Empirical Analysis of Startups

Mitsubishi UFJ Trust Investment Technology Institute Co., Ltd. (MTEC)[1]

Abstract

This study constructed a model to predict the likelihood of startups advancing to their subsequent funding round and quantitatively analyzed the factors contributing to this success. Utilizing FactSet data, a logit regression model was developed with the firm's total assets, the investors' success rate, the elapsed time since the previous financing round, and the transition period from the prior round serving as explanatory variables. The results revealed that larger firm size, support from investors with a proven track record, and a shorter duration for progressing through growth stages were statistically significant factors in successfully advancing to the subsequent funding round.

1. Introduction

Private companies, particularly early-stage startups, are indispensable drivers of innovation, contributing significantly to economic growth and employment generation in modern economies. However, many of these companies exhibit limited financial disclosure and lack objective valuation metrics, such as stock prices available in public markets. Consequently, assessing their credit risk and predicting their future growth potential is a challenging task for investors and financial institutions alike. This inherent difficulty in valuation significantly

---

[1] The information contained in this report is provided "as is" and does not carry any warranties or guarantees of accuracy or completeness. The information is provided solely for informational purposes and should not be considered as professional advice. Any actions or decisions based on the information in this report are at the reader's own risk. This report is intended to provide general information and does not constitute specific advice or endorsements for any particular situation or requirements. Mitsubishi UFJ Trust Investment Technology Institute Co., Ltd. shall bear no responsibility or liability for any actions taken based on this analysis or for any loss or damage incurred based on this report. None from among Mitsubishi UFJ Trust Investment Technology Institute Co., Ltd., its suppliers, nor any of their respective directors or employees accept any liability towards any third party who has obtained this document with respect to any loss or damage directly or indirectly arising out of the use of all or any part of this report.

escalates uncertainty in both investment and lending decisions, underscoring the pressing need for effective solutions.

The growth of startups is highly dependent on the successful completion of sequential funding rounds, such as Seed, Series A, and Series B. These rounds not only represent opportunities for capital acquisition but also serve as crucial milestones for business progress and significant signals indicating a transition to the subsequent growth stage. However, it is widely recognized that not all companies successfully advance to the subsequent round, with many stagnating or shutting down at various stages. Therefore, predicting the probability of a company successfully transitioning to a subsequent funding round and quantitatively identifying the contributing factors is indispensable for enhancing risk management in investment and lending to startups. This paper attempts to address this critical challenge by undertaking a comprehensive analysis using an empirical approach.

Research concerning the growth and exit potential of startups has been conducted from various perspectives. Conventionally, financial information constitutes the primary subject of analysis. Specifically, a study by Hosono et al. (2015) focusing on Japanese firms demonstrated that companies with high size, Return on Assets (ROA), and Total Factor Productivity (TFP), along with low debt ratios and expense ratios, exhibit a higher probability of an Initial Public Offering (IPO). This finding is consistent with the hypothesis that a stable business foundation and robust financial health enhance the likelihood of a future exit.

In recent years, the importance of non-financial factors, in addition to conventional financial information, has been highlighted. Specifically, this perspective posits that the Investor acumen influences the growth potential of their portfolio companies. A study by Sasaki et al. (2024), which analyzed US companies, developed an indicator for investor discernment. They demonstrated that investment by investors possessing strong discernment can serve as a signal for a startup's exit or subsequent funding round.

Furthermore, the "temporal factor" in fundraising has emerged as a crucial consideration. Venture capitalists (VCs) are obligated to return funds to their investors within a predetermined timeframe. Consequently, in their pursuit of high-risk, high-reward opportunities, they prioritize and expect 'high growth' – the generation of substantial returns within a short period – rather than gradual, incremental growth. Therefore, the "momentum" generated by a company consistently achieving its milestones within short timeframes and subsequently securing successive rounds of funding serves as a powerful signal to VCs. This indicates that the business is progressing as planned, or even exceeding expectations. On the other hand, a prolonged inability to secure new funding suggests business stagnation or the presence of underlying issues. This situation contributes significantly to VCs perceiving the investment risk as high.

This paper, drawing upon the findings suggested in these previous studies, utilizes FactSet's comprehensive data to construct a model that predicts the probability of private companies transitioning from each funding round to the subsequent round, and analyzes the determinants.

Specifically, the explanatory variables will include firm size (total assets) as a financial indicator, investor quality (success rate) as a non-financial indicator, and growth momentum as a temporal factor. Growth momentum will be operationalized by the elapsed time from the last funding round and the transition period from the previous round. Through this analysis, we aim to quantitatively identify important factors that contribute to the likelihood of securing funding in subsequent rounds and to gain novel insights into the credit risk assessment of private companies. By constructing a predictive model that integrates traditional financial indicators with capital policy-related information—such as investor quality and fundraising history—and growth momentum, this study aims to quantitatively capture the multifaceted nature of private company valuation and make significant academic and practical contributions.

2. Data and Method

2.1. Data

This analysis develops a model to predict the success of private companies in securing capital in a subsequent funding round, following the completion of a particular prior round. Data for this study were obtained from FactSet's PEVC (Private Equity and Venture Capital) and FPC (Financial Private Company) databases. Specifically, the PEVC database encompasses investor information, transaction history, and data pertaining to company exits.

- ✓ The scope of the analysis was restricted to companies meeting the following criteria:
- ✓ Possess funding transaction data within the period from January 1, 2000, to December 31, 2024.
- ✓ All transaction currencies are denominated in USD.
- ✓ Total asset figures were obtainable from the financial data of their most recent fiscal year as of the valuation date.
- ✓ Their transaction history can be tracked starting from the seed round.

2.1. Model building and characteristics of variables

This study used a logit regression model due to the binary dependent variable: successful (1) or unsuccessful (0) funding in the subsequent round. Logit regression models offer the advantage of providing an intuitive understanding of both the direction and magnitude of variables' effects on the outcome, establishing them as a practical and well-established econometric method.

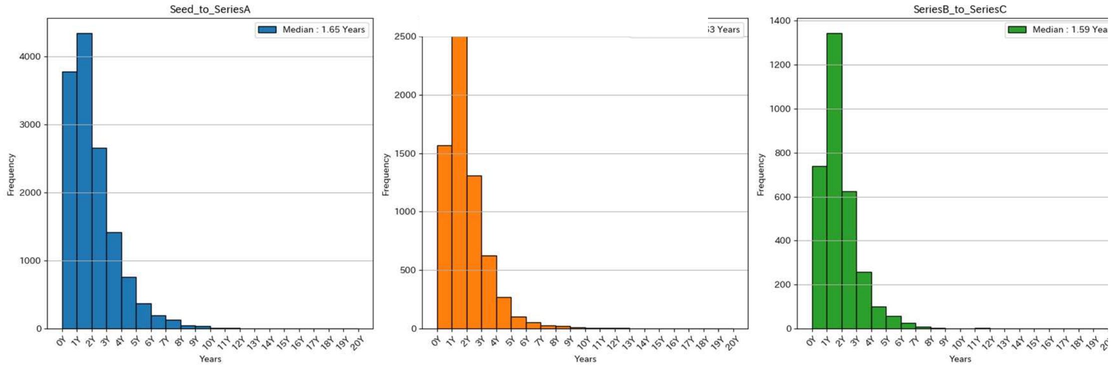
Six models were constructed to evaluate each firm's progress within specific funding rounds (Seed, Series A, Series B) at two separate time points: one year and three years after their initial funding within that round.

- ✓ For companies currently in the Seed round: the probability of securing Series A funding.
- ✓ For companies currently in the Series A round: the probability of securing Series B funding.
- ✓ For companies currently in the Series B round: the probability of securing Series C funding.

Figure 1 illustrates the proportion of companies that successfully advance from one funding round to the subsequent round. Approximately 40-50% of firms progress to the subsequent round, while only about 10% manage to advance from the Seed to the Series C. Figure 2 presents the duration, in days, required for companies to transition between rounds. The median transition period is approximately 600 days. A general trend observed is that the number of companies successfully advancing tends to decrease as the inter-round transition period lengthens.

Figure 1: The proportion of companies advancing to the subsequent round.

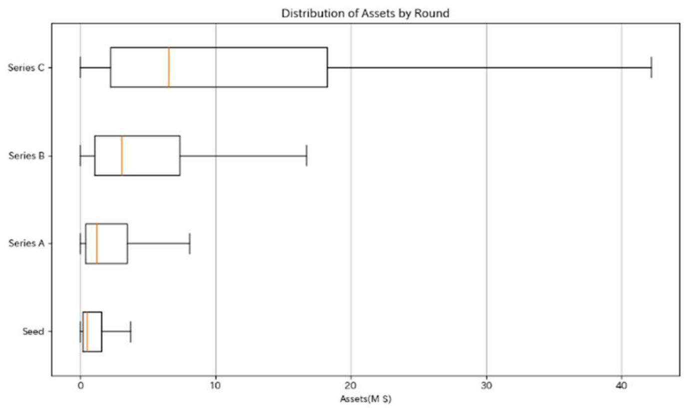
Figure 2: The period for advancing to the subsequent round.



Four explanatory variables were established to proxy for corporate growth potential and investor evaluation, as follows.

- ✓ Log10Assets: The variable represents the common logarithm of the most recent total assets at the time of evaluation, serving as a proxy for the firm's absolute size. Figure 3 presents the distribution of total assets recorded immediately prior to the initial fundraising within each round. (Outliers were removed using the 1.5 times interquartile range rule.) A comparison of medians across Seed, Series A, Series B, and Series C rounds reveals a clear trend: firm size tends to increase progressively with the advancement of funding rounds.

Figure 3: The distribution of total assets



- ✓ InvestorSuccessRate: We define the “investor success rate” as the success rate of the most successful investor among all investors who have invested in a given company by the evaluation point. Here, "success" is defined as the investee company advancing to the subsequent funding round. The success rate was calculated using all available data, irrespective of the specific evaluation point. This approach, by utilizing data across different calendar years, prevents data attrition at the evaluation point and aims to analyze the influence of both temporal and non-financial factors. To mitigate the issue of small sample sizes potentially leading to disproportionately high success rates (e.g., 80% or 100%) for investors with fewer than five investments, their success rate was set to 0%. Follow-on investments, where an investor provides additional capital to the same company, were not counted towards the total number of investments. This variable quantifies the extent to which private companies are able to secure funding from high-performing investors. Figure 4 illustrates the distribution of the number of investments per investor. As shown, most investors have made 1-4 investments, and the number of investors decreases significantly with an increasing number of investments. Figure 5 presents the distribution of investor success rates. Given the high frequency of investors with 1-4 investments, a substantial proportion of investors exhibit a 0% success rate.

Figure 4; The distribution of the number of investments per investor

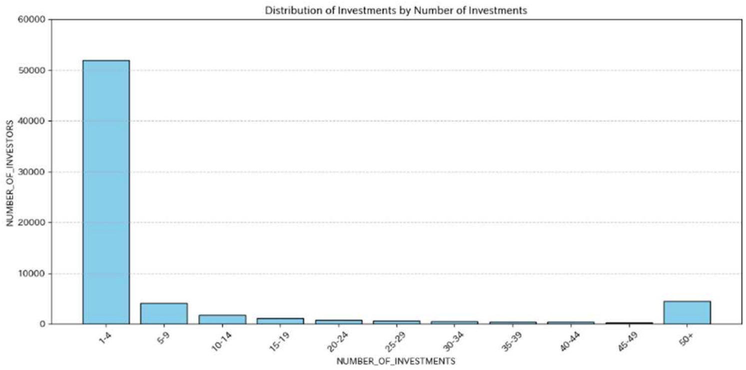
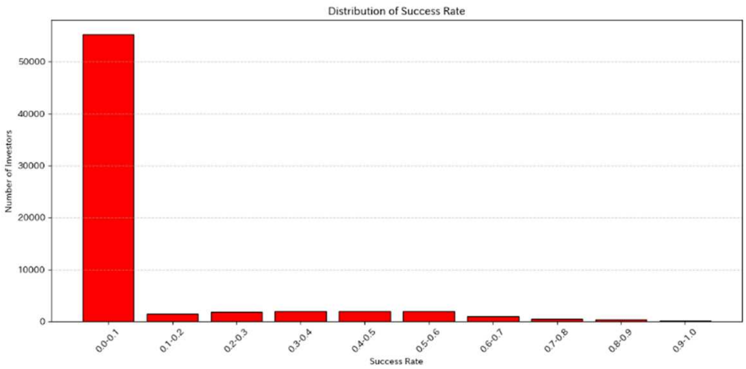
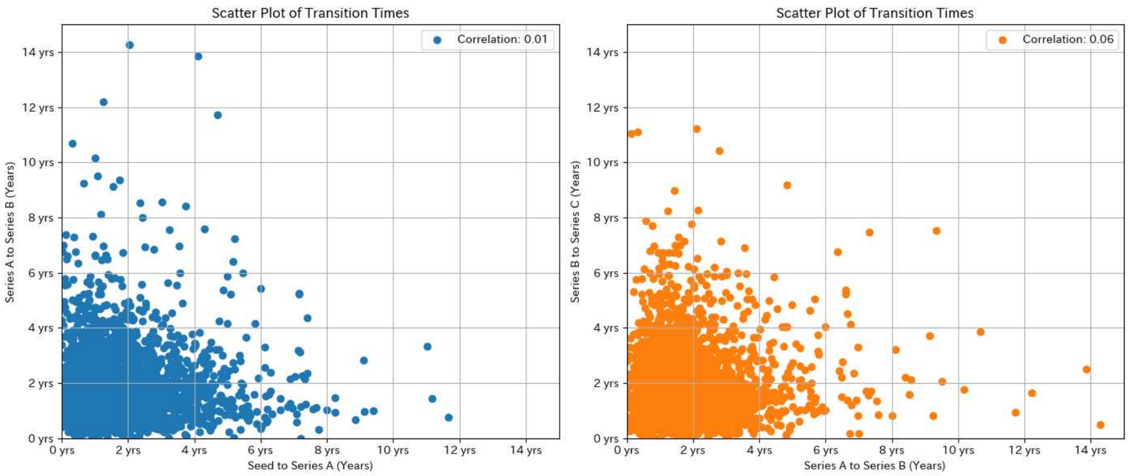


Figure 5: The distribution of investor success rates



- ✓ TimeSinceLastFundraising: The number of days elapsed from the most recent procurement date to the evaluation date is calculated. This duration, divided by 365, is then utilized.
- ✓ TimeSincePreviousRound(SeedToSeriesATime, SeriesAToSeriesBTime): The elapsed days from the previous to the current funding round were calculated, and this value, normalized by dividing by 365, was adopted. Figure 6 comprises two panels: the left panel illustrates the number of days from the Seed round to Series A, and from Series A to Series B. The right panel presents a scatter plot depicting the number of days from Series A to Series B and from Series B to Series C. The correlation coefficients for these respective relationships were 0.01 and 0.06, indicating that no strong correlation was observed.

Figure 6: The distribution of the elapsed days from the previous to the current funding round



3. Result and Evaluation

3.1. Result

The fundamental equation of the developed model is expressed by the following logit regression equation. At the evaluation point, the duration from Seed to Series A is incorporated as a variable for Series A. For Series B, the variables considered include both the duration from Seed to Series A and the duration from Series A to Series B.

$$\ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 \times Log10Assets + \beta_2 \times InvestorSuccessRate + \beta_3 \times TimeSinceLastFundraising + \beta_4 \times SeedToSeriesATime + \beta_5 \times SeriesAToSeriesBTime \quad (1)$$

Logit regression analysis was performed on the six constructed models, which cover 1-year and 3-year horizons for Seed, Series A, and Series B rounds. Table 1 indicated that the impact of each explanatory variable on the likelihood of securing funding in the subsequent round was as follows. The numerical values presented are the regression coefficients, with the z-statistics provided in parentheses. *** and ** denote statistical significance at the 1% and 5% levels, respectively. For logistic regression models, a z-statistic of approximately 2 or greater generally suggests statistical significance at the 5% level.

The coefficient for "Log10Assets" was statistically significant and positive in all models. This indicates that larger firms have a higher probability of succeeding in fundraising in the subsequent round, which is consistent with findings from previous research. Firm size is considered to function as a signal of business foundation stability and financial health.

Table 1: The results of the regression

the time of evaluation	1-year after Seed	3-year after Seed	1-year after Series A	3-year after Series A	1-year after Series B	3-year after Series B
	coef (z-statistic)	coef (z-statistic)	coef (z-statistic)	coef (z-statistic)	coef (z-statistic)	coef (z-statistic)
Log10Assets	*** 0.3044 (8.353)	*** 0.4472 (9.802)	*** 0.5777 (13.015)	*** 0.4062 (5.8)	*** 0.5218 (9.325)	*** 0.3965 (4.289)
InvestorSuccessRate	*** 2.0237 (13.875)	*** 1.6145 (8.232)	*** 2.4472 (11.239)	*** 2.2238 (6.422)	*** 1.3752 (4.751)	*** 1.2939 (2.641)
TimeSinceLastFundraising	13.584 (0.642)	*** -0.57 (-18.689)	*** -0.1636 (-8.877)	*** -0.1825 (-5.926)	-0.1036 (-0.592)	*** -0.6367 (-9.099)
SeedToSeriesATime			*** -0.7345 (-5.943)	*** -0.7371 (-16.184)	*** -0.118 (-4.384)	*** -0.2236 (-4.184)
SeriesAToSeriesBTime					*** -0.1806 (-6.208)	-0.0534 (-1.173)
	N	N	N	N	N	N
successful	2519	1444	1814	604	1046	274
unsuccessful	4683	6315	3735	4459	2198	2478
total	7202	7759	5549	5063	3244	2752

For “InvestorSuccessRate”, the coefficient also demonstrated a statistically significant positive value across all models. This result suggests that firms raising capital from investors with extensive successful track records exhibit higher subsequent growth potential. This quantitatively supports that investors' discernment (or "ability to identify promising ventures") is a crucial factor in assessing firm value.

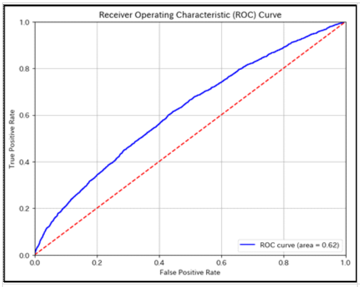
The coefficient for “TimeSinceLastFundraising” showed a statistically significant negative value in many models. This implies that a shorter period since the last fundraising increases the likelihood of progressing to the subsequent round. Startups that have not secured new funding for an extended period tend to be perceived as having stalled growth.

Regarding “SeedToSeriesATime”, “SeriesAToSeriesBTime”, in models evaluating Series A and Series B, the coefficient for this variable showed a significant negative effect. This indicates that firms transitioning to the previous round in a shorter timeframe were more likely to advance to the subsequent round.

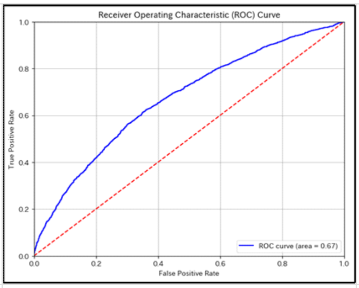
3.2. Evaluation

To evaluate the classification performance of the model, Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves were constructed, and their respective Area Under the Curve (AUC) values were assessed.

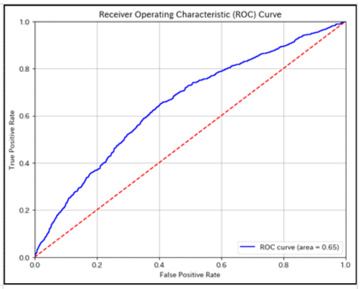
Figure 7-A ROC-curves



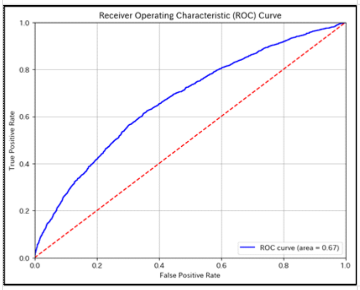
1-year after Seed



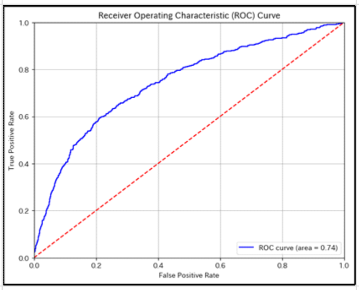
1-year after Series A



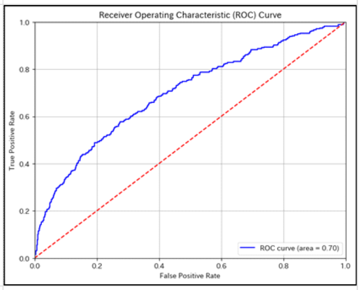
1-year after Series B



3-year after Seed



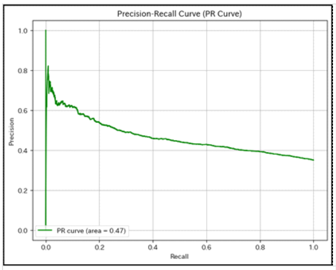
3-year after Series A



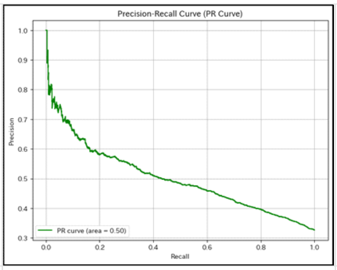
3-year after Series B

The AUC of the model's ROC curve was found to range from 0.62 to 0.74, consistently exceeding random prediction (0.5) and thus confirming a demonstrable level of classification accuracy. This suggests that the present model is valuable as a screening tool for evaluating the future prospects of companies.

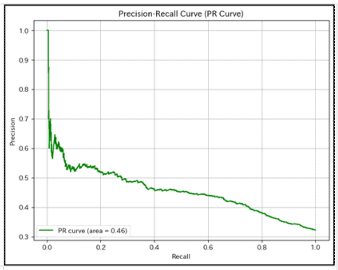
Figure 7-B PR-curves



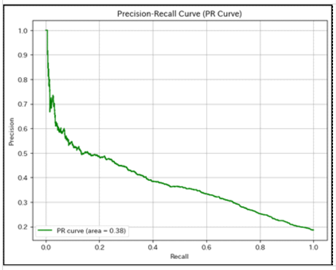
1-year after Seed



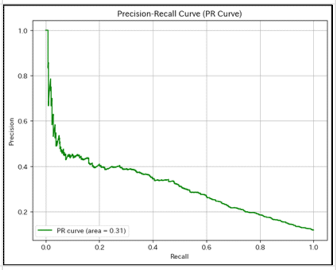
1-year after Series A



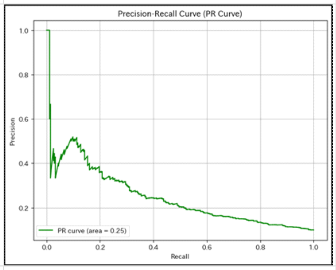
1-year after Series B



3-year after Seed



3-year after Series A



3-year after Series B

Conversely, the AUC of the PR curve exhibited relatively low values, ranging from 0.25 to 0.50. This can be attributed to the imbalanced nature of the dataset, where the proportion of positive instances (companies advancing to the subsequent round) is small within the overall data. With imbalanced data, accurately predicting minority positive instances becomes challenging. Consequently, PR curves, which assess the trade-off between Precision and Recall, typically yield lower AUC values. Furthermore, this inherent difficulty implies that the present model struggles to precisely predict, with high accuracy, companies that will successfully secure funding.

4 . Discussion

4.1. Interpretation of the key analysis results

The results of this analysis quantitatively support the critical factors enabling private firms to advance to subsequent funding rounds

First, we confirm that firm size (total assets) and investor quality (success rate) consistently exert a positive influence on future fundraising potential. Larger companies serve as a robust signal to the market, indicating a stable business foundation and financial credibility. Conversely, investment from investors with a proven track record functions as a multi-faceted evaluation of the firm's prospects. VCs, in their investment decisions, emphasize not only business plans and financial forecasts but also qualitative factors such as marketability, business model, and particularly "management quality" and passion for the business. These qualitative factors, while not directly addressed in our model, may be indirectly reflected through investment from proven investors.

Secondly, temporal factors, specifically "the duration since the last funding" and "the transition period from the previous round," were found to have a negative impact on the probability of successful fundraising. This finding suggests the crucial role of "momentum" in start-up growth. VCs operate under time constraints for investment recovery, prioritizing high-growth potential that promises significant returns in shorter periods over gradual growth. Considering this investment psychology, achieving milestones rapidly and consistently securing funding at a swift pace serves as a powerful signal that the startup is growing at the anticipated pace. Conversely, a prolonged inability to secure follow-on funding indicates operational stagnation or the presence of underlying issues, which can contribute to VCs perceiving the investment as high-risk. Consequently, the "temporal factors" in this model, defined based on the observed outcomes of new investment decisions by VCs, are considered to serve as a proxy variable for growth expectations within the VC investment decision-making process.

4.2. Practical Implications and Model Limitations

Given that the model's ROC-AUC achieves a satisfactory level, it demonstrates utility as a screening tool for evaluating the future potential of private companies. This enables efficient identification and prioritization of companies with relatively high growth potential from a large pool of investment candidates.

However, the low PR-AUC score also indicates limitations in solely relying on this model for final investment decisions. In practical investment decision-making and credit evaluations, a comprehensive assessment is essential, necessitating an integration of the model's probabilistic predictions with qualitative factors such as business content, market environment, and the quality of the management team. VCs often emphasize factors such as "management quality" and "passion for the business," which are not fully captured by purely quantitative data. This highlights both the inherent limitations of the model and reaffirms that practical credit evaluation is a multifaceted process integrating both quantitative and qualitative analyses. Consequently, this model can be positioned as a valuable initial step for efficient candidate screening within this broader, comprehensive evaluation process.

5. Conclusion and Future work

This paper constructed models to predict the likelihood of private companies progressing to their subsequent funding round, leveraging FactSet data. It also quantitatively identifies the factors contributing to this success. The analysis revealed three significant factors:

- ✓ Larger firm size: This serves as a signal of a stable foundation and financial health.
- ✓ Support from investors with a proven track record: Investment from experienced investors can signal the company's growth potential.
- ✓ Rapid progression through growth stages: The presence of "momentum" in the company's growth serves as a crucial signal for assessing future growth expectations.

These findings underscore the importance of considering not only traditional financial metrics but also capital structure-related information, such as funding history, and corporate growth momentum when assessing the credit risk and growth potential of private companies.

This research is expected to contribute to the establishment of more advanced risk management methodologies for investments and financing in private companies. To further enhance the predictive accuracy of the model, future research challenges include the following:

- ✓ Variable Expansion: Beyond financial indicators such as profitability and cash flow metrics, consider incorporating non-financial data, specifically human capital information like the management team's professional backgrounds.
- ✓ Application of Advanced Methodologies: In addition to logit regression models, the application of advanced machine learning models should be investigated, such as Graph Neural Networks, which can analyze the intricate network structures between investors and startups in greater detail. This approach is anticipated to enhance predictive accuracy.
- ✓ Broadening of Scope: While this study focused on USD-denominated transactions, it is crucial to expand the analysis to include private company data from other regions, including Japan, and to verify the generalizability of the finding.

References

[1] Sasaki, W., et al. "The impact of investors with discernment on the subsequent financing of startup companies." *Securities Analysts Journal*, 62(9), pp. 88–98 (2024) (in Japanese).
[2] Hosono, Kaoru, and Miho Takizawa. "Motives for IPOs by Unlisted Firms and Post-Listing Corporate Performance." Research Institute of Economy, Trade and Industry

(Japan), RIETI Discussion Paper Series (2015) (in Japanese).