# Venture Selection Via Machine Learning*

Luiz Bissoto†

June 21, 2024

**Abstract**

I study the predictability of startup outcomes using machine learning (ML) methods under rigorous data availability conditions. Selecting out-of-sample investable ventures chosen by ML models leads to 55% more money being correctly invested in successful startups and reduces the amount invested in failing startups by 52% compared to a similar-sized average investor. These gains from ML are achieved by identifying non-linearities and relevant predictive variables ignored by linear models and are robust to several data stringency conditions and selection bias. I also provide evidence that some investors actively use these technologies, potentially contributing to a predictability decline in recent years.

**Keywords:** Venture Capital, Institutional Investors, Investment Decisions, Machine Learning, Artificial Intelligence

**JEL Codes:** G11, G23, G24

*I thank Alexandre Azoulay, Francesco Celentano, Luise Eisfeld, Andreas Fuster, Benjamin Gabay, Jean Herelle, Federico Baldi Lanfranchi, Semyon Malamud, Giuseppe Matera, Darius Nik Nejad, Florian Perusset, and Mohammad Pourmohammadi for their feedback and helpful comments.

†EPFL and Swiss Finance Institute, Extranef (Unil-Dorigny), Lausanne 1015. Email: luiz.bissoto@epfl.ch

# 1 Introduction

Venture capital (VC) investments are typically understood as high-risk and high-return investments. With incredibly high failure rates, frequently above 90%[1], investors returns' tend to be highly correlated with the frequency of investment exits (Cumming and MacIntosh (2003), Cochrane (2005), Phalippou and Gottschalg (2009)). In the last two decades, several technological advances, particularly in machine learning (ML), a subfield of artificial intelligence (AI), facilitated the development and popularization of tools that outperform previously standard econometrics techniques in forecasting tasks. A significant and growing application of these tools is VC investment screening and classification (Krishna et al. (2016), Ross et al. (2021), Kiss (2022), Bonelli (2022)).

The prevailing view from the existing literature is that investors can make better choices using these tools. Therefore, the central question I seek to answer is whether these potential improvements imply the existence of startup performance predictability, how large it is, and whether (or how) investors act upon it. This question entails three important challenges. The first is whether differentiating successes from failures more efficiently entails higher financial returns: AI-empowered investors may choose to make less risky investments, increasing their success rate at the expense of lower returns (a central result in Bonelli (2022)[2]). The second challenge is that these strategies may not be implementable because collecting data is impossible or costly, or the investor lacks the skills to add value to the startups. Finally, another challenge is to conciliate the existence of performance predictability with a highly competitive environment with "relatively free entry" (Cochrane (2005)).

In this paper, I apply popular ML algorithms to a large sample of startup financing rounds whose outcomes (IPO, acquisition, etc.) are known and test their out-of-sample predictability, comparing their results to a standard econometric tool, logit, and random selection as benchmarks. I find that even with limited data availability, ML algorithms can

[1] In the sample I use, less than 10% of the startups present eventually do an IPO.

[2] *"VCs that adopt AI become better at identifying good quality startups (...) [but] become less likely to invest in startups that achieve major success".*

perform significantly better, particularly at early stages where data is scarcer. Using only characteristics of the startup *name* as explanatory variables allows the investor to obtain a precision[3] at least 42 percent higher (an increase from 18 to 26 percent for startups raising for a Series D round) in selecting financing rounds for startups that eventually do an IPO when compared to the average investor. Second, I apply a financial return proxy, the ratio of future money raised over the past money raised by the startup[4] and find that these ML algorithms also considerably improve the precision in identifying potentially high-return startups. For instance, for startups raising at a pre-seed round, this improvement represents a 35 percent higher precision (an increase from 41 to 55 percent). Third, this predictability seems to flatten or decline only in recent years toward the end of the sample. Together, these results suggest that a reasonable level of predictability exists, is slow to revert, and does not arise at the expense of missing highly successful startup investments.

I measure the total economic benefit of applying ML algorithms by comparing the total amount of money raised by startups correctly predicted to eventually do an IPO (referred to as "successful" for brevity) when compared to an "average investor" selecting startups randomly at each financing round based on the frequency of successful startups. For instance, if around 10% of the startups raising money in Series A rounds are successful, then this average investor selects 10% of these startups randomly[5]. Using random forests, a popular ML algorithm (Ho (1998), Breiman (2001)), I find that the total money invested in startups predicted to succeed and that actually succeed is 55% higher than the money invested by this investor across successful startups on average. Repeating the exercise for failures, I find that the money invested in failing startups is 51% lower when using this same algorithm.

I motivate these predictability exercises by exploring the impact of a plausibly exogenous

[3]The fraction of investments that are actually successful among those predicted to be successful.

[4]The correlation between money raised and pre-money valuation, which is available for about 4 percent of the sample observations and fairly distributed across financing rounds, is around 77 percent.

[5]Naturally, an investor selecting ventures randomly will achieve the same success rate, on average, irrespective of the number of ventures invested. However, limiting the quantity by the frequency of successful investments properly scales the number of investments for a fair comparison with the proposed ML methods and mimics an investor trying to invest only in the successful ventures but with a "placebo" selection skill.

technological shock in the identification and subsequent investment of potentially successful startups: the release of AI and ML libraries TensowFlow by Alphabet in 2015, PyTorch by Meta in 2016, and the Microsoft Cognitive Toolkit (CNTK), also in 2015. While previous technological shocks also profoundly impacted VC activity[6], the release of these libraries considerably reduced the cost and replicability of several ML and AI tools. More importantly, they evidenced the "deep reach" of these three firms, which I refer to as "AI-heavy" in these domains. Post-2016, a startup that eventually does an IPO is significantly more likely to receive investment from these three firms: their presence in the financing rounds of these startups increases to around 6.25 percent from 1.2 percent pre-2016 (Figure 1). While their investments are also significantly not in new sectors, consistent with Bonelli (2022) in that VCs adopting AI tend to avoid "breakthrough" investments, they are also in industries other than software, these investors' main domain of expertise. Together, these results provide evidence that AI-heavy investors use their AI and ML technologies to avoid investment failures and that these investments tend to be highly successful rather than simply less risky.

I construct several variables using the information available from startups' past financing rounds. I focus on three outcomes: an eventual IPO, subsequent financing ("upround"), and the presence of potentially high returns. Naturally, adding information about past financing rounds increases the precision in identifying successful startups. At early rounds like Series A and Series B, the precision of random forests in predicting IPOs is higher by a factor of nearly four when compared to randomly selecting startups or logit[7]. I also compare the F1 score[8], which balances the ability of an algorithm to make correct predictions and not to miss successful investment opportunities, finding similar results. In fact, the relative superiority of F1 score ML algorithms over random selection and logit is typically higher than that of the precision ones. Finally, I also study the linear relationship between startup outcomes and the

[6]For instance, the launch of cloud computing services by Amazon, Google, and Microsoft from 2006 to 2010, a key event in Ewens et al. (2018), Dalla Fontana (2023) and Bonelli (2022).

[7]In a different context, Fuster et al. (2022) also finds that random forests perform better than logit in credit-screening mortgage applicants.

[8]The harmonic mean between precision and recall (Sasaki (2007)).

explanatory variables. While many have a strong statistical relationship with the dependent variables, the $R^2$ is relatively low, ranging from 4 to 10 percent. The key advantage of ML is that its models can capture highly nonlinear relationships and variable interactions without them having to be pre-specified, allowing for substantially higher predictive performance. Additionally, the results unveil and quantify the importance of variables ignored in linear models. I illustrate these aspects through different predictability exercises.

In the last part of the paper, I investigate whether (and how fast) predictability declines over time. I evaluate how the incremental gains from ML vary by reproducing the exercise of predicting startup outcomes for every year of the sample under two conditions: using five and one-year rolling windows, including using a fixed sample size by picking random observations when the sample is too large. These approaches ensure that the results reproduce a setting where the information is certainly available to the investor and that far-in-the-past events like the dot-com bubble period (1999-2000) do not impact results too long after their occurrence. I find that the superior performance of ML algorithms persists under nearly all stringency conditions, along with only limited evidence that investors aggressively exploit this predictability. For instance, the performance in predicting Series A investments' outcomes flattens and slightly declines around 2015-2016. This is consistent with the idea that predictability should decline, albeit slowly, if investors are aware of and capable of exploiting it. In general, they are consistent with the hypothesis that technology-empowered investors can partially exploit, but not eliminate, performance predictability.

The main contributions of this paper are threefold. First, I show that ML algorithms can be used to make investments that are less prone to failure and with a higher likelihood of *major* success, quantifying the economic benefit their use for venture selection could potentially bring. Additionally, I provide evidence on *how* this superior performance is achieved: through the identification of nonlinearities and the importance of variables ignored by linear models, consistent with the "virtuous complex model" hypothesis of Kelly et al. (2024). In particular, I show that outperformance is possible even under *severe* data limitations and potential selection bias. In fact, ML algorithms tend to outperform precisely where selection bias and class imbalance are highest[9] – early-stage rounds – having a potentially substantial role in capital allocation efficiency. Second, I provide evidence that the *recent* use of these technologies *has* contributed to an increased performance by investors that is *not* simply the result of risk-shifting[10]: investors do make investments that fail less often while also achieving major successes more often. This result is underpinned both by the performance of AI-heavy firms, which increased in recent years, and the estimated performance of ML algorithms I provide. Together, they add to Bonelli (2022) by showing that, while in the aggregate, investors might focus on reducing risk and making less innovative or "breakthrough" investments, this behavior, and its consequent impact, seem to arise more due to *how* these technologies are used than from what they *can* achieve[11]. Third, I conciliate the literature centered on startup performance predictability (Krishna et al. (2016), Ross et al. (2021), Kiss (2022)), which usually overlaps with domains outside finance and economics, with two key issues that are often overlooked: the existence and persistence of predictability. I do that by reproducing a setting close to what a VC investor has when making investment decisions and evaluating how predictability has evolved in the face of recent technological advances.

This paper contributes primarily to three streams of literature. First, it adds to the literature focused on the use of ML and AI to assist VC investors in making better investment choices (da Silva Ribeiro Bento (2017), Krishna et al. (2016), Ross et al. (2021), Kiss (2022), Bonelli (2022)). Specifically, I add to it by evaluating and quantifying the performance of mainstream ML algorithms in identifying successful investments and showing how they perform in a realistic setting, with particular attention to forward-looking and selection bias. Second, it relates more broadly to the literature investigating the use of ML and AI technology to address financial economics problems (Gu et al. (2020), Acemoglu (2021),

[9]Before working on the sample used in this paper, I worked with a proprietary institutional dataset containing information on 2.2 million realized or potential investments, finding similar results.

[10]Achieving more successes by picking less risky investments.

[11]The view that how AI is used has far wider implications than what it can achieve is an important argument in Acemoglu (2021).

Cengiz et al. (2021), Kaniel et al. (2023), Leippold et al. (2022), Fuster et al. (2022), Obaid and Pukthuanthong (2022), DeMiguel et al. (2023), Kelly et al. (2024), Murray et al. (2024), Fausch et al. (2024)). In most cases, the consensus is that these technologies are helpful with predictions, but the improvement typically comes at a cost for some economic agents. I add to it by showing that, when directed toward identifying successful startups, their use does not seem to come at a capital allocation efficiency cost and seems not to discriminate certain types of successful investments. In particular, I show that the behavior of investors empowered with AI tools seems to be more a result of *how* they use it rather than what it can achieve, adding nuance to Bonelli (2022) and consistent with the view of Acemoglu (2021). Last, it adds to the literature on performance predictability and persistence in private equity (Kaplan and Schoar (2005), Phalippou and Gottschalg (2009), Harris et al. (2014), Korteweg and Sorensen (2017), Robinson and Sensoy (2013), Ewens and Rhodes-Kropf (2015), Nanda et al. (2020)). I add to this literature by showing that, to the extent that investors dominate certain technologies, they can improve and maintain their performance over time, with predictability taking a long time to decay.

## 2 Hypothesis Development

I investigate whether machine learning (ML) can be used to predict startup performance and its implications. A central underlying question is, therefore, whether predictability exists in the first place. One possibility is that the observed efficiency increases in classifying startups between successes and failure are simply risk-shifting: investors can use machine learning to identify less risky startups and then invest in them, but they will earn lower returns, missing "breakthrough" or highly innovative investments. A competing possibility is that there is at least some level of predictability in VC investments using machine learning, entailing two sub-hypotheses: The first is that investors can predict certain aspects of startup performance but cannot exploit them due to "structural" reasons like leverage or short sale constraints.

The second possibility is that investors can predict certain aspects and exploit them. I summarize these hypotheses below.

$H_0$: **"No Predictability"**: *Machine learning can at best classify VC investments according to risk, with lower risk arising at the expense of missing high(er) returns.*
This hypothesis implies that the predictability of investment outcomes through ML cannot be exploited to achieve better risk-adjusted returns. Under it, while apparent efficiency gains are observed through the use of these algorithms, they come at the expense of worse investment returns: ML merely allows investors to select a desired level of risk more efficiently. Ex-post, there are no incremental risk-adjusted financial returns consistent with the "zero cost" of implementing these technologies. In a related manner, many ML predictions are also impossible to implement, given the scarcity of data at the time of investment and the lack of specific or unique skills required from investors to achieve successful outcomes.

The main result of this hypothesis is that ML algorithms either do not outperform standard econometric tools or are limited to predicting successes or failures only, with no ability to predict success variability. While outperformance by ML algorithms might be observed in some periods, it unpredictably disappears under more stringent and realistic investing conditions, resulting in *de facto* no predictability.

$H_{A1}$: **"Limited Predictability"**: *Machine learning can improve the predictability of VC investments, but the strategies are impossible or too complex to execute.*
This hypothesis allows for the existence of predictability of VC investments but explains it through a "limits to arbitrage" (Shleifer and Vishny (1997)) argument: investors can predict which startups will succeed or fail but, for various reasons, are limited from exploiting this predictability. Key examples are leverage or short sale constraints (SEC (2020)), geographical distance (Bernstein et al. (2016)), along with entrepreneur preferences (Hsu (2004)). Under this hypothesis, predictability through ML is observable and persistent over time.

$H_{A2}$: **"Regular Predictability"**: *Machine learning can improve the predictability of VC investments, and investors do take advantage of it.*

This hypothesis allows for the existence of predictability of VC investments: investors can better predict which startups will succeed or fail and make active use of ML tools for it. While a certain level of predictability is observed by applying ML algorithms to screen for VC investments, this predictability flattens or decays over time as the VC activity of ML-empowered investors increases and as implementing ML strategies becomes cheaper.

# 3 Data and Methodology

## 3.1 Sample

My main data source is Crunchbase, an online database focused on venture capital and startups more broadly. I extract from it all the data available on financing rounds and investors. The sample I use covers the period from 1995 to 2020. I filter out all "non-standard" financing rounds (not classified as angel, pre-seed, seed, or Series A to J), together with rounds outside the U.S., without information on the money raised or the investors, and without U.S. dollars being the funding currency. For the main results (Section 4), I drop the last five years of the sample to mitigate truncation bias[12].

For each startup receiving financing, I identify whether it eventually did an IPO. I focus on IPOs as the main proxy for a "major success", although all the estimations can also be reproduced using acquisitions (see an example in the Online Appendix, Table 9), for which the outperformance of ML methods is equally persistent. I refrain from working with acquisitions because its correlation with success and high returns is ambiguous: acquisitions can occur because the startup is successful, because it *may* be successful (Cunningham et al. (2021)), or because it failed and had some salvage value. The resulting effect is that, in general, acquisitions are associated with lower returns (Brau et al. (2003) reports a 22 percent "premium" in an IPO vs. acquisition).

In Appendix A, I provide an example of the information available about a financing round:

[12]Startup outcomes are frequently unknown at the end of the sample.

The Series C round of Stripe was announced on July 31, 2015, with the startup receiving 100 million dollars for a pre-money valuation of 4.9 billion dollars. It had eight investors, of which two were corporations, and the rest were institutional VC firms. The startup is classified as a finance, fintech, mobile payments, and Software-as-a-Service company.

After excluding the last five years, the resulting sample contains 34,902 financing rounds, of which the most prevalent is Series A, with 9,644 rounds. The average money raised ranges from 0.2 million dollars in pre-seed rounds to 49.9 million in Series F. The fraction of rounds that contain startups that eventually IPO is 7 percent[13], and those that contain startups that are eventually acquired is 38 percent. For 53 percent of the rounds, the startup receives subsequent financing. In about 7 percent of the rounds, the startup raises ten times more money than it has raised in all previous rounds up to the current round. I use this metric as an alternative proxy for high financial returns. The underlying idea is that startups that manage to raise substantially more money in the future tend to provide a large "exit window" for their early investors and at comparatively higher valuations[14], both characteristics implying high returns. A summary of all the dataset properties is provided in Appendix B.

## 3.2 Problem Background

Consider the case of a venture capital investor presented with a representative investment opportunity with a return of $R$ in case of success and zero otherwise, which succeeds with probability $p$. The expected return differential is given by $d(pR) = pdR + Rdp$, and a simple rearrangement of terms shows that for a zero differential, that is, no change in returns, $dR = -\frac{R}{p}dp$. This simple identity illustrates that reducing investment failure rates entails increasing total returns as long as the returns in case of success do not decrease by more than a factor $\frac{R}{p}$. For instance, in my sample, the success rate (the rate of an eventual IPO) for a Series A investment is around 7 percent. Assuming a typical expected return in VC of

[13] The actual *fraction of startups* that do an IPO is lower as startups that eventually do an IPO tend to have many financing rounds, leading to a larger *fraction of rounds* of startups that do an IPO.

[14] Money raised and pre-money valuations are highly correlated, as pointed out in the Introduction.

around 25 percent (Zider (1998)), $R$ needs to be close to $\frac{0.25}{0.07} = 3.57 = 257$ percent for a 25 percent expected return. Therefore, by reducing the failure rate so that investments succeed about 8 percent of the time instead, the investor can afford to pursue investments that yield 213 percent (44 percent less than before) while maintaining its portfolio's expected return.

This simplified problem description illustrates three important facts about venture capital investments. The first is that reducing failure rates, typically easier to measure and observe than returns, has a very high impact on returns. The second is that, for this same reason, one cannot simply conclude that investors making less risky investments (hence with a lower failure rate) will earn lower *total* returns because, while that may be the case for individual investments, *total* returns can still be higher. Finally, the third is that having at least a proxy for returns is necessary to make inferences about the impact on total returns because, after all, reducing failure rates may come at the expense of a very large reduction in returns in case of success. I detail the variables I use to mitigate this issue in the next subsection.

## 3.3 Construction of variables

I define three indicator variables for the outcome of a startup. The first, $IPO$, is defined as one for a financing round provided that the startup being financed eventually goes public. $HasNext$ is equal to one if the startup has a subsequent financing round, and $HighRet10X$ if the startup raises at least ten times more money in its future rounds relative to all the previous rounds, including the current. The idea of using both $IPO$ and $HighRet10X$ as proxies for highly successful startups is to address the concern that some startups might offer very high returns for their early investors but never do an IPO, while some startups might do an IPO and not offer particularly high returns to its investors. While both groups are rare (the frequency of $HighRet10X$ and $IPO$ are 7 and 8 percent, as shown in Appendix B), they do not considerably overlap. Most $HighRet10X$ observations equal to one are concentrated in early investment stages (and, even for a firm that does an IPO, it tends to be zero for late stages), while $IPO$ equal to one tends to be concentrated in late stages.

As potential explanatory variables that help predict success, I provide two categories of variables. The first contains variables related to the *name* of the startup: name length, word count, presence of numerals, name complexity (the number of unique characters), and the presence of English words. The average startup name has 11 characters and one word; only two percent have numerals. The average name complexity is 8, and about 35 percent of names contain English words. The second category includes variables related to the startup financing history: the total number of past investors across all previous rounds[15], which averages two per startup, the total number of past lead investors, which averages less than one per startup, the total past money raised (excluding the current round), which averages 11 million dollars and the money raised in the current round, which averages 13 million dollars. Finally, I also use indicator variables for the U.S. state in which the startup is based.

A key concern when computing the variables is forward-looking bias; that is, the variables must be available for the investor when making the investment decision. In general, I attempt to replicate as closely as possible the actual data and overall information that an investor has available to conclude whether there is any observable predictability. When estimating the performance of ML models, I make tests under different stringency conditions.

Finally, the third category includes variables related to the investors' skills, experience, and reputation. The VC literature provides extensive evidence that entrepreneurs tend to prefer experienced and reputable investors (Hsu (2004)), grounded in the *belief* that they have a superior ability to add value to firms (Barry et al. (1990), Bernstein et al. (2016), Nanda et al. (2020)). The variables I use to proxy for these characteristics is the total amount of unique investments[16] that all the investors participating in a financing round made in the ten years preceding (current plus nine last years) the financing round along with the average combined money raised in all these rounds. The average number of past investments made in the preceding decade is 260, each raising 13.9 million dollars on average. I provide the

[15] An investor participating in $n$ past rounds is counted $n$ times. The idea is that successful startups will tend to have more investors, including more repeating investors, and unsuccessful ones will have fewer.

[16] That is, if a round has two investors and they participated in two financing rounds each in the past, one of them together, the unique amount of rounds is three.

summary statistics for all variables in Table 1 and all variable definitions in Appendix C.

### 3.4 The Classification Problem

The problem described in Section 3.2 is a classification problem where each investment has a binary outcome (startup does an IPO vs. does not, has high return vs. has not, etc.). Given the nature of venture capital investments described, the outcomes are frequently heavily unbalanced: most investments fail, and a few succeed. Two important concerns arise from this fact. The first is that the sample used suffers severely from selection bias, in the sense that there is typically an even larger amount of investments that fail and are not included in the sample[17]. The second is that when applying ML models, which will try to learn what characterizes each outcome, the "minority class" (the term used to denote the least frequent outcome), by being underrepresented, will be more difficult to learn about. I mitigate the first concern by showing that the ML algorithms' performance, in fact, increases when the minority class is smaller, provided that the sample is large enough: the ML algorithms used tend to perform better in early rounds, where successes are rarer. As such, mitigating selection bias by having more complete data should strengthen my results. The second concern is mitigated by applying "class weights", a way of rebalancing the importance of the characteristics of the minority class, allowing for a better learning performance.

The two main ML algorithms that I use are Random Forests (Ho (1998), Breiman (2001)) and Extreme Gradient Boosting (XGBoost) (Friedman (2001), Chen and Guestrin (2016)). The main reason for choosing these two methods in particular is availability: even though more recent methods exist, one cannot claim investors could have done better using ML by applying a method the investor could not have used because it did not exist yet. Both algorithms use "decision trees" in multiple configurations to achieve an optimal classification prediction. The first, random forests, can be thought of as a group of interconnected,

[17] Metinko (2022): "*The rate at which startups are founded is nearly impossible to keep up with, and investors are turning more to data to help manage information on the flood of new companies and opportunities.*"

simplified decision-making processes. Each "decision maker", called a tree, evaluates different aspects of the data to make predictions. Combining the insights from these multiple decision-makers enhances the accuracy and robustness of the predictions. XGBoost is an alternative and more recent algorithm specialized in learning complex patterns within data. It continually refines its predictions until its objective function is maximized (contrasted with random forests, which have no explicit objective function). I provide the specific *hyperparameters*[18] used to configure each algorithm in Appendix D. In the Online Appendix (Figure 9), I provide details on how the performance of random forests varies with the number of decision trees and how that of XGBoost varies with the depth of the decision trees.

The efficiency of each model is assessed using the precision and the F1 score. Precision[19] measures how accurately the model identifies successful investments among those it predicts to be successful. The F1 score, on the other hand, is the harmonic mean of precision and recall[20], given by $\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$. Recall assesses the model's ability to identify successful investments among all actual successful investments, serving as a measure of completeness. In particular, the recall addresses the concern that, while a model might make correct predictions, it may miss several successful investments. Combined, the F1 score encompasses both the models' ability to make the right predictions and not miss successful investments.

For the main results, I split the dataset into two windows: 1995-2011 for training and 2012-2015 for testing, so all performance metrics are *out-of-sample*, and only *past* data is used to train the models. I also verify the robustness of my results using one-year and five-year rolling windows, including a one-year window of fixed size every year[21]. When assessing whether predictability declines toward the end of the sample, I train the models with 2016-2018 data and test them with 2019-2020 data. I provide in Appendix E the number of observations available per time window per financing round.

I compare both models with two benchmarks. The first is logit. Since logit outputs

[18] Named as such because they are set *before* the model is estimated.
[19] True Positives/(True Positives + False Positives)
[20] True Positives/(True Positives + False Negatives)
[21] By always picking a fixed number $N$ of random observations from the previous year to train the models.

are continuous, a threshold must be defined to differentiate successes from failures. For each estimation, I find the threshold for which the F1 score is the highest using *test* data. This approach ensures I compare the ML models' performances with the best achievable performance for logit rather than artificially low values. The second is a simple random selection model based on population frequencies. For instance, if the success metric is an eventual IPO and the frequency of IPOs among the financing rounds tested is 5 percent, the model randomly attributes success and failures with a 5 percent frequency. This benchmark mimics an investor trying to invest only in successful ventures (as the number of investments he makes, that is, the predicted successes, is limited by the frequency of successes) but with a "placebo" selection skill: as he invests randomly, his perform is equivalent to that of an average investor. Limiting the number of investments in this benchmark allows for a fair comparison in terms of money invested with the other models and benchmarks[22]. In general, using these two benchmarks allows me to assess whether the ML algorithms are better than i) an efficient and robust econometric alternative and ii) the average investor.

# 4 Main results

## 4.1 Motivating Evidence: AI-heavy VCs' Performance

I investigate whether the performance of three "AI-heavy" firms – Meta, Alphabet, and Microsoft – distinctively improved after 2016. This cutoff year represents the period after the release of three major ML and AI libraries – PyTorch, TensorFlow, and the Microsoft Cognitive Toolkit (CNTK) – each developed by the firms just mentioned. I assume that these events represent a plausibly exogenous technological shock, marking a period of distinctive dominance of these firms in the AI and ML domains. In particular, I examine whether the

[22]For instance, an investor investing in all ventures instead would also obtain an average performance, allowing for a relative performance comparison with other alternatives. However, it is guaranteed to also invest more money in successful ventures because it invests in all ventures (and hence in all successful ones), making comparisons about money quantities difficult.

financing round of a to-be-successful startup is more likely to contain any of these firms as investors post-2016. The key hypothesis is that their capability to identify highly successful startups has increased distinctively. Using financing rounds as observations, I estimate:

$$\text{AIFirm} = \alpha + \beta_1 \text{Post2016} + \beta_2 \text{M} + \beta_3 \text{Post2016} \times \text{M} + \gamma \text{X} + \varepsilon \tag{1}$$

Where AIFirm is an indicator variable equal to one for the presence of the aforementioned "AI-heavy" firms, Post2016 is equal to one for 2016 and later years, M are selected startup characteristics, and X are control variables. The selected startup characteristics are an eventual IPO, having software as one of its categories ($Softw$), and both belonging to a category under three years old ($NewSector$) and having an eventual IPO. In this latter case, I also add the respective interactions of each individual variable with Post2016.

The coefficient $\beta_1$ measures the impact of this technological shock on a startup being financed or not by an AI-heavy firm, $\beta_2$ measures whether these selected characteristics are associated with receiving financing from them, and $\beta_3$, the main coefficient of interest, measures the differential or incremental impact of these characteristics[23] on receiving financing from an AI-heavy firm.

The results are in Table 2. In all specifications, the coefficient on Post2016 is not significant. This result highlights that there is no increased investment activity from AI-heavy firms in the period: a financing round is equally likely to have these firms listed as investors either before or after 2016. For this reason, whatever change in performance they have cannot be explained by changes in their *level* of investment activity. The coefficient on $IPO$ is negative in both the first and third specifications, suggesting that, generally, a startup that eventually does an IPO is less likely to receive investment from an AI-heavy firm. This effect changes after 2016: the coefficient on $Post2016 \times IPO$ ($\beta_3$) is statistically significant

[23] An eventual IPO is, naturally, an unobservable characteristic when the investment is made. This reduced-form model tests precisely whether this unobservable characteristic explains investment from an AI-heavy firm, evidencing that these firms can somehow identify this characteristic with a distinctive precision.

and of higher magnitude than that of $IPO$ ($\beta_2$), implying that startups that eventually do an IPO become distinctively more likely to receive investment from these firms in this period. This increase is substantial in terms of sample frequencies. Before 2016, the share of financing rounds in which the startup eventually does an IPO and these firms participate is 1.2 percent ($\frac{34}{2789}$), while from 2016 onwards, this share is 6.25 percent ($\frac{37}{592}$). I provide in Figure 1 the evolution of the success rate of these firms in the 2006-2020 period[24]. Notice that the improved performance after 2015-2016 is even higher for early investment stages, suggesting that performance improvement arises from more efficiently identifying successful startups earlier rather than increased participation in later stages (less risky) rounds[25].

I also verify whether the financed startups being software businesses or belonging to a new sector alters these results. The coefficient on $Post2016 \times IPO$ ($\beta_3$) remains significant, with the coefficient on the triple interacted term being negative in both cases. This implies that, while successful startups are more likely to receive funding from these AI-heavy firms, this effect does not persist if they are software businesses or if they belong to new sectors. This result adds nuance to Bonelli (2022) in showing that VC investors empowered with AI tools can achieve major success, although indeed by avoiding certain "breakthrough" investments (at least if these investments are interpreted as new sector startups). At the same time, these investors also avoid successful software businesses, adding evidence that their improved performance is more plausibly related to improved screening capabilities rather than their domain knowledge as software businesses themselves.

Altogether, these results are consistent with AI-heavy firms having substantially and distinctively improved their performance in recent years, coinciding with the release of the aforementioned major AI and ML libraries. Provided that these firms have been using these technologies to better screen potential investments, these findings evidence that VC investors can not only make investments that fail less often but also achieve major success more often.

[24]There are no recorded investments in the sample pre-2006 for these firms.

[25]Participating more actively in later rounds typically results from improved access to deal flow, a competitive advantage that only certain successful investors can enjoy (Nanda et al. (2020)).

## 4.2 Predictability under limited information

In this section, I evaluate the performance of random forests and XGBoost in predicting startup outcomes. The main hypothesis I test is whether ML algorithms *can* improve startup screening and by *how much*. Naturally, this hypothesis implies that VC returns have at least some predictability. In a later section, I aim to conciliate this predictability with the nearly costless availability of ML tools that allow for better investment decision-making.

Important characteristics of venture capital investments like illiquidity, small firm size (Cochrane (2005)), leverage and short sale constraints (SEC (2020)), and entrepreneurs' preferences (Hsu (2004)) impact the ability of investors to exploit predictable outcomes: even if an investor knows a certain startup will succeed, it might not be able to invest in it, and if it knows it will fail, it cannot short its stock. Together, these characteristics make the question regarding predictability existence fundamentally different than that in public markets, where it tends to be controversial for different reasons (Cochrane (2008)). Additionally, other factors like costly information acquisition (Celentano and Rempel (2020)) and irreplaceable VC investment skills (Sosyura and Ewens (2023)) can contribute to predictability regarding startup outcomes to arise and persist for prolonged periods.

A second and equally important factor regarding predictability's existence is data availability. For instance, Kiss (2022) and Krishna et al. (2016) specifically mention data limitations and few explanatory variables as limiting factors for performance ML algorithms can obtain. A consequent related problem that profoundly difficult performance evaluation is forward-looking bias, along with the associated experiment design: it is nearly impossible to know *when* data was available to the investor. For instance, da Silva Ribeiro Bento (2017) and Ross et al. (2021) achieve out-of-sample accuracy metrics on classifying startups by outcomes above 80 percent, way above those reported in this paper. Two important issues limit the interpretation of these results. The first is that they classify startups, not financing rounds. Classifying financing rounds usually ensures that all financed startups are in the same investment stage, and only information up to that stage is used, a similar context to that an investor faces when investing. Not treating for time in any way simply transforms the problem into a firm classification problem rather than a predictability exercise: the model can differentiate successes from failures, but only ex-post[26].

The second issue is that even if one looks only at financing rounds, many variables they use, like social media presence, are still potentially forward-looking. Unless one knows the exact date the startup joined a social media platform (e.g., LinkedIn), using this characteristic (and related metrics like the number of followers) to predict success is still problematic: successful startups will tend to have large follower counts, but that does not necessarily imply the count was high when the startup was seeking to raise funds.

I start my analysis by using a purportedly limited set of explanatory variables: variables derived exclusively from the *name* of the startup (see Section 3.3 for their definition). Except in cases where the startup changes its name, it is known across all rounds. I start by verifying how certain name characteristics are associated with the selected outcome variables: $HasNext$, $IPO$, and $HighRet10X$. The results are in Table 3. $NameLength$ is associated with a higher likelihood of subsequent rounds and future money raised, while the word count, the presence of numerals, and name complexity, the number of unique characters in a name, have either negative or no statistical association with the outcome. Names having an English word are positively associated with all three outcomes. While most of these characteristics are statistically significant in explaining outcomes, the $R^2$ in all three specifications is low, varying from 4.6 to 7.4 percent, implying that their power in explaining the variation across outcomes is limited.

Next, I apply Random Forests and XGBoost using only these name-related explanatory variables as predicting variables. As mentioned earlier, I use data from 1995 to 2011 for training and from 2012 to 2015 for testing, so all the results reported are out-of-sample. The underlying idea is that, given that these algorithms can capture highly nonlinear relationships, their power in predicting outcomes should be higher than that of ordinary least

[26] A predictive model should be able to differentiate, for instance, Google and a failed startup when both were young firms rather than simply differentiating them *after* their outcomes were realized.

squares. The results are in Table 4 and constitute a central result of this paper: how, even with very limited data, ML algorithms can better identify and classify to-be-successful startups, particularly where selection bias and class imbalance[27] are highest. The fact that performance is better for early-stage rounds, where successes are rarer, is evidence that mitigating selection bias should improve the performance of the ML algorithms: having more data about failed startups will increase data completeness and make successes even rarer, leading to a better gain from using ML.

In Panel A, I provide the F1 and precision scores for each financing round with a subsequent financing round ($HasNext$) as the main outcome variable. The precision rates vary from 41.5 percent with random forests at Series F rounds to 63.8 percent in pre-seed rounds with XGBoost. In all cases, at least one between random forests or XGBoost performs better than random selection, and logit performs better than both. Additionally, performance tends to decrease toward later stage rounds. I find a similar hierarchy of results concerning F1 scores. Random forests and XGBoost perform better for earlier stages, with both F1 and precision scores better than random selection. However, logit also performs better than both across all financing rounds. These results suggest that when it comes to predicting whether a startup will receive financing in the future, ML algorithms are better than the average investor but worse than existing and widespread standard econometric alternatives.

The performance of random forests and XGboost is comparatively superior when the outcomes analyzed are IPOs. While logit is still superior in terms of F1 scores in all cases, the precision of random forests and XGBoost is superior to logit from Series C and beyond investment rounds. Apart from Series F rounds, the precision of ML algorithms is also better than random selection, and for F1 scores, this is the case in all rounds. When the outcome variable is HighRet10X, the performance of ML algorithms tends to decline for later rounds, being better than random selection in all cases. As with the other outcome variables, ML performs worse than logit in terms of F1 score and better in precision.

[27] Low frequency (<10 percent) for the outcomes of interest ($IPO$ or $HighRet10X$).

Overall, my results suggest that ML algorithms can be used to identify startups that receive subsequent financing, eventually do an IPO, and have potentially high returns, even with extremely limited data. The higher F1 scores achieved by ML algorithms indicate that this improvement does not come at the expense of missing highly successful startups. Additionally, the increased efficiency is typically more pronounced in early rounds, where selection bias and class imbalance are the highest. In fact, the relative efficiency increases with the latter. Again, this result adds nuance to Bonelli (2022) in that "AI algorithms"[28] *can* assist VC investors in making less risky investments without that being at the expense of avoiding major successes. Finally, another piece of evidence regarding that is that the comparatively higher efficiency of ML algorithms arises precisely in identifying these major successes ($IPO$ and $HighRet10X$) rather than the occurrence of subsequent financing ($HasNext$).

## 4.3 Predictability with publicly-available information

Next, I seek to understand to what extent additional data can help predict startup outcomes. I add as explanatory variables the count of investors in previous rounds, including leading investors, the total money raised in previous rounds, the money raised by the startup, along with the investment count of the current investors and the average money raised in each of these investments, both over the decade preceding the investment. Notably, most of this data comes from publicly available sources or has a relatively low acquisition cost, making it plausibly available to any investor. The central hypothesis I test is whether, with this additional data, ML algorithms can "decisively" outperform the benchmarks.

The results on how these variables relate to the outcomes are shown in Table 5. The $R^2$ increases from the previous 4.6 to 7.4 percent range when using only name characteristics as explanatory variables to 6.3 to 9.8 percent, highlighting that these variables add explanatory power to the outcomes variation. Both the money raised and the count of previous lead investors are strongly significant variables in all cases. The remaining variables are strongly

[28] ML is typically understood as a subfield of AI, so here the terms are used interchangeably.

significant for at least one of the three outcomes. Finally, apart from $HasNumerals$, the name-related variables mostly maintain their statistical significance levels, suggesting that their contribution in predicting startup outcomes is uncorrelated to that of other variables.

The results for random forests, XGBoost, and logit under this new set of variables are in Table 6. Importantly, for logit, I also test the results under different regularization techniques[29] and without location (U.S. state) variables (also with regularization). The results are only marginally better, and the main results are qualitatively and quantitatively similar.

The precision in predicting a subsequent round increases (Panel A), being higher than 50 percent for all rounds (for at least one between random forests and XGBoost) except Series F. In general, both ML algorithms can produce better precision scores. The performance of logit is marginally superior to random selection and marginally inferior to the ML algorithms in all cases except in Series F rounds, where only XGBoost is better. This hierarchy changes for the F1 scores. While ML algorithms perform better than random selection, logit has better F1 scores in all cases. Given its worse precision scores, this suggests that logit is better at not missing successful investments at the cost of less accurate predictions.

In Panel B, I compare the performances of random forests and XGBoost in predicting IPOs. Apart from very early stages (pre-seed, seed, and angel), where data is frequently insufficient for estimation, the performance of both ML algorithms is vastly superior to the benchmarks in nearly all cases. For instance, in the case of Series A rounds, random forests produce both F1 and precision scores that are higher than the best of the benchmarks by a factor of nearly five ($\frac{0.249}{0.053}$ and $\frac{0.228}{0.045}$). This result again highlights that ML algorithms are helpful (and hence predictability is highest) precisely when data is scarce and selection bias and class imbalance are highest. In later rounds, several explanatory variables cease to be important, and predictability declines. I identify two potential reasons for this happening. The first is the low number of data points, as evidenced in Appendix B: the number of Series E and F rounds totals 986, vs. 9,644 Series A rounds alone. Having a low number

[29] I apply a grid search over three regularization types, L1, L2, and Elastic Net, with C = [0.01, 0.1, 1, 10, 100] and the L1 ratio (for elastic net) at 0.1, 0.5 and 0.9.

of data points makes the training and the testing of the model more difficult, as both the estimated parameters and the predictions are subject to higher errors. The second is the lower heterogeneity across the "classes" ($IPO = 0$ or 1). For later rounds, the percentage of rounds in which the venture eventually does an IPO is above 20 percent (versus 7 percent at Series A). Similarly to the case of $HasNext$, a more equal distribution between the classes tends to be associated with a lower comparative performance for the ML algorithms. To mitigate this issue, more explanatory variables are required, as evidenced by the performance improvement when adding the current explanatory variables from name-related ones only.

In Panel C, I provide the performance of ML algorithms for the outcome variable HighRet10X. Their performance is generally better than the benchmarks for the precision scores (in all cases except Series D, where logit has a 7.7% precision vs. 6.2% for random forests) and close to that of logit for F1 scores (marginally smaller in all cases except seed rounds), which is better than random selection. In the Online Appendix, I provide the average precision scores across investors from the top performance quartile in each financing round[30]. While the ML algorithms marginally underperform them in most (but not all) rounds in selecting correctly HasNext and HighRet10X outcomes, they perform better across nearly all rounds in correctly selecting IPOs: only in pre-seed, seed, and series F rounds do the top quartile investors perform better than both ML algorithms.

I extend the analysis of these results by comparing the amount of money raised by startups depending on the predictions of each model and its benchmarks. Specifically, I compare the amount of money raised correctly, that is, in startups that actually do succeed (which I refer to as "correct investments" for brevity), with the money raised incorrectly, that is, in startups that are predicted to succeed but do not ("wrong investments"). The results for Series B or earlier rounds are in Figure 5. Compared to the benchmarks, both random forests and XGBoost avoid suggesting wrong investments by a large margin, with wrong investments being close to 10B USD and 17B USD, respectively, vs. more than 25B for both benchmarks

[30]Table 8.

for the test period 2012-2015. When looking at correct investments, random forests provide a total that is slightly worse than the average investor (4.6B vs. 5.2B USD, -13%), while XGB is larger than both logit and the average investor (7.8B vs. 1.2B and 5.2B USD, respectively). Overall, the ratio between the money raised in correct and wrong investments is more than twice as large with random forests and XGBoost vs. the two benchmarks. The comparison looking only at late-stage rounds (Series C or later) is shown in Figure 6. Both ML methods, random forests and XGBoost, make correct investments that are nearly twice as large as the benchmarks (21.8B and 25.6 vs. 14.8 and 11.7B USD, respectively). Wrong investments are lower for random forests compared to the two benchmarks (16.8B vs. 31.2B and 23B USD, respectively, but larger for XGBoost when compared to the average investor (26.5B vs. 23B). Similarly as in the case for early stage investments, the ratio between correct and wrong money invested is larger by a factor of two for the two ML methods vs. the benchmarks. In aggregate, the money raised correctly when following the ML methods is around 55% larger than that of the average investor and 52% lower for wrong investments[31]. For comparison, these values correspond to around 15 and 13 billion dollars respectively, in a period where the total money raised by startups was around 160 billion dollars. Together, these figures suggest that investors following ML methods can make both better investments *and* fewer failures. I also provide in the Online Appendix[32] an extended version of Figures 5 and 6 comparing the *future* money raised by startups corresponding to correct and wrong investments. Similarly, the ML methods largely outperform the benchmarks, suggesting that they allow for better investment *returns* rather than simply a reduction in wrong investments.

Overall, the results are consistent with ML algorithms assisting investors in better identifying successful outcomes, particularly when these outcomes are rare. Performance typically (but not necessarily) improves when more explanatory variables are added. Additionally, the results highlight how the lack of observations, together with a more distributed repre-

[31] The precise numbers for random forests for instance are $\frac{\text{Random Forests Correct Early Stage + Correct Late Stage}}{\text{Avg. Investor Correct Early Stage + Correct Late Stage}} = \frac{4.55+21.81}{5.24+11.72} \approx 1.55$ and $\frac{\text{Random Forests Wrong Early Stage + Wrong Late Stage}}{\text{Avg. Investor Wrong Early Stage + Wrong Late Stage}} = \frac{8.31+16.76}{28.7+23.07} \approx 0.48$.

[32] Figures 10 and 11.

sentation between the classes, harms performance, with the performance of ML algorithms being nearly as bad as random when the outcome studied is more prevalent[33].

## 4.4 Interpretability

I study in this section *how* the superior performance of ML algorithms is achieved. In Figure 2, I provide the result of a simple exercise: predicting which startups that raised money in a Series B round in 2017 would eventually do an IPO by training all the models using only data from 2016 or earlier with a single explanatory variable, the count of previous investors.

With logit, a linear classifier, predictions follow a single decision tree: as the relationship with the outcome is negative, all predictions below a threshold are classified as successes and above as failures. The top-left figure shows the correct logit precisions, which have a precision of 6.7 percent and an F1 score of 12.6 percent. In the case of random forest and XGBoost predictions, predictions can be made across the entire explanatory variable range. Random forest makes both correct success and failure predictions across the $x = 0$ to 10 range, while XGBoost makes correct predictions up to $x = 15$. As a linear classifier, logit considers anything below $x = 20$ as a success, missing failures in this range. This ability from random forests and XGBoost to fit the data into a highly nonlinear function and make correct out-of-sample predictions is a key reason *why* it achieves a superior performance. Notice that the precision is about 67 percent higher for XGBoost than logit (11.2 vs. 6.7 percent), with an F1 score roughly 50 percent higher (18.7 vs. 12.6 percent). Together, these results highlight that ML algorithms achieve a superior performance primarily by identifying successful outcomes that linear models ignore.

I investigate next the *importance* of the explanatory variables in explaining the outcome. Differently from linear models, with random forests and XGBoost, one cannot assign a single linear relationship to each variable to understand how they contribute to the outcome. My approach is similar to that in DeMiguel et al. (2023). I compute the SHAP values

[33]This is the case for $HasNext$ in all rounds and $IPO$ and $HighRet10X$ for later rounds.

(Lundberg and Lee (2017)). This method (SHapley Additive exPlanations (SHAP)) is based on cooperative game theory (Shapley (1953)) and aims to arrive at a measure of the "fair contribution" of each explanatory variable toward the prediction of the outcome.

The results are in Figure 3. Random Forests and XGBoost can derive predictions using several explanatory variables: startup name characteristics, investor skills (InvTotal10), and location having an important contribution toward predicting successful outcomes, evidenced by the several U.S. states that appear as having a significant average impact on model output. In contrast, logit cannot identify and measure the importance of any explanatory variable beyond past and current money raised. This finding remains unchanged even after regularization techniques (the same described in Section 4.3).

Overall, my results illustrate the two main aspects explaining how ML algorithms can classify startups more efficiently than linear models: identifying and making correct predictions using highly nonlinear functions and identifying and assigning importance to a wider set of variables. Together, they achieve a higher precision in identifying success that does not come at the expense of missing successful startups. On the contrary, many correct predictions come precisely from sets of observations that linear models completely ignore.

## 4.5 Predictability decay and persistence

Next, I analyze whether predictability decays over time. First, I examine how the relative fraction of the outcomes studied and the number of data points have varied over time, shown in Figure 4. The number of financing rounds in which the startup has a subsequent round increases over time, with local peaks in 2000, 2007, and 2014-2015. The decline in 2019-2020 does not necessarily reflect a reduction in startups getting subsequent financing; it can also result from truncation bias. The number of rounds in which the startup eventually does an IPO is relatively flat and slightly declines in relative terms over time. This finding is consistent with the literature pointing to structural changes in the U.S. market that made IPOs a less attractive exit option for VC investors (Gao et al. (2013), Doidge et al. (2013), Doidge et al. (2017)). As I show next, the period of higher investment activity, namely from 2011 to 2019, coincides with the period where the improvement obtained from ML algorithms is highest. This finding reinforces the idea already discussed previously, which is that larger samples improve predictability. Additionally, it also suggests higher investment activity may provide more opportunities for investors empowered with technologies to exploit predictability to do so, as it also coincides with the period where the performance of AI-heavy firms increased[34].

I also compute the incremental F1 and precision scores obtained for each year in the sample using past five-year rolling windows. The results are in Figure 7 and Figure 8. Each point in the curve is computed as $I_{model} - I_{benchmark}$, where $I$ is either the indicated (in the subplot) algorithm F1 or precision score, depending on the figure, and the benchmark is indicated in each subplot. The values are computed only for Series A, Series B, and Series C investment rounds for two reasons. First, the data for rounds earlier than Series A and later than Series C is frequently insufficient for an estimation (I set a minimum of 100 observations for the test dataset). Second, these rounds are precisely where potential gains from using ML to predict outcomes were shown to be the highest.

The incremental F1 scores mostly increase over the years in the sample period for all algorithms, particularly for XGBoost, which tends to perform better than the alternatives. More importantly, however, is that post-2016, the "gains from ML"[35] tend to flatten or decline. A $t$-test rejects the hypothesis at the 5% significance level that the slope of the curves, representing growing gains from ML, continues to increase after 2016 in 8 out of the 12 curves plotted for the incremental F1 score (Figure 7) and in 11 out of 12 curves plotted for the incremental precision score (Figure 8). The incremental precision score results are qualitatively and quantitatively similar: from around 2010 onwards, the gains from ML increase and then flatten toward the end of the sample, with similar magnitudes for the

[34] See Section 4.1.

[35] Measured by the incremental precision and F1 obtained with Random Forests and XGBoost over logit and random selection.

average incremental score.

These findings are consistent with investors taking advantage of predictability across investments at early stages, particularly because Series A rounds tend to be where gains from predictability are the highest on average (both in these figures and when looking at the entire sample). In the Online Appendix, I reproduce this exercise using one-year rolling windows[36], first with all data available for each past year and then with a randomly selected number of observations from the past year to ensure a fixed same sample size over time. In all cases, the results are consistent with the hypothesis that predictability increases over time but flattens or declines toward the end of the sample period.

Importantly, truncation bias is also a potential reason for this observed predictability decline or flattening. I investigate this hypothesis by restricting the sample to 2016 and later only and comparing the F1 and precision scores obtained with those before 2016. The models are trained using data from 2016 to 2018 and tested on data from 2019 and 2020. In that way, the 2016 or later sample consists predominantly of observations subject to truncation bias, with the bias impacting both the train and test samples. For brevity, I only discuss the results for $IPO$ as the outcome variable, presented in Table 7. I provide the tables for $HasNext$ and $HighRet10X$ in the Online Appendix (Table 10 and Table 11), which have qualitatively and quantitatively similar properties.

The results for $IPO$ indicate that there is a considerable increase in predictability for early-stage rounds (Series A to C), with the F1 and precision scores typically increasing compared to the sample pre-2016 and the scores for logit and random selection decreasing. This finding is consistent with the incremental F1 and precision scores increasing when the "minority class" is smaller (rare outcomes), as discussed at the end of the previous section (Section 4.3). For later rounds, the incremental F1 and precision scores decline more than that for logit and random selection, suggesting that at these investment rounds, there are fewer gains in using ML algorithms, consistent with the pre-2016 sample where ML

[36] Figures 12, 13, 14 and 15.

algorithms perform worse in these investment rounds.

My results highlight that the incremental performance of ML algorithms in correctly classifying potential VC investments has improved throughout the sample period, having flattened or declined – but not disappeared – toward the end of the sample. The mixed results in terms of performance improvement (good for early-stage, bad for late-stage) in this restricted sample (2016-2020) exercise, less subject to truncation bias, suggests that this bias alone is unable to explain the overall predictability decline observed earlier. Combined with the results where the performance of AI-heavy firms is studied (Section 4.1), my results are consistent with the hypothesis that ML methods can and do provide a substantial level of improvement in the capital allocation process, albeit with a potential predictability decline. Overall, the actual use of these technologies seems limited (to date) to a small group of investors that have not yet been able to eliminate predictability entirely, evidenced by the higher F1 and precision scores in the range of 0.10-0.20 on average in all post-2010 years.

# 5 Conclusion

Machine learning (ML) algorithms substantially improve the prediction of successful startup outcomes. In general, their performance is better than logit and random selection and is not far from that of the top quartile of investors. This improved efficiency is preserved through a series of rigorous robustness checks. It arises from ML models identifying several nonlinearities and variable relationships that linear models miss, along with important predictive variables that are ignored in linear models. In particular, this improved efficiency is higher precisely when data is scarce and class imbalance is high, so investors can predict rare outcomes more efficiently.

I investigate whether firms behind the development of widely used ML and AI libraries[37] became more able to identify highly successful startups themselves. I provide evidence that these firms have significantly increased their presence in financing rounds of startups that

[37] Google, Facebook, and Microsoft.

eventually do an IPO. This effect persists after controlling for aspects that could impact the investment choice and outcome, like investor skill and local investment activity. These findings contradict previous empirical evidence suggesting that VCs adopting AI and relying on past data may become more prone to overlook highly successful investment opportunities.

The two main ML algorithms I apply are random forests and extreme gradient boosting (XGBoost), using data from 2011 or earlier to train the models and 2012-2015 to test them. Both perform considerably better than the benchmarks, particularly at early investment stages. Using only characteristics about the name of a startup, I show that these algorithms provide two to three times higher F1 and precision scores than the average investor in identifying startups that eventually do an IPO. Therefore, even when data is scarce or costly, the efficiency gains in identifying successful investments are tangible. Additional explanatory variables provide further strength to the models, increasing the F1 and precision scores by 20 to 30 percent on average. I show that an investor following the predictions of a random forests model using these additional characteristics could increase the money invested in successful startups by about 55%, while investing around 52% less in failing ones. In terms of actual money invested, this corresponds to around 15 and 13 billion dollars respectively in a period where the total money raised by startups was around 160 billion dollars.

Finally, I provide substantial evidence that the improved predictions do not arise at the expense of missing highly successful investments through several additional robustness checks. All estimations compare startups at the same investment stage and are robust to forward-looking bias, consistent with the investor available information set, and all performance metrics are out-of-sample. My investigation on whether predictability declines shows that investors have been able to exploit it: in the last years of the sample, the gains from ML, both in terms of F1 and precision scores, significantly declined in many cases. However, they remain positive, evidencing that ML methods can have an important role in improving capital allocation for prolonged periods.

## References


Daron Acemoglu. Harms of AI. Working Paper 29247, National Bureau of Economic Research, September 2021.

Christopher B. Barry, Chris J. Muscarella, John W. Peavy, and Michael R. Vetsuypens. The role of venture capital in the creation of public companies: Evidence from the going-public process. *Journal of Financial Economics*, 27(2):447–471, 1990. ISSN 0304-405X.

Shai Bernstein, Xavier Giroud, and Richard R. Townsend. The impact of venture capital monitoring. *The Journal of Finance*, 71(4):1591–1622, 2016.

Maxime Bonelli. The adoption of artificial intelligence by venture capitalists ("Data-driven Investors" in more recent versions). *Working Paper*, 2022.

James C. Brau, Bill Francis, and Ninon Kohers. The choice of IPO versus takeover: Empirical evidence. *The Journal of Business*, 76(4):583–612, 2003.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Francesco Celentano and Mark Rempel. Public listing choice with persistent hidden information. *Swiss Finance Institute Research Paper No. 23-28*, 2020.

Doruk Cengiz, Arindrajit Dube, Attila S Lindner, and David Zentler-Munro. Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes. Working Paper 28399, National Bureau of Economic Research, January 2021.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322.

John H. Cochrane. The risk and return of venture capital. *Journal of Financial Economics*, 75(1):3–52, 2005. ISSN 0304-405X.

John H. Cochrane. The dog that did not bark: A defense of return predictability. *The Review of Financial Studies*, 21(4):1533–1575, 2008. ISSN 08939454, 14657368.

Douglas J Cumming and Jeffrey G MacIntosh. A cross-country comparison of full and partial venture capital exits. *Journal of Banking Finance*, 27(3):511–548, 2003. ISSN 0378-4266. Markets and Instututions: Global perspectives.

Colleen Cunningham, Florian Ederer, and Song Ma. Killer acquisitions. *Journal of Political Economy*, 129(3):649–702, 2021. doi: 10.1086/712506.

Francisco Ramadas da Silva Ribeiro Bento. Predicting start-up success with machine learning. *Masters' Thesis, Nova Information Management School*, 2017.

Silvia Dalla Fontana. Entry and specialization in the VC industry. 2023.

Victor DeMiguel, Javier Gil-Bazo, Francisco J. Nogales, and André A.P. Santos. Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, 150(3):103737, 2023. ISSN 0304-405X.

Craig Doidge, G. Andrew Karolyi, and René M. Stulz. The U.S. left behind? financial globalization and the rise of ipos outside the u.s. *Journal of Financial Economics*, 110(3):

546–573, 2013. ISSN 0304-405X.

Craig Doidge, G. Andrew Karolyi, and René M. Stulz. The U.S. listing gap. *Journal of Financial Economics*, 123(3):464–487, 2017. ISSN 0304-405X.

Michael Ewens and Matthew Rhodes-Kropf. Is a VC partnership greater than the sum of its partners? *The Journal of Finance*, 70(3):1081–1113, 2015.

Michael Ewens, Ramana Nanda, and Matthew Rhodes-Kropf. Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, 128(3):422–442, 2018. ISSN 0304-405X.

Jürg Fausch, Moreno Frigg, Stefan Ruenzi, and Florian Weigert. Machine learning mutual fund flows. *Working Paper*, 2024.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001.

Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47, 2022.

Xiaohui Gao, Jay R. Ritter, and Zhongyan Zhu. Where have all the IPOs gone? *The Journal of Financial and Quantitative Analysis*, 48(6):1663–1692, 2013. ISSN 00221090, 17566916.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273, 02 2020. ISSN 0893-9454. doi: 10.1093/rfs/hhaa009.

Robert S. Harris, Tim Jenkinson, and Steven N. Kaplan. Private equity performance: What do we know? *The Journal of Finance*, 69(5):1851–1882, 2014. ISSN 00221082, 15406261.

Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998. doi: 10.1109/34.709601.

David H. Hsu. What do entrepreneurs pay for venture capital affiliation? *The Journal of Finance*, 59(4):1805–1844, 2004. ISSN 00221082, 15406261.

Ron Kaniel, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh. Machine-learning the skill of mutual fund managers. *Journal of Financial Economics*, 150(1):94–138, 2023. ISSN 0304-405X.

Steven N. Kaplan and Antoinette Schoar. Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance*, 60(4):1791–1823, 2005. ISSN 00221082, 15406261.

Bryan Kelly, Semyon Malamud, and Kangying Zhou. The virtue of complexity in return prediction. *The Journal of Finance*, 79(1):459–503, 2024. doi: https://doi.org/10.1111/jofi.13298.

Barnabás Kiss. Using big data in startup selection: Exploring machine learning as a tool to predict successful startups in the age of social media. *Masters' Thesis, Católica Lisbon School of Business  Economics*, 2022.

Arthur Korteweg and Morten Sorensen. Skill and luck in private equity performance. *Journal* *of Financial Economics*, 124(3):535–562, 2017. ISSN 0304-405X.

Amar Krishna, Ankit Agrawal, and Alok Choudhary. Predicting the outcome of startups: Less failure, more success. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 798–805, 2016. doi: 10.1109/ICDMW.2016.0118.

Markus Leippold, Qian Wang, and Wenyu Zhou. Machine learning in the chinese stock market. *Journal of Financial Economics*, 145(2, Part A):64–82, 2022. ISSN 0304-405X.

Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

Chris Metinko. As venture capital continues to change, what will 2022 bring? *Crunchbase News*, 2022.

Scott Murray, Yusen Xia, and Houping Xiao. Charting by machines. *Journal of Financial Economics*, 153:103791, 2024. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2024.103791.

Ramana Nanda, Sampsa Samila, and Olav Sorenson. The persistent effect of initial success: Evidence from venture capital. *Journal of Financial Economics*, 137(1):231–248, 2020. ISSN 0304-405X.

Khaled Obaid and Kuntara Pukthuanthong. A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, 144(1):273–297, 2022. ISSN 0304-405X.

Ludovic Phalippou and Oliver Gottschalg. The performance of private equity funds. *The Review of Financial Studies*, 22(4):1747–1776, 2009. ISSN 08939454, 14657368.

David T. Robinson and Berk A. Sensoy. Do Private Equity Fund Managers Earn Their Fees? Compensation, Ownership, and Cash Flow Performance. *The Review of Financial Studies*, 26(11):2760–2797, 11 2013. ISSN 0893-9454.

Greg Ross, Sanjiv Das, Daniel Sciro, and Hussain Raza. Capitalvx: A machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science*, 7:94–114, 2021. ISSN 2405-9188.

Yutaka Sasaki. The truth of the F-measure. *Lecture Notes*, 2007.

U.S. SEC. Modernizing the SEC's definition of venture capital fund. *U.S. Securities and Exchange Commission*, 2020.

L. Shapley. A value for n-person games. In H. Kuhn and A. Tucker, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton, 1953. doi: 10.1515/9781400881970-018.

Andrei Shleifer and Robert W. Vishny. The limits of arbitrage. *The Journal of Finance*, 52 (1):35–55, 1997. ISSN 00221082, 15406261.

Denis Sosyura and Michael Ewens. Irreplaceable venture capitalists. *Working Paper*, 2023.

Bob Zider. How venture capital works*. *Harvard Business Review*, 1998.

Figure 1: Evolution of the Success Rate of AI-heavy Firms (2006-2020)

This figure shows how the relative presence of three firms – Google, Facebook, and Microsoft – among the investors in financing rounds of startups that eventually do an IPO varies in the period. The years 2015 and 2016 mark a period where these three firms released AI and ML libraries that considerably impacted the cost of implementing these tools in any sort of application, as they are free for anyone to use.

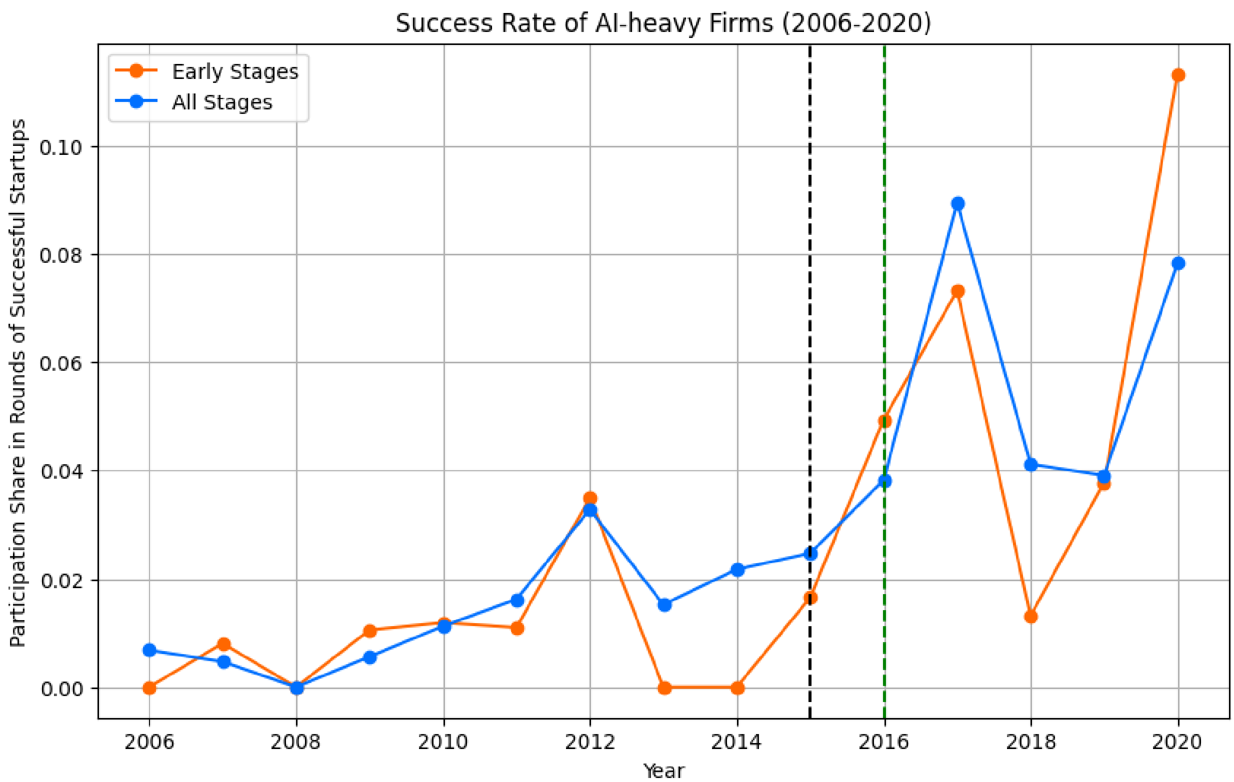


Figure 2: Identifying Non-Linearities with Random Forests and XGBoost

These figures illustrate how random forests and XGBoost differ from a linear-in-parameters model like logit. In each figure, I show correct positive and negative predictions for IPO as a startup outcome ($IPO = 1$ or $IPO = 0$), using a single "feature" (explanatory variable): the count of investors in previous financing rounds. This exercise uses information exclusively on Series B rounds before 2017 to predict which startups raising for Series B rounds in 2017 will eventually do an IPO. Notice that while logit classifies as success or failure depending on a single threshold, random forests and XGBoost can identify correct predictions across the entire range of variable values, achieving higher precision and F1 scores.

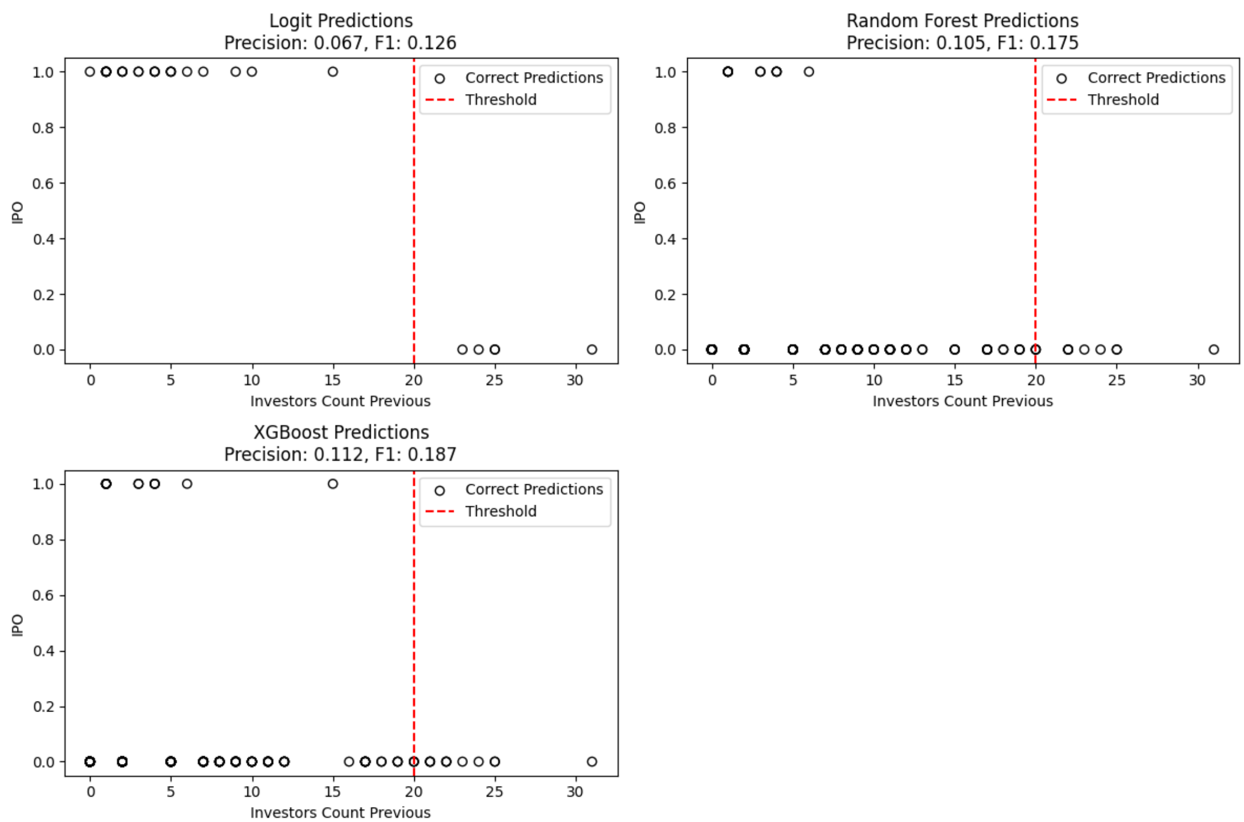


Figure 3: Characteristic Importance in a Prediction Exercise

This figure shows the mean Shapley value (Shapley (1953), Lundberg and Lee (2017)) for each characteristic used in the prediction exercise. The idea is to understand how much each feature contributes to the magnitude of the different models' predictions. This exercise uses all Series B rounds information available before 2017 to predict which startups raising for Series B rounds in 2017 will eventually do an IPO. The hyperparameter configuration is identical to that described in Appendix D.

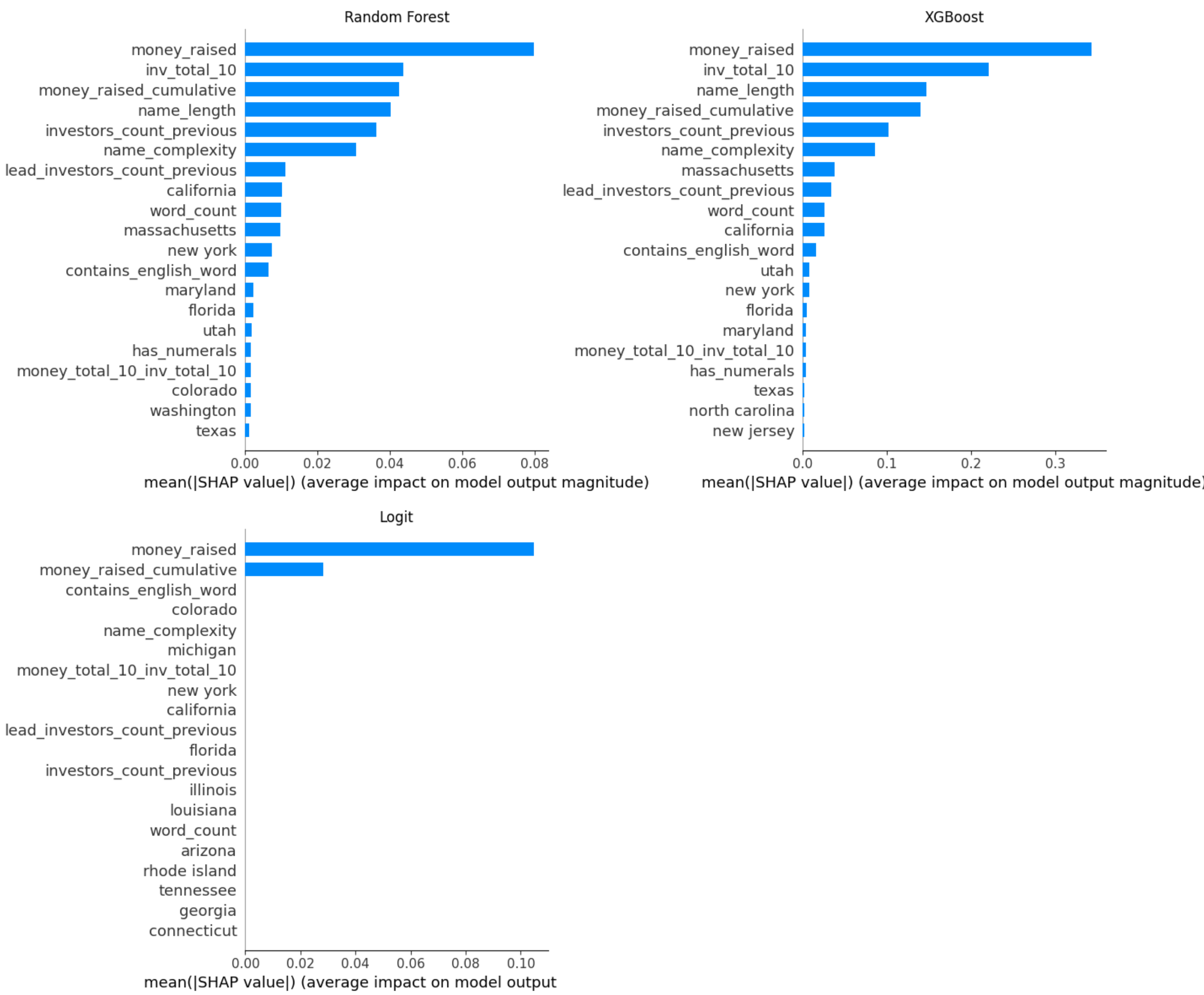


Figure 4: Startup Outcomes Distribution (1999-2020)

This figure shows the total count and relative fractions of the indicated startup outcomes: IPO, $HasNext$ (occurrence of a subsequent financing round), and ten times more future money raised than past and present money raised ($HighRet10X$), for each financing round and year in the sample. The last few years of the sample are highly affected by truncation bias, as for many of the financing rounds, the outcomes of the startups are still undefined or were not known when the data was collected (mid-2021).

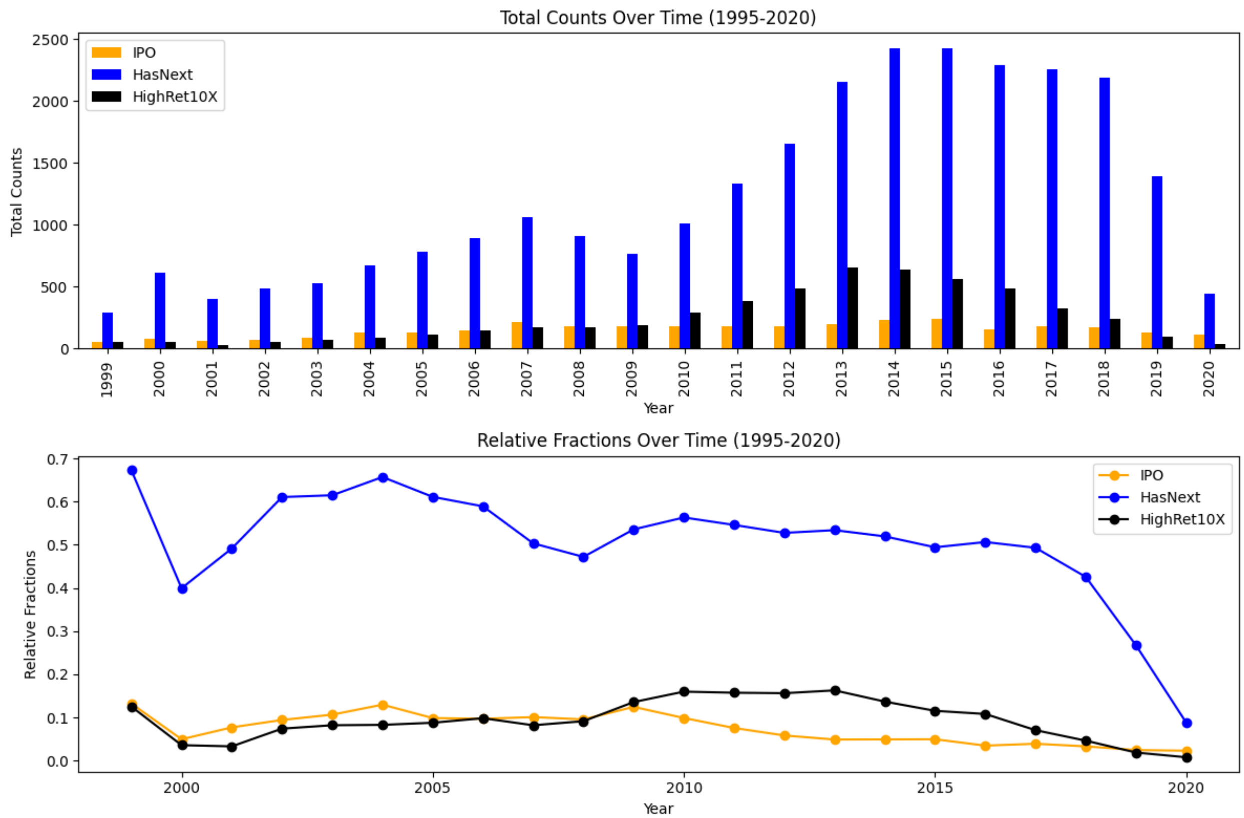


Figure 5: Comparative Money Raised By Model and Prediction Correctness (Early Stage)

The left figure shows the total money raised across early-stage rounds whose predictions of success were either correct or wrong, with success defined as an eventual IPO, by model: random forests (RF), XGBBoost (XGB), logit (LR), and random selection (Random (avg. investor)), in the test sample (2012-2015). These amounts correspond to the aggregate money raised for each category of rounds (with correct success predictions and with wrong success predictions). I define early-stage rounds as pre-seed, seed, angel, Series A, and Series B. A round with a correct success prediction is defined by a prediction of an eventual IPO that materializes, while a round with a wrong success prediction is defined by a non-materialized prediction of an eventual IPO. The right figure shows the ratios between the correct and wrong success prediction monetary amounts by model.

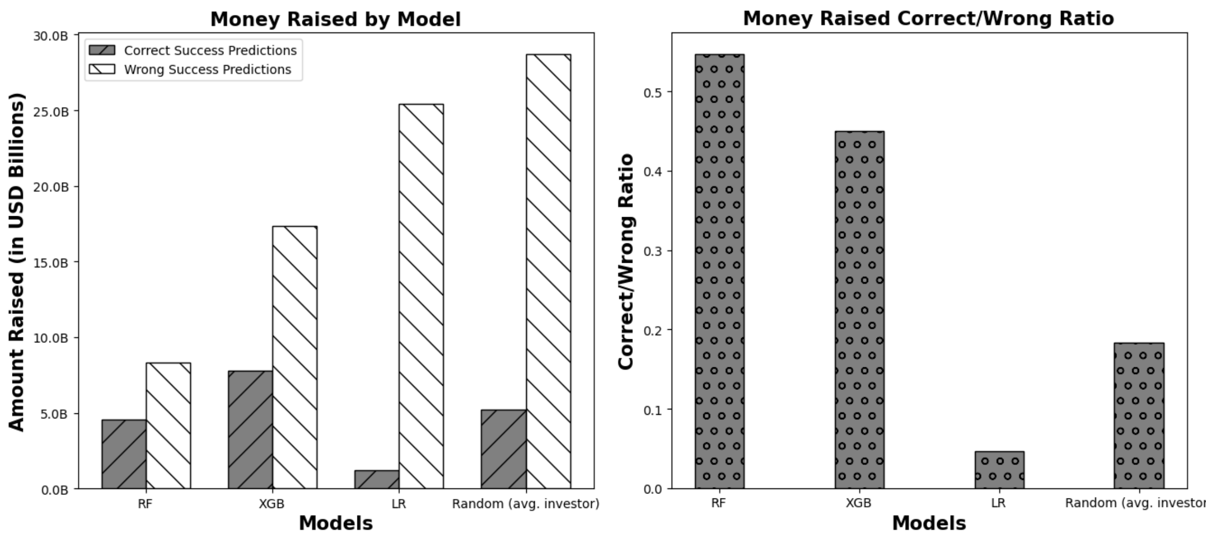


Figure 6: Comparative Money Raised By Model and Prediction Correctness (Late Stage)

The left figure shows the total money raised across late-stage rounds whose predictions of success were either correct or wrong, with success defined as an eventual IPO, by model: random forests (RF), XGBBoost (XGB), logit (LR), and random selection (Random (avg. investor)), in the test sample (2012-2015). These amounts correspond to the aggregate money raised for each category of rounds (with correct success predictions and with wrong success predictions). I define late-stage rounds as Series C or later rounds. A round with a correct success prediction is defined by a prediction of an eventual IPO that materializes, while a round with a wrong success prediction is defined by a non-materialized prediction of an eventual IPO. The right figure shows the ratios between the correct and wrong success prediction monetary amounts by model.

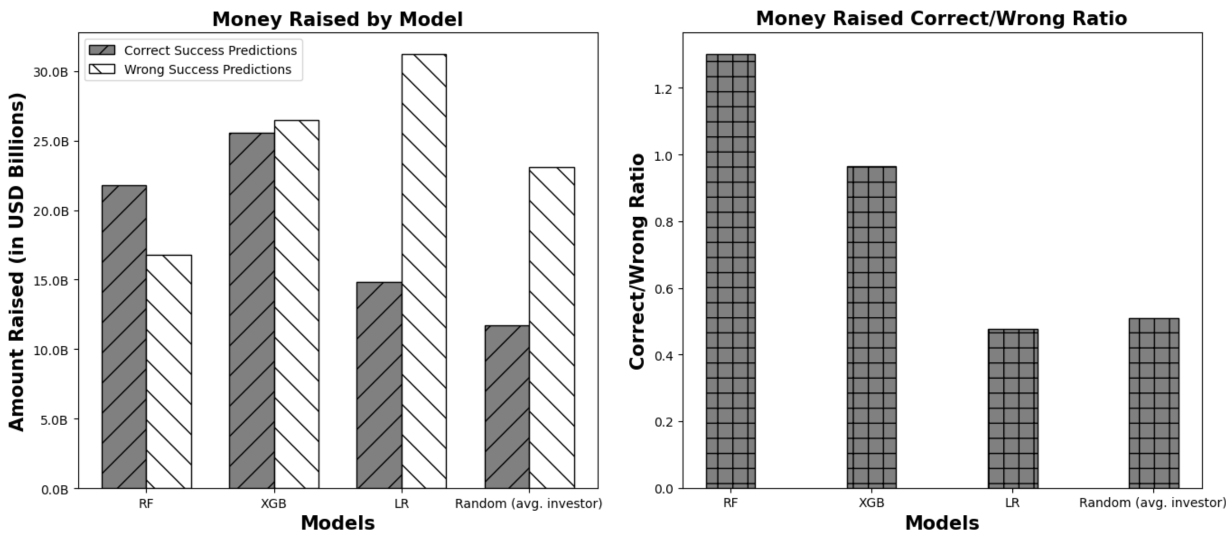


Figure 7: Incremental F1 Score Evolution by Model and Financing Round

This figure shows how each machine learning (ML) algorithm performs regarding their resulting F1 score when predicting IPOs compared to random selection and logistic regression, using a sample limited to data available within the five years preceding each year indicated in the x-axis. For instance, for the year 2005, only data from 2000 to 2004 is used to train the model, and then it is tested using 2005 data so that all results are always out-of-sample. Each curve represents the difference between the F1 score obtained by the indicated algorithm (Random Forest of XGBoost) and the benchmarks (random selection and logistic regression). The percentages correspond to the fraction of years the algorithm is strictly better than the benchmark and to the fraction of years in which the algorithm is at least as good as the benchmark. The average indicated refers to the average incremental F1 score in the entire period (2000-2020).

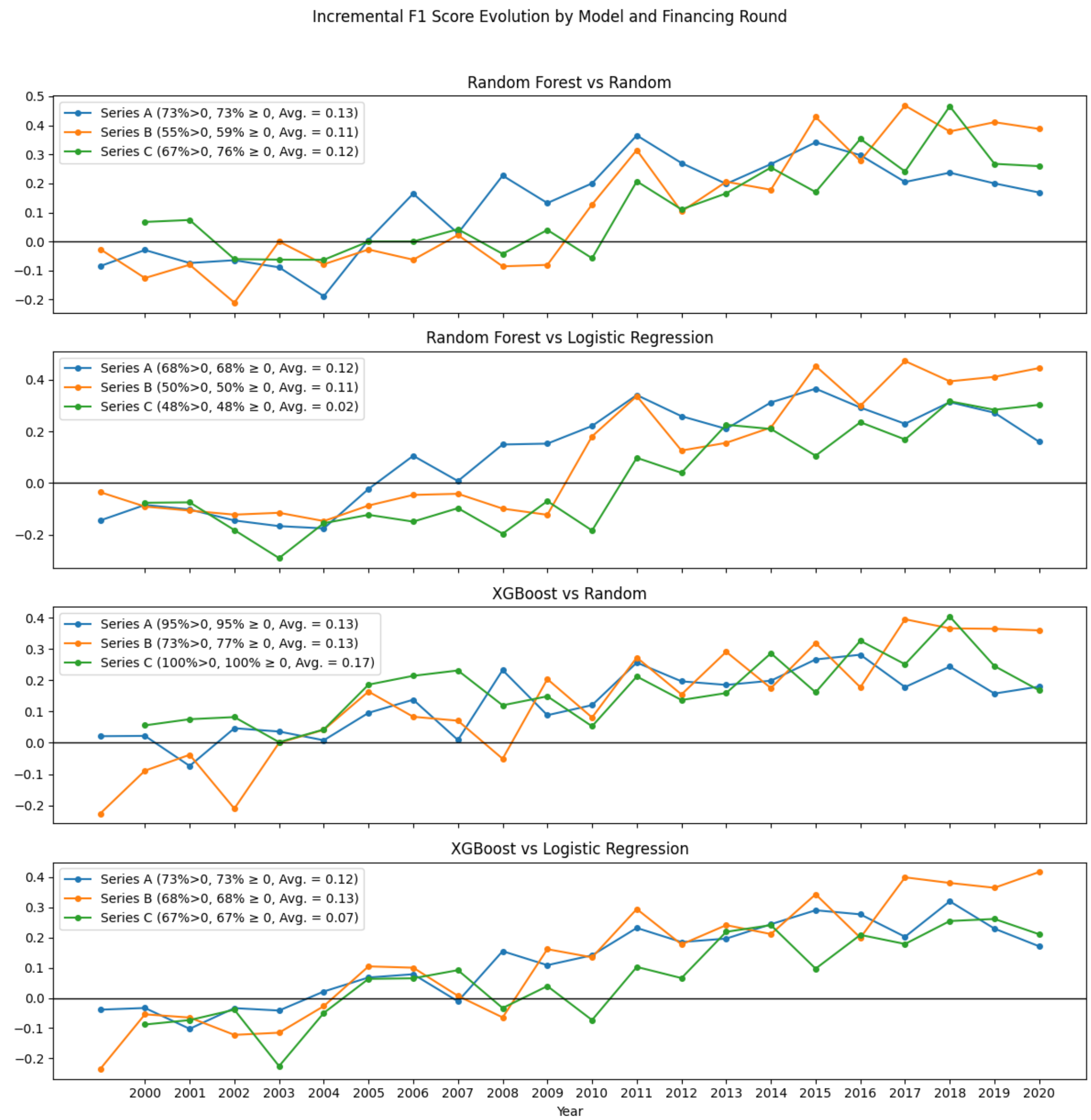


Figure 8: Incremental Precision Score Evolution by Model and Financing Round

This figure shows how each machine learning (ML) algorithm performs regarding their resulting precision score when predicting IPOs compared to random selection and logistic regression, using a sample limited to data available within the five years preceding each year indicated in the x-axis. For instance, for the year 2005, only data from 2000 to 2004 is used to train the model, and then it is tested using 2005 data so that all results are always out-of-sample. Each curve represents the difference between the precision score obtained by the indicated algorithm (Random Forest of XGBoost) and the benchmarks (random selection and logistic regression). The percentages indicate the fraction of years the algorithm is strictly better than the benchmark and the fraction of years the algorithm is at least as good as the benchmark. The average indicated refers to the average incremental precision score in the entire period (2000-2020).

Appendix A: Financing Round Data Example

| Entry | Value |
|---|---|
| Venture Name | Stripe |
| Investment Round | Series C |
| Investors | American Express, FJ Labs, Kleiner Perkins, Paua Ventures, Playfair Capital, Sequoia Capital, Square Peg Capital, Visa |
| Date | 31.07.2015 |
| Funding Currency | USD |
| Location | San Francisco, California, United States |
| Money Raised (USD millions) | 100 |
| Pre-Money Valuation (USD millions) | 4,900 |
| Sectors | Finance, FinTech, Mobile Payments, SaaS |

Appendix B: Dataset Summary (1995-2015)

This table summarizes the main properties of the dataset. The dataset is constructed using data from financing rounds and investors from Crunchbase. I drop all “non-standard” financing rounds (not classified as angel, pre-seed, seed, or Series A to J), together with financing rounds not taking place in the U.S., without information on the money raised or investors, and without USD being the funding currency. For all estimations, financing rounds of the last five years of the sample (2016-2020) are not included to mitigate truncation bias.

| Round | Quantity | Avg. Money Raised (Millions) | IPOs | Acquisitions | HasNext | HighReturn10X |
|---|---|---|---|---|---|---|
| Pre-Seed | 530 | 0.2 | 9 (2%) | 83 (16%) | 327 (62%) | 120 (23%) |
| Seed | 9,920 | 1.0 | 112 (1%) | 2,427 (24%) | 4,743 (48%) | 1,075 (11%) |
| Angel | 1,101 | 0.8 | 16 (1%) | 245 (22%) | 530 (48%) | 130 (12%) |
| Series A | 9,644 | 8.0 | 631 (7%) | 3,958 (41%) | 5,878 (61%) | 886 (9%) |
| Series B | 7,008 | 15.4 | 688 (10%) | 3,279 (47%) | 3,931 (56%) | 190 (3%) |
| Series C | 3,885 | 21.1 | 524 (13%) | 1,933 (50%) | 1,961 (50%) | 36 (1%) |
| Series D | 1,754 | 28.2 | 337 (19%) | 859 (49%) | 781 (45%) | 4 (0%) |
| Series E | 730 | 41.9 | 197 (27%) | 343 (47%) | 288 (39%) | 1 (0%) |
| Series F | 256 | 49.9 | 85 (33%) | 99 (39%) | 83 (32%) | 0 (0%) |
| Total | 34,828 | 10.65 | 2,624 (7%) | 13,252 (38%) | 18,546 (53%) | 2,442 (7%) |

## Appendix C: Variable Definitions

Each variable is calculated based on the information available for each financing round.

| Variable | Description |
|---|---|
| Investors Count Previous | The number of investors that participated in all past financing rounds of a startup. This variable does not include investors in the current financing round. |
| Lead Investors Count Previous | The number of lead investors that participated in all past financing rounds of a startup. This variable does not include investors in the current financing round. |
| InvTotal10 | The total amount of investments (in thousands) the investors participating in a financing round made in the last ten years (current year plus previous nine). |
| MoneyPerAvgInvTotal10 | The average money raised across all investments the investors participating in a financing round made in the last ten years (current year plus previous nine). |
| LogAvgIPO | The total share of investments the investors participating in a financing round made in the last ten years (current year plus previous nine) in which the startup eventually did an IPO. This variable is used as a control only in Table 2. I avoid it as a control in the prediction results to mitigate forward-looking bias. |
| Money Raised Cumulative | The amount of money the startup raised in all past financing rounds. This variable does not include money raised in the current financing round. |
| HasNext, IPO | Indicator variables equal to one if the startup has a subsequent financing round ($HasNext$) or eventually does an IPO ($IPO$). |
| HighRet10X | The ratio between the future money raised by a startup across every future financing round over the past money raised across every past financing round plus the current round, that is $\frac{future}{past+current}$. |
| Has Numerals | Indicator variables equal to if the startup name has numerals. |
| Name Length | Startup name length in characters. |
| Name Complexity | The amount of unique characters in a startup name. |
| Word Count | The number of words a startup name contains. |
| $AIFirm$ | An indicator variable equal to one for Google, Microsoft, and Facebook, including its corporate venture capital entities. |
| $NewSector$ | The count of how many categories among the "company categories" a startup being financed belongs to that is a "new" sector. I define a "new sector" as any sector whose first financing round of a startup belonging to it was later than two years before the year of the financing round of the startup. |
| $Post2016$ | An indicator variable equal to one for 2016 and later years. |
| $Softw$ | An indicator variable equal to one if the startup being financed has "software" among any of its categories. |

Appendix D: Machine Learning Algorithms Hyperparameters

I provide in this table the main details required to replicate the machine learning algorithms used in this paper. The original code, written in Python, can be provided by the author upon request. The specific libraries used were RandomForestClassifier and xgboost.

| | |
|---|---|
| **Random Forests** | |
| n_estimators | 5000 |
| class_weight | {0: class_weights[0], 1: class_weights[1]} |
| max_depth | None |
| min_samples_split | 4 |
| min_samples_leaf | 2 |
| max_features | 'auto' |
| bootstrap | True |
| criterion | 'gini' |
| random_state | 12 |
| **Extreme Gradient Boosting** | |
| max_depth | 500 |
| min_child_weight | 5 |
| gamma | 1.5 |
| subsample | 0.8 |
| colsample_bytree | 0.8 |
| learning_rate | 0.00025 |
| colsample_bytree | 0.8 |
| objective | binary:logistic' |
| n_estimators | 5000 |
| random_state | 12 |
| scale_pos_weight | len(y_train[y_train == 0]) / len(y_train[y_train == 1]) |

Appendix E: Train and Test Datasets Length

This table shows the number of observations available for each train and test dataset for each financing round.

| Period<br>Round | 1995-2011<br>$N_{\text{train}}$ | 2012-2015<br>$N_{\text{test}}$ | 2016-2018<br>$N_{\text{train}}$ | 2019-2020<br>$N_{\text{test}}$ |
|---|---|---|---|---|
| Pre Seed | 105 | 392 | 1024 | 1254 |
| Seed | 1922 | 6353 | 4577 | 2880 |
| Angel | 417 | 593 | 300 | 81 |
| Series A | 5621 | 3298 | 2923 | 2022 |
| Series B | 4598 | 1835 | 1627 | 1120 |
| Series C | 2588 | 983 | 800 | 604 |
| Series D | 1119 | 512 | 387 | 300 |
| Series E | 430 | 252 | 166 | 152 |
| Series F | 143 | 100 | 71 | 56 |

Table 1: Summary Statistics

This table provides summary statistics for the main variables used in the paper. See Variables Definitions (Appendix C) for their detailed description. For the main section, I report the summary statistics based on the sample without the last five years (2016, 2017, 2018, 2019, and 2020). The reason is that the sample without these years is used for all the results reported, apart from Table 2.

| **Main Section Variables (1995-2015)** | Mean | Std. Dev. | p(25) | p(50) | p(75) |
|---|---|---|---|---|---|
| Investors Count Previous | 2.40 | 4.50 | 0.00 | 0.00 | 3.00 |
| Money Raised Cumulative (In Millions) | 11.00 | 70.00 | 0.00 | 0.00 | 5.00 |
| Lead Investors Count Previous | 0.42 | 0.72 | 0.00 | 0.00 | 1.00 |
| InvTotal10 (In Thousands) | 0.26 | 0.62 | 0.01 | 0.08 | 0.31 |
| MoneyPerAvgInvTotal10 (In Millions) | 13.90 | 12.60 | 4.80 | 13.70 | 19.70 |
| Money Raised (In Millions) | 13.00 | 44.00 | 1.00 | 4.10 | 12.00 |
| Name Length | 11.00 | 9.10 | 7.00 | 9.00 | 13.00 |
| Word Count | 1.40 | 0.68 | 1.00 | 1.00 | 2.00 |
| Has Numerals | 0.02 | 0.15 | 0.00 | 0.00 | 0.00 |
| Name Complexity | 8.80 | 3.60 | 6.00 | 8.00 | 11.00 |
| Contains English Word | 0.35 | 0.48 | 0.00 | 0.00 | 1.00 |
| IPO | 0.06 | 0.23 | 0.00 | 0.00 | 0.00 |
| HasNext | 0.46 | 0.50 | 0.00 | 0.00 | 1.00 |
| HighRet10X | 0.09 | 0.29 | 0.00 | 0.00 | 0.00 |
| Observations | 38,718 | | | | |
| **Regression Table 2 (1995-2020)** | | | | | |
| $AIFirm$ | .012 | .11 | 0 | 0 | 0 |
| $Post2016$ | .41 | .49 | 0 | 0 | 1 |
| $NewSector$ | .063 | .32 | 0 | 0 | 0 |
| $IPO$ | .057 | .23 | 0 | 0 | 0 |
| $Softw$ | .33 | .47 | 0 | 0 | 1 |
| Observations | 54,559 | | | | |

Table 2: Financing Round Characteristics Explaining Investment From "AI-heavy" Investors

This table shows the results for ordinary OLS regressions studying whether certain financing rounds and startup characteristics can explain the startup receiving investment from certain "AI-heavy" investors. The dependent variable $AIFirm$, is equal to one if among the investors in a financing round Google, Microsoft, Facebook, or any of its corporate venture capital entities are present. The explanatory variables are indicator variables for software startups ($Softw$), 2016 or later rounds ($Post2016$), rounds of startups that eventually do an IPO ($IPO$), and rounds of startups in categories that are less than three years old ($NewSector$). The controls are detailed in Section 3.3 and Appendix C. I apply U.S. state, investment round, and investment round-U.S. state fixed effects in all specifications. Standard errors are clustered at the state level in all regressions. t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

| | (1) $AIFirm$ | (2) $AIFirm$ | (3) $AIFirm$ |
|---|---|---|---|
| $Post2016$ | -0.00156 | -0.00211 | -0.00130 |
| | (-0.67) | (-0.80) | (-0.59) |
| $IPO$ | -0.00731*** | -0.00610 | -0.00814*** |
| | (-2.72) | (-1.59) | (-2.71) |
| $Post2016 \times IPO$ | 0.0363** | 0.0453*** | 0.0374** |
| | (2.48) | (3.46) | (2.59) |
| $Softw$ | | 0.000359 | |
| | | (0.24) | |
| $Post2016 \times Softw$ | | 0.00148 | |
| | | (0.72) | |
| $Softw \times IPO$ | | -0.00458 | |
| | | (-0.98) | |
| $Post2016 \times Softw \times IPO$ | | -0.0659*** | |
| | | (-6.16) | |
| $NewSector$ | | | 0.00186 |
| | | | (1.52) |
| $Post2016 \times NewSector$ | | | -0.00770*** |
| | | | (-2.87) |
| $NewSector \times IPO$ | | | 0.00673** |
| | | | (2.51) |
| $Post2016 \times NewSector \times IPO$ | | | -0.0658*** |
| | | | (-4.69) |
| Control Variables | Yes | Yes | Yes |
| State, Round, State × Round FE | Yes | Yes | Yes |
| $N$ | 54559 | 54559 | 54559 |
| $R^2$ | 0.0309 | 0.0314 | 0.0310 |

$t$ statistics in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 3: Startup Outcome Relationship With Startup Name Characteristics

This table shows the results for ordinary OLS regressions having startup name characteristics as explanatory variables for startup outcomes: an "upround" ($HasNext$), an indicator variable equal to one if the startup has a subsequent financing round, an eventual IPO ($IPO$), and ten times more future money raised than past and present money raised ($HighRet10X$). As explanatory variables, I use only characteristics related to the name of the startup: name length, word count, presence of numerals, presence of English words, and name complexity. I apply U.S. state, investment round, and U.S. state-investment round fixed effects and clustered standard errors at the state level in all regressions. t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

| | (1) $HasNext$ | (2) $IPO$ | (3) $HighRet10X$ |
|---|---|---|---|
| $NameLength$ | 0.000986** | 0.000154 | 0.000354*** |
| | (2.59) | (0.70) | (2.68) |
| $WordCount$ | -0.0552*** | -0.0249** | -0.0218*** |
| | (-6.31) | (-2.45) | (-6.57) |
| $HasNumerals$ | -0.0298* | 0.0127 | -0.00159 |
| | (-1.80) | (0.88) | (-0.19) |
| $NameComplexity$ | -0.00580** | 0.00390 | -0.000602 |
| | (-2.55) | (1.13) | (-0.85) |
| $ContainsEnglishWord$ | 0.0777*** | 0.0137* | 0.0199*** |
| | (11.17) | (1.95) | (10.30) |
| $N$ | 34871 | 34871 | 34871 |
| $R^2$ | 0.0457 | 0.0744 | 0.0468 |

$t$ statistics in parentheses

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

Table 4: Startup Outcome Prediction Using Only Startup Name Characteristics

This table compares the F1 and precision scores across four methods (Random Forests, Logit, XGBoost, and Random) concerning their ability to correctly predict an "upround" ($HasNext$, Panel A), an indicator variable equal to one if the startup has a subsequent financing round, and an eventual IPO ($IPO$, Panel B). As explanatory variables, I use only characteristics related to the name of the startup: name length, word count, presence of numerals, presence of English words, and name complexity. The models are trained using data from 1995 to 2011 and tested on data from 2012 to 2015.

**Panel A: HasNext**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.649 | 0.634 | 0.757 | 0.567 | 0.631 | 0.638 | 0.609 | 0.579 |
| Seed | 0.517 | 0.542 | 0.617 | 0.459 | 0.482 | 0.487 | 0.447 | 0.435 |
| Angel | 0.498 | 0.514 | 0.635 | 0.481 | 0.471 | 0.474 | 0.467 | 0.466 |
| Series A | 0.612 | 0.631 | 0.734 | 0.606 | 0.607 | 0.613 | 0.580 | 0.583 |
| Series B | 0.634 | 0.602 | 0.726 | 0.554 | 0.594 | 0.592 | 0.571 | 0.565 |
| Series C | 0.595 | 0.605 | 0.700 | 0.522 | 0.560 | 0.576 | 0.539 | 0.565 |
| Series D | 0.463 | 0.520 | 0.668 | 0.469 | 0.479 | 0.507 | 0.502 | 0.538 |
| Series E | 0.524 | 0.463 | 0.645 | 0.394 | 0.546 | 0.473 | 0.489 | 0.500 |
| Series F | 0.420 | 0.560 | 0.584 | 0.281 | 0.415 | 0.467 | 0.412 | 0.375 |

**Panel B: IPO**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Seed | 0.016 | 0.018 | 0.000 | 0.010 | 0.008 | 0.009 | 0.000 | 0.007 |
| Angel | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series A | 0.109 | 0.122 | 0.123 | 0.034 | 0.066 | 0.071 | 0.200 | 0.028 |
| Series B | 0.187 | 0.199 | 0.213 | 0.096 | 0.119 | 0.122 | 0.173 | 0.097 |
| Series C | 0.234 | 0.250 | 0.249 | 0.133 | 0.158 | 0.168 | 0.147 | 0.154 |
| Series D | 0.262 | 0.297 | 0.344 | 0.184 | 0.208 | 0.229 | 0.208 | 0.186 |
| Series E | 0.400 | 0.406 | 0.497 | 0.233 | 0.398 | 0.331 | 0.331 | 0.319 |
| Series F | 0.472 | 0.541 | 0.619 | 0.267 | 0.457 | 0.441 | 0.448 | 0.471 |

**Panel C: HighRet10X**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.503 | 0.539 | 0.591 | 0.413 | 0.458 | 0.440 | 0.431 | 0.392 |
| Seed | 0.325 | 0.329 | 0.345 | 0.222 | 0.225 | 0.225 | 0.212 | 0.194 |
| Angel | 0.250 | 0.248 | 0.285 | 0.195 | 0.194 | 0.181 | 0.168 | 0.165 |
| Series A | 0.239 | 0.253 | 0.258 | 0.144 | 0.151 | 0.154 | 0.155 | 0.137 |
| Series B | 0.117 | 0.112 | 0.071 | 0.021 | 0.067 | 0.063 | 0.078 | 0.024 |
| Series C | 0.043 | 0.071 | 0.000 | 0.000 | 0.024 | 0.039 | 0.000 | 0.000 |
| Series D | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series E | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series F | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 5: Startup Outcome Relationship With Name and Financing Rounds Characteristics

This table shows the results for ordinary OLS regressions having startup name together with past financing rounds characteristics as explanatory variables for startup outcomes: an "upround" ($HasNext$), an indicator variable equal to one if the startup has a subsequent financing round, an eventual IPO ($IPO$), and ten times more future money raised than past and present money raised ($HighRet10X$). As explanatory variables, I use characteristics related to the name of the startup: name length, word count, presence of numerals, presence of English words, and name complexity, along with variables related to the past financing history of the startup. The precise definition of these additional explanatory variables is provided in Appendix B. I apply U.S. state, investment round, and U.S. state-investment round fixed effects and clustered standard errors at the state level in all regressions. t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

| | (1)<br>$HasNext$ | (2)<br>$IPO$ | (3)<br>$HighRet10X$ |
|---|---|---|---|
| $NameLength$ | 0.000804** | 0.000119 | 0.000396*** |
| | (2.15) | (0.58) | (2.76) |
| $WordCount$ | -0.0467*** | -0.0242** | -0.0233*** |
| | (-5.38) | (-2.58) | (-7.23) |
| $HasNumerals$ | -0.0259 | 0.0148 | -0.00269 |
| | (-1.56) | (1.04) | (-0.31) |
| $NameComplexity$ | -0.00436** | 0.00430 | -0.00102 |
| | (-2.32) | (1.33) | (-1.29) |
| $ContainsEnglishWord$ | 0.0641*** | 0.0115* | 0.0226*** |
| | (11.02) | (1.71) | (12.00) |
| $InvestorsCountPrevious$ | 0.00611*** | 0.0000393 | -0.00261*** |
| | (2.99) | (0.08) | (-6.31) |
| $MoneyRaisedCumulative$ (in billions) | 0.0470 | 0.0943*** | 0.116*** |
| | (1.16) | (3.54) | (3.69) |
| $LeadInvestorsCountPrevious$ | 0.0183*** | 0.0190*** | -0.00911*** |
| | (4.73) | (4.37) | (-10.87) |
| $InvTotal10$ | 0.0716*** | 0.00773 | -0.0137*** |
| | (8.87) | (1.24) | (-5.93) |
| $MoneyPerAvgInvTotal10$ | 0.166*** | -0.00430 | -0.00421 |
| | (23.57) | (-1.39) | (-0.93) |
| $MoneyRaised$ | 0.205*** | 1.376*** | -0.239*** |
| | (2.89) | (5.77) | (-3.33) |
| $N$ | 34871 | 34871 | 34871 |
| $R^2$ | 0.0634 | 0.0984 | 0.0499 |

$t$ statistics in parentheses

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

Table 6: Startup Outcome Prediction Using Multiple Characteristics

This table compares the F1 and precision scores across four methods (Random Forests, Logit, XGBoost, and Random) concerning their ability to correctly predict an "upround" ($HasNext$, Panel A), an indicator variable equal to one if the startup has a subsequent financing round, an eventual IPO ($IPO$, Panel B), and ten times more future money raised than past and present money raised ($HighRet10X$, Panel C). As explanatory variables, I use, in addition to characteristics related to the name of the startup, the count of previous common investors and lead investors, the total past money raised, the money raised, and proxies for the investment experience of the investors (number of investments made and average money raised per investment over the last decade). The models are trained using data from 1995 to 2011 and tested on data from 2012 to 2015.

**Panel A: HasNext**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.708 | 0.631 | 0.756 | 0.590 | 0.626 | 0.637 | 0.607 | 0.579 |
| Seed | 0.622 | 0.617 | 0.617 | 0.456 | 0.540 | 0.538 | 0.446 | 0.433 |
| Angel | 0.528 | 0.527 | 0.632 | 0.477 | 0.510 | 0.491 | 0.462 | 0.452 |
| Series A | 0.700 | 0.681 | 0.734 | 0.598 | 0.625 | 0.649 | 0.579 | 0.578 |
| Series B | 0.699 | 0.692 | 0.725 | 0.551 | 0.624 | 0.639 | 0.568 | 0.553 |
| Series C | 0.632 | 0.644 | 0.700 | 0.513 | 0.577 | 0.583 | 0.539 | 0.527 |
| Series D | 0.588 | 0.628 | 0.665 | 0.479 | 0.593 | 0.586 | 0.500 | 0.545 |
| Series E | 0.560 | 0.593 | 0.617 | 0.373 | 0.514 | 0.519 | 0.455 | 0.474 |
| Series F | 0.154 | 0.377 | 0.533 | 0.300 | 0.333 | 0.448 | 0.400 | 0.450 |

**Panel B: IPO**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.000 | 0.000 | 0.032 | 0.000 | 0.000 | 0.000 | 0.016 | 0.000 |
| Seed | 0.000 | 0.034 | 0.011 | 0.000 | 0.000 | 0.021 | 0.006 | 0.000 |
| Angel | 0.167 | 0.059 | 0.667 | 0.000 | 0.100 | 0.031 | 1.000 | 0.000 |
| Series A | 0.249 | 0.204 | 0.047 | 0.053 | 0.228 | 0.139 | 0.025 | 0.045 |
| Series B | 0.317 | 0.316 | 0.082 | 0.079 | 0.323 | 0.235 | 0.046 | 0.082 |
| Series C | 0.312 | 0.333 | 0.150 | 0.156 | 0.287 | 0.236 | 0.088 | 0.159 |
| Series D | 0.358 | 0.366 | 0.298 | 0.272 | 0.316 | 0.259 | 0.182 | 0.299 |
| Series E | 0.495 | 0.549 | 0.454 | 0.281 | 0.429 | 0.414 | 0.306 | 0.391 |
| Series F | 0.591 | 0.655 | 0.578 | 0.390 | 0.578 | 0.537 | 0.424 | 0.441 |

**Panel C: HighRet10X**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.551 | 0.484 | 0.583 | 0.407 | 0.453 | 0.444 | 0.416 | 0.386 |
| Seed | 0.341 | 0.353 | 0.332 | 0.223 | 0.267 | 0.250 | 0.200 | 0.197 |
| Angel | 0.205 | 0.280 | 0.281 | 0.167 | 0.204 | 0.212 | 0.166 | 0.153 |
| Series A | 0.153 | 0.222 | 0.236 | 0.160 | 0.159 | 0.167 | 0.144 | 0.149 |
| Series B | 0.039 | 0.073 | 0.114 | 0.022 | 0.062 | 0.055 | 0.073 | 0.026 |
| Series C | 0.000 | 0.038 | 0.000 | 0.057 | 0.000 | 0.033 | 0.000 | 0.077 |
| Series D | 0.000 | 0.000 | 0.333 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 |
| Series E | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series F | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 7: IPO Predictability Decline Post-2016

This table compares the F1 and precision scores across four methods (Random Forests, Logit, XGBoost, and Random) concerning their ability to correctly predict an eventual IPO after 2016 (Panel A), along with the relative differences from the pre-2016 sample (Panel B). As explanatory variables, I use, in addition to characteristics related to the name of the startup, the count of previous common investors and lead investors, the total past money raised, the money raised, and proxies for the investment experience of the investors (number of investments made and average money raised per investment over the last decade). The models are trained using data from 2016 to 2018 and tested on data from 2019 and 2020.

**Panel A: 2016 or later**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Seed | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 |
| Angel | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series A | 0.267 | 0.232 | 0.008 | 0.000 | 0.171 | 0.139 | 0.004 | 0.000 |
| Series B | 0.465 | 0.370 | 0.020 | 0.106 | 0.353 | 0.238 | 0.011 | 0.100 |
| Series C | 0.414 | 0.329 | 0.085 | 0.087 | 0.333 | 0.219 | 0.046 | 0.068 |
| Series D | 0.290 | 0.348 | 0.156 | 0.145 | 0.290 | 0.262 | 0.088 | 0.132 |
| Series E | 0.269 | 0.346 | 0.263 | 0.145 | 0.209 | 0.225 | 0.156 | 0.129 |
| Series F | 0.292 | 0.320 | 0.320 | 0.133 | 0.175 | 0.190 | 0.235 | 0.091 |

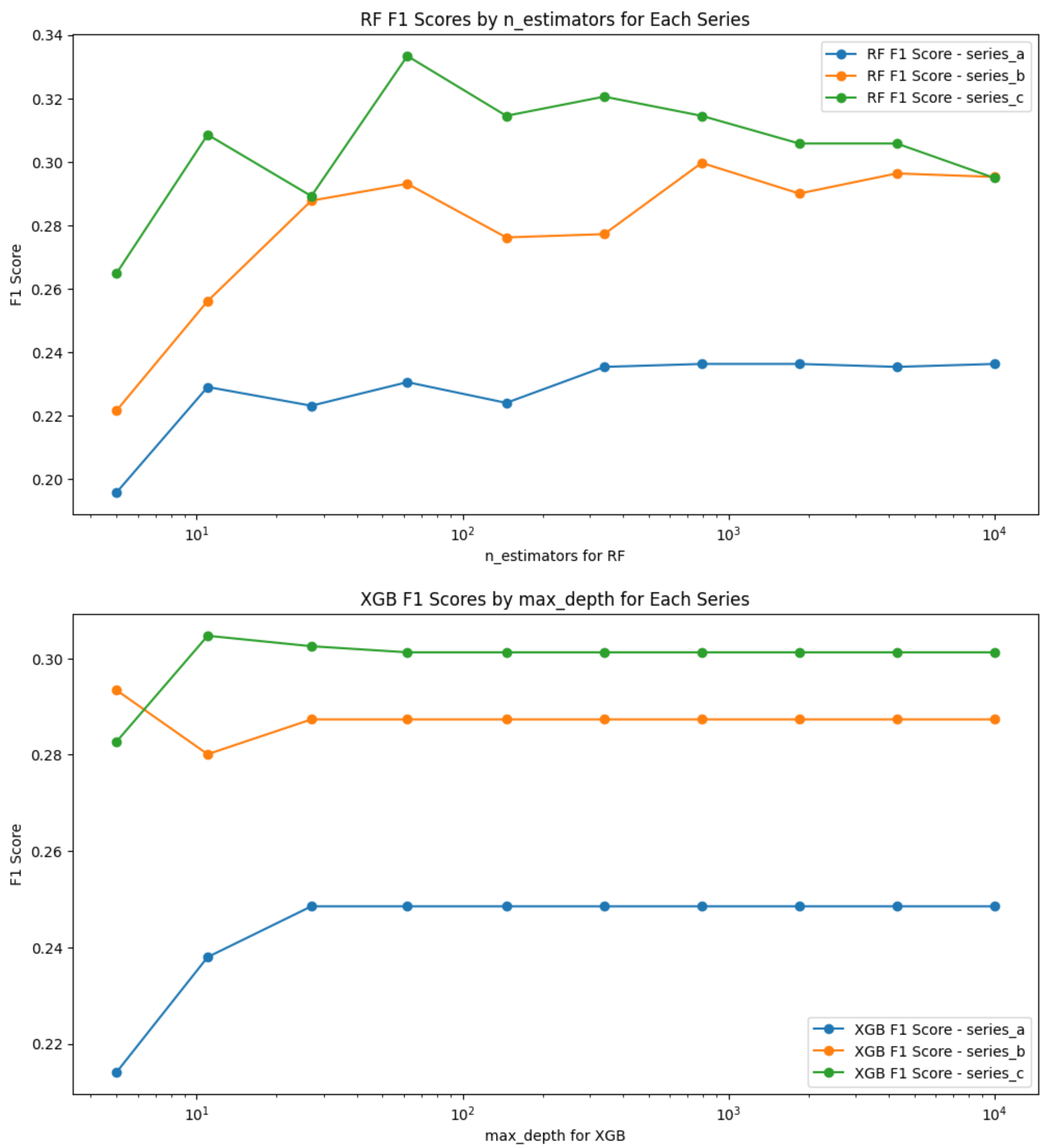
**Panel B: Relative Differences (%)**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0 | 0 | -100.0 | 0 | 0 | 0 | -100.0 | 0 |
| Seed | 0 | -100.0 | -63.6 | 0 | 0 | -100.0 | -66.7 | 0 |
| Angel | -100.0 | -100.0 | -100.0 | 0 | -100.0 | -100.0 | -100.0 | 0 |
| Series A | 7.2 | 13.7 | -83.0 | -100.0 | -25.0 | 0.0 | -84.0 | -100.0 |
| Series B | 46.7 | 17.1 | -75.6 | 34.2 | 9.3 | 1.3 | -76.1 | 22.0 |
| Series C | 32.7 | -1.2 | -43.3 | -44.2 | 16.0 | -7.2 | -47.7 | -57.2 |
| Series D | -19.0 | -4.9 | -47.7 | -46.7 | -8.2 | 1.2 | -51.6 | -55.9 |
| Series E | -45.7 | -37.0 | -42.1 | -48.4 | -51.3 | -45.7 | -49.0 | -67.0 |
| Series F | -50.6 | -51.1 | -44.6 | -65.9 | -69.7 | -64.6 | -44.6 | -79.4 |

## Online Appendix

Figure 9: Random Forest and XGB F1 Scores Based on "n_estimators" and "max_depth"

The upper figure shows how the F1 score obtained with the random forests algorithm varies for a wide range of different numbers of decision trees ("n_estimators"), and the lower figure shows how the F1 score of XGBoost varies with the maximum depth of an individual decision tree within the ensemble of decision trees ("max_depth"). All other parameters are identical to those indicated in Appendix D. Both figures are computed for a selection of financing rounds (Series A, Series B, and Series C) only.



## Online Appendix

Table 8: Precision Rates of the Top Investors

This table provides the average precision scores across investors from the top performance quartile in each financing round. To be considered in the sample, the investor must have made at least 20 investments at a given investment stage during the entire sample period (1995-2016) in the case of Series D or earlier rounds and at least 10 investments for Series E and Series F rounds.

| | IPO | HasNext | HighRet10X |
|---|---|---|---|
| Angel | 0.150 | 0.600 | 0.350 |
| Pre Seed | 0.065 | 0.839 | 0.516 |
| Seed | 0.052 | 0.784 | 0.263 |
| Series A | 0.160 | 0.785 | 0.182 |
| Series B | 0.198 | 0.735 | 0.069 |
| Series C | 0.224 | 0.683 | 0.055 |
| Series D | 0.350 | 0.613 | 0.005 |
| Series E | 0.527 | 0.680 | 0.000 |
| Series F | 0.800 | 0.600 | 0.000 |

Figure 10: Comparative Current and Future Money Raised By Model and Prediction Correctness (Early Stage)

Both figures on the left are the same as shown in Figure 5. The top left figure shows the total money raised across late-stage rounds whose predictions of success were either correct or wrong, with success defined as an eventual IPO, by model: random forests (RF), XGBBoost (XGB), logit (LR), and random selection (Random (avg. investor)), in the test sample (2012-2015). These amounts correspond to the aggregate money raised for each category of rounds (with correct success predictions and with wrong success predictions). I define late-stage rounds as Series C or later rounds. A round with a correct success prediction is defined by a prediction of an eventual IPO that materializes, while a round with a wrong success prediction is defined by a non-materialized prediction of an eventual IPO. The bottom left figure shows the ratios between the correct and wrong success prediction monetary amounts by model. The figures on the right are analogous to the results on the left but use the *future* money raised by the startups instead, serving as a proxy for investment returns obtained by the investment strategy implied by the model.

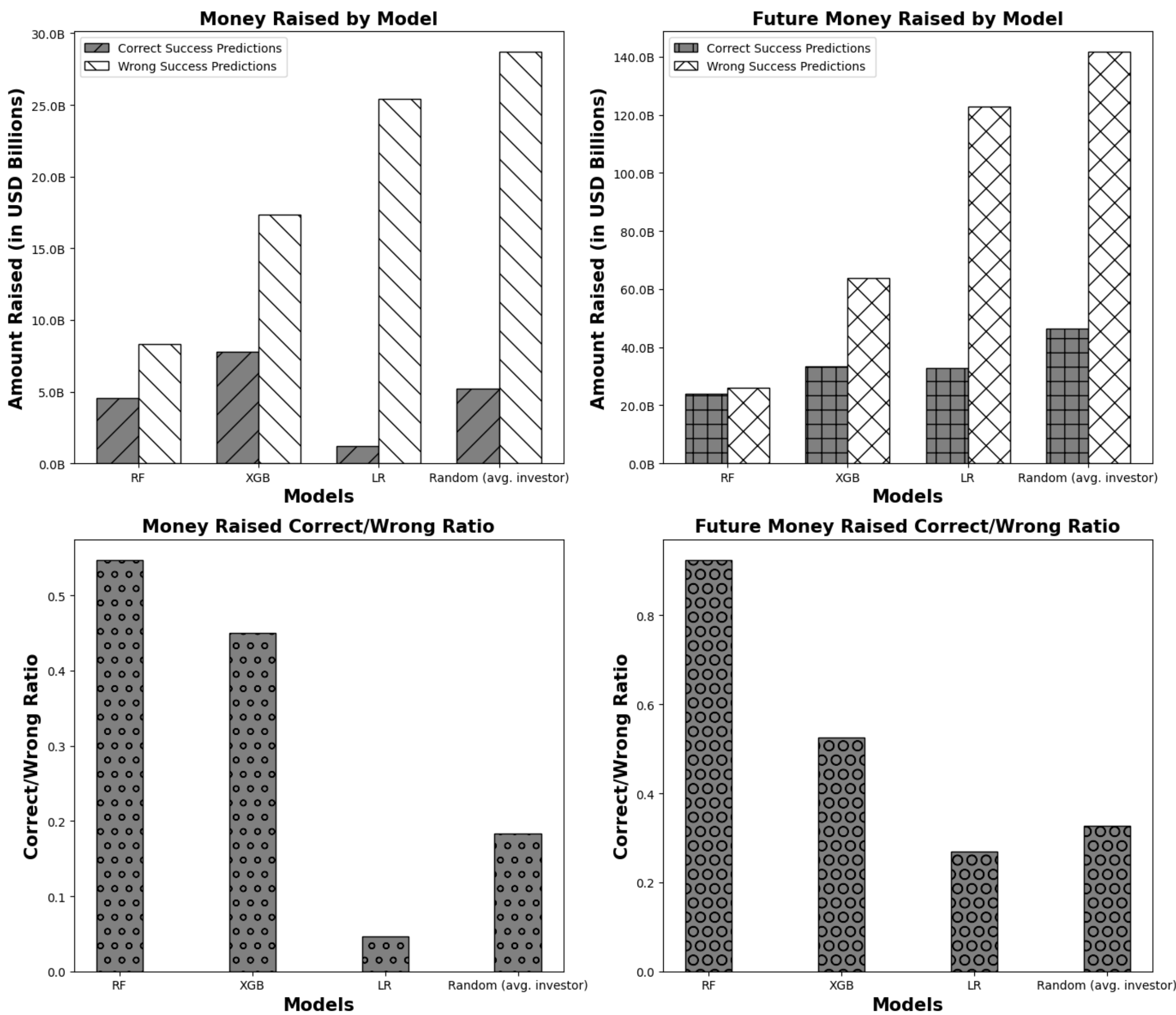
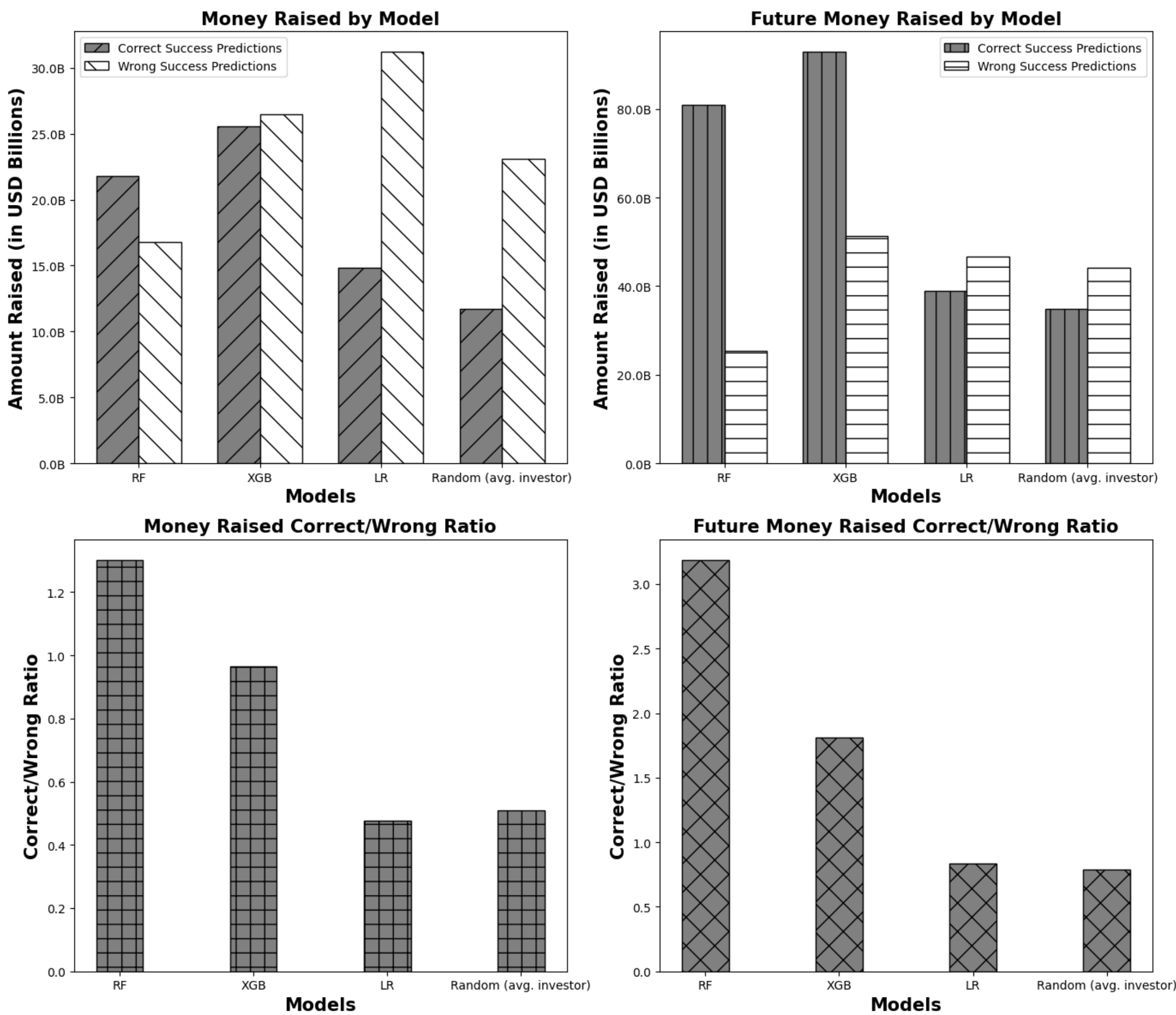


Figure 11: Comparative Current and Future Money Raised By Model and Prediction Correctness (Late Stage)

Both figures on the left are the same as shown in Figure 6. The top left figure shows the total money raised across late-stage rounds whose predictions of success were either correct or wrong, with success defined as an eventual IPO, by model: random forests (RF), XGBBoost (XGB), logit (LR), and random selection (Random (avg. investor)), in the test sample (2012-2015). These amounts correspond to the aggregate money raised for each category of rounds (with correct success predictions and with wrong success predictions). I define late-stage rounds as Series C or later rounds. A round with a correct success prediction is defined by a prediction of an eventual IPO that materializes, while a round with a wrong success prediction is defined by a non-materialized prediction of an eventual IPO. The bottom left figure shows the ratios between the correct and wrong success prediction monetary amounts by model. The figures on the right are analogous to the results on the left but use the *future* money raised by the startups instead, serving as a proxy for investment returns obtained by the investment strategy implied by the model.



## Online Appendix

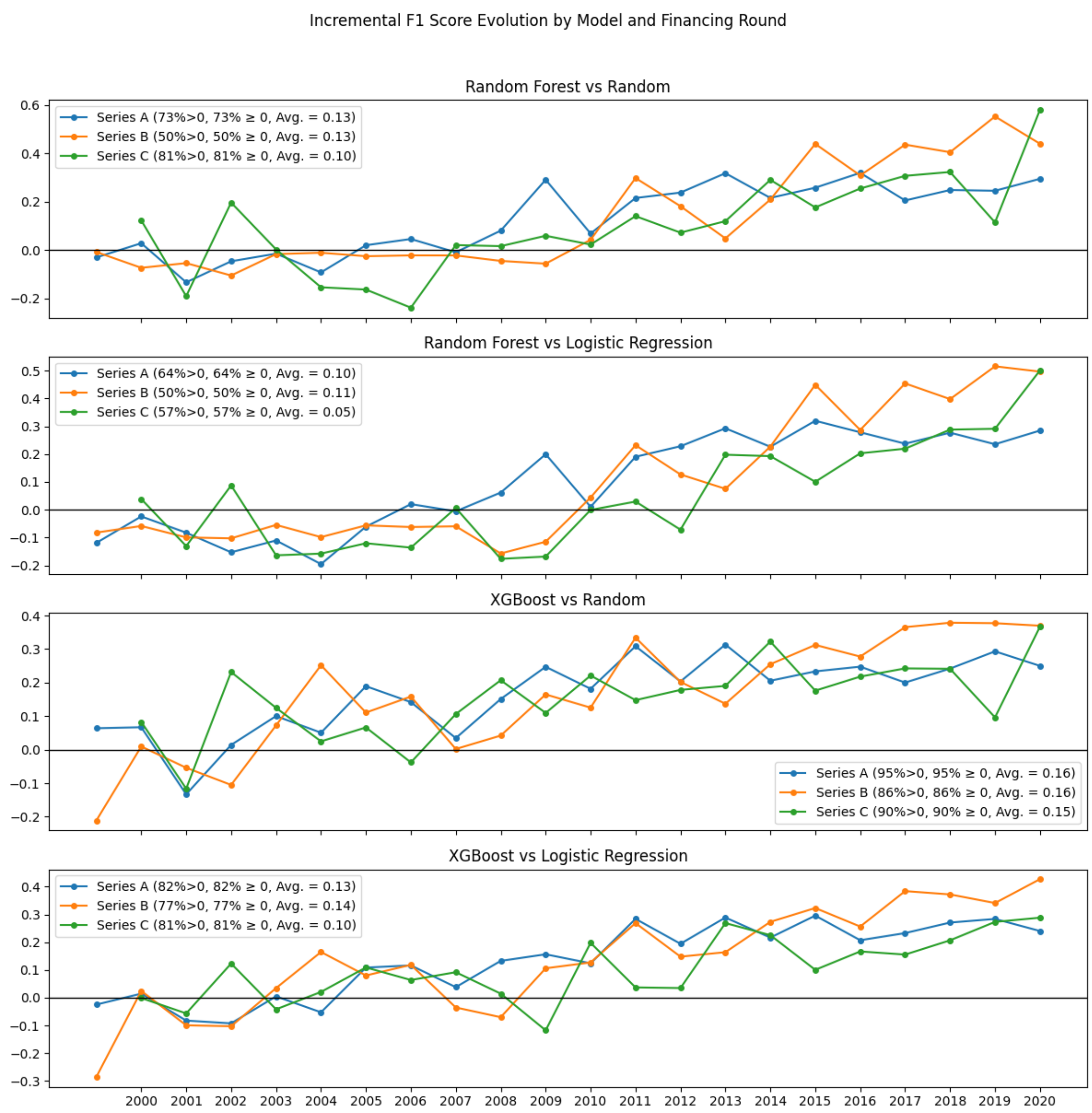
Figure 12: Incremental Precision Score Evolution Using Only Previous Year Data

This figure shows how each machine learning (ML) algorithm performs regarding their resulting precision score when predicting IPOs compared to random selection and logistic regression, using a sample limited to data available in the **previous year** only. For instance, for the year 2005, only data from 2004 is used to train the model, which is then tested on 2005 data. Each curve represents the difference between the precision score obtained by the indicated algorithm (Random Forest of XGBoost) and the benchmarks (random selection and logistic regression). The percentages indicate the fraction of years the algorithm is strictly better than the benchmark and the fraction of years the algorithm is at least as good as the benchmark. The average indicated refers to the average incremental precision score in the entire period (2000-2020).

## Online Appendix

Figure 13: Incremental F1 Score Evolution Using Only Previous Year Data

This figure shows how each machine learning (ML) algorithm performs regarding their resulting F1 score when predicting IPOs compared to random selection and logistic regression, using a sample limited to data available in the **previous year** only. For instance, for the year 2005, only data from 2004 is used to train the model, which is then tested on 2005 data. Each curve represents the difference between the precision score obtained by the indicated algorithm (Random Forest of XGBoost) and the benchmarks (random selection and logistic regression). The percentages indicate the fraction of years the algorithm is strictly better than the benchmark and the fraction of years the algorithm is at least as good as the benchmark. The average indicated refers to the average incremental F1 score in the entire period (2000-2020).



## Online Appendix

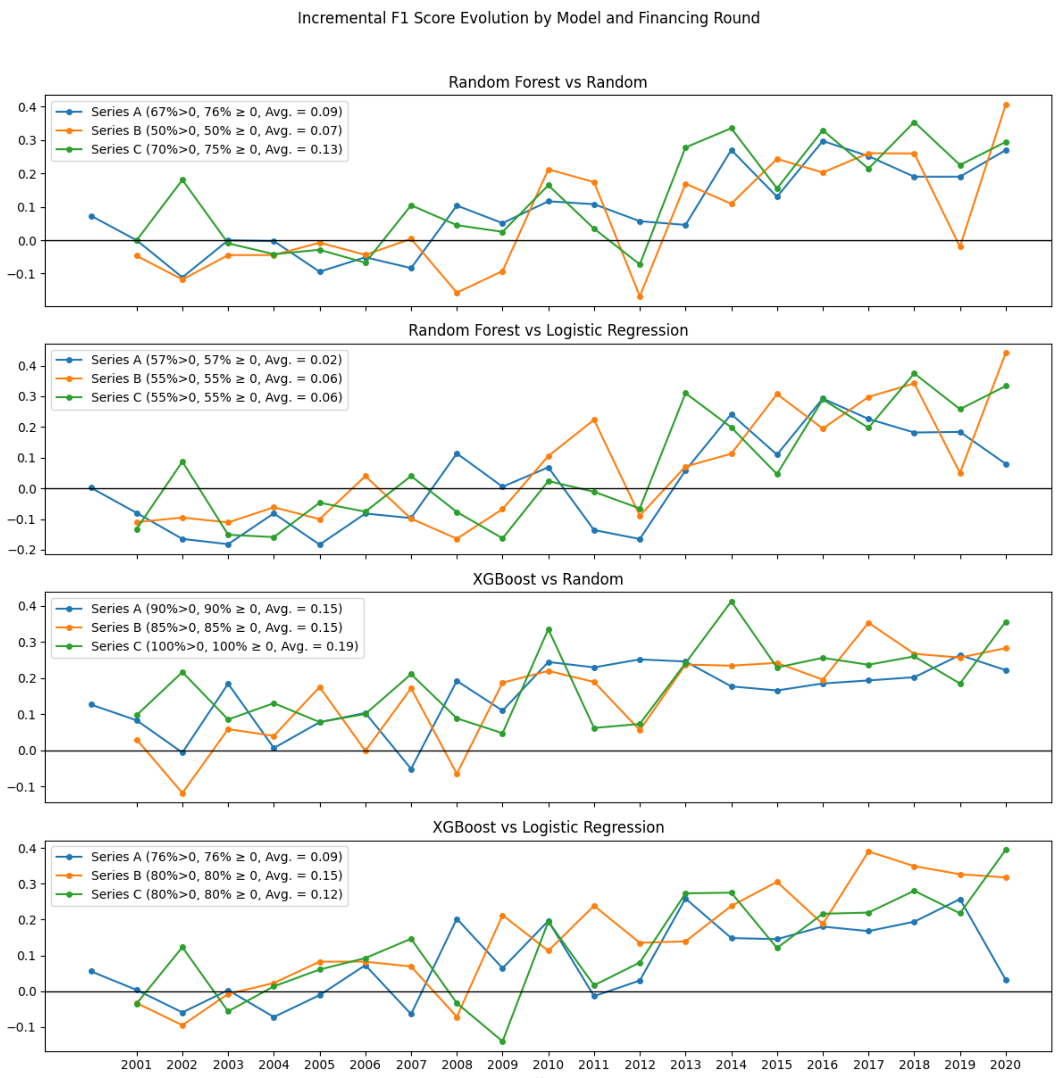
Figure 14: Incremental Precision Score Evolution Using Previous Year Data and Fixed Sample Size

This figure shows how each machine learning (ML) algorithm performs regarding their resulting precision score when predicting IPOs compared to random selection and logistic regression, using a sample limited to data available in the **previous year** and only N randomly selected observations, **where N is the smallest sample size available for all years in the period**. For instance, for the year 2005, only a random fixed N amount of observations from 2004 is used to train the model, which is then tested on 2005 data. Each curve represents the difference between the precision score obtained by the indicated algorithm (Random Forest of XGBoost) and the benchmarks (random selection and logistic regression). The percentages indicate the fraction of years the algorithm is strictly better than the benchmark and the fraction of years the algorithm is at least as good as the benchmark. The average indicated refers to the average incremental precision score in the entire period (2000-2020).

## Online Appendix

Figure 15: Incremental F1 Score Evolution Using Previous Year Data and Fixed Sample Size

This figure shows how each machine learning (ML) algorithm performs regarding their resulting F1 score when predicting IPOs compared to random selection and logistic regression, using a sample limited to data available in the **previous year** and only N randomly selected observations, **where N is the smallest sample size available for all years in the period**. For instance, for the year 2005, only a random fixed N amount of observations from 2004 is used to train the model, which is then tested on 2005 data. Each curve represents the difference between the precision score obtained by the indicated algorithm (Random Forest of XGBoost) and the benchmarks (random selection and logistic regression). The percentages indicate the fraction of years the algorithm is strictly better than the benchmark and the fraction of years the algorithm is at least as good as the benchmark. The average indicated refers to the average incremental F1 score in the entire period (2000-2020).



## Online Appendix

Table 9: Startup Outcome Prediction Using Multiple Characteristics

This table compares the F1 and precision scores across four methods (Random Forests, Logit, XGBoost, and Random) concerning their ability to correctly predict acquisitions, an indicator variable equal to one if the startup is eventually acquired. As explanatory variables, I use, in addition to characteristics related to the name of the startup, the count of previous common investors and lead investors, the total past money raised, the money raised, and proxies for the investment experience of the investors (number of investments made and average money raised per investment over the last decade). The models are trained using data from 1995 to 2011 and tested on data from 2012 to 2015.

| **Predicted Outcome: Acquisitions** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **F1** | | | | **Precision** | | | |
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.000 | 0.000 | 0.020 | 0.111 | 0.000 | 0.000 | 0.010 | 0.067 |
| Seed | 0.000 | 0.050 | 0.013 | 0.029 | 0.000 | 0.030 | 0.007 | 0.019 |
| Angel | 0.167 | 0.057 | 0.667 | 0.111 | 0.100 | 0.030 | 1.000 | 0.062 |
| Series A | 0.266 | 0.223 | 0.076 | 0.064 | 0.259 | 0.156 | 0.040 | 0.055 |
| Series B | 0.295 | 0.322 | 0.150 | 0.113 | 0.290 | 0.237 | 0.082 | 0.109 |
| Series C | 0.314 | 0.333 | 0.239 | 0.162 | 0.280 | 0.237 | 0.136 | 0.171 |
| Series D | 0.361 | 0.370 | 0.332 | 0.150 | 0.310 | 0.263 | 0.200 | 0.156 |
| Series E | 0.497 | 0.529 | 0.483 | 0.233 | 0.432 | 0.400 | 0.321 | 0.319 |
| Series F | 0.575 | 0.655 | 0.592 | 0.216 | 0.568 | 0.537 | 0.424 | 0.258 |

## Online Appendix

Table 10: Upround Predictability Decline Post-2016

This table compares the F1 and precision scores across four methods (Random Forests, Logit, XGBoost, and Random) concerning their ability to correctly predict an "upround" ($HasNext$) after 2016 (Panel A), along with the relative differences from the pre-2016 sample (Panel B). As explanatory variables, I use, in addition to characteristics related to the name of the startup, the count of previous common investors and lead investors, the total past money raised, the money raised, and proxies for the investment experience of the investors (number of investments made and average money raised per investment over the last decade).

| **Panel A: 2016 or later** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **F1** | | | | **Precision** | | | |
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.407 | 0.412 | 0.425 | 0.316 | 0.290 | 0.310 | 0.295 | 0.228 |
| Seed | 0.281 | 0.267 | 0.262 | 0.200 | 0.176 | 0.169 | 0.158 | 0.136 |
| Angel | 0.500 | 0.469 | 0.469 | 0.314 | 0.364 | 0.341 | 0.311 | 0.220 |
| Series A | 0.291 | 0.279 | 0.281 | 0.218 | 0.182 | 0.174 | 0.176 | 0.143 |
| Series B | 0.256 | 0.270 | 0.269 | 0.240 | 0.160 | 0.171 | 0.160 | 0.156 |
| Series C | 0.253 | 0.286 | 0.262 | 0.230 | 0.162 | 0.177 | 0.151 | 0.151 |
| Series D | 0.255 | 0.216 | 0.233 | 0.244 | 0.152 | 0.136 | 0.132 | 0.157 |
| Series E | 0.393 | 0.403 | 0.320 | 0.329 | 0.273 | 0.266 | 0.192 | 0.255 |
| Series F | 0.304 | 0.279 | 0.303 | 0.276 | 0.194 | 0.182 | 0.179 | 0.211 |
| **Panel B: Relative Differences (%)** | | | | | | | | |
| | **F1** | | | | **Precision** | | | |
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | -42.5 | -34.7 | -43.8 | -46.4 | -53.7 | -51.3 | -51.4 | -60.6 |
| Seed | -54.8 | -56.7 | -57.5 | -56.1 | -67.4 | -68.6 | -64.6 | -68.6 |
| Angel | -5.3 | -11.0 | -25.8 | -34.2 | -28.6 | -30.5 | -32.7 | -51.3 |
| Series A | -58.4 | -59.0 | -61.7 | -63.5 | -70.9 | -73.2 | -69.6 | -75.3 |
| Series B | -63.4 | -61.0 | -62.9 | -56.4 | -74.4 | -73.2 | -71.8 | -71.8 |
| Series C | -60.0 | -55.6 | -62.6 | -55.2 | -71.9 | -69.6 | -72.0 | -71.3 |
| Series D | -56.6 | -65.6 | -65.0 | -49.1 | -74.4 | -76.8 | -73.6 | -71.2 |
| Series E | -29.8 | -32.0 | -48.1 | -11.8 | -46.9 | -48.7 | -57.8 | -46.2 |
| Series F | 97.4 | -26.0 | -43.2 | -8.0 | -41.7 | -59.4 | -55.2 | -53.1 |

## Online Appendix

Table 11: High-Returns Predictability Decline Post-2016

This table compares the F1 and precision scores across four methods (Random Forests, Logit, XGBoost, and Random) concerning their ability to correctly identify a "high return" startup ($HighRet10X$) after 2016 (Panel A), along with the relative differences from the pre-2016 sample (Panel B). As explanatory variables, I use, in addition to characteristics related to the name of the startup, the count of previous common investors and lead investors, the total past money raised, the money raised, and proxies for the investment experience of the investors (number of investments made and average money raised per investment over the last decade).

**Panel A: 2016 or later**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | 0.162 | 0.096 | 0.139 | 0.094 | 0.102 | 0.059 | 0.076 | 0.058 |
| Seed | 0.159 | 0.107 | 0.156 | 0.049 | 0.103 | 0.059 | 0.088 | 0.029 |
| Angel | 0.000 | 0.069 | 0.182 | 0.143 | 0.000 | 0.037 | 0.111 | 0.083 |
| Series A | 0.000 | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 0.007 | 0.000 |
| Series B | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series C | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series D | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series E | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Series F | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Panel B: Relative Differences (%)**

| | **F1** | | | | **Precision** | | | |
|---|---|---|---|---|---|---|---|---|
| Round | RF | XGB | Logit | Random | RF | XGB | Logit | Random |
| Pre-Seed | -70.6 | -80.2 | -76.2 | -76.9 | -77.5 | -86.7 | -81.7 | -85.0 |
| Seed | -53.4 | -69.7 | -53.0 | -78.0 | -61.4 | -76.4 | -56.0 | -85.3 |
| Angel | -100.0 | -75.4 | -35.2 | -14.4 | -100.0 | -82.5 | -33.1 | -45.8 |
| Series A | -100.0 | -100.0 | -93.6 | -100.0 | -100.0 | -100.0 | -95.1 | -100.0 |
| Series B | -100.0 | -100.0 | -100.0 | -100.0 | -100.0 | -100.0 | -100.0 | -100.0 |
| Series C | 0 | -100.0 | 0 | -100.0 | 0 | -100.0 | 0 | -100.0 |
| Series D | 0 | 0 | -100.0 | 0 | 0 | 0 | -100.0 | 0 |
| Series E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Series F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |