The Value of Financial Statements in Predicting the Innovation Potential of High-Tech Startups

Bin Ke[1], Xinjie Ma[2], and Yanbo Wang[3]

June 13, 2025

ABSTRACT

Using a proprietary dataset of grant applications submitted to China's Innovation Fund for Small Technology-based Firms (Innofund), this study evaluates the usefulness of financial statements in predicting the innovation potential of high-tech startup firms. We find that an XGBoost model trained solely on raw accounting statement data outperforms a counterpart model based on the evaluation scores of human experts using the full grant applications. Furthermore, combining accounting data with expert scores yields only modest improvement in predictive accuracy, suggesting limited incremental value from expert judgment. Shapley value analysis reveals significant differences in the evaluation criteria employed by human experts and the XGBoost model and helps identify two categories of theoretical constructs—related to financial constraints and innovation investment—that can closely mimic the predictive performance of the accounting data-based XGBoost model. Overall, our findings underscore the value of financial statement data in identifying high-tech startups with strong innovation prospects.

Key words: machine learning; high-tech startup firms; predictive value of financial statements; innovation potential

We thank Dmitri Byazlov, Daniel Rabetti, Guoman She, Detian Yang (discussant), and workshop participants at Central University of Finance and Economics, Shenzhen University, Sun Yat-sen University (Shenzhen), University of Hong Kong, University of Macau, EAA Annual Congress, Jiangxi University of Finance and Economics, AAA annual meeting, INFORMS annual meeting, SMU conference on INNOVATION IN THE AI ERA, and National University of Singapore for their helpful comments. Ma acknowledges financial support from National University of Singapore's Start-Up Grant (R-521-000-051-133) and Wang acknowledges financial support from Hong Kong University Grant Committee (GRF No. 17504422). This article is written with the assistance of ChatGPT 4.0.

[1] Department of Accounting, Business School, National University of Singapore, Mochtar Riady Building, BIZ 1, #07-30, 15 Kent Ridge Drive, Singapore 119245. Tel: +65 6601 3133. Email: bizk@nus.edu.sg.
[2] Department of Accounting, Business School, National University of Singapore, Mochtar Riady Building, BIZ 1, #07-19, 15 Kent Ridge Drive, Singapore 119245. Tel: +65 6601 1476. Email: xinjiema@nus.edu.sg.
[3] Area of Management and Strategy, HKU Business School, University of Hong Kong, HK SAR. Tel: +852 6845 4567. Email: yanbo.wang@hku.hk.

1. Introduction

This study utilizes a proprietary dataset of grant applications from high-tech startups submitted to China's Innovation Fund for Small Technology-based Firms (Innofund) to investigate whether historical financial statements can predict future innovation performance of high-tech startups. We develop a machine learning model trained solely on historical financial statement data and compare its predictive accuracy to a benchmark model based on overall innovation potential evaluation scores assigned by human evaluators who reviewed the full applications. Since our model draws on only a limited subset of the information available to human experts, it is ex ante unclear whether it can match or outperform their assessments.

We focus on historical financial statements because a fundamental strength of financial statements is that they meticulously encapsulate a firm's financing, investing, and operating activities using uniform financial terminology. Furthermore, financial statements stand out as the only structured data of the grant application amenable to quantitative analysis without incurring significant expenses. Hence, our approach, if successful, could offer a cost-effective yet insightful evaluation tool for funding agencies.

Extensive accounting research underscores the utility of publicly listed firms' historical financial statements in various decision-making contexts. Yet, their relevance for privately held firms, especially high-tech startup firms, remains underexplored due to the scarcity of startup firm data in the public domain (Ball 2024). Both academic and practitioner circles tend to view the financial statements of privately held firms with suspicion, a perception exacerbated in markets characterized by weak institutional frameworks (e.g., Stuart and Wang 2016). Previous research on private firms primarily focuses on the role of financial statement information in equity valuation and debt contracting using in-sample analysis (e.g., Beuselinck et al. 2023). No existing study has explored the *predictive* value of historical financial statements in assessing the *innovation potential* of high-tech startups.

Our research question is important because startup firms around the world operating within the realms of science and technology have emerged as pivotal players in driving innovations (Sorensen and

Stuart 2000; Kerr and Nanda 2015; Looze and Goff 2022). Their nimble structures, specialized focus, and keen insights into niche markets often position them at the forefront of technological advancements and scientific breakthroughs. However, as the broader benefits of innovation, often termed as positive externalities, extend far beyond the confines of the originating enterprise, private enterprises exhibit a natural tendency to under-invest in innovation activities (Hall and Lerner 2010).

Recognizing this gap and its implications for national growth and competitiveness, governments around the world have taken the mantle of facilitating and accelerating such endeavors.[1] Every year, billions are earmarked and funneled through specialized agencies with a singular aim: to spur research and innovation in science and technology sectors dominated by early-stage entrepreneurial companies (David, Hall, and Toole 2000; Lerner 2009; Bloom, Reenen, and Williams 2019). These allocations serve as both an acknowledgment of the critical role early-stage high-tech ventures play and a tool to level the playing field, enabling these entities to partake in high-risk, high-reward innovation ventures. China's Innofund is one such example whose primary mission is to catalyze the fast growth of strong early-stage high-tech ventures so that they can attract more private capital from venture capital and private equity firms.

Yet, one of the most contentious aspects of such state-driven initiatives is the mechanism by which these funds are awarded (NAS 2024). The traditional modus operandi involves high-tech startups submitting detailed grant proposals, which are then subjected to rigorous scrutiny by a panel of independent human experts (Howell 2017; Bian et al. 2021). Based on a myriad of factors, ranging from the feasibility of the proposed innovation to its potential impact, these proposals are ranked for their innovation potential (see Section 2 for the details). Funding is then allocated to the top-tier projects, though discretionary adjustments by the awarding agency are not uncommon in practice. However, this process, as rigorous as it may seem, is fraught with challenges. A salient issue is the intrinsic

[1] Prominent examples include the US Small Business Innovation Research (SBIR) grant program, China's Innovation Fund for Small Technology-based Firms (Innofund), Singapore's Agency for Science, Technology and Research (A*STAR) grant programs, and European Union's Framework Programmes for Research and Technological Development (later renamed to Horizon Europe).

unpredictability of innovation. By its very nature, Research and Development (R&D) is a long-haul endeavor. There's a substantial time lag between initial investments and the eventual manifestation of tangible outcomes. Predicting which projects will bear fruit and which will falter is a Herculean task that even the most adept experts grapple with (Scott et al. 2020).

Furthermore, as with any system that involves human judgment, there's the specter of biases casting a shadow over the process. Prior research shows that even human experts have to face the challenges of limited information processing power and cognitive biases (Wilke and Mata 2012; Bian et al. 2021; Lane et al. 2022). Critics also highlight instances of nepotism or favoritism, where certain entities, based on personal affiliations rather than merit, are given undue preference (e.g., Fang et al. 2023; Wang, Siegel, and Li 2024). Such biases, whether real or perceived, undermine the legitimacy of the award process and raise questions about the effective utilization of taxpayer funds (Lerner 2009; Wang, Furman, and Li 2017; Howell 2024).

Bao et al. (2020) show that raw accounting data outperform financial ratios in detecting accounting fraud. Motivated by their findings, we use all available raw financial statement data items as inputs to our machine learning model. As a benchmark, we employ a prediction model that applies the same machine learning technique but substitutes raw financial data with human evaluators' overall innovation potential evaluation scores, which are based on the full content of each grant application.

High-tech startups' innovation potential is challenging to quantify and only partially visible (David et al. 2000; Lerner 2009). Building on past research, we gauge this potential using two types of proxies. The first type uses each applicant's ex post patent applications, which is one important performance indicator specifically emphasized by the Chinese government (Wang et al. 2017; Ke and Wang 2021). We utilize multiple proxies. Our first measure is a dummy variable, PATENT, which indicates if a firm submits at least one patent application to China's Patent and Trademark Office within five years after its grant application. Over half of the grant applicants in our sample haven't filed any patents within this window, prompting our reliance on a patent dummy variable instead of counting patent applications. As not all patents are created equal, we also develop two alternative dummy

measures of patent-based innovation performance, including whether a firm has any invention patent (INVENTION_PATENT) and whether a firm has any blockbuster patent based on forward patent citations (BLOCKBUSTER_PATENT), over the same specified five-year time horizon.[2]

Patents aren't exhaustive indicators of innovation (Mezzanotti and Simcoe 2023). Some innovations are either non-patentable or intentionally left unpatented due to theft concerns. Therefore, we introduce a second type of innovation performance proxy, VCPED, a dummy variable indicating whether a firm secures new venture capital and/or private equity (VC/PE) funding post their grant proposal, measured as of 2023, assuring sufficient time for innovation investments to bear fruit and attract VC/PE attention. Given that VC/PEs evaluate both patented and unpatented innovations in their investment decisions and that the Innofund regards an applicant's ability to attract subsequent VC/PE funds as a credible signal of the firm's innovative potential, this metric offers a robust, supplementary measure of a firm's innovation performance. This performance measure has also been prevalently used in the state subsidy and the innovation literature (Howell 2017; Wang et al. 2017)

We use XGBoost as our machine learning model due to its proven efficacy in various applications. Unlike human experts, XGBoost effectively captures nonlinearity and interactions among input variables, aiding in understanding intricate relations between inputs and future innovation outputs. We randomly allocate 80% of these applications for training and the remaining 20% for testing.[3] Our innovation outcome proxies have varying degrees of class imbalance. Hence, we follow Bao et al. (2020) by using NDCG@k as our primary performance evaluation metric, even though we obtain similar inferences using the Area Under the Receiver Operating Characteristic curve (AUC-ROC), a common performance evaluation metric used in prior studies.

Our sample consists of 6,648 firm-innovation project applications submitted between 2005 and 2010 across two provinces and one municipality. The Ministry of Science and Technology (MOST)

[2] The Chinese patent system categorizes patents into three types: invention, utility model, and design patents. Among these, only invention patents represent substantive technological advancements and significant commercial potential.
[3] For robustness, we also adopt a temporal approach by using 2005-2008 grant applications as training dataset and 2009-2010 applications as test dataset and obtain similar inferences.

appoints human experts to evaluate the innovation potential of each application. Given the multi-dimensional nature of innovation potential, human evaluators are asked to assign both a technical score—based on disclosed product technologies, product market conditions, and business models—and a financial score—based on the submitted financial statements. The overall evaluation score of an applicant's innovation potential, which serves as a key input for funding decisions, is a composite of the technical and financial scores calculated using a proprietary algorithm developed by MOST.

What are our findings? For both types of innovation performance proxies, an XGBoost model based on expert evaluations predicts the applicants' future innovation performance better than random guessing. More importantly, an XGBoost model using financial statement data alone outperforms the benchmark using expert evaluation scores only, suggesting that the financial statements capture predictive details overlooked by human experts. The performance of our proposed XGBoost model versus the benchmark model is also economically significant. However, we find mixed evidence that combining expert scores with financial statement data offers additional advantages, suggesting that third-party human experts bring little incremental information to the prediction task.

The observed disparities in performance between the XGBoost model and the benchmark model likely stem from the superior data processing capabilities of XGBoost compared to human experts. To test this hypothesis, we replicate the same machine learning models using Logit, which lacks the ability to handle nonlinearities and interactions among predictors. Using human experts' evaluation scores as model input, we obtain similar performance for the Logit model versus the XGBoost model. However, when using the financial statement data items as model input, the performance of the Logit model significantly deteriorates compared to the XGBoost model. It even underperforms compared to the Logit based on human expert scores. These outcomes highlight the advantage of XGBoost in handling complex predictive tasks compared to human experts bound by limited cognitive capabilities.

The ages of Innofund applicants range from 0 (freshly born) to 31 years, with a median of 4 years. Younger firms typically have a limited performance track record, leading to higher levels of uncertainty (Caod, Segarra, and Teruel 2016). To evaluate whether the predictive value of financial statements

applies universally, we divided the firms into young and old categories using the median age as a cutoff. Our results indicate that the XGBoost model based on financial statements consistently outperforms the benchmark model based on expert evaluation scores, with one exception: for the dependent variable BLOCKBUSTER_PATENT, the benchmark model performs better for young applicants. Overall, we conclude that the predictive value of financial statements for innovation potential holds across different firm ages.

To enhance the interpretability of our XGBoost models predicting outcomes such as PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED, we apply Shapley values (Shapley 1953) to identify the contribution of each of the raw financial statement variables. Recognizing that human experts assign a financial score based on these same statements, we also conduct a parallel Shapley value analysis on an XGBoost model that uses an identical set of financial variables to predict the experts' financial scores. This allows us to compare how machine learning and expert evaluations weigh the same financial indicators differently.

Our analysis reveals notable disparities in the evaluative criteria between human experts and our machine learning models. In assessing applicants' financial statements, experts predominantly focus on principal financial metrics such as total revenue, profit, total liabilities, and cash and cash equivalents. In contrast, our XGBoost models prioritize variables associated with cash flows as well as operating and non-operating expenses. In addition, the directions of the influence exerted by the independent variables are often inconsistent across the two models. A comparison of the top ten most influential predictors used by human experts against those utilized by the accounting data-based XGBoost models indicates minimal overlap, underscoring the unique patterns our machine learning models leverage, which human evaluators typically overlook. These insights underscore the advanced pattern-recognition capabilities of our machine learning models, highlighting their potential to complement traditional human assessments.

The Shapley value analysis is not readily interpretable within a standard causal inference framework. To address this limitation, we construct an alternative XGBoost model for each innovation

outcome proxy—referred to as the XGBoost_CAUSAL model—based on seventeen proxies for two theoretical constructs derived from the same set of raw accounting variables. The selection of these variables is guided by the Shapley value analysis that finds that raw accounting variables related to cash flows and operating and non-operating expenses are among the most influential predictors. Specifically, we identify twelve financial ratios related to cash flows, which proxy for the extent of a firm's financial constraints, and five ratios related to operating and non-operating expenses, which serve as proxies for innovation investment. The XGBoost_CAUSAL models approximate the predictive performance of the original XGBoost_ACCOUNTING models, suggesting that the two categories of theoretical construct proxies capture much of the key variation associated with innovation outcomes.

The rest of the paper is organized as follows. The next section discusses the prior literature and our contributions. Section 3 provides details on the review process of China's Innofund grant applications. Section 4 describes the sample selection procedures. Section 5 explains the details of our machine learning modeling. Section 6 discusses the machine learning results and related robustness checks. Section 7 analyzes the model performance by firm age. Section 8 sheds light on how human experts and machine learning differ in processing historical financial statement data. Section 9 concludes.

2. Prior literature and contributions

Our study contributes to multiple strands of literature. First, we extend the growing body of research on the informativeness of private firm financial statements. Prior work documents that private firms typically exhibit lower earnings quality than their public counterparts (Beatty et al. 2002; Ball and Shivakumar 2005; Katz 2009; Hope et al. 2013). Nonetheless, financial statement provision is associated with improved outcomes in equity financing (Baik et al. 2023; Kisseleva et al. 2023), bank credit access (Donelson et al. 2017), and debt contracting (Allee and Yohn 2009; Minnis and Sutherland 2017). Chen et al. (2011) further link reporting quality to investment efficiency. Complementing these findings, value relevance studies show that specific financial statement items explain variation in private firm valuations (Hand 2005; Armstrong et al. 2006; Kisseleva et al. 2023). Most of this literature adopts

an outsider perspective, focusing on how accounting information correlates with firm-level outcomes. However, less is known about the mental models that insiders—such as managers and analysts—rely on to extract actionable insights from raw accounting data. Moreover, the quality of these mental models remains largely unexamined, and it is unclear whether they can be systematically improved using machine learning.

Our study distinguishes itself from the previous research through three key deviations. First, we directly show the *predictive* value of private firms' financial statements using machine learning (i.e., out-of-sample analysis), setting our work apart from the majority, which has largely concentrated on their *explanatory* value of private firms' financial statements across various scenarios (i.e., in-sample analysis). Moreover, we demonstrate that our machine learning approach solely based on accounting data can beat a model based on the evaluation scores of human experts who have access to the full grant application. Apart from Fuertes-Callén et al. (2022), who demonstrate financial statements' predictive value for bankruptcy, few prior studies have explored the predictive utility of financial statements as highlighted by Shmueli (2010) and Kleinberg et al. (2015). Such predictive analysis is critical to business insiders who aim to improve their decision making quality and efficiency.

Second, our study is the first to examine the predictive value of financial statements for forecasting *innovation outcomes* in high-tech startup firms—a sector critical to national economic growth. We show that raw financial statement data can be used to develop effective machine learning models that predict innovation potential. This approach offers a scalable tool for government agencies and venture capitalists considering data-driven methods in research grant allocation. We also provide a conceptual rationale for why accounting data are informative in this prediction context.

Third, our research is one of the first studies to present concrete evidence on the predictive power of financial statements for high-tech startup firms within China's challenging institutional environment. This focus is particularly noteworthy given the predominance of research on European private firms, attributed to better data accessibility (see Bar-Yosef et al. 2019; Beuselinck et al. 2023 for reviews). Considering the recognized frailties in China's financial reporting standards and investor protection

institutions (Lennox and Wu 2022; Lu et al. 2023), even among publicly listed companies, our study is significant because we shed light on the nuanced and previously underexplored predictive value of financial statements for early-stage high-tech ventures in a large financial market with weak institutions.

Our second contribution relates to the economics and management literature on the efficacy of macro-level R&D subsidy policies. Most studies in this literature assess the direct effects of these policies on innovation (David et al. 2000; Howell 2017; Klette et al. 2000; Myers and Lanahan 2022; Wang et al. 2017) or, to a lesser extent, describe the policy selection process (Antonelli and Crespi 2013; Bian et al. 2021; Thomas Falk and Svensson 2020; Fang et al. 2023; Wang et al. 2024). However, a fundamental question remains unaddressed: how should the allocation of government subsidies be optimized? We enhance this discourse by introducing a machine learning model that can outperform human experts in predicting the innovation performance of early-stage high-tech ventures.[4] Our prediction model lays the necessary foundation for the subsequent optimization of innovation grant allocation.

Finally, our research complements the management literature on integrating technology with organizational decision-making (Haenlein and Kaplan 2019; Huber 1990). While recent AI advancements have demonstrated accurate predictions in areas such as investment recommendations (Cheng et al. 2020; Coleman et al. 2022), bail decisions (Kleinberg et al. 2017), patent screening (Choudhury et al. 2021; Zheng 2021), and venture capitalist decisions (Lyonnet and Stern 2022), other studies emphasize the symbiotic relationship between human experts and machines in decision-making contexts like financial analysis (Cao et al. 2024) and bank lending (Costello et al. 2020; Liu 2022; Lu and Zhang 2025). Our study adds nuance to this discourse by examining human-machine complementarity in government grant evaluations. Contrary to findings underscoring the discretionary value of human experts in these prior studies, our results indicate that in predicting future innovation

[4] Yoo et al. (2023) employ machine learning to predict the performance of startup firms' R&D projects funded by the South Korean government. However, due to the lack of data on applicants rejected by the government, their sample is limited to those funded by the government, leaving the prediction outcomes of the rejected applications unknown. Additionally, their analysis does not specifically examine the predictive value of financial statements.

outcomes, human experts add much limited predictive power over our machine learning model. A significant differentiator in our study is that our human experts, temporary third-party experts recruited by government agencies, contrast with the permanently employed experts by for-profit organizations in previous research who possess more profound institutional insights and motivation. This disparity suggests a distinct agency problem and information advantage in the government sector, differentiating it from the private sector. More importantly, we show that it is possible to develop data-driven algorithms that can help mitigate such agency problems.

3. Institutional background

Established in 1999 by China's State Council, the primary goal of the Innovation Fund for Small Technology-based Enterprises (Innofund) is to support technological innovation of early-stage high-tech ventures (Wang et al. 2017). The fund is administered by the Ministry of Science and Technology (MOST)'s Innofund Center. Modeled after the U.S. Small Business Innovation Research Grant (SBIR) program, Innofund served as a national-level flagship innovation subsidy program targeting early-stage high-tech ventures. The competition for Innofund grants is fierce due to the non-dilutive nature of the funding. The typical grant size for Innovation Projects ranges from RMB 0.5 million to RMB 1.0 million (approximately $68,334 to $136,668 USD as of October 2023). Successful applicants not only receive financial support from the central government but are also eligible for matching grants from local governments. Furthermore, MOST's approval of a grant confers state endorsements on recipient firms. This recognition facilitates access to additional state-controlled advantages, including rent-free office space, tax incentives, and the prestigious label of a "high-tech enterprise," which offers numerous benefits, both tangible and intangible.

Firms interested in applying for Innofund grants must first secure a recommendation from a local MOST branch or a prestigious science park, as direct submissions to MOST are not permitted. The Innofund has two application tracks: while most firms apply through the "Innovation Track," a small number of firms apply through the "Startup Track." Startup-track firms are usually younger (less than 2 years old at grant application) and lack substantive performance records. In contrast, innovation-track

firms are typically more established and possess performance records such as revenue and profit. Our study focuses on this latter group as they are required to submit detailed financial statements. It is worth noting that the applicants under the "Innovation Track" remain small early-stage startups (see Section 5.2).

MOST employs external third-party experts on a temporary basis to assess application materials. These experts typically possess extensive experience in the technology sector, either in technological development or venture operations. MOST has access to a substantial array of resources, including research grants allocated to universities, state labs, and state-owned enterprises (SOEs). Furthermore, MOST oversees the distribution of "high-tech enterprise" certification, which is accompanied by financial perks such as rent-free office space, tax benefits, and preferred access to bank loans. Given these incentives, controlled by MOST, invitations from MOST to serve as experts are seldom declined, particularly by individuals affiliated with SOEs, state labs, public universities, incubators, accelerators, science parks, or former Innofund recipients. Prior research shows that China's Innofund appears to enlist more private sector experts than the US SBIR program (Bian et al. 2021: 314).[5]

The applications of the Innofund applicants are required to include a comprehensive set of information categorized, including (i) entity background; (ii) details of technological innovation; (iii) competitive positioning of the product; (iv) business model; (v) financial forecasts; and (vi) recent historical financial statements (for innovation-track applicants). Experts are instructed to evaluate the innovation potential of the applications using computers at a MOST-designated facility, without prior knowledge of the firms they will assess. Experts are explicitly instructed not to consider fairness or equity in their evaluation of the assigned applications. Before the evaluation, experts undergo a 20-

[5] Bian et al. (2021) studied a Beijing-based funding program modelled after the Innofund. They find that among the experts recruited by the funding program, about 9.25% of them were professional VC/PE investors. Most others are high-tech entrepreneurs, mid- to high-level executives from large private firms or state owned enterprises, incubator and accelerator managers, and research scientists from universities and state research institutes. The authors also demonstrate that the proportion of VC/PE investors in the Beijing program significally exceeds that of typical U.S. funding programs such as the SBIR program.

minute training session where Innofund officials outline essential disciplinary rules, such as prohibiting contact with applicants, and familiarize them with the evaluation form.

Given the multi-dimensional nature of innovation potential, experts are required to provide their scores on multiple sub-dimensions of a firm's innovation potential, including team capability, technological capability, business model innovation, and financial capability.[6] The first three sub-dimensions are combined to construct a technical score while the financial capability sub-dimension is used to represent an applicant's financial score. The final innovation potential evaluation score is a composite, derived from the technological score and the financial score, calculated through a proprietary algorithm not disclosed to the experts.

For innovation-track applicants, the evaluation process involves four experts independently assigning technical scores to assess the applicant's innovation potential and one expert responsible for assessing the applicant's overall financial health based on submitted financial statements. In contrast, the younger and smaller startup-track applicants are not required to submit historical financial statements. As a result, these enterprises are evaluated solely based on the technical aspects of their products and the innovativeness of their business models, as detailed in the application materials, without a financial health assessment. As we focus on the value of financial statements, we exclude startup-track applicants from our study.

The financial score primarily relies on external specialists' expertise to evaluate applicants' overall financial health. MOST also leverages these experts to identify potential financial irregularities within a firm. If financial discrepancies are suspected, experts may assign a zero, a notably low score, or withhold the financial score entirely. Discussions with anonymous MOST officials indicate that these scores are also used to assess a firm's financial capacity to initiate and sustain innovation activities.

After experts submit their individual evaluation scores, MOST aggregates these into a final composite score for each applicant using its proprietary algorithm. However, certain applicants may be

[6] These four sub-dimensions align with key factors commonly considered by VCs in their investment evaluations.

excluded from receiving a composite score under specific circumstances, including: 1) demonstrable non-compliance with environmental regulations, 2) operation in a sector featured with overcapacity, 3) pending completion of previous grants, 4) non-selection for review due to grant rule violations or application material incompleteness, or 5) guaranteed grants due to corrupt practices.

MOST utilizes the final composite scores to rank applicants, awarding grants to the highest scorers within its budget constraints. Applications can be disqualified by MOST for various reasons, such as detected fraud, environmental violations, a financial score below 60 (out of a scale of 0 to 100), or adverse findings during the public notice period (e.g., unfinished project from previous funding rounds, contract, intellectual property or equity ownership disputes). MOST may fund applicants with lower scores if higher-ranking candidates become ineligible or in situations involving corruption.

4. Sample

Table 1 outlines the sample selection process for the study, beginning with an initial pool of 6,648 Innofund grant applications from the Beijing, Jiangxi, and Jiangsu provinces, spanning the years 2005 to 2010. Beijing serves as the capital city of the country, while Jiangsu, a coastal province, and Jiangxi, a hinterland province are both situated along the Yangze River. The first step in the selection process involves retaining only applications with complete expert evaluation scores, which reduces the sample size to 6,336. Applications from the startup track are explicitly excluded, as these do not require financial data and consequently lack a financial expert score. In addition, we include only applications that received a non-missing financial score, reducing the count to 5,199 observations. The final refinement step requires the inclusion of applications with non-missing financial statements from the previous year, resulting in a final sample of 4,866 grant applications, representing 4,374 unique firms. Some firms submitted multiple applications over the sample period. From this refined sample, 2,416 (49.7%) applications were awarded Innofund grants.

5. Research design

5.1. Proxies for innovation outcomes

The Innofund program is designed to stimulate the development of nascent high-tech industries by supporting the innovative endeavors of startup firms in the science and technology sectors. It serves as a "priming pump," intended to attract additional investments from local governments, businesses, and financial institutions. Given the absence of direct data following each applicant's grant application, we assess post-application innovation performance using established proxies from academic research. These include patenting activities and funding received from venture capital or private equity sources (Howell 2017; Wang et al. 2017).

To evaluate post-application innovation performance, we analyze patent applications filed by firms with China's Patent and Trademark Office within five years following their initial grant applications, following Bernstein (2015). Patent applications are an appropriate measure of a firm's innovation performance in China because the Chinese government has been adopting a variety of proactive policies and incentives to encourage Chinese firms and VC funds to innovate, especially in the form of filing patent applications (Ke and Wang 2021). The Chinese government even made an explicit call for local Chinese firms to apply for two million patents by 2015 (The Economist 2014).

This analysis includes both utility and invention patents. Our findings indicate that over half of the firms in our sample did not apply for any patents during the five-year window in the post-application period. To quantify patenting activity, we use a dummy variable PATENT, which identifies whether a firm has engaged in any patenting activity. Appendix 1 provides all variable definitions used in the study.

Recognizing that PATENT does not assess the quality or impact of a patent, we introduce two additional, more restrictive dummies: INVENTION_PATENT and BLOCKBUSTER_PATENT. INVENTION_PATENT is assigned a value of one for patent applications that fall into the more innovative category of invention patents, and zero otherwise. BLOCKBUSTER_PATENT denotes a high-impact patent, determined by whether the number of forward citations it receives ranks within the

top 10% of all patents in the patent database. We use this second dummy variable to address the possibility that MOST cares about blockbuster innovations only in its grant award decisions.

While patent filings are a common measure of innovation, not all firms utilize patents to protect their innovations, particularly in industries where patents are less applicable (Mezzanotti and Simcoe 2023). This situation may result in measurement errors when using patent-related proxies to assess innovation performance. To address this limitation, we introduce an additional dummy variable, VCPED, which is assigned a value of one if a firm receives new equity investments from venture capital (VC) or private equity (PE) firms following their application to the Innofund program. Since VC and PE firms typically evaluate a firm's innovation potential in their investment decisions—regardless of whether the innovations are patented—VCPED serves as an alternative measure that may capture the innovation activity more accurately across different industries. It is also worth noting that Innofund views an applicant's subsequent VC or PE funding as an indicator of the firm's innovation performance.

We sourced data on Innofund applicants' VC and PE investments from multiple databases as of June 2023, including the website of the Asset Management Association of China (AMAC) and Qichacha.com.[7] These sources provide comprehensive information on equity investments, helping to ensure the reliability of VCPED in representing post-application innovation activity. Given that the last year for our Innofund applications is 2010 and the variable VCPED is defined as of 2023, we should have ample time to capture a firm's innovation outcomes. This timeframe allows for the realization of innovation investments, even if these take many years to yield results.

5.2. Predictors of future innovation performance outcomes

We analyze two types of predictors in our study. First, we use the experts' evaluation scores as the primary input for our machine learning models, designating these as our benchmark models. These experts are tasked with assessing the absolute innovation potential of Innofund applicants, making their

[7] A self-regulatory organization under the supervision of the China Securities Regulatory Commission (CSRC), AMAC plays a pivotal role in overseeing and guiding the asset management industry in China. Qichacha is a widely used Chinese business information platform that provides detailed data on companies registered in mainland China.

scores the most obvious predictor of future innovation performance outcomes. We have access to both the composite score and the individual technical and financial scores for each Innofund applicant.

Our primary analysis utilizes the composite score. However, to provide a more nuanced understanding of the factors influencing innovation performance, we also explore the results derived from the technical and financial component scores. This dual approach allows us to assess the distinct contributions of technical prowess and financial health to the overall innovation potential of the firms.

Second, we use the historical financial statements from Innofund applications to develop a machine learning prediction model for innovation performance. Typically, Innofund applicants submit financial data for the preceding month as well as annual financial reports for the last two fiscal years. In order to simplify our approach, we focus exclusively on the annual data from year t-1 for model training. This ensures a comprehensive snapshot of each firm's financial health while streamlining the dataset. The dataset encompasses all 55 financial statement items from the application, including 27 items from balance sheets, 15 from income statements, and 13 from cash flow statements. This model aims to test the predictive power of detailed financial statement data items against the qualitative assessments made by expert evaluators.[8]

Table 2 details the statistics for the 55 financial statement items utilized in our analysis. It is important to note that not all applicants consistently provided complete data for every item. Instead of imputing missing values or excluding these entries, we opted to retain records with incomplete data, relying on the robustness of our machine learning model to effectively manage and interpret such discrepancies. For firms with complete data sets, the average total assets are RMB 17.42 million, with a median value of RMB 10.28 million. Likewise, the average net income is registered at RMB 1.81 million, with a median value of RMB 0.90 million.[9] These descriptive statistics provide clear evidence that the Innofund applicants are indeed small early-stage firms.

[8] Inferences are qualitatively the same if we use two years of financial statements (untabulated).
[9] The exchange rate between Chinese yuan and the U.S. dollar was around 1:6.60 in December, 2010.

5.3. Method and performance evaluation metrics

We leverage the power of a cutting-edge machine learning method known as Extreme Gradient Boosting (XGBoost) to predict a firm's future innovation performance in the post-application era. XGBoost, which is a decision-tree-based ensemble machine learning algorithm employing a gradient boosting framework, was originally introduced by Chen and Guestrin (2016). Thanks to its exceptional capabilities in addressing regression, classification, and ranking tasks, as well as its remarkable processing speed, XGBoost outperforms other prominent machine learning algorithms like GBRT and LASSO, as demonstrated in recent studies (Li and Zheng, 2023; Tantri, 2021; Zheng, 2021).

Due to the limited time series data available from our data sources, our main analyses adopt a cross-sectional approach in machine learning modeling. Specifically, we randomly allocate 80% of these applications for training and the remaining 20% for testing. Appendix 2 provides a detailed description of our model training procedures. Some firms submitted multiple Innofund applications during our sample period. To avoid potential overstatement of model performance caused by the same firm showing up in both the training and test samples, we perform the random sampling at the unique firm level for all analyses.

The Area Under the Receiver Operating Characteristic curve (AUC-ROC) is a widely utilized performance assessment metric in the realm of machine learning model evaluation. However, its suitability diminishes when dealing with imbalanced distributions of the dependent variable. Bao et al. (2020) introduce the Normalized Discounted Cumulative Gain at k (NDCG@k) as an alternative metric tailored for assessing performance under such circumstances. Given the varying degrees of class imbalance in our dependent variables, we follow Bao et al. (2020) and employ both AUC-ROC and NDCG@k as our chosen performance evaluation metrics. The variable 'k' signifies a predefined number of top results to be considered during the calculation (Bao et al., 2020). This metric takes into careful consideration the positioning of correctly predicted instances within the model-generated ranked list. In simpler terms, it assigns greater significance to accurate predictions situated nearer to the top of

the list, recognizing the common practice where decision-makers predominantly focus on the leading predictions.

NDCG@k holds particular relevance for our scenario, as the Innofund's grant allocation process can be conceptualized as a ranking problem. According to the Innofund guidelines, MOST allocates funds to the highest-ranked applicants first and then goes down the score list until the budgeted capital limit is reached. By incorporating NDCG@k into our evaluation, we can assess the precision and relevance of the model in ranking potential successful innovators, thereby introducing an additional layer of performance assessment. We set the value of 'k' as the realized percentage of the outcome dummy variables in the test sample, which represents the expected outcome under random guessing multiplied by the size of the test sample. The performance evaluation metric NDCG@k aligns with the practical focus on identifying top performers.

5.4. The predictive power of machines versus human experts

Given that our sample consists primarily of high-tech ventures, it raises an empirical question as to whether machine learning models based solely on historical financial statements can outperform benchmark models derived from human expert evaluation scores. Several factors suggest why our proposed machine learning models might not surpass the benchmark models. First, human experts assess a wider array of information included in the applications compared to our machine learning models that use only financial statements. Since a firm's future innovation performance may heavily depend on the technical details presented in the application materials, expert evaluations are likely to gain an advantage by considering these technical details. Nuances such as the feasibility of the technology, market potential, and team competence are likely to be evaluated by human experts but may not be captured by the financial statement data. In addition, experts can bring their valuable external knowledge that is not present in the application material to the evaluation task.

Second, the reliability of financial statements from privately owned Chinese firms is a concern, potentially affecting the predictive accuracy of our models. As Stuart and Wang (2016) note, Innofund applicants often provide overly optimistic self-reported financial data, particularly among firms with

relatively weak innovation capabilities. Even with accurate financial statements, their predictive ability for future innovation outcomes might be constrained due to the small size and relatively unestablished market presence of the firms. As a result, these firms' financial statements may not consistently provide insightful data regarding their future innovative potential.

Third, there exists an inherent mechanical bias favoring the benchmark model based on expert evaluation scores. Our data shows a strong and positive correlation (0.72 in our final sample) between Innofund applicants' expert scores and their subsequent likelihood of receiving an Innofund grant. Since grant recipients generally have access to more resources for innovation, they are more likely to achieve superior innovation outcomes compared to non-recipients in the post-application period.

On the flip side, there are compelling reasons to believe that our proposed machine learning model could outperform the benchmark models based on human expert evaluations. First, the predictive relationship between the 55 financial statement items and future innovation outcomes is likely to be complex and non-linear. Owing to limited information processing power, human experts may face challenges in discerning the intricate interactions among these financial variables, which could be critical for predicting innovation performance. In contrast, machine learning models excel in managing such complexity, identifying patterns and relationships that are not immediately obvious to human reviewers. By effectively capturing and analyzing these non-linear interactions, our machine learning approach might uncover significant predictive insights that may be missed in human evaluations.

Second, human experts may have conflicts of interest or subjective biases that influence their judgment, leading to evaluations that do not accurately reflect the true potential of the grant applications (Bertomeu 2020). In contrast, a machine learning model operates without such biases, provided it is well-calibrated and trained on a robust dataset. It assesses applications based solely on the data provided, potentially leading to more objective and consistent evaluations.

Third, while each human expert is asked to evaluate only a small number of grant applications, our machine learning models can learn from all available application materials in the training sample.

As a result, our machine learning models could be more effective than experts in learning a firm's innovation potential from all application materials.

6. Empirical results

6.1. Descriptive statistics

Table 3 reports the descriptive statistics of the key regression variables using the entire final sample. Out of the 4,866 grant applications analyzed, exactly half were awarded the Innofund grant, as indicated by the mean value of the variable FUNDED. The composite expert evaluation score, denoted as SCORE_COMP, has a mean and standard deviation of 71.62 and 9.70, respectively. Similarly, the mean and standard deviation of SCORE_TECH are 69.31 and 10.44, respectively. In contrast, the mean and standard deviation of SCORE_FIN stand at 70.26 and 10.40, respectively.

The variable PATENT_RAW serves as a continuous version of PATENT, with a mean value of 5.14 patents and a relatively high standard deviation of 16.64. Notably, the mean value of the dummy variable PATENT is 0.42, suggesting that, on average, around 42% of Innofund applicants filed at least one patent application with China's Patent and Trademark Office within the five-year period following their grant application. Shifting our focus to the more innovative invention patents, the mean value of INVENTION_PATENT_RAW is 2.66, accompanied by a standard deviation of 11.04. As expected, the distribution of INVENTION_PATENT displays a greater class imbalance compared to PATENT, with a mean value of 0.36. The mean value of BLOCKBUSTER_PATENT is 0.14, considerably smaller than the mean values of PATENT and INVENTION_PATENT, aligning well with the notion that developing blockbuster patents is challenging.

The distribution of VCPED exhibits the most pronounced class imbalance. The average value of VCPED is only 0.11, implying that in the post-application era, 11% of Innofund applicants received investments from venture capitalists or private equity firms as of June 2023, marking the conclusion of our data collection. This finding underscores the selective nature of VC and PE funding decisions.

6.2. Out-of-sample performance analysis

6.2.1. The benchmark model based on human experts' evaluation scores

Panel A of Table 4 showcases the out-of-sample performance of the benchmark model, which is an XGBoost model relying solely on human experts' composite evaluation scores (SCORE_COMP). We refer to this model as XGBOOST_BENCHMARK. While we report both ROC-AUC and NDCG@k as performance metrics, our discussion will predominantly focus on NDCG@k. This preference is due to the limitations of ROC-AUC in effectively handling imbalanced datasets. We have set the value of k at the unconditional mean of each dependent variable in the test sample, mirroring the expected outcome under random guessing. The unconditional mean values for our dependent variables—PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED— in the test sample are 0.441, 0.365, 0.153, and 0.119, respectively.

As the standalone value of NDCG@k can be difficult to interpret, we complement this metric with PRECISION@k and RECALL@k to provide a more comprehensive understanding of our model's performance. As k is set at the unconditional mean of the dependent variable in the test sample, the values of PRECISION@k and RECALL@k are identical. For the dependent variable PATENT, PRECISION@k is 0.452, implying that 45.2% of the top k ranked firms by our benchmark model are PATENT=1 firms. As this value exceeds the unconditional mean of PATENT (0.441), this comparison suggests that the human experts' composite evaluation scores offer a modest predictive advantage over random guessing, reflecting some ability to discern firms with higher innovation potential.

For the other dependent variables—INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED—the performance metrics, as measured by NDCG@k, show a declining trend, which correlates with increasing class imbalance among these variables. Despite this challenge, the benchmark model's performance still surpasses that of random guessing across these categories. This outcome highlights the model's robustness, particularly in identifying firms with significant innovation outcomes, even amid pronounced data imbalance.

Given that SCORE_COMP is a composite of SCORE_TECH and SCORE_FIN, it is reasonable to consider whether a more effective prediction model could be developed by using these components either collectively or individually. The outcomes of these exploratory benchmark models are detailed in Panels B-D of Table 4. When focusing on NDCG@k as our primary performance metric, we find no evidence that these alternative models surpass the performance of the benchmark model presented in Panel A. Consequently, we will continue to utilize the model from Panel A as our benchmark for subsequent discussions.

6.2.2. XGBoost model based on historical financial statement data

Table 5 displays the out-of-sample predictive performance of our XGBoost model, referred to as XGBoost_ACCOUNTING, which utilizes 55 annual financial statement data items sourced from the Innofund application. Using the NDCG@k metric, we find that XGBoost_ACCOUNTING outperforms XGBoost_BENCHMARK in Table 4 substantially. The improvements in NDCG@k are 25.4% for PATENT, 38.4% for INVENTION_PATENT, 10.5% for BLOCKBUSTER_PATENT, and 58.9% for VCPED.

To shed better light on the economic implications of these results, we examine the changes in PRECISION@k and RECALL@k between XGBoost_ACCOUNTING and XGBoost_BENCHMARK. As these two variables' values are identical, we focus on PRECISION@k. We find that PRECISION@k shows an increase of 20.4% for PATENT, 27.3% for INVENTION_PATENT, 23.7% for BLOCKBUSTER_PATENT, and 15.5% for VCPED. These enhancements are economically significant.

Table 5 provides two key insights. First, it underscores the value of historical financial statements as predictors of high-tech startup firms' future innovative outcomes. This is noteworthy considering the generally perceived lower quality of financial reporting among Chinese private firms. Second, the comparison between Tables 4 and 5 highlights the effectiveness of our machine learning model, XGBoost, over human experts in harnessing the predictive power of historical financial data for innovation performance.

6.2.3. Combined model

Recent studies (e.g., Costello et al. 2020; Liu 2022; Cao et al. 2024) have illuminated the synergistic relationship between human experts and intelligent machines. While machines excel in efficiently deciphering intricate connections among variables and forecasting future outcomes based on existing data sources, human experts hold a distinct competitive edge when it comes to swiftly assimilating nuanced, context and time-sensitive information—commonly known as human discretion. Furthermore, experts often wield a wealth of domain-specific knowledge that remains unaccounted for in available data repositories. Consequently, the assessments provided by human experts may encapsulate this invaluable reservoir of soft information, which could be absent in historical financial statements.

To test this hypothesis, we construct a new machine learning model that incorporates not just the 55 financial statement data items used in Table 5, but also the composite evaluation score (SCORE_COMP) from Table 4 (referred to as XGBoost_COMBINED). The outcomes of this analysis are presented in Table 6. The change in NDCG@k for XGBoost_COMBINED relative to XGBoost_ACCOUNTING in Table 5 is 0.8% for PATENT, -1.4% for INVENTION_PATENT, 13% for BLOCKBUSTER_PATENT, and 12% for VCPED. These results show no consistent evidence that XGBoost_COMBINED outperforms XGBoost_ACCOUNTING. Therefore, we conclude that there is no clear evidence that human experts' discretion contributes significantly to the prediction of future innovation outcomes. While this study does not delve into the definitive reasons behind the disparities in our findings compared to those of previously cited research, it is crucial to highlight a significant institutional difference between our study and the earlier ones. In our specific setting, the experts involved are temporary independent third-party consultants enlisted by government agencies to evaluate Innofund applications. In contrast, the experts examined in prior studies are permanent employees working in for-profit organizations with more profound institutional insights and potentially greater motivation. This distinction reveals a notable agency problem and information disadvantage within the government sector, setting it apart from the private sector.

6.2.4. Uncovering the performance difference between XGBoost_ACCOUNTING and XGBoost_BENCHMARK

Given that both XGBoost_ACCOUNTING and XGBoost_BENCHMARK utilize the same Innofund grant applications, we posit that the observed performance differences partially stem from the varying capabilities of XGBoost versus human experts in processing information. As discussed in Section 4.3, we suggest that advanced machine learning models, like XGBoost, are better equipped to manage nonlinearity and interactions among many predictors than human experts who face limited cognitive capabilities. To empirically test this hypothesis, we rerun both models using the Logit approach, which lacks the ability to handle nonlinearity and predictor interactions. These models are designated as Logit_ACCOUNTING and Logit_BENCHMARK, respectively.

Table 2 highlights the presence of varying degrees of missing values among the 55 data items. Unlike the XGBoost model, which can directly accommodate missing values, the Logit model requires pre-processing of the missing values. We recode the missing values using the mean values in the training sample prior to training the prediction models.[10]

Table 7 presents the regression outcomes for the Logit models. Panel A of Table 7 details the performance of Logit_BENCHMARK, and Panel B reports the performance of Logit_ACCOUNTING. The results indicate that the performance of Logit_BENCHMARK aligns closely with that of XGBoost_BENCHMARK in Table 4. However, there is a significant decline in the performance of Logit_ACCOUNTING when compared to XGBoost_ACCOUNTING, as documented in Table 5. Additionally, Logit_ACCOUNTING no longer outperforms Logit_BENCHMARK; in fact, the NDCG@k values for Logit_ACCOUNTING are slightly lower than those for Logit_BENCHMARK. These outcomes support our hypothesis that the superior ability of XGBoost to manage nonlinearity

[10] Inferences are qualitatively similar if each data item's missing values are replaced by zero (untabulated). In addition, the performance of XGBoost_ACCOUNTING in Table 5 is not negatively affected if we replace the missing values with zero or training sample mean (untabulated).

and interactions among predictors is pivotal in explaining the performance discrepancy between XGBoost_ACCOUNTING and XGBoost_BENCHMARK.[11]

6.2.5. Temporal analysis

Due to the limited time series data available from our sources, we have primarily relied on a cross-sectional approach when constructing our machine learning models thus far. However, it raises a pertinent question: would our findings hold if we employ a temporal approach? To address this question, we split our data into training and testing sets based on time periods. Specifically, we utilize all applications submitted between 2005 and 2008 for our training dataset, and applications from 2009 to 2010 for our testing dataset.[12]

One complexity we encounter in our dataset is the presence of startup firms submitting multiple Innofund grant applications over time. To ensure that our out-of-sample performance results are not influenced by the inclusion of the same firms in both the training and testing datasets, we remove applicants from the testing dataset if they are found to have appeared in the training dataset. This approach helps to safeguard the integrity of our analysis and ensures that our model's performance is not biased by the repeated presence of certain startups in our data.

Table 8 presents the out-of-sample performance results, comparing the benchmark model from Panel A in Table 4, our proposed model from Table 5, and the combined model in Table 6. The NDCG@k values continue to be consistently higher for XGBoost_ACCOUNTING than for XGBoost_BENCHMARK. The differences in NDCG@k are also economically significant. In addition, we continue to find no evidence that the combined model can beat the XGBoost_ACCOUNTING model.

[11] As noted before, experts' conflicts of interest could also explain partially the lower performance of experts' model. Unfortunately, we cannot directly test this hypothesis because we have no information about the recruited experts.

[12] It is important to note that the temporal model cannot be implemented in real time as VCPED is defined as of 2023.

7. Model performance by firm age

Although the median age of Innofund applicants, defined as the difference between the firm's founding year and the Innofund application year, is only 4 years, their ages range from 0 to 31 years. We analyze whether performance differences between the XGBoost_ACCOUNTING model and the XGBoost_BENCHMARK model hold for young versus old firms, using the median firm age as a cutoff.

Table 9 presents the NDCG@k performance statistics for the test sample. The value of 'k' is set as the realized percentage of the outcome dummy variables in each subsample, representing the expected outcome under random guessing, multiplied by the size of the test subsample. Our analysis yields two main conclusions. First, the XGBoost_ACCOUNTING model consistently outperforms the XGBoost_BENCHMARK model for both young and old firms. The only exception is that the XGBoost_ACCOUNTING model underperforms the XGBoost_BENCHMARK model for young firms when the dependent variable is BLOCKBUSTER_PATENT. Thus, we conclude that the superiority of the XGBoost_ACCOUNTING model over XGBoost_BENCHMARK model holds for both young and old Innofund applicants.

Second, the percentage increase in prediction performance of the XGBoost_ACCOUNTING model compared to the XGBoost_BENCHMARK model is consistently larger for old firms across all four dependent variables. This indicates that the superiority of the XGBoost_ACCOUNTING model over the XGBoost_BENCHMARK model is less pronounced for young Innofund applicants.

8. Uncovering the black box behind Table 5's XGBoost_ACCOUNTING models

To understand the drivers of our XGBoost_ACCOUNTING models—specifically those analyzing PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED in Table 5 - we employ Shapley values (Shapley 1953) to elucidate how these models process the 55 financial statement data items. Shapley values, based on game theory, are the average marginal contribution of a feature considering all possible permutations of other features. We standardize these values to ensure that the total sum of normalized Shapley values for all features within a model equals 100%, facilitating the interpretation of feature importance across models.

Given that human experts also process the 55 financial statement data items to derive their financial scores (SCORE_FIN), we also build an XGBoost model that uses the same 55 financial statement data items to predict SCORE_FIN (referred to as XGBoost_SCORE_FIN) to understand how human experts process the financial statement data items to derive their financial scores. We will use this latter model to illuminate the disparities between human experts and XGBoost_ACCOUNTING when it comes to integrating 55 financial statement data elements into the forecast of future innovation outcomes. The unreported findings indicate that the out-of-sample R^2 (OOS R2) for XGBoost_SCORE_FIN stands at 0.141. This result suggests that the set of 55 financial statement data items provides a reasonable prediction of experts' financial scores.

To make the Shapley value analysis tractable, we focus on each model's top 10 most significant input variables, measured by the magnitude of Shapley values. Table 10 reports the top 10 most important predictors for XGBoost_SCORE_FIN versus the XGBoost_ACCOUNTING models for dependent variables PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED. Rank '1' means the most important input variable.

In addition to ranking each input variable, we also report the signs of each input variable. Following Chen and Ke (2024), the sign of each input variable is determined in the following steps: First, we run the linear regression of $SHAP = \beta_0 + \beta_1 x + \varepsilon$, where x represents one of the 55 financial statement data items. We replace the missing values of x using the training sample mean for this analysis. Second, we check the sign of the estimated coefficient $\hat{\beta}_1$. However, if the absolute value of $\hat{\beta}_1$ is smaller than 1% of the unconditional mean of predicted dependent variable, we consider the pertinent variable to have no meaningful economic effect and thus assign a value of zero in the sign column.

Table 10, Panel A reports the Shapley value results for XGBoost_SCORE_FIN versus XGBoost_ACCOUNTING for PATENT. We draw the following conclusions. First, the XGBoost_ACCOUNTING model tends to focus on data items related to cash flows and operating and

non-operating expenses. In contrast, the XGBoost_SCORE_FIN model tends to concentrate on the data items related to total revenue, total liabilities, and aggregate profit.

Second, the direction of influence exerted by the independent variables often differs between the two models. For instance, in Panel A, the XGBoost_SCORE_FIN model assigns a strong positive weight to "Main operating revenue," ranking it first in importance. In contrast, the XGBoost_ACCOUNTING model ranks the same variable 23rd in importance and assigns a negative effect. Conversely, the XGBoost_ACCOUNTING model assigns a positive weight to first ranked data item, "Non-operating expenses", while the XGBoost_SCORE_FIN model ranks it 19th and does not assign it a consistent directional impact.

Third, we observe little overlap in the top 10 most important input variables between the two models. These results suggest fundamental differences in information processing between our XGBoost_ACCOUNTING model and human experts' mental model.

We reach similar conclusions when we compare the XGBoost_SCORE_FIN model against the XGBoost_ACCOUNTING models for INVENTION_PATENT in Panel B, BLOCKBUSTER_PATENT in Panel C, and VCPED in Panel D.

Although the Shapley value analysis in Table 10 identifies the most influential predictors among the 55 raw accounting variables, these results are not easily interpretable within a standard causal inference framework. To address this limitation, we conduct an exploratory analysis by constructing an alternative XGBoost model based on the theoretical constructs derived from the same set of raw accounting data, referred to as the XGBoost_CAUSAL model. If this alternative model achieves predictive performance comparable to that of the XGBoost_ACCOUNTING models in Table 5, we can interpret the included theoretical construct proxies as offering a meaningful economic interpretation of the relationships captured by the original XGBoost_ACCOUNTING models.

The Shapley value analysis in Table 10 indicates that raw accounting variables related to cash flows as well as operating and non-operating expenses are among the most influential predictors. Guided by this finding, we identify candidate theoretical construct proxies that closely correspond to

these two categories. Our selection process is informed by prior innovation literature and established frameworks in financial statement analysis, such as the DuPont decomposition.

First, we identify a set of theoretical construct proxies linked to the cash flow–related raw accounting items, which captures a firm's degree of financial constraints. Prior finance research highlights that firms with lower financial constraints—due to either strong internal cash flows or access to external financing—are better positioned to sustain R&D investment and realize innovation outcomes. For instance, Brown and Petersen (2011) show that cash reserves help startups smooth R&D expenditures amid earnings volatility or credit market disruptions. Similarly, Brown, Fazzari, and Petersen (2009) find that R&D spending in young, high-tech U.S. firms is closely tied to internal cash flows and equity financing, whereas mature firms' R&D is not. To proxy for financial constraints, we construct twelve economically meaningful financial ratios using the 55 raw accounting variables. Because no single measure fully captures financial constraints, we rely on multiple indicators. Panel A of Table 11 presents these ratios along with their definitions.

Second, we identify several candidate theoretical construct proxies related to operating and non-operating expense items, which represent a firm's flow and stock in innovation investments. Prior research emphasizes the role of research and development (R&D) spending as a key driver of innovation outcomes (e.g., Chan, Lakonishok, and Sougiannis 2001). However, under GAAP, R&D expenditures are typically expensed as incurred (Lev and Sougiannis 1996). Furthermore, Koh and Reeb (2015) document that many firms engaged in substantial innovation do not report R&D expenses separately, further complicating their identification. Based on these findings, we hypothesize that certain expense-related items—such as non-operating expenses and wages payable—may serve as proxies for unreported or intangible innovation investments, potentially explaining their predictive relevance.[13] To capture both the flow and stock aspects of such expenditures, we construct five economically motivated financial ratios from the 55 raw accounting variables, as detailed in Panel A of Table 11.

[13] Notably, the 55 raw accounting variables do not include an explicit line item for R&D expenses, suggesting that the sample firms likely embed their R&D outlays within broader expense categories.

We retrain the XGBoost models from Table 5 using the 17 candidate theoretical construct proxies and report their predictive performance on the test sample in Panel B of Table 11. Using NDCG@k as our primary evaluation metric, we find that the ratios of NDCG@k in Table 11 relative to those in Table 5 are 0.99 for PATENT, 1.02 for INVENTION_PATENT, 1.05 for BLOCKBUSTER_PATENT, and 0.79 for VCPED. These results indicate that the XGBoost_CAUSAL models approximate the predictive capacity of the original XGBoost_ACCOUNTING models reasonably well, suggesting that the selected theoretical construct proxies capture much of the relevant variation associated with innovation outcomes.

9. Conclusion

This study utilizes a proprietary database of Innofund grant applications submitted by Chinese high-tech startup firms to examine whether their innovation performance can be predicted using an advanced machine learning model, XGBoost, that is solely based on historical financial statements from their grant application proposals. We compare the performance of our proposed financial statement-based model with a benchmark XGBoost model based on human experts' comprehensive evaluation scores of the same grant applications. While the value of publicly listed firms' historical financial statements has been well documented, the value of private firms' historical financial statements is subject to debate among academic and practitioner circles, especially for financial markets with weak institutional environments. No prior research has provided direct evidence on the value of historical financial statements in predicting high-tech startup firms' innovation performance.

Our research question holds significance because governments worldwide allocate billions to stimulate research and innovation through high-tech startup firms. Government agencies typically enlist third-party experts to evaluate the merits of such grant applications. However, a core challenge lies in the substantial time required for R&D investments to manifest outcomes, making innovation performance predictions intricate, even for skilled experts who also face the challenges of limited information processing capacity and cognitive biases. Moreover, critics often highlight potential biases, such as nepotism.

Our analysis of out-of-sample performance reveals that an XGBoost model utilizing solely financial statement data consistently outperforms a benchmark model employing XGBoost but utilizing expert evaluation scores as sole model input. We show that this performance discrepancy is due to the superior capacity of XGBoost over human experts in managing nonlinearity and interactions among predictors. Furthermore, our findings indicate little evidence that integrating expert evaluation scores with financial statement data does not enhance predictive performance. We also find that the superiority of the XGBoost_ACCOUNTING model over the XGBoost_BENCHMARK model holds for both young and old Innofund applicants.

A Shapley value analysis of the XGBoost_ACCOUNTING models identifies two key categories of predictors: raw accounting items related to cash flows and those related to operating and non-operating expenses. To better understand the causal mechanisms underlying the predictive power of these variables, we construct an alternative XGBoost model for each innovation outcome proxy using seventeen theoretical construct proxies. These ratios, derived from the same set of 55 raw accounting items and informed by the Shapley analysis, capture dimensions of financial constraints and innovation investment. We find that the alternative models closely replicate the predictive performance of the original XGBoost_ACCOUNTING models, underscoring the central role of financial constraints and innovation investment in shaping future innovation outcomes.

Our results challenge the prevalent skepticism concerning the quality of financial statements from Chinese private firms in the science and technology sector, demonstrating that these documents indeed possess significant predictive value for future innovation outcomes. However, realizing this potential requires the application of advanced machine learning techniques. Moreover, our analysis reveals that human judgment contributes minimally to the accuracy of our predictive tasks, marking a departure from previous studies that have highlighted the beneficial role of human discretion in for-profit organizational contexts. These insights carry substantial implications for how government agencies should distribute the extensive research funds dedicated to fostering innovation. Our findings

underscore the potential promise of machine learning as a governance mechanism in overcoming the challenges faced by government entities in managing and allocating public financial resources.

References

Allee, K. D., and T. L. Yohn. 2009. The Demand for Financial Statements in an Unregulated Environment: An Examination of the Production and Use of Financial Statements by Privately Held Small Businesses. The Accounting Review 84 (1): 1–25.

Antonelli, C., and F. Crespi. 2013. The "Matthew effect" in R&D public subsidies: The Italian evidence. *Technological Forecasting and Social Change* 80 (8): 1523–1534.

Armstrong, C., A. Davila, and G. Foster. 2006. Venture-Backed Private Equity Valuation and Financial Statement Information. Review of Accounting Studies 11 (1): 119–154.

Baik, B. K., N. Berfeld, and R. S. Verdi. 2023. Do Public Financial Statements Influence Venture Capital and Private Equity Financing? SSRN Scholarly Paper. Rochester, NY.

Ball, R. 2024. By What Criteria Do We Evaluate Accounting? Some Thoughts on Economic Welfare and the Archival Literature. *Journal of Accounting Research* 62 (1): 7–54.

Ball, R., and L. Shivakumar. 2005. Earnings Quality in UK Private Firms: Comparative Loss Recognition Timeliness. Journal of Accounting and Economics 39 (1): 83–128.

Bao, Y., B. Ke, B. Li, Y. J. Yu, and J. Zhang. 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58 (1): 199–235.

Bar-Yosef, S., C. D'Augusta, and A. Prencipe. 2019. Accounting Research on Private Firms: State of the Art and Future Directions. The International Journal of Accounting 54 (02): 1950007.

Beatty, A.L., B. Ke, K.R. Petroni. 2002. Earnings management to avoid earnings declines across publicly and privately held banks. *The Accounting Review*, 77 (3), 547-570.

Bian, J., Greenberg, J., Li, J., & Wang, Y. 2022. Good to go first? Position effects in expert evaluation of early-stage ventures. *Management Science*, 68(1), 300-315.

Bernstein, S. 2015. Does Going Public Affect Innovation? *The Journal of Finance* 70 (4): 1365–1403.

Bertomeu, J. 2020. Machine learning improves accounting: discussion, implementation and research opportunities. *Review of Accounting Studies* 25 (3): 1135–1155.

Beuselinck, C., F. Elfers, J. Gassen, and J. Pierk. 2023. Private Firm Accounting: The European Reporting Environment, Data and Research Perspectives. Accounting and Business Research 53 (1): 38–82.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives* 33 (3): 163–84.

Brown, J. R., & Petersen, B. C. (2011). Cash holdings and R&D smoothing. *Journal of Finance*, 66(1), 151-185

Brown, J. R., Fazzari, S. M., & Petersen, B. C. (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom. *Journal of Finance*, 64(1), 151-185.

Cao, S., W. Jiang, J. L. Wang, and B. Yang. 2024. From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses. *Journal of Financial Economics* 160: 103910

Chan, L. K., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6), 2431-2456.

Chen, C. and B. Ke. 2024. Machine Learning as a Management Control Mechanism: The Case of Employee Selection. National University of Singapore working paper.

Chen, F., O.-K. Hope, Q. Li, and X. Wang. 2011. Financial Reporting Quality and Investment Efficiency of Private Firms in Emerging Markets. The Accounting Review 86 (4): 1255–1288.

Chen, T., and C. 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. KDD '16. New York, NY, USA: Association for Computing Machinery.

Cheng, S., R. Lu, and X. Zhang. 2020. What Should Investors Care About? Mutual Fund Ratings by Analysts vs. Machine Learning Technique. *SSRN Electronic Journal*.

Choudhury, P., R. T. Allen, and M. G. Endres. 2021. Machine learning for pattern discovery in management research. *Strategic Management Journal* 42 (1): 30–57.

Coad, A., Segarra, A., & Teruel, M. 2016. Innovation and firm growth: does firm age play a role?. *Research Policy*, 45(2), 387-400.
Coleman, B., K. Merkley, and J. Pacelli. 2022. Human Versus Machine: A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations. *The Accounting Review* 97 (5): 221–244.
Costello, A. M., A. K. Down, and M. N. Mehta. 2020. Machine + man: A field experiment on the role of discretion in augmenting AI-based lending models. *Journal of Accounting and Economics* 70 (2–3): 101360.
David, P. A., Hall, B. H., & Toole, A. A. 2000. Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research Policy*, *29*(4-5), 497-529.
Donelson, D. C., R. Jennings, and J. Mcinnis. 2017. Financial Statement Quality and Debt Contracting: Evidence from a Survey of Commercial Lenders. *Contemporary Accounting Research* 34 (4): 2051–93.
The Economist, 2014. Intellectual property in China: patent fiction. *The Economist*. December 13, 2014.
Fang L, Lerner J, Wu C, Zhang Q. 2023. Anticorruption, government subsidies, and innovation: Evidence from China. *Management Science*. 69(8):4363-88.
Fuertes-Callén, Y., B. Cuellar-Fernández, and C. Serrano-Cinca. 2022. Predicting Startup Survival Using First Years Financial Statements. Journal of Small Business Management 60 (6): 1314–1350.
Haenlein, M., and A. Kaplan. 2019. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review* 61 (4): 5–14.
Hall, B. H., & Lerner, J. 2010. The financing of R&D and innovation. In Handbook of the Economics of Innovation (Vol. 1, pp. 609-639). North-Holland.
Hand, J. R. M. 2005. The Value Relevance of Financial Statements in the Venture Capital Market. The Accounting Review 80 (2): 613–648.
Hope, O.-K., W. B. Thomas, and D. Vyas. 2013. Financial Reporting Quality of U.S. Private and Public Firms. The Accounting Review 88 (5): 1715–1742.
Howell, S. 2017. Financing Innovation: Evidence from R&D Grants. *American Economic Review* 107 (4): 1136–1164.
Howell, S. 2024. Government Intervention in Innovation. *Annual Review of Financial Economics*, *16*: 367-390.
Huber, G. P. 1990. A Theory of the Effects of Advanced Information Technologies on Organizational Design, Intelligence, and Decision Making. *Academy of Management Review* 15 (1): 47–71.
Katz, S. P. 2009. Earnings Quality and Ownership Structure: The Role of Private Equity Sponsors. The Accounting Review 84 (3): 623–658.
Ke, B., and L.F. Wang. 2021. Mixed ownership and firm performance: Evidence from the Chinese venture capital industry. *Emerging Markets Review* 49,100759.
Kerr, W. R., & Nanda, R. 2015. Financing innovation. Annual Review of Financial Economics, 7(1), 445-462.
Kisseleva, K., A. Mjøs, and D. T. Robinson. 2023. On the Importance of Accounting Information in Early-Stage Financing. SSRN Scholarly Paper. Rochester, NY.
Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. 2017. Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*.
Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer. 2015. Prediction Policy Problems. American Economic Review: Papers & Proceedings, 105(5),: 491–495.
Klette, T. J., J. Møen, and Z. Griliches. 2000. Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies. *Research Policy* 29 (4–5): 471–495.
Koh, P.-S., Reeb, D.M., 2015. Missing R&D. *J. Account. Econ.* 60 (1), 73–94.
Lane, J. N., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E. C., & Lakhani, K. R. 2022. Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Management Science*, 68(6), 4478-4495.

Lennox, C., and J. S. Wu. 2022. A review of China-related accounting research in the past 25 years. *Journal of Accounting and Economics* 74 (2): 101539.
Lerner, J. 2009. Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed--and What to Do about It. Princeton, NJ: Princeton University Press.
Lev, B., Sougiannis, T., 1996. The capitalization, amortization, and value-relevance of R&D. *J. Account. Econ.* 21 (1), 107–138.
Li, F., and X. Zheng. 2023. Carbon emission and asset prices: new evidence from machine learning. Working Paper.
Liu, M. 2022. Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. *Journal of Accounting Research* 60 (2): 607–651.
Looze, J. and Goff, T. 2022. Job Creation by Firm Age: Recent Trends in the United States. Ewing Marion Kauffman Foundation, Kansas City.
Lu, H., J.-E. Shin, and M. Zhang. 2023. Financial reporting and disclosure practices in China. *Journal of Accounting and Economics*: 101598.
Lu, T. and Y. Zhang. 2025. 1 + 1 > 2? Information, Humans, and Machines. *Information Systems Research* 2025 36:1, 394-418.
Lyonnet, V., and L. H. Stern. 2022. Venture Capital (Mis)Allocation in the Age of AI. SSRN Scholarly Paper. Rochester, NY.
Mezzanotti, F., & Simcoe, T. 2023. Innovation and appropriability: Revisiting the role of intellectual property (No. w31428). National Bureau of Economic Research.
Minnis, M., and A. Sutherland. 2017. Financial Statements as Monitoring Mechanisms: Evidence from Small Commercial Loans. Journal of Accounting Research 55 (1): 197–233.
Myers KR, Lanahan L. Estimating spillovers from publicly funded R&D: Evidence from the US Department of Energy. 2022. *American Economic Review*. Jul 1;112(7):2393-423.
National Academies of Sciences, Engineering, and Medicine. 2024. Experimental Approaches to Improving Research Funding Programs: Proceedings of a Workshop.
Scott, E. L., P. Shu, and R. M. Lubynsky. 2020. Entrepreneurial Uncertainty and Expert Evaluation: An Empirical Analysis. *Management Science* 66 (3): 1278–1299.
Shapley, L. 1953. A value for n-person games. *Contributions to the Theory of Games,* 28, 307-317.
Shmueli, G. 2010. To explain or to predict. *Statistical Science* 25: 289-310.
Shrestha, Y. R., V. Krishna, and G. von Krogh. 2021. Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research* 123: 588–603.
Sørensen, J. B., & Stuart, T. E. 2000. Aging, obsolescence, and organizational innovation. Administrative Science Quarterly, 45(1), 81-112.
Stuart, T., and Y. Wang. 2016. Who cooks the books in China, and does it pay? Evidence from private, high-technology firms. *Strategic Management Journal* 37 (13): 2658–2676.
Tantri, P. 2021. Fintech for the Poor: Financial Intermediation Without Discrimination. *Review of Finance* 25 (2): 561–593.
Thomas Falk, M., and R. Svensson. 2020. Evaluation criteria versus firm characteristics as determinants of public R&D funding. *Science and Public Policy*: scaa032.
Wang, Y., J. Li, and J. L. Furman. 2017. Firm performance and state innovation funding: Evidence from China's Innofund program. *Research Policy* 46 (6): 1142–1161.
Wang, Y., Siegel, J., & Li, J. 2024. Who captures the state? Evidence from irregular awards in a public innovation grant program. *Strategic Management Journal*. In print.
Wilke, A., R. Mata. 2012. Cognitive bias. Encyclopedia of Human Behavior (Second Edition), 531-535.
Yoo, H.S., Y.L. Jung, and S.P. Jun. 2023. Prediction of SMEs' R&D performances by machine learning for project selection. *Scientific Reports* 13:7598.
Zheng, X. 2021. How can Innovation Screening be Improved? A Machine Learning Analysis with Economic Consequences for Firm Performance. Working Paper.

Appendix 1. Variable definition

Variable	Definition
SCORE_COMP	composite expert evaluation score for each Innofund applicant
SCORE_TECH	expert technical evaluation score for each Innofund applicant
SCORE_FIN	expert financial evaluation score for each Innofund applicant
FUNDED	An indicator variable that equals 1 for successful applicants and 0 otherwise
PATENT_RAW	The number of patents applied in the 5-year period post Innofund application
INVENTION_PATENT_RAW	The number of invention patents applied in the 5-year period post Innofund application
BLOCKBUSTER_PATENT_RAW	The number of invention patents applied in the 5-year period post Innofund application that forward citations being the top 10% of all patents in the patent database
PATENT	An indicator variable that equals 1 if Innofund applicants applied for at least one patent in the 5-year period post-application and 0 otherwise
INVENTION_PATENT	An indicator variable that equals 1 if Innofund applicants applied for at least one invention patent in the 5-year period post-application and 0 otherwise
BLOCKBUSTER_PATENT	An indicator variable that equals 1 if Innofund applicants applied for at least one patent in the 5-year period post-application with the patent's forward citations falling into the top 10% (10^{th} decile) of all patents in the patent database and 0 otherwise
VCPED	An indicator variable that equals 1 if Innofund applicants had at least one new VC/PE shareholder in the post-application period until 2023 and 0 otherwise

Appendix 2. A description of XGBoost's training

Like most other machine learning techniques, XGBoost relies on a specific set of parameters that require optimization to achieve optimal predictive performance. These parameters can be broadly categorized into two groups: those that define the characteristics of individual regression trees and those that govern the boosting process. However, due to computational constraints, we focus our fine-tuning efforts on three pivotal parameters within the XGBoost algorithm: the number of trees, the maximum tree depth, and the learning rate.

To fine-tune our parameters, we adopt the well-established practice of n-fold cross-validation. Despite the inclusion of applicants spanning from 2005 to 2010 within our dataset, each applicant is subject to individual evaluation, ensuring minimal overlap in their assessments.[14] Therefore, we adopt a cross-sectional approach, aggregating applicants from various years and treating each application as an independent entity. We then employ a random 80-20 partitioning scheme with a group shuffle split, ensuring that multiple applications from the same firm are grouped either in the training or test dataset.

Furthermore, we enhance our evaluation process by utilizing the stratified 5-fold cross-validation technique. This involves splitting our training sample into five equally sized subsets, or folds. During each iteration of this process, we employ four of these folds for training our model while reserving one for validation. This iterative approach guarantees that each subset serves as the validation set at least once, providing a comprehensive evaluation of our model's predictive performance across diverse data segments.

We employ a rigorous optimization process known as grid search cross-validation. This method entails a thorough examination of all potential parameter combinations within a predefined search range. With each unique combination, we assess the model's predictive capabilities through the application of the stratified 5-fold cross-validation technique noted above with the Area Under the Receiver Operating Characteristic curve (AUC-ROC) as the performance evaluation metric. From this

[14] In exceptional instances, MOST would decline an applicant if the company had an Innofund-supported project that remained incomplete. Conversely, external experts assess projects solely on the information furnished within the application form.

exhaustive evaluation, we identify the set of parameters that consistently produces the highest mean validation score across the five folds, ultimately selecting it as the optimal parameter configuration. Table A1 reports the search range of XGBoost parameter values and the cross-validation results for the main model XGBoost_ACCOUNTING in Table 5.

In binary classification scenarios characterized by imbalanced class distributions, it's common for algorithms to prioritize the prediction of the majority class, resulting in subpar performance when it comes to the minority class. To effectively address the data imbalance issue in the dependent variables BLOCKBUSTER_PATENT and VCPED, we employ a technique that augments the penalties for classification errors on the minority class. By doing so, we compel the model to allocate greater attention to this underrepresented class.[15]

[15] We do so by setting the scale_pos_weight parameter in XGBoost algorithms to the ratio between the count of negative instances to the count of the positive instances of the train/validation sample.

Table A1. Tuning Details of XGBoost Parameters for the XGBoost_ACCOUNTING models

Panel A reports the search range of XGBoost parameter values for the XGBoost_ACCOUNTING models reported in Table 5. Panel B shows the chosen parameter values for the main XGBoost models for each dependent variable for the model with financial statement items (XGBoost_Accounting).

Panel A: Search range of XGBoost parameter values

Number of trees	50, 100, 200, 300, 400, 500
Maximum depth of the tree	1, 3, 5, 7
Learning rate	0.01, 0.1, 0.3

Panel B: Cross-validation results

Dependent Variable	Best parameters	Performance Evaluation Metric	Cross Validation Mean Score	Cross Validation Std
PATENT	n_tree=50, max_depth=3, lr=0.1	ROC_AUC	0.666	±0.014
		NDCG@k	0.900	±0.009
INVENTION_PATENT	n_tree=100, max_depth=1, lr=0.1	ROC_AUC	0.658	±0.011
		NDCG@k	0.886	±0.011
BLOCKBUSTER_PATENT	n_tree=400, max_depth=3, lr=0.01	ROC_AUC	0.634	±0.017
		NDCG@k	0.598	±0.015
VCPED	n_tree=50, max_depth=1, lr=0.01	ROC_AUC	0.658	±0.028
		NDCG@k	0.516	±0.020

Table 1. Sample selection procedures

Selection criteria	No. of unique firms
Innofund administrative data from 2005 to 2010	6,648
Require non-missing expert total scores	6,336
Require non-entrepreneurial applicants/non-missing financial scores	5,199
Require non-missing financial statement in t-1 (i.e., not brand-new firms)	4,866

Table 2. Descriptive statistics of the 55 financial statement line items

This table provides the descriptive statistics of the 55 financial statement line items tabulated in the Innofund application forms during our sample period. The measurement unit is in million RMB.

	count	mean	std	25%	50%	75%
Monetary funds	4,851	429.65	29,757.31	0.42	1.08	2.56
Short term investment	1,564	0.17	1.66	0.00	0.00	0.00
Accounts receivable	4,384	3.51	6.59	0.40	1.55	4.18
Other accounts receivable	4,428	1.90	9.64	0.09	0.44	1.62
Inventory	4,385	3.14	10.76	0.28	1.25	3.56
Prepaid account	947	1.65	11.89	0.12	0.45	1.34
Total working capital	4,802	11.91	71.04	2.27	6.19	13.30
Original value of fixed assets	4,723	5.50	16.12	0.38	1.68	6.11
Net fixed assets	1,510	4.53	10.78	0.37	1.73	5.75
Total fixed assets	4,740	4.49	8.34	0.28	1.35	5.39
Construction in progress	394	2.24	4.35	0.26	0.82	2.33
Intangible assets	2,990	2.27	18.67	0.00	0.47	2.24
Deferred assets	2,000	0.47	11.74	0.00	0.00	0.08
Total assets	4,826	17.42	30.80	3.68	10.28	21.54
Short-term borrowing	2,558	3.34	9.68	0.00	0.75	3.30
Accounts payable	4,192	2.11	6.47	0.14	0.85	2.51
Salary payable	2,865	0.20	0.52	0.00	0.04	0.20
Other payables	4,436	1.57	4.06	0.09	0.48	1.58
Unearned revenue	3,000	1.23	4.94	0.00	0.21	1.00
Total current liabilities	4,771	6.54	13.72	0.75	2.84	7.58
Total liabilities	4,694	7.38	17.33	0.77	3.05	8.40
Net assets	2,102	10.41	13.21	2.84	6.79	13.52

Paid-in capital	4,806	6.41	15.26	1.00	3.00	8.06
Capital reserve	2,698	1.37	3.91	0.00	0.10	1.01
Surplus reserve	3,067	1.10	3.34	0.00	0.26	1.01
Retained Earnings	4,689	3.24	6.36	0.35	1.38	3.78
Total owners' equity	4,812	10.21	15.19	2.21	6.25	12.79
Sales revenue	1,913	17.17	21.98	4.12	12.10	23.69
Main operating revenue	4,602	15.86	26.98	2.50	9.35	20.80
Costs of main operations	4,465	11.29	20.39	1.28	6.05	14.90
Main operating profit	4,558	4.62	9.62	0.81	2.60	5.73
Other business profits	2,006	0.19	0.80	0.00	0.00	0.06
Administrative expenses	4,720	1.84	2.96	0.45	1.08	2.21
Operating expenses	2,704	0.73	3.50	0.00	0.08	0.48
Selling expenses	3,717	1.14	5.98	0.03	0.35	1.08
Financing expenses	4,626	0.17	0.53	0.00	0.01	0.13
Investment income	1,623	0.11	0.83	0.00	0.00	0.00
Operating profit	4,673	2.15	4.27	0.25	1.01	2.78
Non-operating revenue	2,352	0.13	0.52	0.00	0.00	0.04
Non-operating expenses	3,005	0.25	10.59	0.00	0.00	0.04
Total profit	4,738	2.93	47.82	0.28	1.07	2.86
Net profit	4,726	1.81	3.48	0.23	0.90	2.27
Cash received from selling goods and providing services	4,429	17.66	55.80	2.31	9.41	22.41
Subtotal of cash inflows from operating activities	121	2.77	4.64	0.61	1.69	3.17
Cash paid for goods and services	4,431	11.61	21.56	1.10	5.57	15.34
Cash paid to and for employees	4,465	1.59	2.66	0.31	0.87	1.92
Net cash flow from operating activities	4,463	2.65	10.68	0.30	0.99	2.62

Subtotal of cash inflows from investing activities	1,865	0.86	10.41	0.00	0.00	0.02
Cash paid to build fixed assets, intangible assets and other long-term investments	4,209	1.70	3.93	0.08	0.45	1.72
Subtotal of cash outflows from investing activities	4,050	2.45	17.48	0.09	0.50	2.00
Net cash flow generated from investing activities	3,967	1.96	4.46	0.09	0.52	2.05
Subtotal of cash inflows from financing activities	3,032	4.59	20.26	0.00	1.00	4.42
Subtotal of cash outflows from financing activities	2,755	3.42	28.86	0.00	0.14	1.66
Net cash flow from financing activities	3,450	2.42	5.67	0.06	0.73	2.51
Net increase in cash and cash equivalents	4,517	1.70	17.97	0.17	0.49	1.35

Table 3. Descriptive statistics of key regression variables

This table provides the descriptive statistics of key independent and dependent variables. Variables are defined in Appendix 1.

	count	mean	std	25%	50%	75%
SCORE_COMP	4,866	71.62	9.70	64.95	72.08	79.23
SCORE_TECH	4,866	69.31	10.44	62.36	69.98	77.36
SCORE_FIN	4,866	70.26	10.40	66.03	71.30	76.38
FUNDED	4,866	0.50	0.50	0	0	1.00
PATENT_RAW	4,866	5.14	16.64	0	0	4.00
INVENTION_PATENT_RAW	4,866	2.66	11.04	0	0	2.00
BLOCKBUSTER_PATENT_RAW	4,866	0.31	1.53	0	0	0
PATENT	4,866	0.42	0.49	0	0	1.00
INVENTION_PATENT	4,866	0.36	0.48	0	0	1.00
BLOCKBUSTER_PATENT	4,866	0.14	0.35	0	0	0
VCPED	4,866	0.11	0.31	0	0	0

Table 4. Out-of-sample prediction performance: XGBoost models based on expert evaluation scores (XGBoost_BENCHMARK)

This table reports performance metrics of XGBoost models that use expert evaluation scores to predict innovation performance. Our test sample includes 20% (i.e., 989) randomly selected applicants during the sample period. The unconditional mean values for our dependent variables—PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED— in the test sample are 0.441, 0.365, 0.153, and 0.119, respectively. See Appendix 1 for variable definitions.

Panel A. using SCORE_COMP as the predictor

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.517	0.510	0.480	0.548
NDCG@k	0.449	0.365	0.153	0.158
PRECISION@k	0.452	0.374	0.139	0.161
RECALL@k	0.452	0.374	0.139	0.161

Panel B. using SCORE_TECH and SCORE_FIN as predictors

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.527	0.510	0.490	0.525
NDCG@k	0.447	0.349	0.148	0.146
PRECISION@k	0.450	0.360	0.132	0.186
RECALL@k	0.450	0.360	0.132	0.186

Panel C. using SCORE_FIN as the predictor

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.521	0.518	0.465	0.515
NDCG@k	0.450	0.371	0.128	0.126
PRECISION@k	0.438	0.355	0.113	0.102
RECALL@k	0.438	0.355	0.113	0.102

Panel D. using SCORE_TECH as the predictor

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.529	0.506	0.490	0.538
NDCG@k	0.463	0.362	0.148	0.145
PRECISION@k	0.447	0.349	0.132	0.153
RECALL@k	0.447	0.349	0.132	0.153

Table 5. Out-of-sample prediction performance: XGBoost models based on the 55 financial statement items (XGBoost_ACCOUNTING)

This table reports performance metrics of the XGBoost models that use financial statement items to predict innovation performance. Our test sample includes 20% (i.e., 989) randomly selected applicants during the sample period. The unconditional mean values for our dependent variables—PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED— in the test sample are 0.441, 0.365, 0.153, and 0.119, respectively. See Appendix 1 for variable definitions.

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.633	0.635	0.574	0.629
NDCG@k	0.563	0.505	0.169	0.251
PRECISION@k	0.544	0.476	0.172	0.186
RECALL@k	0.544	0.476	0.172	0.186

Table 6. Out-of-sample prediction performance: XGBoost models based on the 55 financial statement data items plus SCORE_COMP (XGBoost_COMBINED)

This table reports performance metrics of XGBoost models that use the 55 financial statement data items and the composite evaluation score to predict innovation performance. Our test sample includes 20% (i.e., 989) randomly selected applicants during the sample period. The unconditional mean values for our dependent variables—PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED— in the test sample are 0.441, 0.365, 0.153, and 0.119, respectively. See Appendix 1 for variable definitions.

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.636	0.630	0.556	0.648
NDCG@k	0.568	0.498	0.191	0.281
PRECISION@k	0.544	0.476	0.205	0.220
RECALL@k	0.544	0.476	0.205	0.220

Table 7. Out-of-sample prediction performance using logit models

This table reports performance metrics of logit models that use the composite evaluation score or the 55 financial statement data items to predict innovation performance. Missing values are filled with the unconditional mean in the training sample. Our test sample includes 20% (i.e., 989) randomly selected applicants during the sample period. The unconditional mean values for our dependent variables—PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED— in the test sample are 0.441, 0.365, 0.153, and 0.119, respectively. See Appendix 1 for variable definitions.

Panel A. Using SCORE_COMP as the predictor (Logit_BENCHMARK)

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.512	0.507	0.483	0.528
NDCG@k	0.446	0.368	0.172	0.160
PRECISION@k	0.462	0.375	0.180	0.168
RECALL@k	0.440	0.363	0.159	0.144

Panel B. Using the 55 financial statement data items as predictors (Logit_ACCOUNTING)

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.5	0.5	0.5	0.5
NDCG@k	0.441	0.365	0.153	0.119
PRECISION@k	0.435	0.361	0.120	0.109
RECALL@k	0.415	0.349	0.106	0.093

Table 8. Out-of-sample prediction performance using the temporal approach

This table reports the performance metrics of XGBoost models that use the composite evaluation score or the 55 financial statement items or both to predict innovation performance. Our test sample includes 1,316 applicants applied during the later period of our sample (i.e., 2009 and 2010). See Appendix 1 for variable definitions.

Panel A. XGBoost based on SCORE_COMP

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.551	0.553	0.531	0.555
NDCG@k	0.554	0.472	0.148	0.13
PRECISION@k	0.559	0.487	0.157	0.132
RECALL@k	0.559	0.487	0.157	0.132

Panel B. XGBoost based on the 55 financial statement data items

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.652	0.656	0.626	0.592
NDCG@k	0.644	0.58	0.244	0.133
PRECISION@k	0.621	0.554	0.241	0.147
RECALL@k	0.620	0.554	0.241	0.147

Panel C. XGBoost based on the 55 financial statement data items and SCORE_COMP

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.653	0.654	0.624	0.587
NDCG@k	0.644	0.581	0.248	0.165
PRECISION@k	0.626	0.554	0.246	0.147
RECALL@k	0.625	0.554	0.246	0.147

Table 9. Model performance by firm age

This table compares the performance of XGBoost models using expert evaluation scores (XGBoost_BENCHMARK) and those using 55 financial statement items (XGBoost_ACCOUNTING), categorized by young and old firms in the test sample. Our test dataset consists of 989 applicants, representing 20% of the total sample, randomly selected during the sample period. Firms are classified as young or old based on the median firm age of 4 years. Firm age (firm_age) is defined as the difference between the founding year and the Innofund application year. We focus on the performance metric NDCG@k due to the skewness of our dependent variables. The value of k corresponds to the mean values of our dependent variables in each test subsample. Since the values of PRECISION@k and RECALL@K are identical, we report the former only in the table. See Appendix 1 for variable definitions.

		XGBoost_BENCHMARK (1)	XGBoost_ACCOUNTING (2)	Percentage change from (1) to (2)
DV = PATENT				
firm_age > median (k=0.46)	NDCG@k	0.463	0.603	30.2%
	PRECISION@k	0.481	0.568	18.1%
firm_age < median (k=0.33)	NDCG@k	0.361	0.384	6.4%
	PRECISION@k	0.356	0.386	8.4%
DV = INVENTION_PATENT				
firm_age > median (k=0.38)	NDCG@k	0.400	0.524	31.0%
	PRECISION@k	0.409	0.509	24.4%
firm_age < median(k=0.29)	NDCG@k	0.328	0.376	14.6%
	PRECISION@k	0.330	0.393	19.1%
DV = BLOCKBUSTER_PATENT				
firm_age > median (k=0.15)	NDCG@k	0.200	0.239	19.5%
	PRECISION@k	0.179	0.238	33.0%
firm_age < median (k=0.10)	NDCG@k	0.121	0.094	-22.3%
	PRECISION@k	0.132	0.105	-20.5%
DV = VCPED				
firm_age > median (k=0.11)	NDCG@k	0.092	0.244	165.2%
	PRECISION@k	0.125	0.234	87.2%
firm_age < median (k=0.11)	NDCG@k	0.098	0.193	96.9%
	PRECISION@k	0.071	0.214	201.4%

Table 10. Shapley values from various XGBoost models: the top 10 most important features

This table reports the XGBoost model's 10 most important input variables for the following dependent variables: *SCORE_FIN*, *PATENT*, *INVENTION_PATENT*, *BLOCKBUSTER_PATENT*, and *VCPED*. Feature importance is measured by the magnitude of SHAP values of each feature in the test sample. Rank '1' means the most important input variable. Our testing sample includes 20% (i.e., 989) randomly selected applicants during the sample period. The sign of each input variable is determined in the following steps: First, we run the linear regression of $SHAP = \beta_0 + \beta_1 x + \varepsilon$. Second, we check the sign of the estimated coefficient β_1. To gauge the economic significance, we assign a value of zero in the sign column for a given variable if the absolute value of the corresponding β_1 is smaller than 1% of the unconditional mean of the predicted dependent variable. See Appendix 1 for variable definitions.

Panel A. The top 10 most important features for the XGBoost_*SCORE_FIN* model and the XGBoost_ACCOUNTING model for PATENT

Feature name	XGBoost_*SCORE_FIN*	Sign of XGBoost_*SCORE_FIN*	XGBoost_ACCOUNTING	Sign of XGBoost_ACCOUNTING	Difference in ranking	Same sign
Main operating revenue	1	+	23	-	-22	FALSE
Total liabilities	2	-	4	+	-2	FALSE
Monetary funds	3	+	18	-	-15	FALSE
Capital reserves	4	+	20	+	-16	TRUE
Administrative expenses	5	-	10	-	-5	TRUE
Total profit	6	+	32	+	-26	TRUE
Gross profit	7	+	48	0	-41	FALSE
Total fixed assets	8	+	6	0	2	FALSE
Cash received from selling goods and providing services	9	+	50	0	-41	FALSE
Total current assets	10	+	31	+	-21	TRUE
Non-operating expenses	19	0	1	+	18	FALSE
Wages payable	15	0	2	+	13	FALSE
Net profit	32	0	3	-	29	FALSE

Net increase in cash and cash equivalents	36	0	5	+	31	FALSE
Retained Earnings	30	0	7	-	23	FALSE
Cash paid for goods and services	34	0	8	+	26	FALSE
Net assets	45	0	9	+	36	FALSE

Panel B. The top 10 most important features for the XGBoost_*SCORE_FIN* model and the XGBoost_ACCOUNTING model for *INVENTION_PATENT*

Feature name	XGBoost_*SCORE_FIN*	Sign of XGBoost_*SCORE_FIN*	XGBoost_ACCOUNTING	Sign of XGBoost_*ACCOUNTING*	Difference in ranking	Same sign
Main operating revenue	1	+	40	0	-39	FALSE
Total liabilities	2	-	1	+	1	FALSE
Monetary funds	3	+	30	0	-27	FALSE
Capital reserves	4	+	19	+	-15	TRUE
Administrative expenses	5	-	15	-	-10	TRUE
Total profit	6	+	49	0	-43	FALSE
Main operating profit	7	+	41	0	-34	FALSE
Total fixed assets	8	+	11	+	-3	TRUE
Cash received from selling goods and providing services	9	+	50	0	-41	FALSE
Total current assets	10	+	33	0	-23	FALSE
Non-operating expenses	19	0	2	+	17	FALSE
Net increase in cash and cash equivalents	36	0	3	+	33	FALSE
Cash paid to and for employees	43	0	4	+	39	FALSE
Net cash flows from operating activities	44	0	5	+	39	FALSE
Wages payable	15	0	6	+	9	FALSE
Net assets	45	0	7	+	38	FALSE
Net profit	32	0	8	-	24	FALSE
Surplus reserves	20	0	9	+	11	FALSE
Accounts payable	31	0	10	+	21	FALSE

Panel C. The top 10 most important features for the XGBoost_*SCORE_FIN* model and the XGBoost_ACCOUNTING model for *BLOCKBUSTER_PATENT*

Feature name	XGBoost_*SCORE_FIN*	Sign of XGBoost_*SCORE_FIN*	XGBoost_*BLOCKBUSTER_PATENT*	Sign of XGBoost_*BLOCKBUSTER_PATENT*	Difference in ranking	Same sign
Main operating revenue	1	+	52	0	-51	FALSE
Total liabilities	2	-	26	-	-24	TRUE
Monetary funds	3	+	32	+	-29	TRUE
Capital reserves	4	+	33	+	-29	TRUE
Administrative expenses	5	-	43	0	-38	FALSE
Total profit	6	+	16	-	-10	FALSE
Main operating profit	7	+	35	+	-28	TRUE
Total fixed assets	8	+	23	-	-15	FALSE
Cash received from selling goods and providing services	9	+	50	0	-41	FALSE
Total current assets	10	+	45	0	-35	FALSE
Net cash flows from operating activities	44	0	1	+	42	FALSE
Net increase in cash and cash equivalents	36	0	2	+	33	FALSE
Paid-in capital	28	0	3	+	24	FALSE
Surplus reserves	20	0	4	-	15	FALSE
Selling expenses	24	0	5	-	18	FALSE
Operating expenses	25	0	6	0	18	TRUE
Net cash flows from financing activities	35	0	7	+	27	FALSE
Unearned revenue	23	0	8	0	14	TRUE
Cost of goods sold	26	0	9	-	16	FALSE
Retained Earnings	30	0	10	0	30	TRUE

Panel D. The top 10 most important features for the XGBoost_*SCORE_FIN* model and the XGBoost_ACCOUNTING model for *VCPED*

Feature name	XGBoost_*SCORE_FIN*	Sign of XGBoost_*SCORE_FIN*	XGBoost_*VCPED*	Sign of XGBoost_*VCPED*	Difference in ranking	Same sign
Main operating revenue	1	+	35	0	-34	FALSE
Total liabilities	2	-	12	-	-10	TRUE
Monetary funds	3	+	14	+	-11	TRUE
Capital reserves	4	+	30	0	-26	FALSE
Administrative expenses	5	-	4	+	1	FALSE
Total profit	6	+	45	0	-39	FALSE
Gross profit	7	+	37	0	-30	FALSE
Total fixed assets	8	+	7	-	1	FALSE
Cash received from selling goods and providing services	9	+	47	0	-38	FALSE
Total current assets	10	+	22	0	-12	FALSE
Unearned revenue	23	0	1	+	22	FALSE
Net cash flows from operating activities	44	0	2	+	42	FALSE
PPE gross	11	0	3	-	8	FALSE
Accounts payable	31	0	5	-	26	FALSE
Net increase in cash and cash equivalents	36	0	6	+	30	FALSE
Selling expenses	24	0	8	+	16	FALSE
Cash paid for goods and services	34	0	9	+	25	FALSE
PPE net	39	0	10	-	29	FALSE

Table 11. Out-of-sample prediction performance: XGBoost models based on the 17 candidate theoretical construct proxies derived from the 55 financial statement items (XGBoost_CAUSAL)

This table reports performance metrics of the XGBoost models that use 17 candidate theoretical construct proxies derived from the 55 raw financial statement items to predict innovation performance. Our test sample includes 20% (i.e., 989) randomly selected applicants during the sample period. The unconditional mean values for our dependent variables—PATENT, INVENTION_PATENT, BLOCKBUSTER_PATENT, and VCPED— in the test sample are 0.441, 0.365, 0.153, and 0.119, respectively. See Appendix 1 for variable definitions.

Panel A. List of candidate theoretical construct proxies

Variables related to cash flows	**Definitions**
Cash holdings ratio	Monetary funds / Total assets
Net cash increase ratio	Net increase in cash and cash equivalents / Total assets
Cash outflows for goods/services	Cash paid for goods and services / Total assets
Operating cash flow ratio	Net cash flow from operating activities / Total assets
Short-term debt ratio	Short-term borrowing / Total assets
Accrued wages ratio	Salary payable / Total assets
Current liabilities ratio	Total current liabilities / Total assets
Total liabilities ratio	Total liabilities / Total assets
Financing cash flow ratio	Net cash flow from financing activities / Total assets
Capital reserve ratio	Capital reserve / Total assets
Retained earnings ratio	Retained earnings / Total assets
Net asset base ratio	Net assets / Total assets
Variables related to operating and non-operating expenses	**Definitions**
Management-related expense intensity	Administrative expenses / Total assets
Operating expense intensity	Operating expenses / Total assets
Financial expense intensity	Financing expenses / Total assets
Non-operating expense intensity	Non-operating expenses / Total assets
Intangible asset intensity	Intangible assets / Total assets

Panel B. test sample performance statistics

Values	***PATENT***	***INVENTION_PATENT***	***BLOCKBUSTER_PATENT***	***VCPED***
ROC_AUC	0.611	0.616	0.540	0.603
NDCG@k	0.558	0.515	0.177	0.199
PRECISION@k	0.537	0.488	0.185	0.186
RECALL@k	0.537	0.488	0.185	0.186