whose SEOs are underwritten by more central lead underwriters (both in terms of the number of institutional investors holding SEO firms' equity as well as the proportion of SEO firms' equity held by institutional investors). These results are also economically significant. For example, a one-standard-deviation increase in *Degree* increases the number of institutional investors

[13] We also run the same set of regressions using the change in institutional ownership (from before to after SEO issues) as dependent variables and find qualitatively similar results as in Table 8.

holding SEO firms' equity by 35.6% (about 36 more institutional investors) and increases the proportion of SEO firms' equity held by institutional investors by 6.2 percentage points (which corresponds to approximately 9.7% increase relative to the average proportion of SEO firm's equity held by institutional investors of 54.2%). These results provide further support for our hypothesis **H7**.

7.7 Lead SEO underwriter centrality and post-SEO long-run stock returns of issuing firms

In this subsection, we test our hypothesis **H8** which predicts a positive relationship between lead SEO underwriter centrality and post-SEO long-run stock returns of SEO firms. We measure the post-SEO long-run stock return performance of SEO firms by computing their buy-and-hold abnormal stock returns over 252 trading days after the SEO (*BHAR*). We estimate abnormal returns using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO issue date. We then regress the above variable on our six lead SEO underwriter centrality measures while controlling for underwriter reputation, SEO offer size, firm size, one-month stock market return prior to the SEO issue date, and offer year × two-digit SIC industry code fixed effects.

We report our findings on the relationship between lead SEO underwriter centrality and SEO firm post-issue long-run stock returns in Table 9. All six lead SEO underwriter centrality measures have positive coefficient estimates and four of them are statistically significant (except for *Indegree* and *2-StepReach*). This suggests that firms whose SEOs are underwritten by more central lead SEO underwriters realize better post-SEO long-run stock performance. These findings are also economically significant. For example, a one-standard-deviation increase in *Degree* increases *BHAR* by 14.0 percentage points. These results provide support for our hypothesis **H8**.

7.8 Identification results

7.8.1 Instrumental variable analysis

The results of our IV analysis are shown in Table 10: we report the results of our first-stage regressions where we regress *Degree* (our lead SEO underwriter centrality measure) on our instrument and a set of control variables, and our second-stage regressions where we regress various SEO and firm characteristics (the forecast error of financial analysts, the dispersion in analysts' forecasts, the average daily bid-ask spread, announcement period returns, absolute value of SEO offer price revision, SEO discount, SEO underpricing, industry-adjusted Q ratio, post-SEO long-run stock returns, and the number of institutional investors holding SEO firm shares) on the predicted values of *Degree* from first-stage regressions and a set of control variables.[14] The control variables in this analysis are the same as in our baseline OLS regressions reported in previous sections.[15]

Our first-stage regressions in Table 10 demonstrate that our instrumental variable is positively correlated with lead SEO underwriter centrality measured by *Degree* confirming our expectations discussed above and validating the relevance of our instrument. Our first-stage regressions also present the F-statistics of the weak instruments test (or the test of excluded instruments).[16] Given that the F-statistic reported for the first-stage regressions in Table 10 are well above the critical value of 8.96 (except when we use the industry-adjusted Q ratio as our dependent variable), the null hypothesis that our instrument is weak is strongly rejected.

Our second-stage regressions in Table 10 show that the coefficient estimates of predicted values of *Degree* from first-stage regressions have the same signs as reported in our baseline OLS regressions and they are all statistically significant except for *QAdj* and *BHAR*. This indicates that, even after controlling for potential endogeneity of lead SEO underwriter cen-

[14] We conduct our IV analysis using other five measures of lead SEO underwriter centrality as well. The results of such IV analyses are similar to those reported here using *Degree*. For the sake of brevity we do not report these additional results in this paper; however, these results are available to interested readers upon request.

[15] We do not control for industry fixed effects in our 2SLS regressions given that our instrumental variable is an industry level variable.

[16] This test is used to determine whether instrumental variables used in first-stage regressions are strong. In their survey of the literature on weak instruments, Stock, Wright, and Yogo (2002) develop benchmarks for the necessary magnitude of the F-statistic. They indicate that if the number of instruments is equal to one, then the critical value of the F-statistic is 8.96.

trality, SEOs underwritten by more central lead SEO underwriters (as measured by *Degree*) are associated with higher (less negative) announcement period returns, smaller absolute value of SEO offer price revisions, lower SEO discounts, and lower SEO underpricing; and firms whose SEOs are underwritten by more central lead SEO underwriters (as measured by *Degree*) are associated with less information asymmetry (proxied by analysts? forecast error, analysts' forecast dispersion, and average bid-ask spread) and a larger number of institutional investors holding their equity post-SEO. Overall, our IV analysis demonstrates the robustness of our findings in previous subsections.

7.9 Analysis of lead SEO underwriter compensation

In this subsection, we test whether more central lead SEO underwriters are able to create more value in underwriting SEOs by studying the effect of lead SEO underwriter centrality on various measures of underwriter compensation. We measure underwriter compensation using three different proxies: gross spread, management fees, and underwriting fees. *GrossSpread* is the total compensation for the underwriting syndicate and is measured in millions of US dollars. *MgmtFee* is the management fee paid to the lead managers for their managing service and is measured in millions of US dollars. *UndwrtFee* is the underwriting fee paid to the lead and co-managers for their underwriting service and is measured in millions of US dollars. We control for the same set of control variables as above in all of our regressions (*xMktShare*, *OfferSize*, *FirmSize*, and *PriorMktRetIss*). Lastly, we also include issue year $\times$ two-digit SIC industry fixed effects to control for time-varying unobservables across different industries.

Panel A of Table 11 presents our empirical results on lead SEO underwriter compensation using *GrossSpread* as the dependent variable. All lead SEO underwriter centrality measures, except for *Indegree*, have positive and significant coefficient. These results are also robust to alternative measures of lead underwriters' compensation as shown in Panels B (for *MgmtFee*) and C (for *UndwrtFee*) of Table 11. All results are also economically significant. For example, a one-standard-deviation increase in *Degree* increases the dollar amount of gross spread, management fees, and underwriting fees by \$550,622, \$44,298, and \$66,982, respectively. These findings indicate that more central lead SEO underwriters are able to create more value since they are rewarded with higher compensation in the competitive market of underwriting business.

8 Conclusion

Using various "centrality" measures from Social Network Analysis (SNA), we analyze, for the first time in the literature, how the location of a lead SEO underwriter in its network of investment banks affects various aspects of seasoned equity offerings (SEOs). We hypothesize that investment banking networks perform an important economic role in the underwriting process for SEOs, namely, that of information dissemination, where the lead underwriter uses its investment banking network to disseminate information about the SEO firm to institutional investors. Consistent with the above information dissemination role, we show that firms whose SEOs are underwritten by more central lead underwriters are associated with a smaller extent of information asymmetry in the equity market. We then develop testable hypotheses based on the information dissemination role of underwriter networks for the relationship between SEO underwriter centrality and various SEO characteristics, which we test in our empirical analysis. Consistent with the above hypotheses, we find that more central lead SEO underwriters are associated with less negative SEO announcement effects; smaller SEO offer price revisions; smaller SEO discounts and underpricing; higher immediate post-SEO equity valuations of issuing firms; and greater post-SEO long-run stock returns of issuing firms. We also find that SEOs with more central lead underwriters are associated with greater institutional investor participation. Our instrumental variable (IV) analysis using the industry-average bargaining power of underwriters relative to issuers as the instrument shows that the above results are causal. Consistent with greater value creation by more central lead SEO underwriters, we find that such lead SEO underwriters receive greater compensation.

References

Altınkılıç, Oya and Robert S Hansen (2003), "Discounting and underpricing in seasoned equity offers." *Journal of Financial Economics*, 69, 285–323.

Asquith, Paul and David W Mullins (1986), "Equity issues and offering dilution." *Journal of Financial Economics*, 15, 61–89.

Bajo, Emanuele, Thomas J Chemmanur, Karen Simonyan, and Hassan Tehranian (2016), "Underwriter networks, investor attention, and initial public offerings." *Journal of Financial Economics*, 122, 376–408.

Benveniste, Lawrence M and Paul A Spindt (1989), "How investment bankers determine the offer price and allocation of new issues." *Journal of Financial Economics*, 24, 343–361.

Bernard, Victor L and Jacob K Thomas (1989), "Post-earnings-announcement drift: Delayed price response or risk premium?" *Journal of Accounting Research*, 27, 1–36.

Booth, James R and Richard L Smith (1986), "Capital raising, underwriting and the certification hypothesis." *Journal of Financial Economics*, 15, 261–281.

Brennan, Michael J and Avanidhar Subrahmanyam (1995), "Investment analysis and price formation in securities markets." *Journal of Financial Economics*, 38, 361–381.

Carhart, Mark M (1997), "On persistence in mutual fund performance." *Journal of Finance*, 52, 57–82.

Chemmanur, Thomas J and Paolo Fulghieri (1994), "Investment bank reputation, information production, and financial intermediation." *Journal of Finance*, 49, 57–79.

Chemmanur, Thomas J, Shan He, and Gang Hu (2009), "The role of institutional investors in seasoned equity offerings." *Journal of Financial Economics*, 94, 384–411.

Chemmanur, Thomas J and Yawen Jiao (2011), "Institutional trading, information production, and the SEO discount: A model of seasoned equity offerings." *Journal of Economics & Management Strategy*, 20, 299–338.

Chemmanur, Thomas J, Imants Paeglis, and Karen Simonyan (2009), "Management quality, financial and investment policies, and asymmetric information." *Journal of Financial and Quantitative Analysis*, 44, 1045–1079.

Chuluun, Tuugi (2015), "The role of underwriter peer networks in IPOs." *Journal of Banking & Finance*, 51, 62–78.

Clarke, Jonathan and Kuldeep Shastri (2000), "On information asymmetry metrics." *Available at SSRN 251938.*

Corwin, Shane A (2003), "The determinants of underpricing for seasoned equity offers." *Journal of Finance*, 58, 2249–2279.

El-Khatib, Rwan, Kathy Fogel, and Tomas Jandik (2015), "CEO network centrality and merger performance." *Journal of Financial Economics*, 116, 349–382.

Engelberg, Joseph, Pengjie Gao, and Christopher A Parsons (2012), "Friends with money." *Journal of Financial Economics*, 103, 169–188.

Fama, Eugene F and Kenneth R French (1993), "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33, 3–56.

Foster, George, Chris Olsen, and Terry Shevlin (1984), "Earnings releases, anomalies, and the behavior of security returns." *Accounting Review*, 574–603.

Gao, Xiaohui and Jay R Ritter (2010), "The marketing of seasoned equity offerings." *Journal of Financial Economics*, 97, 33–52.

Gerard, Bruno and Vikram Nanda (1993), "Trading and manipulation around seasoned equity offerings." *Journal of Finance*, 48, 213–245.

Giammarino, Ronald M and Tracy Lewis (1988), "A theory of negotiated equity financing." *Review of Financial Studies*, 1, 265–288.

Gibson, Scott, Assem Safieddine, and Ramana Sonti (2004), "Smart investments by smart money: Evidence from seasoned equity offerings." *Journal of Financial Economics*, 72, 581–604.

Gustafson, Matthew T (2018), "Price pressure and overnight seasoned equity offerings." *Journal of Financial and Quantitative Analysis*, 53, 837–866.

Hochberg, Yael V, Alexander Ljungqvist, and Yang Lu (2007), "Whom you know matters: Venture capital networks and investment performance." *Journal of Finance*, 62, 251–301.

Huang, Rongbing and Donghang Zhang (2011), "Managing underwriters and the marketing of seasoned equity offerings." *Journal of Financial and Quantitative Analysis*, 46, 141–170.

Jeon, Jin Q, Cheolwoo Lee, Tareque Nasser, and M Tony Via (2015), "Multiple lead underwriter IPOs and firm visibility." *Journal of Corporate Finance*, 32, 128–149.

Krishnaswami, Sudha and Venkat Subramaniam (1999), "Information asymmetry, valuation, and the corporate spin-off decision." *Journal of Financial Economics*, 53, 73–112.

Larcker, David F, Eric C So, and Charles CY Wang (2013), "Boardroom centrality and firm performance." *Journal of Accounting and Economics*, 55, 225–250.

Lazarsfeld, Paul Felix, Bernard Berelson, and Hazel Gaudet (1968), *The people's choice: How the voter makes up his mind in a presidential campaign*, volume 3. Columbia University Press New York.

Loderer, Claudio F, Dennis P Sheehan, and Gregory B Kadlec (1991), "The pricing of equity offerings." *Journal of Financial Economics*, 29, 35–57.

Loughran, Tim and Jay R Ritter (1995), "The new issues puzzle." *Journal of Finance*, 50, 23–51.

Masulis, Ronald W and Ashok N Korwar (1986), "Seasoned equity offerings: An empirical investigation." *Journal of Financial Economics*, 15, 91–118.

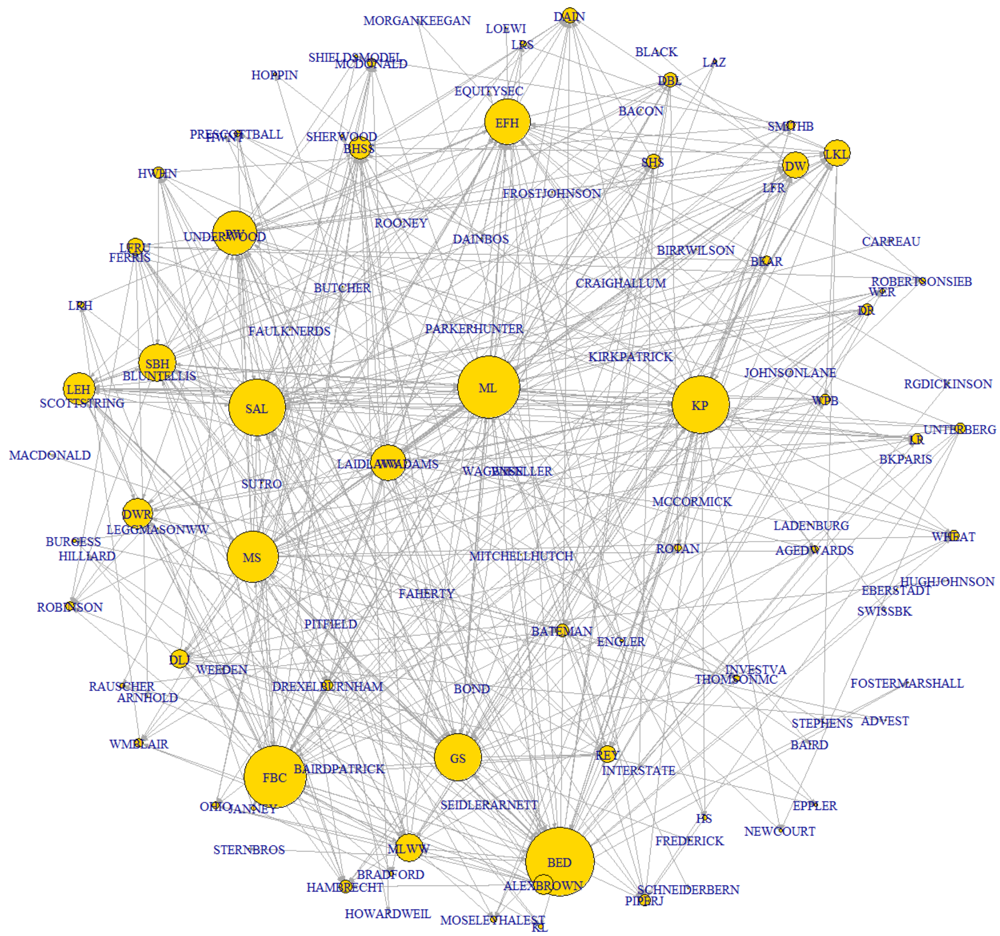
Myers, Stewart C and Nicholas S Majluf (1984), "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics*, 13, 187–221.

Ritter, Jay R (2003), "Investment banking and securities issuance." In *Handbook of the Economics of Finance*, volume 1, 255–306, Elsevier.

Stock, James H, Jonathan H Wright, and Motohiro Yogo (2002), "A survey of weak instruments and weak identification in generalized method of moments." *Journal of Business & Economic Statistics*, 20, 518–529.

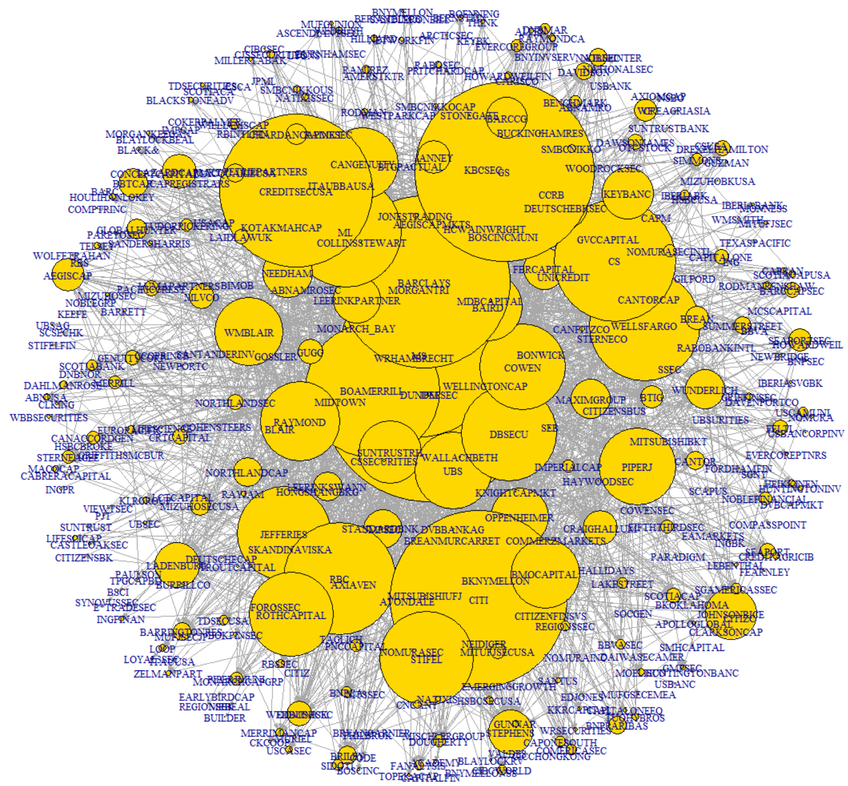
Titman, Sheridan and Brett Trueman (1986), "Information quality and the valuation of new issues." *Journal of Accounting and Economics*, 8, 159–172.

Wasserman, Stanley and Katherine Faust (1994), *Social network analysis: Methods and applications*, volume 8. Cambridge university press.



Gray arrows between pairs of underwriters indicate that the pair was a part of an IPO or an SEO syndicate in the previous five-year period (1975 – 1979). Arrows originate from lead underwriters and point in the direction of non-lead members of underwriting syndicates. The size of each circle represents the size of degree centrality for each underwriter.

Figure 1: **Network of SEO underwriters in 1980**



Gray arrows between pairs of underwriters indicate that the pair was a part of an IPO or an SEO syndicate in the previous five-year period (2012 – 2016). Arrows originate from lead underwriters and point in the direction of non-lead members of underwriting syndicates. The size of each circle represents the size of degree centrality for each underwriter.

Figure 2: **Network of SEO underwriters in 2017**

Table 1: **Summary Statistics**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years prior to the SEO issue year as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years prior to the SEO issue year. *CAR [-1: 1]*, and *CAR [-2: 2]* are the cumulative abnormal returns on SEO firms' equity cumulated over 3 days (from day -1 to day +1), and 5 days (from day -2 to day +2) around SEO announcement dates, respectively. The abnormal return is estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO announcement date. *ForError* is the absolute difference between the mean EPS forecasted by the financial analysts following the SEO firm and the actual EPS divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *Dispersion* is the standard deviation in EPS estimates forecasted by the financial analysts following the SEO firm divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *B-S Spread* is the average daily bid-ask spread over one-month (21-trading-day) period after the SEO issue date, where the daily bid-ask spread is equal to the difference between ask price and bid price divided by the mean of ask price and bid price. *Ln(InstNum)* is the natural logarithm of the number of institutional investors holding SEO firm shares at the end of the first fiscal quarter after the SEO. *InstProp* is the proportion of SEO firm shares held by institutional investors at the end of the first fiscal quarter after the SEO. *AbsPriceRev* is the absolute value of the percentage difference between the SEO offer price and the midpoint of initial filing range. *Discount* is the percentage difference between SEO offer price and the closing price on the day prior to the SEO issue day. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *QAdj* is the industry-adjusted Q ratio of SEO firms. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and SEO issue day closing price. *QAdj* is constructed by subtracting contemporaneous 2-digit SIC code industry median Q ratio from the SEO firm Q ratio. *BHAR* is the buy-and-hold abnormal return on SEO firms' equity over 252 trading days post-SEO (starting from the first day after the SEO issue day). The abnormal return is estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO issue date. *GrossSpread* is the total compensation for the underwriting syndicate and is measured in millions of US dollars. *MgmtFee* is the management fee paid to the lead managers for their managing service and is measured in millions of US dollars. *UndwrtFee* is the underwriting fee paid to the lead and co-managers for their underwriting service and is measured in millions of US dollars. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. *PriorMktRetFile* and *PriorMktRetIss* are the returns on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO filing date and prior to the SEO issue date, respectively. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels.

Panel A: Summary statistics of underwriter centrality measures and market share

	N	Mean	Median	Min.	Max.	S.D.
Degree	8,201	0.187	0.168	0.004	0.460	0.107
Outdegree	8,201	0.170	0.154	0.003	0.457	0.118
Indegree	8,201	0.053	0.052	0	0.119	0.021
Betweenness	8,201	0.053	0.035	0	0.211	0.051
Eigenvector	8,201	0.169	0.186	0	0.331	0.067
2-StepReach	8,201	0.700	0.649	0.009	0.962	0.194
MktShare	8,201	0.050	0.033	0	0.191	0.050

Panel B: Summary statistics of SEO announcement effects

	N	Mean	Median	Min.	Max.	S.D.
CAR [-1:1]	6,658	-0.024	-0.021	-0.293	0.264	0.068
CAR [-2:2]	6,653	-0.025	-0.024	-0.349	0.371	0.085

Panel C: Summary statistics of SEO and firm characteristics						
	N	Mean	Median	Min.	Max.	S.D.
ForError	6,707	0.112	0.001	0	12.683	1.008
Dispersion	6,122	0.037	0.001	0	4.151	0.331
B-S Spread	6,832	0.009	0.004	0	0.056	0.011
Ln(InstNum)	8,013	4.191	4.369	0	6.633	1.121
InstProp	7,981	0.557	0.558	0	1.202	0.292
AbsPriceRev	7,449	10.522	7.051	0	79.710	11.879
Discount	7,519	4.501	2.693	-9.212	42.514	6.930
Underpricing	7,561	2.736	1.558	-13.638	30.503	5.054
QAdj	7,617	-0.009	0	-6.007	11.869	1.866
BHAR	7,529	-1.340	-0.485	-36.520	2.470	3.733
GrossSpread	7,762	5.201	3.030	0.115	48.774	6.663
UndwrtFee	4,749	1.068	0.646	0.046	9.331	1.302
MgmtFee	4,749	1.077	0.617	0.045	10.688	1.398
OfferSize	8,201	0.142	0.065	0.002	1.782	0.230
FirmSize	7,596	1.987	0.237	0.004	48.255	5.718
PriorMktRetIss	8,197	0.015	0.017	-0.098	0.131	0.036
PriorMktRetFile	7,617	0.015	0.017	-0.131	0.144	0.038

Table 2: **Relation between lead SEO underwriter centrality and information asymmetry facing SEO firms**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *ForError* is the absolute difference between the mean EPS forecasted by the financial analysts following the SEO firm and the actual EPS divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *Dispersion* is the standard deviation in EPS estimates forecasted by the financial analysts following the SEO firm divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *B-S Spread* is the average daily bid-ask spread over one-month (21-trading-day) period after the SEO issue date, where the daily bid-ask spread is equal to the difference between ask price and bid price divided by the mean of ask price and bid price. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relation between lead SEO underwriter centrality and analyst forecast errors at the end of the first post-SEO fiscal quarter

Dependent Variable	ForError					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	-2.177***					
	(-7.33)					
Indegree		-3.439**				
		(-2.16)				
Outdegree			-1.880***			
			(-7.12)			
Betweenness				-2.456***		
				(-4.41)		
Eigenvector					-4.226***	
					(-8.25)	
2-StepReach						-3.011***
						(-11.29)
xMktShare	1.613*	-2.512***	1.684*	-1.350*	1.179	-0.204
	(1.81)	(-4.37)	(1.83)	(-1.70)	(1.53)	(-0.34)
OfferSize	-0.298*	-0.391**	-0.283*	-0.354**	-0.319**	-0.317**
	(-1.87)	(-2.44)	(-1.77)	(-2.21)	(-2.00)	(-2.00)
PriorMktRetIss	-0.448	-0.483	-0.430	-0.463	-0.419	-0.448
	(-0.58)	(-0.62)	(-0.55)	(-0.59)	(-0.54)	(-0.58)
FirmSize	7.709	8.491	7.866	8.386	7.484	7.361
	(1.14)	(1.25)	(1.16)	(1.23)	(1.10)	(1.09)
Constant	-0.492	-0.202	-0.525	-0.244	-0.587	-0.253
	(-0.18)	(-0.07)	(-0.19)	(-0.09)	(-0.21)	(-0.09)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.081	0.075	0.081	0.075	0.084	0.095
Observations	6205	6205	6205	6205	6205	6205

Panel B: Relation between lead SEO underwriter centrality and analyst forecast dispersion at the end of the first post-SEO fiscal quarter

Dependent Variable	Dispersion					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	-0.513***					
	(-5.26)					
Indegree		0.874				
		(1.63)				
Outdegree			-0.477***			
			(-5.51)			
Betweenness				-0.759***		
				(-4.23)		
Eigenvector					-0.840***	
					(-4.73)	
2-StepReach						-0.588***
						(-5.79)
xMktShare	0.163	-0.651***	0.276	-0.254	-0.100	-0.317
	(0.57)	(-3.51)	(0.93)	(-0.99)	(-0.40)	(-1.61)
OfferSize	-0.081	-0.087*	-0.075	-0.087*	-0.087*	-0.089*
	(-1.61)	(-1.73)	(-1.50)	(-1.74)	(-1.73)	(-1.79)
PriorMktRetIss	-0.294	-0.296	-0.288	-0.295	-0.292	-0.299
	(-1.16)	(-1.16)	(-1.13)	(-1.16)	(-1.15)	(-1.18)
FirmSize	1.632	1.788	1.635	1.703	1.608	1.681
	(0.76)	(0.83)	(0.76)	(0.79)	(0.75)	(0.79)
Constant	0.241	0.263	0.223	0.281	0.243	0.288
	(0.28)	(0.31)	(0.26)	(0.33)	(0.28)	(0.34)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.142	0.140	0.142	0.140	0.141	0.145
Observations	5669	5669	5669	5669	5669	5669

Panel C: Relation between lead SEO underwriter centrality and average post-SEO one-month daily bid-ask spread

Dependent Variable	B-S Spread					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	-0.018***					
	(-16.90)					
Indegree		-0.032***				
		(-5.71)				
Outdegree			-0.017***			
			(-17.62)			
Betweenness				-0.026***		
				(-12.31)		
Eigenvector					-0.035***	
					(-20.39)	
2-StepReach						-0.015***
						(-19.90)
xMktShare	-0.004	-0.030***	-0.000	-0.021***	-0.002	-0.017***
	(-1.10)	(-13.98)	(-0.06)	(-7.03)	(-0.60)	(-7.58)
OfferSize	-0.004***	-0.005***	-0.004***	-0.005***	-0.004***	-0.004***
	(-7.36)	(-8.49)	(-7.09)	(-7.99)	(-7.39)	(-7.54)
PriorMktRetIss	-0.010***	-0.011***	-0.010***	-0.011***	-0.010***	-0.010***
	(-3.62)	(-3.66)	(-3.61)	(-3.68)	(-3.40)	(-3.47)
FirmSize	0.035	0.038	0.035	0.039	0.029	0.030
	(1.34)	(1.47)	(1.37)	(1.50)	(1.14)	(1.17)
Constant	0.010	0.012	0.010	0.010	0.010	0.011
	(0.84)	(0.98)	(0.82)	(0.86)	(0.90)	(0.99)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.653	0.648	0.654	0.647	0.662	0.666
Observations	6308	6308	6308	6308	6308	6308

Table 3: **Relation between lead SEO underwriter centrality and SEO announcement effect**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *CAR [-1: 1]* and *CAR [-2: 2]* are the cumulative abnormal returns on SEO firms' equity cumulated over 3 days (from day -1 to day +1) and 5 days (from day -2 to day +2) around SEO announcement dates, respectively. The abnormal return is estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO announcement date. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years prior to the SEO filing year as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetFile* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO filing date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO filing year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Filing year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relation between lead SEO underwriter centrality and 3-day cumulative abnormal returns around the SEO announcement date

Dependent Variable	CAR [-1:1]					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	0.039***					
	(3.68)					
Indegree		0.038				
		(0.73)				
Outdegree			0.034***			
			(3.61)			
Betweenness				0.059***		
				(2.92)		
Eigenvector					0.046***	
					(2.79)	
2-StepReach						0.018**
						(2.14)
xMktShare	-0.029	0.044**	-0.032	0.006	0.009	0.030
	(-0.90)	(2.12)	(-0.98)	(0.23)	(0.34)	(1.34)
OfferSize	0.009	0.011*	0.009	0.010*	0.010*	0.010*
	(1.53)	(1.84)	(1.47)	(1.68)	(1.69)	(1.70)
PriorMktRetFile	0.088***	0.090***	0.088***	0.089***	0.088***	0.089***
	(3.38)	(3.43)	(3.37)	(3.39)	(3.39)	(3.40)
FirmSize	0.499**	0.482*	0.496*	0.496*	0.491*	0.489*
	(1.97)	(1.91)	(1.96)	(1.96)	(1.94)	(1.93)
Constant	0.324***	0.318***	0.326***	0.324***	0.314***	0.310***
	(2.98)	(2.92)	(3.00)	(2.98)	(2.89)	(2.84)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.207	0.206	0.207	0.206	0.206	0.206
Observations	6337	6337	6337	6337	6337	6337

Panel B: Relation between lead SEO underwriter centrality and 5-day cumulative abnormal returns around the SEO announcement date						
Dependent Variable	CAR [-2:2]					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	0.039***					
	(2.93)					
Indegree		-0.020				
		(-0.31)				
Outdegree			0.037***			
			(3.07)			
Betweenness				0.061**		
				(2.41)		
Eigenvector					0.050**	
					(2.41)	
2-StepReach						0.011
						(1.00)
xMktShare	0.005	0.061**	-0.004	0.035	0.031	0.055**
	(0.13)	(2.35)	(-0.09)	(0.99)	(0.89)	(1.98)
OfferSize	0.009	0.010	0.009	0.010	0.010	0.010
	(1.24)	(1.39)	(1.18)	(1.34)	(1.33)	(1.37)
PriorMktRetFile	0.138***	0.139***	0.138***	0.139***	0.138***	0.139***
	(4.20)	(4.24)	(4.20)	(4.22)	(4.21)	(4.23)
FirmSize	0.497	0.482	0.496	0.494	0.492	0.486
	(1.56)	(1.51)	(1.56)	(1.55)	(1.54)	(1.53)
Constant	0.404***	0.407***	0.407***	0.405***	0.396***	0.397***
	(2.95)	(2.97)	(2.97)	(2.96)	(2.89)	(2.90)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.194	0.194	0.194	0.194	0.194	0.194
Observations	6333	6333	6333	6333	6333	6333

Table 4: **Relation between lead SEO underwriter centrality and the absolute value of SEO offer price revision**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *AbsPriceRev* is the absolute value of the percentage difference between the SEO offer price and the midpoint of initial filing range. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	AbsPriceRev					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	-12.043***					
	(-7.30)					
Indegree		-2.858				
		(-0.36)				
Outdegree			-11.456***			
			(-7.73)			
Betweenness				-17.049***		
				(-5.36)		
Eigenvector					-20.300***	
					(-7.94)	
2-StepReach						-9.423***
						(-7.81)
xMktShare	-2.105	-19.118***	1.081	-13.737***	-3.579	-11.082***
	(-0.43)	(-5.90)	(0.21)	(-3.13)	(-0.84)	(-3.22)
OfferSize	-4.391***	-4.792***	-4.263***	-4.676***	-4.489***	-4.484***
	(-4.74)	(-5.18)	(-4.60)	(-5.05)	(-4.87)	(-4.87)
PriorMktRetIss	-14.282***	-14.650***	-14.231***	-14.579***	-13.812***	-13.804***
	(-3.30)	(-3.38)	(-3.29)	(-3.36)	(-3.19)	(-3.19)
FirmSize	-20.753	-15.411	-20.528	-17.002	-20.951	-21.523
	(-0.53)	(-0.39)	(-0.52)	(-0.43)	(-0.54)	(-0.55)
Constant	27.921	29.055*	27.656	29.059*	27.674	29.095*
	(1.62)	(1.68)	(1.60)	(1.68)	(1.60)	(1.69)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.280	0.277	0.280	0.277	0.281	0.283
Observations	6892	6892	6892	6892	6892	6892

Table 5: **Relation between lead SEO underwriter centrality and SEO discount**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *Discount* is the percentage difference between SEO offer price and the closing price on the day prior to the SEO issue day. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	Discount					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	-10.008***					
	(-10.47)					
Indegree		-3.042				
		(-0.64)				
Outdegree			-9.372***			
			(-10.95)			
Betweenness				-14.127***		
				(-7.69)		
Eigenvector					-16.005***	
					(-10.39)	
2-StepReach						-7.935***
						(-10.73)
xMktShare	6.498**	-11.549***	9.526***	-3.435	1.915	-4.918**
	(2.30)	(-6.16)	(3.24)	(-1.36)	(0.78)	(-2.48)
OfferSize	-4.147***	-4.606***	-4.015***	-4.419***	-4.318***	-4.347***
	(-7.83)	(-8.68)	(-7.57)	(-8.33)	(-8.17)	(-8.25)
PriorMktRetIss	-1.488	-1.608	-1.440	-1.559	-1.258	-1.315
	(-0.60)	(-0.64)	(-0.58)	(-0.62)	(-0.51)	(-0.53)
FirmSize	41.747*	48.512**	41.930*	45.644**	42.809**	43.966**
	(1.93)	(2.23)	(1.94)	(2.10)	(1.98)	(2.04)
Constant	6.769	7.737	6.849	7.013	6.817	9.294
	(0.73)	(0.83)	(0.74)	(0.75)	(0.73)	(1.00)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.320	0.311	0.322	0.314	0.320	0.323
Observations	7011	7011	7011	7011	7011	7011

Table 6: **Relation between lead SEO underwriter centrality and SEO underpricing**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	SEO Underpricing					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	-5.181***					
	(-6.94)					
Indegree		-15.845***				
		(-4.30)				
Outdegree			-4.491***			
			(-6.71)			
Betweenness				-7.770***		
				(-5.43)		
Eigenvector					-9.568***	
					(-7.99)	
2-StepReach						-4.330***
						(-7.50)
xMktShare	1.557	-7.028***	1.590	-2.910	1.193	-3.304**
	(0.71)	(-4.83)	(0.69)	(-1.48)	(0.62)	(-2.13)
OfferSize	-1.934***	-2.216***	-1.910***	-2.057***	-1.979***	-2.010***
	(-4.67)	(-5.37)	(-4.60)	(-4.97)	(-4.80)	(-4.88)
PriorMktRetIss	3.712*	3.679*	3.721*	3.669*	3.890**	3.829**
	(1.91)	(1.90)	(1.92)	(1.89)	(2.01)	(1.98)
FirmSize	9.943	11.346	10.482	11.687	9.706	10.695
	(0.59)	(0.67)	(0.62)	(0.69)	(0.57)	(0.63)
Constant	22.473***	23.158***	22.589***	22.527***	22.297***	23.720***
	(3.09)	(3.19)	(3.11)	(3.10)	(3.07)	(3.27)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.201	0.200	0.201	0.199	0.204	0.204
Observations	7036	7036	7036	7036	7036	7036

Table 7: **Relation between lead SEO underwriter centrality and post-SEO market valuation of issuing firms**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *QAdj* is the industry-adjusted Q ratio of SEO firms. Q ratio is defined as the market value of assets over the book value of assets, where the market value of assets is equal to the book value of assets minus the book value of equity plus the product of the number of shares outstanding and SEO issue day closing price. *QAdj* is constructed by subtracting contemporaneous 2-digit SIC code industry median Q ratio from the SEO firm Q ratio. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	QAdj					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	0.593**					
	(2.07)					
Indegree		-0.909				
		(-0.65)				
Outdegree			0.586**			
			(2.28)			
Betweenness				1.205**		
				(2.19)		
Eigenvector					0.945**	
					(2.08)	
2-StepReach						0.045
						(0.21)
xMktShare	1.798**	1.787***	1.744**	1.788**	1.608**	1.885***
	(2.12)	(3.20)	(1.97)	(2.37)	(2.18)	(3.16)
OfferSize	0.546***	0.541***	0.545***	0.546***	0.543***	0.549***
	(3.42)	(3.40)	(3.40)	(3.42)	(3.41)	(3.45)
PriorMktRetIss	-1.095	-1.090	-1.095	-1.095	-1.102	-1.088
	(-1.47)	(-1.46)	(-1.47)	(-1.47)	(-1.48)	(-1.46)
FirmSize	-45.305***	-45.417***	-45.288***	-45.302***	-45.232***	-45.364***
	(-6.94)	(-6.96)	(-6.94)	(-6.94)	(-6.93)	(-6.96)
Constant	0.142	0.262	0.160	0.188	0.138	0.167
	(0.05)	(0.09)	(0.06)	(0.07)	(0.05)	(0.06)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.051	0.052	0.051	0.051	0.051	0.051
Observations	7084	7084	7084	7084	7084	7084

Table 8: **Relation between lead SEO underwriter centrality and institutional equity ownership**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *Ln(InstNum)* is the natural logarithm of the number of institutional investors holding SEO firm shares at the end of the first fiscal quarter after the SEO. *InstProp* is the proportion of SEO firm shares held by institutional investors at the end of the first fiscal quarter after the SEO. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relation between lead SEO underwriter centrality and the number of institutional investors holding SEO firm shares at the end of the first post-SEO fiscal quarter

Dependent Variable	Ln(InstNum)					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	3.328***					
	(27.01)					
Indegree		3.284***				
		(5.62)				
Outdegree			3.169***			
			(28.75)			
Betweenness				4.585***		
				(18.99)		
Eigenvector					5.663***	
					(30.23)	
2-StepReach						2.650***
						(28.51)
xMktShare	1.192***	5.451***	0.358	4.357***	1.306***	3.263***
	(3.34)	(22.51)	(0.96)	(13.56)	(4.24)	(13.14)
OfferSize	1.133***	1.254***	1.099***	1.213***	1.166***	1.174***
	(16.51)	(17.99)	(16.06)	(17.38)	(17.28)	(17.63)
PriorMktRetIss	0.181	0.215	0.165	0.219	0.075	0.119
	(0.58)	(0.68)	(0.53)	(0.69)	(0.24)	(0.39)
FirmSize	33.825***	32.589***	33.832***	32.709***	33.693***	33.488***
	(12.32)	(11.65)	(12.38)	(11.70)	(12.45)	(12.53)
Constant	-0.906	-0.674	-0.828	-0.839	-0.428	-1.024
	(-0.73)	(-0.54)	(-0.67)	(-0.67)	(-0.35)	(-0.85)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.550	0.532	0.554	0.532	0.562	0.572
Observations	7426	7426	7426	7426	7426	7426

Panel B: Relation between lead SEO underwriter centrality and the proportion of SEO firm shares held by institutional investors at the end of the first post-SEO fiscal quarter

Dependent Variable	InstProp					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	0.904***					
	(26.65)					
Indegree		0.355**				
		(2.18)				
Outdegree			0.858***			
			(28.32)			
Betweenness				1.267***		
				(18.97)		
Eigenvector					1.324***	
					(25.19)	
2-StepReach						0.598***
						(22.72)
xMktShare	-0.250**	1.186***	-0.526***	0.628***	0.166*	0.690***
	(-2.54)	(17.56)	(-5.15)	(7.06)	(1.92)	(9.82)
OfferSize	0.204***	0.241***	0.193***	0.226***	0.222***	0.225***
	(10.80)	(12.43)	(10.25)	(11.73)	(11.73)	(11.96)
PriorMktRetIss	-0.073	-0.057	-0.078	-0.063	-0.092	-0.079
	(-0.85)	(-0.64)	(-0.92)	(-0.72)	(-1.07)	(-0.92)
FirmSize	-1.997***	-2.493***	-1.997***	-2.297***	-2.170***	-2.243***
	(-2.64)	(-3.20)	(-2.66)	(-2.97)	(-2.86)	(-2.97)
Constant	-0.062	-0.006	-0.041	-0.067	0.057	-0.090
	(-0.18)	(-0.02)	(-0.12)	(-0.19)	(0.17)	(-0.27)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.494	0.462	0.501	0.470	0.490	0.494
Observations	7412	7412	7412	7412	7412	7412

Table 9: **Relation between lead SEO underwriter centrality and post-SEO long-run stock return of issuing firms**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *BHAR* is the buy-and-hold abnormal return on SEO firms' equity over 252 trading days post-SEO (starting from the first day after the SEO issue day). The abnormal return is estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO issue date. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. *t*-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	BHAR					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	1.309**					
	(2.35)					
Indegree		1.570				
		(0.61)				
Outdegree			1.407***			
			(2.80)			
Betweenness				2.092*		
				(1.94)		
Eigenvector					2.400***	
					(2.79)	
2-StepReach						0.208
						(0.50)
xMktShare	2.590	3.249***	1.901	3.355**	2.012	3.298***
	(1.58)	(2.99)	(1.11)	(2.32)	(1.41)	(2.84)
OfferSize	0.625**	0.649**	0.604*	0.645**	0.619**	0.644**
	(2.00)	(2.08)	(1.92)	(2.06)	(1.98)	(2.06)
PriorMktRetIss	1.334	1.336	1.320	1.348	1.284	1.351
	(0.94)	(0.94)	(0.93)	(0.95)	(0.90)	(0.95)
FirmSize	27.070**	26.953**	27.256**	26.781**	27.265**	26.791**
	(2.18)	(2.17)	(2.20)	(2.16)	(2.20)	(2.16)
Constant	-0.126	0.023	-0.091	-0.040	-0.009	-0.106
	(-0.02)	(0.00)	(-0.02)	(-0.01)	(-0.00)	(-0.02)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.197	0.197	0.197	0.197	0.197	0.197
Observations	7320	7320	7320	7320	7320	7320

Table 10: **Instrumental variable analysis of the relation between lead SEO underwriter centrality and SEO characteristics**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *ForError* is the absolute difference between the mean EPS forecasted by the financial analysts following the SEO firm and the actual EPS divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *Dispersion* is the standard deviation in EPS estimates forecasted by the financial analysts following the SEO firm divided by the SEO firm stock price at the end of the fiscal quarter of the SEO. *B-S Spread* is the average daily bid-ask spread over one-month (21-trading-day) period after the SEO issue date, where the daily bid-ask spread is equal to the difference between ask price and bid price divided by the mean of ask price and bid price. *CAR [-1: 1]* is the cumulative abnormal return on SEO firms' equity cumulated over 3 days (from day -1 to day +1) around SEO announcement dates. The abnormal return is estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO announcement date. *AbsPriceRev* is the absolute value of the percentage difference between the SEO offer price and the midpoint of initial filing range. *Discount* is the percentage difference between SEO offer price and the closing price on the day prior to the SEO issue day. *Underpricing* is the percentage difference between the issue day closing price and the SEO offer price. *QAdj* is the industry-adjusted Q ratio of SEO firms. *Ln(InstNum)* is the natural logarithm of the number of institutional investors holding SEO firm shares at the end of the first fiscal quarter after the SEO. *BHAR* is the buy-and-hold abnormal return on SEO firms' equity over 252 trading days post-SEO (starting from the first day after the SEO issue day). The abnormal return is estimated using the market model with CRSP value-weighted index return as the market return; market model variables (alphas and betas) are estimated over a 150-day period ending 50 days prior to the SEO issue date. *DegreeHat* is the predicted value of *Degree* from first-stage regressions. *Degree* is one of the measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years prior to the SEO issue year as described in Section 6. *IRBP* is the instrumental variable defined as the average number of book runners for IPOs and SEOs in a given industry over three-year period prior to the SEO issue year. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetFile* and *PriorMktRetIss* are the returns on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO filing date and prior to the SEO issue date, respectively. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Year fixed effects are included in all regressions. All variables in Columns (7) and (8) of Panel A are measured prior to the SEO filing year as described in Table 3. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relation between lead SEO underwriter centrality and SEO characteristics (1)

Dependent Variable	Degree	ForError	Degree	Dispersion	Degree	B-S Spread	Degree	CAR [-1:1]	Degree	Abs-PriceRev
	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IRBP	0.004***		0.003***		0.003***		0.002***		0.003***	
	(5.77)		(4.78)		(5.06)		(2.59)		(5.18)	
DegreeHat		-7.743**		-4.602***		-0.041**		0.924**		-116.338***
		(-2.19)		(-2.91)		(-2.56)		(2.07)		(-3.75)
xMktShare	0.585***	4.425**	0.624***	2.608**	0.558***	0.008	0.481***	-0.421*	0.503***	46.438***
	(15.94)	(2.02)	(16.74)	(2.55)	(14.70)	(0.84)	(13.05)	(-1.93)	(14.30)	(2.81)
OfferSize	0.099***	0.316	0.087***	0.283*	0.117***	-0.002	0.112***	-0.095*	0.115***	7.806**
	(15.43)	(0.82)	(13.55)	(1.88)	(17.63)	(-1.02)	(17.10)	(-1.85)	(18.02)	(2.06)
PriorMktRetIss(File)	0.033	-0.092	0.029	-0.079	0.060*	-0.009***	0.085***	0.025	0.065**	-6.059
	(0.99)	(-0.13)	(0.82)	(-0.29)	(1.72)	(-3.02)	(2.77)	(0.47)	(2.03)	(-1.09)

FirmSize	-0.501*	-1.806	-0.455*	-1.519	-0.539*	0.041*	-0.596**	1.044***	-0.580**	-131.916***
	(-1.87)	(-0.33)	(-1.71)	(-0.75)	(-1.89)	(1.77)	(-2.14)	(2.74)	(-2.18)	(-3.01)
Constant	0.147***	1.228*	0.145***	0.708**	0.153*	0.013	0.149***	-0.165*	0.105	19.332
	(6.52)	(1.71)	(5.88)	(2.32)	(1.65)	(1.63)	(2.88)	(-1.85)	(1.17)	(1.30)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		-0.070		-0.414		0.507		-1.331		-0.511
Observations	6134	6134	5608	5608	6236	6236	6256	6256	6799	6799
F Statistics	33.256		22.808		25.571		6.706		26.836	

Panel B: Relation between lead SEO underwriter centrality and SEO characteristics (2)

Dependent Variable	Degree	Discount	Degree	Under-pricing	Degree	QAdj	Degree	Ln(Inst-Num)	Degree	BHAR
	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IRBP	0.003***		0.003***		0.003***		0.003***		0.003***	
	(4.81)		(4.83)		(4.81)		(4.76)		(4.95)	
DegreeHat		-96.113***		-22.675*		2.938		16.553***		10.353
		(-4.16)		(-1.91)		(0.70)		(4.81)		(1.19)
xMktShare	0.495***	44.902***	0.496***	7.708	0.500***	0.256	0.490***	-4.892***	0.477***	-0.498
	(14.41)	(3.75)	(14.44)	(1.25)	(14.56)	(0.12)	(14.92)	(-2.77)	(14.24)	(-0.11)
OfferSize	0.114***	5.722**	0.115***	0.457	0.114***	0.124	0.109***	-0.360	0.114***	-0.579
	(18.33)	(2.05)	(18.36)	(0.32)	(18.31)	(0.24)	(17.96)	(-0.90)	(18.52)	(-0.55)
PriorMktRetIss	0.045	1.843	0.049	4.950***	0.053*	-1.057	0.054*	-0.578	0.065**	0.751
	(1.44)	(0.50)	(1.57)	(2.60)	(1.70)	(-1.57)	(1.80)	(-1.10)	(2.14)	(0.53)
FirmSize	-0.555**	-33.285	-0.545**	-31.249**	-0.505**	-28.844***	-0.353	41.683***	-0.322	45.803***
	(-2.23)	(-1.15)	(-2.19)	(-2.10)	(-2.03)	(-5.57)	(-1.49)	(10.61)	(-1.35)	(4.45)
Constant	0.177***	18.291***	0.177***	3.937*	0.172***	-0.466	0.179***	0.183	0.176***	-2.457
	(10.58)	(3.94)	(10.58)	(1.65)	(10.47)	(-0.57)	(17.15)	(0.28)	(16.84)	(-1.49)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2		-1.000		-0.027		0.002		-0.647		0.015
Observations	6897	6897	6921	6921	6971	6971	7295	7295	7188	7188
F Statistics	23.151		23.287		23.105		22.624		24.455	

Table 11: **Relation between lead underwriter centrality and underwriter compensation**

The sample consists of seasoned equity offerings (SEO) conducted in 1980 - 2017. *GrossSpread* is the total compensation for the underwriting syndicate and is measured in millions of US dollars. *MgmtFee* is the management fee paid to the lead managers for their managing service and is measured in millions of US dollars. *UndwrtFee* is the underwriting fee paid to the lead and co-managers for their underwriting service and is measured in millions of US dollars. *Degree*, *Indegree*, *Outdegree*, *Betweenness*, *Eigenvector*, and *2-StepReach* are measures of lead SEO underwriter centrality using both IPO and SEO participation in the past five years as described in Section 6. *MktShare* is the lead underwriter's share of total proceeds raised in both IPO and SEO markets in the previous five years. *xMktShare* is the residuals from regressing *MktShare* on six lead SEO underwriter centrality measures. *OfferSize* is the SEO offer size and is measured in billions of US dollars. *PriorMktRetIss* is the return on the CRSP value-weighted index over one-month (21-trading-day) period prior to the SEO issue date. *FirmSize* is the book value of total assets at the end of the fiscal year prior to the SEO issue year and is measured in billions of US dollars. All variables except for our six measures of underwriter centrality are winsorized at the 0.5% and 99.5% levels. Issue year × industry (two-digit SIC code) fixed effects are included in all regressions. t-statistics are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Relation between lead underwriter centrality and total compensation for the underwriting syndicate

Dependent Variable	GrossSpread					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	5.146***					
	(10.88)					
Indegree		-3.559				
		(-1.64)				
Outdegree			4.988***			
			(11.75)			
Betweenness				8.962***		
				(9.78)		
Eigenvector					6.753***	
					(9.38)	
2-StepReach						2.300***
						(6.56)
xMktShare	2.519*	8.871***	0.915	5.258***	5.130***	7.203***
	(1.81)	(9.55)	(0.63)	(4.28)	(4.22)	(7.25)
OfferSize	25.105***	25.246***	25.040***	25.186***	25.198***	25.212***
	(92.91)	(93.54)	(92.56)	(93.31)	(93.46)	(93.55)
PriorMktRetIss	2.077*	2.251*	2.051*	2.149*	1.999*	2.083*
	(1.72)	(1.86)	(1.70)	(1.78)	(1.65)	(1.72)
FirmSize	-123.997***	-127.212***	-123.674***	-125.275***	-125.162***	-125.481***
	(-11.04)	(-11.30)	(-11.02)	(-11.14)	(-11.13)	(-11.16)
Constant	-1.703	-1.126	-1.579	-1.716	-1.323	-1.722
	(-0.36)	(-0.24)	(-0.34)	(-0.36)	(-0.28)	(-0.36)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.809	0.808	0.809	0.808	0.808	0.808
Observations	7166	7166	7166	7166	7166	7166

Panel B: Relation between lead underwriter centrality and management fees for the managing service of lead underwriter

Dependent Variable	MgmtFee					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	0.414***					
	(4.56)					
Indegree		-0.195				
		(-0.62)				
Outdegree			0.441***			
			(5.24)			
Betweenness				0.679***		
				(4.40)		
Eigenvector					0.467***	
					(4.24)	
2-StepReach						0.104
						(1.47)
xMktShare	0.612***	0.852***	0.473**	0.764***	0.736***	0.809***
	(3.15)	(6.13)	(2.27)	(4.66)	(4.16)	(5.26)
OfferSize	6.057***	6.059***	6.054***	6.058***	6.058***	6.059***
	(138.69)	(138.78)	(138.59)	(138.72)	(138.71)	(138.70)
PriorMktRetIss	0.533***	0.542***	0.527***	0.538***	0.534***	0.538***
	(3.16)	(3.22)	(3.13)	(3.20)	(3.17)	(3.19)
FirmSize	-20.979***	-21.170***	-20.955***	-21.063***	-21.031***	-21.073***
	(-11.14)	(-11.23)	(-11.14)	(-11.19)	(-11.17)	(-11.19)
Constant	-0.395	-0.329	-0.382	-0.370	-0.400	-0.402
	(-0.77)	(-0.65)	(-0.75)	(-0.72)	(-0.78)	(-0.78)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.951	0.951	0.951	0.951	0.951	0.951
Observations	4386	4386	4386	4386	4386	4386

Panel C: Relation between lead underwriter centrality and underwriting fees for the underwriting service of lead underwriter

Dependent Variable	UndwrtFee					
	(1)	(2)	(3)	(4)	(5)	(6)
Degree	0.626***					
	(6.59)					
Indegree		0.276				
		(0.85)				
Outdegree			0.614***			
			(6.98)			
Betweenness				1.094***		
				(6.77)		
Eigenvector					0.674***	
					(5.85)	
2-StepReach						0.173**
						(2.35)
xMktShare	0.801***	1.193***	0.694***	0.991***	1.030***	1.122***
	(3.94)	(8.19)	(3.18)	(5.77)	(5.56)	(6.97)
OfferSize	5.585***	5.590***	5.582***	5.587***	5.588***	5.588***
	(122.20)	(122.24)	(122.09)	(122.22)	(122.18)	(122.17)
PriorMktRetIss	0.510***	0.522***	0.505***	0.518***	0.514***	0.518***
	(2.89)	(2.96)	(2.87)	(2.94)	(2.91)	(2.94)
FirmSize	-21.486***	-21.633***	-21.495***	-21.600***	-21.585***	-21.636***
	(-10.90)	(-10.96)	(-10.91)	(-10.96)	(-10.95)	(-10.97)
Constant	-0.501	-0.427	-0.480	-0.469	-0.505	-0.522
	(-0.94)	(-0.80)	(-0.90)	(-0.88)	(-0.94)	(-0.97)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.939	0.939	0.939	0.939	0.939	0.939
Observations	4386	4386	4386	4386	4386	4386