

Discussion of “Machine Learning about Venture Capital Choices”

Ramana Nanda

IESE Banking Initiative Workshop

March 20, 2025

Overview

- Unique data on entrepreneur characteristics for over 120,000 startups in France between 1998 and 2010 merged with administrative data to get startups' outcomes (survival, revenue 5 and 7 years out)
- Use supervised machine learning models to predict
 - (1) Successful startups (“predict success model”)
 - (2) Startups that are predicted to be attractive to VCs (“mimic VC model”)

Then compare these to understand focus of VC investment

Overview (2)

- VCs are systematically associated with funding startups that perform predictably poorly and to also miss some predictable successes.
- Consistent with VCs using heuristic that puts disproportionate weight on characteristics that are over-represented in the most successful firms (based in Paris, run by men, from Elite schools)

Contribution of this paper

- Using machine learning to study VC decision making is both important and interesting
 - VCs are gatekeepers who make investment decisions before product market success... and hence can shape realized innovation
- Frictions in how they select and advance new ventures has received relatively little attention compared to work on VC governance
 - This paper focuses on differences in characteristics of firms that receive VC funding and not and the degree to which the characteristics emphasized by VCs are 'optimal'
 - Interesting and intuitive mechanism proposed by the authors
- Also related to literature on firm dynamics
 - How much of firm type is *ex ante* predictable vs. learnt *ex post* through learning

Focus of my comments

- Ideal vs. actual study design
- Data and Analysis
- Mechanism

Ideal study design

- Map VC's objective function to a machine learning model
 - Option value of VC contract might lead them to care about “volatility” not just expected value.
 - Portfolio construction constraints might impact whether a given firm is right for the portfolio
- Observe all firms seeking that VC's funding and those *chosen* by VC (not necessarily that got VC funding as this is a two-sided market)
- Use machine model to study whether decision making could be improved, given the choice set and objective function
 - Note still implicitly assumes “efficient markets” so that all good firms do ultimately get funded, and moreover, similar treatment effects of different VCs

Actual study design

- A 'representative VC' who cares about expected revenue and is the preferred source of funding by entrepreneurs so that actual firms receiving funding are those chosen by VC
 - Any firm that did not get VC was not matched to its preferred source of financing
- Counterfactual feels inconsistent
 - If VC was the dominant / preferred source of finance that is needed for success, then why do we see so many successful firms without VC
 - On the other hand, if VC is not needed for success, then how can we assume the VC got to choose from the best firms

Data and Analysis

- The firms in the sample, although small appear to have characteristics of older, more mature firms (or those that are part of larger firms)
 - 60% survive to age 5 (vs. about 40% of startups in the US)
 - 35% or more have bank loan (vs. 16% of startups in US SBO)
- How do you address acquisition of non-VC backed companies in the data?
- How well do the models work if you work in levels rather than in logs?
 - Investors are targeting outliers in levels
 - *Ex ante* home runs will lead *ex post* to a combination of home runs and failure. I have found it much harder to train machine learning model with such a data generating process, particularly when combined with power-law right tailed distributions

Mechanisms

- VC's simple heuristics lead to mistakes they can learn to avoid
 - Use machine learning to render information for VCs so as to improve allocation and returns
- VCs source deals inefficiently which leads to inefficient outcomes
 - (*“We are no longer taking blind pitches. Instead, we are going to focus exclusively on deals that come to us through our trusted network of friends and colleagues whom we admire” – VC investor in the US*)
- What appears as mistakes with one objective function may not be with another
 - Invest in deals that look *ex ante* like home runs. Abandon those that don't show promise. Double down on those that do (initial investment gives both abandonment and re-investment options that are very valuable)
 - VC compensation model pushes them to have preference for skew (prefer 1% chance of billion to 10 million for sure)
This may be worth it to incentivize VCs with the very best access but inefficient in aggregate

Big picture narrative suggestion

- We know very little about the entrepreneurial characteristics that VCs emphasize in their investment decisions
 - Why not start with a rich description of these including industry and geographic focus, and other entrepreneurial characteristics – this is a “revealed” preference approach that we know little about
 - Compare to tighter comparison groups and demonstrate the “over-emphasis” on certain traits
- Discuss possible mechanisms for this over-emphasis including yours noting that some are frictions that can be addressed
- Suppose your friction was the only one and ML could improve heuristic, what is the potential quantitative size?