Pontifícia Universidade Católica
do Rio de Janeiro

Eduardo Oliveira Marinho

Essays on Finance, Macrofinance, and Machine Learning applied in Finance.

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor : Prof. Márcio Gomes Pinto Garcia
Co-advisor: Prof. Marcelo Cunha Medeiros

Rio de Janeiro
December 2025

Pontifícia Universidade Católica
do Rio de Janeiro

Eduardo Oliveira Marinho

Essays on Finance, Macrofinance, and Machine Learning applied in Finance.

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the Examination Committee.

Prof. Márcio Gomes Pinto Garcia
Advisor
Department of Economics – PUC-Rio

Prof. Marcelo Cunha Medeiros
Co-advisor
Department of Economics – UIUC

Prof. Walter Novaes
Department of Economics – PUC-Rio

Prof. Bernardo Ricca
Department of Economics – Insper

Prof. Alan de Genaro Dario
Department of Finance – FGV EAESP

Prof. Ruy Ribeiro
Department of Economics – Insper

Rio de Janeiro, December 15th, 2025

Eduardo Oliveira Marinho

BA in Economics from the Fluminense Federal University (UFF) in 2007.

MSc in Macroeconomics and Finance from PUC-Rio in 2019.

Bibliographic data

Marinho, Eduardo Oliveira

Essays on Finance, Macrofinance, and Machine Learning applied in Finance. / Eduardo Oliveira Marinho; advisor: Márcio Gomes Pinto Garcia; co-advisor: Marcelo Cunha Medeiros. – Rio de janeiro: PUC-Rio, Departamento de Economia, 2025.

v., 162 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia – Teses. 2. Finanças – Teses. 3. Macro finanças – Teses. 4. Econometria – Teses. 5. Venture Capital. 6. Política Monetária. 7. Investimento. 8. Modelos de escolha discreta. 9. Aprendizado por máquina. 10. Private Equity. 11. Performance. 12. Limited Partners. 13. Inflação implícita. 14. Estrutura a termo da curva de juros. 15. Expectativas. 16. PLN. 17. Processamento de linguagem natural. 18. Pesquisa Focus.
I. Garcia, Márcio Gomes Pinto. II. Medeiros, Marcelo Cunha. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. IV. Título.

CDD: 330

To my daughters.

Acknowledgments

I would like to especially thank Maristela for all her support with our children during my absences, for the material support throughout this journey, and for supporting my decision to resign from my last job to dedicate myself to a dream: to become PhD in economics at one of the most renowned institutions in Brazil;

I would also like to thank my friends Inaiara and Kalany, and my family, especially Nilça, Néia, and Caroline, for all their support with my daughters during my various absences over these almost 5 years;

To my Co-advisor, Marcelo Medeiros, one of those responsible for this journey since an unassuming invitation at the end of my professional master's degree. For all the teachings: from his field, about the academic career, and for all the support during my stay in Illinois for the sandwich doctorate;

To my advisor, Márcio Garcia, for his constant support and necessary demands;

To Professor Walter Novaes, from whom I learned a great deal in my field (Finance), and who always demanded my best, drawing out the best of me;

To Professor Ruy Ribeiro, also one of those responsible for this journey, with his letter of recommendation, and for all the parallel learning I had with him during my doctoral studies;

To my co-authors Sebastião Oliveira, José Carlos Abreu, and Jeronymo Marcondes Pinto, who worked with me on the articles that make up this thesis;

To the entire faculty of Economics at PUC, for the excellent course I took during my doctorate, difficult moments of great dedication, but magical moments of great personal development and learning at the cutting edge of science;

To the administrative team of the economics department, Bianca and Priscilla, for all their promptness and support throughout this journey.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001, and Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ)

Abstract

Marinho, Eduardo Oliveira; Garcia, Márcio Gomes Pinto (Advisor); Medeiros, Marcelo Cunha (Co-Advisor). **Essays on Finance, Macrofinance, and Machine Learning applied in Finance.**. Rio de Janeiro, 2025. 162p. Doctoral dissertation – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation consists of three essays on Finance, Macrofinance, and Machine Learning applied to Finance. The first chapter provides novel evidence on how monetary policy shapes venture capital (VC) markets. We document that contractionary monetary policy shocks lead VCs to cut back on seed-stage investments while increasing funding to mature firms. This capital reallocation lengthens the time between funding rounds and raises failure rates for early-stage firms, suggesting that monetary tightening can hinder firm survival, innovation, and long-term growth. The second chapter employs high-dimensional econometric models to predict the success of Private Equity funds using information available at the time of fundraising. Our models achieve an accuracy of up to 75% in predicting VC funds with abnormal returns, indicating that data-driven approaches can effectively reduce reliance on subjective judgment in due diligence. Finally, the third chapter investigates breakeven inflation as a measure of disbelief in monetary policy, Machine learning techniques and Natural Language Processing (NLP) are used to evaluate the impact of newspapers on the difference in inflation expectations (breakeven inflation observed through the yield curve versus the Focus survey). The Lasso model is used to observe the main news stories that predict the market and interpret the Brazilian economy over the last years. The results suggest that the Brazilian yield curve market is highly efficient, with main information already incorporated in prices, news data providing limited additional power to predict the future direction of interest rate curves.

Keywords

Venture Capital Monetary Policy Investment Discrete choice models Machine Learning Private Equity Performance Limited Partners Breakeven inflation Yield Curve Expectations NLP Natural Language Processing Focus survey

Resumo

Marinho, Eduardo Oliveira; Garcia, Márcio Gomes Pinto; Medeiros, Marcelo Cunha. **Ensaios em Finanças, Macrofinanças, e Aprendizado de Máquina aplicado a Finanças.**. Rio de Janeiro, 2025. 162p. Tese de doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese consiste em três ensaios em Finanças, Macrofinanças e Aprendizado de Máquina aplicado a Finanças. O primeiro capítulo apresenta novas evidências sobre como a política monetária molda os mercados de *Venture Capital* (VC). Demonstramos que choques de política monetária contracionista levam os investidores de VC a reduzir os investimentos em companhias de estágio inicial, ao mesmo tempo que aumentam o investimento em empresas mais maduras. Essa realocação de capital prolonga o tempo entre as rodadas de investimento e eleva o nível de falência de empresas em estágio inicial, sugerindo que o aperto monetário pode prejudicar a sobrevivência, a inovação e o crescimento a longo prazo destas empresas. O segundo capítulo utiliza modelos econométricos de alta dimensão para prever o sucesso de fundos de *Private Equity* usando informações disponíveis no momento da captação de recursos. Nossos modelos alcançam uma acurácia de até 75% na previsão de fundos de VC com retornos anormais, indicando que abordagens baseadas em dados podem reduzir efetivamente a dependência de julgamentos subjetivos no processo de *due diligence*. Por fim, o terceiro capítulo investiga a inflação implícita como medida de descrença na política monetária. Técnicas de aprendizado de máquina e processamento de linguagem natural (PLN) são utilizadas para avaliar o impacto de noticias de jornal na diferença entre as expectativas de inflação (inflação implícita observada por meio da curva de juros versus a pesquisa Focus). O modelo Lasso é utilizado para observar as principais notícias que influenciam o mercado e interpretam a economia brasileira nos últimos anos. Os resultados sugerem que o mercado brasileiro de curvas de juros é altamente eficiente, com as principais informações já incorporadas aos preços. Noticias de jornal oferecem poder adicional limitado para prever a direção futura das curvas de juros.

Palavras-chave

Venture Capital Política Monetária Investimento Modelos de escolha discreta Aprendizado por máquina Private Equity Performance Limited Partners Inflação implícita Estrutura a termo da curva de juros Expectativas PLN Processamento de linguagem natural Pesquisa Focus

Table of contents

List of figures

List of tables

1
The Impact of Monetary Policy on Venture Capital Finance

Abstract. We provide novel evidence on how monetary policy shapes venture capital (VC) markets. Contractionary policy shocks lead VCs to cut back on seed-stage investment while increasing funding to mature firms, suggesting a reallocation of capital toward shorter-duration, more predictable ventures. Tighter conditions lengthen the time between funding rounds and raise early-stage firms' failure rates, highlighting the destructive side of reallocation. Monetary tightening also affects VC fundraising activity, prolonging cycles and reducing new fund launches, which constrains future investment capacity. Overall, our findings reveal a new channel through which monetary policy can hinder firm survival, innovation, and long-term growth.

Keywords: Venture capital, monetary policy, investment.

1.1 Introduction

The transmission of monetary policy to the corporate sector has been a central topic in economics and finance.[1] Most of this literature has emphasized the "debt channel", reflecting the direct link between interest rates and firms' borrowing costs. This focus has been shaped by influential theoretical frameworks that highlight how negative economic shocks tighten financial conditions through credit markets Bernanke *et al.* (1999). However, recent evidence suggests that monetary policy significantly operates through non-debt channels. For instance, Almeida *et al.* (2025) show that constrained firms seeking equity financing, not debt financing, are the most affected by contractionary shocks, suggesting that frictions beyond debt markets play an important role.

This growing evidence motivates our examination of how monetary policy affects private markets such as venture capital (VC). Venture capital is a key source of financing for high-growth, early-stage firms that typically lack access to public markets or traditional debt. These firms often depend heavily on external equity and face significant cash flow uncertainty, making them highly sensitive to macro-financial conditions. Monetary policy can influence VC activity by affecting both the cost of capital and investor risk appetite. Indeed, Figure 1.1 shows a strong negative correlation[2] between VC deals and U.S. 10-year Treasury rates. Despite the importance of this relationship, there is no comprehensive empirical analysis of the impact of monetary policy on venture capital finance.

We provide novel, comprehensive evidence on how monetary policy shocks affect VC markets. We show that, on average, contractionary monetary policy shocks reduce VC investment, with the largest effects in the earliest stages of financing. The decline occurs on both extensive and intensive margins, with fewer startups receiving funding and smaller amounts being invested per deal.

[1]Examples include Gertler & Gilchrist (1994); Ottonello & Winberry (2020); Cloyne *et al.* (2023); Jungherr *et al.* (2024).

[2]The correlation for this window period is -0.74

At the same time, investment in more mature firms increases, consistent with a reallocation of capital away from riskier, longer-duration ventures toward firms with nearer-term and more predictable cash flows, whose valuations are less sensitive to rising discount rates. This reallocation demonstrates that monetary policy not only affects the overall level of VC investment but also its allocation across the firm life cycle, impacting which innovations are financed and shaping the long-term growth trajectory of the startup ecosystem.

To understand these results and derive new insights, we develop a simple theoretical framework in which a risk-averse VC allocates funds between early-stage projects (longer horizon, greater uncertainty) and late-stage projects (shorter horizon, more predictable). Our data show that about 63.8% of VCs in our sample adopt a multi-stage strategy, investing in two or more different stages (see Section 1.3). For specialized funds (single-stage strategy), re-allocations typically occur within a stage (e.g., from Pre-Seed to Seed), which, aggregating across single-stage funds, gives rise to shifts from early- to late-stage activity. Our model abstracts from these finer distinctions by considering two stages that we call "early-stage" and "late-stage" projects, capturing the main reallocation margin in a tractable way. The static allocation in the model can also be interpreted dynamically: a shift toward late-stage projects corresponds to either (i) terminating or not initiating early-stage projects, or (ii) extending approval to later-stage projects.

In the model, a rise in interest rates lowers present values—i.e., reduces entry prices—more strongly for long-horizon investments, a duration effect. After a contractionary shock, VCs face a timing trade-off: they can invest immediately at prevailing valuations, assuming some rigidity in entry prices, or wait for valuations to adjust, incurring deployment costs in the meantime. If the expected post-shock price correction is large relative to the cost of waiting, investors optimally delay early-stage deals ("wait-and-see" behavior); if not, capital shifts toward later-stage opportunities. Because these responses only partially offset the rate-induced decline in valuations, the model predicts a net reduction in the number and value of VC deals, longer intervals between funding rounds, and a compositional shift from seed and early-stage toward later-stage activity.

A fundamental challenge with testing these predictions is the lack of detailed deal-level data and the endogeneity of monetary policy to macroeconomic conditions. To overcome these challenges, we construct a granular dataset of VC deals from 1990 to 2023, covering 30,805 unique firms. The data include information on financing stage, deal volume, general partners (GPs), and firm industry. To address endogeneity, we employ high-frequency monetary policy shocks (MPS), measured as changes in federal funds rate futures in the 30-minute window around FOMC announcements. Building on Bauer & Swanson (2023), we refine these shocks by removing components correlated with macroeconomic and financial data, ensuring that they capture exogenous monetary surprises.

A natural concern is that VC deals often take several months to move from initial conversation to final closing Gompers *et al.* (2020). This might raise doubts about whether high-frequency monetary policy shocks can meaningfully affect observed deal activity in the short run. It is important to note, however, that the relevant margin is not only the time to close conditional on a deal being completed, but also whether initial conversations ultimately progress to a closing at all. Contractionary monetary policy shocks can stall or prevent deals from being finalized, thereby affecting the extensive margin of investment. In this sense, VC activity should be viewed as a continuous process, with some potential deals advancing to completion and others being terminated along the way. Our empirical design captures this dynamic by focusing on how monetary policy shocks shift both the likelihood and timing of deal completion, not just the conditional duration of negotiations.[3]

We start by showing that, on average, a contractionary monetary policy

[3]Because we observe completed deals at the quarterly level, a contractionary MPS that prevents ongoing conversations from culminating in a closing will be reflected as a decline in both the number and volume of deals. Our estimates therefore capture not only the timing of negotiations but also the extensive margin of whether deals close.

shock significantly reduces both the volume and the number of VC deals. In response to a 25 bps contractionary shock, the number of deals declines by 2.33%. The results for the total volume of VC deals are consistent with our findings for the number of investments: in response to the same shock, the total volume of VC deals decreases by 5.4%. These magnitudes are comparable to the estimated reduction in traditional investment in physical assets, such as capital expenditures, which falls by 5–6% following a 25 bps contractionary shock.[4] These findings indicate that the impact of monetary policy on VC investment—and, consequently, on innovation—is economically significant.

We next examine the heterogeneous responses of VC investment across financing stages. Using Preqin's classification system, we categorize deals into Seed (seed, angel, grant), Early Stage (Series A), Middle Stage (Series B), Late Stage (Series C–L, add-ons, growth capital), Exit (merger, PIPE, pre-IPO), and Venture Debt. We show that the decline in VC investment volume is concentrated entirely in seed-stage deals, while middle-stage transactions experience an increase in funding. In the VC industry, capital is raised through closed-end vehicles based on commitments secured during fundraising, which represent guaranteed available capital that fund managers (GPs) can deploy when attractive opportunities arise. Because limited partners (LPs) cannot withdraw these commitments during the fund's life, the total pool of capital remains available for deployment regardless of macroeconomic shocks. Therefore, when investment in early-stage startups declines, it is natural to see a corresponding increase in later-stage funding as part of a reallocation toward ventures with shorter durations and more predictable cash flows.

Monetary policy can also affect the timing of follow-on investments. In venture capital, startups typically rely on a sequence of funding rounds to reach key development milestones, and the time between rounds—the inter-round duration—can be critical for survival. Delays in follow-on funding can raise

[4]See Almeida *et al.* (2025) and others.

the risk of firm failure, force cutbacks in operations, or push startups into less favorable financing arrangements. If VCs reallocate scarce capital away from early-stage firms toward later-stage firms during tightening periods, we would expect inter-round durations to lengthen for the former and shorten for the latter. Consistent with this prediction, we find that contractionary shocks delay follow-on investments for early-stage firms while accelerating them for later-stage firms, underscoring the role of monetary policy in shaping not only the scale but also the pace of VC capital deployment.

We next quantify the destructive side of capital reallocation and show that monetary policy, beyond influencing the flow of investment, also shapes the survival of young companies. We document that contractionary monetary policy shocks significantly increase the probability of failure for early-stage firms, whereas expansionary shocks have little effect.[5] A 25 bps contractionary shock increases the probability of firm failure one year later by over 4 percentage points, an increase of 30.7% relative to the baseline probability of failure. This asymmetry underscores the destructive role of monetary tightening in the venture ecosystem, revealing that young firms are far more sensitive to adverse shocks than they are responsive to positive ones. The finding is consistent with theories emphasizing financing frictions, where tightening quickly constrains capital supply and forces failures, while easing restores funding only gradually. These results point to important real consequences of monetary tightening for firm survival, innovation, and long-term economic growth.

Another margin through which monetary policy can influence VC activity is via its effects on the fundraising of VC funds, which is a key channel shaping future investment capacity. Fundraising represents the supply side of the VC market: it determines the pool of capital that funds can allocate to new deals. When monetary conditions tighten, investors may reduce their commitments to

[5] Firm Failure is an indicator equal to one if a firm ceases to appear in the venture capital deal database for at least one year following its last observed funding round.

new funds, either because of higher discount rates, portfolio rebalancing toward safer assets, or limited liquidity. This translates into longer fundraising cycles. Our estimates show that a 25 bps contractionary monetary policy shock increases fundraising duration by roughly 93 days (about three months). This extension lengthens the typical fundraising cycle from 15 months to approximately 18 months, representing a sizable economic effect in an industry where timely capital commitments are critical.

We conclude our analysis by subjecting our baseline results to several robustness tests. First, we re-estimate our main specification using monetary policy shocks that move long-term interest rates, which are arguably more relevant for VC valuations and investment flows. Second, we examine asymmetric effects of monetary policy shocks and show that both contractionary and expansionary shocks affect VC investment, with the former having a stronger impact. Third, we estimate our results using alternative specifications and include fixed effects that absorb industry-time shocks, allowing identification from within-industry variation across firms. Overall, our results are robust to these tests, highlighting the importance of monetary policy in shaping VC markets and its potential to reduce innovation and slow long-run economic growth.

1.2 Related Literature

Our paper contributes to several strands of the corporate finance, monetary policy and private equity (PE) literature. First, our work relates to the literature that examines how financial frictions, risk, and monetary policy affect venture capital and investment decisions. Nanda & Rhodes-Kropf (2013) find that VCs invest in riskier and more innovative startups in hot markets (rather than just worse firms). Ewens *et al.* (2018) study an increased prevalence of a "spray and pray" investment approach – where investors provide a little funding and limited governance to an increased number of startups that they are more likely to abandon, but where initial experiments significantly inform beliefs about the future potential of the venture.

We contribute to the PE literature that emphasizes the role of macroeconomic conditions in the industry Gompers *et al.* (1998); Gompers & Lerner (2000); Kaplan & Schoar (2005); Gompers *et al.* (2016); Robinson & Sensoy (2016). These studies focus on understanding market liquidity and its relationship with performance. More related to our work, Robinson & Sensoy (2016) identify a procyclical systematic component in capital calls for LPs. We examine the specific role of monetary policy and employ an identification strategy that isolates the real effects of interest rates on VC deals. Specifically, our paper leverages granular VC deal-level data combined with high-frequency monetary policy shocks to provide the first comprehensive empirical evidence on how monetary policy influences VC markets. We also document heterogeneous effects across financing stages, industries, and types of funds—an aspect not addressed by previous research.

Our paper is related to Boyer *et al.* (2023), who demonstrate that discount-rate risk creates a wedge between cash-flow based and market-based measures of private equity performance. We extend this line of reasoning to venture capital, documenting that monetary policy shocks, which operate by shifting discounting rates, affect the allocation of VC capital across stages. Our framework also relates to the recent theory of liquidity in private equity developed by Maurin *et al.* (2023). In their model, LP liquidity constraints and GP incentive problems generate inefficiencies in the timing of capital deployment, leading to accelerated drawdowns or delayed investments depending on liquidity conditions. We capture a parallel mechanism in a simplified two-period VC setting: following a contractionary monetary policy shock, funds face a trade-off between investing immediately at lower valuations or delaying deployment at the cost of waiting. Whereas Maurin *et al.* (2023) focus on how illiquidity shapes LP–GP contracts and the cyclicality of private equity activity, our model highlights the macro-financial dimension of this problem by showing how monetary policy shocks alter discount rates and liquidity premia, thereby influencing both the timing and the stage allocation of VC investment.

Many empirical and theoretical studies have studied how monetary policy transmit to firm investment Gertler & Gilchrist (1994); Ottonello & Winberry (2020); Cloyne *et al.* (2023); Jungherr *et al.* (2024); Perez-Orive *et al.* (2024). Close to our work, Almeida *et al.* (2025) document that public constrained firms seeking equity financing are the most affected by monetary policy by showing that equity-focused constrained firms endure more substantial declines in stock prices and implement deeper cuts in capital expenditure and R&D when faced with a contractionary monetary policy shock. Beyhaghi *et al.* (2024) document the effects of monetary policy on private firms, but their analysis primarily reflects debt-financed firms covered in Federal Reserve administrative data and does not fully capture VC-backed firms. Our contribution is to provide direct evidence on how monetary policy affects venture capital markets—a key source of financing for early-stage, high-growth firms with the potential to drive innovation.

Also related to our work, Ma & Zimmermann (2023) use aggregate VC investment as a measure of innovation and document that monetary policy can affect the productive capacity of the economy in the longer term. While their analysis provides important aggregate-level evidence, it does not capture potential heterogeneity across the VC market. The granularity of our data allows us to move beyond aggregate patterns and uncover novel heterogeneous effects of monetary policy across deal-financing stages, as well as its influence on the survival of young companies and fundraising activity. In particular, our paper is the first to establish that monetary policy plays a unique role in VC markets by disproportionately reducing access to early-stage financing, where firms tend to be riskier, more innovative, and have longer exit horizons. This channel highlights how contractionary policy can shift the allocation of capital toward later-stage and lower-risk ventures, with important real consequences for the survival of early-stage firms.

We also relate to recent literature that studies the interactions of monetary policy and innovation. Döttling & Ratnovski (2023) find that intangible investment responds less to monetary policy than tangible investment. Similarly, Caggese & Pérez-Orive (2022) show that lower interest rates are less stimulating for high-intangible firms than to high-tangible firms. We show that monetary policy can significantly affect firms' innovation and, potentially, long-term growth by reducing VC investment, with an effect that is quantitatively similar to the reduction in traditional physical investments.

1.3
A Simple Theoretical Framework

1.3.1
Setup and VC features

Monetary policy shocks affect venture capital investment decisions through both a duration channel, longer-horizon projects suffer greater declines in present value when discount rates rise, and a timing channel, funds can choose whether to invest immediately or wait for entry valuations to adjust to new financing conditions.

To capture these mechanisms, we develop a simple two-period framework in which a capital-constrained, risk-averse VC fund allocates investment across early-stage (long duration, higher uncertainty) and late-stage (short duration, lower uncertainty) projects. Following a contractionary monetary policy shock, VCs face a trade-off: invest immediately at current valuations, or delay deployment in anticipation of a price adjustment, incurring a potential cost of waiting. This structure allows us to model both the possibility of delayed investment and firms' stage reallocation, and to generate empirical predictions about the effect of monetary tightening on VC deal activity.

Table 1.3 shows that 63.8% of VCs in our sample adopt a multistage strategy, investing in two or more different stages. Moreover, 43% of VCs from Preqin explicitly invest in both early- and late-stage deals. This figure exceeds survey-based estimates. For example, a report by Different Funds finds that about 26% of U.S. VC funds pursue strategies that span both early- and late-stage investments.[6] For funds that operate across both early and late stages, reallocation across stages is direct. For the remaining stage-specialized funds, reallocation likely occurs within a given stage—for instance, shifting from Pre-Seed to Seed. At the aggregate level, these within-stage reallocations effectively translate into a broader pattern resembling shifts from early- to late-stage investment.

In our model, we abstract from these finer distinctions and assume only two types of projects—early stage and late stage. This simplification captures the essential reallocation margin while keeping the framework tractable. Moreover, the static allocation in the model can be interpreted dynamically: a shift toward late-stage projects corresponds to either (i) terminating or not initiating early-stage projects, or (ii) extending approval to later-stage projects. This interpretation allows the model to reflect the persistence of investment responses to monetary policy shocks while remaining consistent with the empirical facts observed in the data.

1.3.2 The Model

We build on the literature linking duration-based discounting and VC investment dynamics. We adopt a mean–variance utility framework in which risk-averse VCs face a trade-off between expected return and exposure to uncertainty. Following Holmstrom & Tirole (1997), the VC fund operates under a capital constraint, with its size determined by the total capital commitment. Duration-based discounting implies that longer-horizon investments experience larger declines in expected net present value and returns when interest rates rise Campbell & Viceira (2001).

[6] https://differentfunds.com/research/european-vcs/

The fund allocates capital between early-stage investment (E), which are riskier, long-duration projects with uncertain outcomes, and late-stage investment (L), which are safer, short-duration projects with more predictable outcomes. Let T_i be the length of the time it takes for a type $i \in \{E, L\}$ to deliver returns. We assume $T_E > T_L$ to capture the fact that early-stage projects take longer to mature, making them more sensitive to changes in interest rates.

Let I_E and I_L denote capital allocated to early- and late-stage deals, respectively. The capital constraint is:

$$I_E + I_L \leq K,$$

where K is total committed capital. The per-dollar return on early-stage investment is stochastic, $R_E \sim \mathcal{N}(\mu_E(r), \sigma_E^2)$, while for simplicity, we assume late-stage returns are deterministic $R_L = \mu_L(r)$. This assumption is, in indeed, extremely simplistic; the distribution of returns in the venture capital market in general (both early and late stage) is characterized by outliers in the tail. These outliers are the true drivers of fund returns when many assets cease to exist. Given that, our intention is primarily to characterize the duration effect and the choice between reallocation and delay in investment, suppressing the distribution of returns in the late stage is without loss of generality and would introduce extra complexity to the model beyond our intended purpose.

In the VC industry, returns are often reported as the multiple on invested capital (MOI):

$$\text{MOI}_i = \frac{V_{T_i}}{V_{0_i}},$$

where V_{0_i} is the entry value, V_{T_i} is the terminal (exit) value, and T_i is the project duration in years.

The MOI measure does not account for the time value of money in investment. To compare projects with different maturities, we compute a discounted, annualized return that adjusts the exit value to present value using the risk-free rate r and a VC-specific risk premium rp:

$$\mu_i(r) = \frac{\frac{V_{T_i}}{(1+r+rp)^{(T_i-1)}}}{V_{0_i}} - 1, \quad i \in \{E, L\}.$$

Of note, both MOI and normalized return measures account for the characteristic of this market, where gains are typically linked to a liquidity event far in the future and, in most cases, no cash flows are received until that event occurs.

Differentiating with respect to r yields

$$\frac{\partial \mu_i}{\partial r} = -(T_i - 1)\frac{V_{T_i}}{V_{0_i}(1 + r + rp)^{T_i}} < 0,$$

and since $T_E > T_L > 1$,

$$\left|\frac{\partial \mu_E}{\partial r}\right| > \left|\frac{\partial \mu_L}{\partial r}\right|.$$

Thus, early-stage returns are more sensitive to interest rate changes.

We next extend the model to include an option to delay investment. In our setting, even subtle shocks to short-term interest rates influence decisions in a long-term-oriented market such as venture capital. Changes in short-term rates affect the entire yield curve and shape investors' perception of risk, since risk-free assets yield higher returns. Our theoretical model captures this through two channels: first, rigidity in the market, as entry prices for riskier assets do not adjust immediately; and second, the absence of waiting flexibility for some participants, since funds surprised at the end of the investment period face high waiting costs in our framework. Consequently, an interest rate shock should reset asset prices, but due to rigid adjustments, riskier assets effectively become more expensive. To continue investing in these assets while respecting the risk–return tradeoff, waiting becomes necessary. For some funds, however, waiting is too costly, leading them to reallocate capital toward safer assets that are less affected by the interest rate shock.

Consider a contractionary monetary policy shock at time $t = 0$. The VC fund can invest immediately in period 1 at cost V_1, or wait for the market to adjust entry valuations to the new interest rate and invest in period 2 at cost V_2. The entry price adjustment is therefore defined by $PA = V_1 - V_2 > 0$. However, waiting can be costly. After fundraising, VC funds typically operate in three phases: investment, value creation, and disinvestment. Although the investment phase often spans several years, delaying investment may incur in non-pecuniary costs (reputational damage from slow deployment), or pecuniary costs (if management fees are charged only on invested assets rather than committed capital).It is common to see management fees charged on committed capital at the beginning of the fund, and, after the investment period, charged on the fund's invested equity. The per-dollar cost of waiting for a type j GP is $c_j \geq 0$.[7]

Let $M_i(t) = T_i - t$ denote the remaining years to maturity for project type $i \in \{E, L\}$ if the investment occurs at time $t \in \{1, 2\}$. The annualized, discounted net return from investing in project i at time t is:

$$\mu_i(r_t) = \frac{\frac{V_{T_i}}{(1+r_t+rp)^{M_i(t)}}}{V_{t_i} + c_j \cdot \mathbb{1}_{\{t=2\}}} - 1, \tag{1.1}$$

where $\mathbb{1}_{\{t=2\}}$ is an indicator function equal to 1 if the investment occurs in period 2 (i.e., the GP waited) and 0 otherwise. Intuitively, if the GP invests immediately ($t = 1$), the project is discounted over $T_i - 1$ years and no waiting cost is incurred. If the GP waits ($t = 2$), the project is discounted over $T_i - 2$ years, potentially at a lower entry price V_{2_i}, but the GP pays the waiting cost c_j. Because $T_E > T_L$, early-stage projects lose more value when r_t increases, making them more sensitive to interest rate changes (duration channel). Waiting improves returns only if the price adjustment $(V_{1_i} - V_{2_i})$ exceeds the cost of waiting c_j.

[7]Given the rigidity proposed for early-stage assets, the impact of interest rates on a balance sheet effect mechanism could yield a similar result to the duration effect. However, companies in this segment are not characterized by having significant liabilities. We believe that the most plausible channel explaining the empirical results is the combination proposed here.

The VC is risk-averse ($\gamma > 0$) and chooses allocations to maximize mean–variance utility:

$$\max_{I_E, I_L, I_{E2}, I_{L2}} \quad I_E E[\mu_E(r_1)] - \frac{\gamma}{2} I_E^2 \sigma_E^2 + I_L \mu_L(r_1) + I_{E2} E[\mu_E(r_2)] - \frac{\gamma}{2} I_{E2}^2 \sigma_E^2 + I_{L2} \mu_L(r_2)$$

subject to:

$$I_E + I_L + I_{E2} + I_{L2} \leq K.$$

With deterministic late-stage returns, the first-order conditions yield:[8]

$$I_E^* = \frac{\mathbb{E}[\mu_E(r_1)] - \mu_L}{\gamma \sigma_E^2}, \qquad I_{E2}^* = \frac{\mathbb{E}[\mu_E(r_2)] - \mu_L}{\gamma \sigma_E^2}. \tag{1.2}$$

Define the entry price adjustment $PA \equiv V_{1_i} - V_{2_i} > 0$, so $V_{2_i} = V_{1_i} - PA$. Consider the early-stage project ($i = E$). Waiting dominates investing immediately if and only if:

$$\mu_E(r_2) \ > \ \mu_E(r_1).$$

Using (1.1) with $t = 1$ and $t = 2$, and cancelling the common terminal payoff V_{T_E}, this is equivalent to

$$\frac{1}{(1 + r_2 + rp)^{M_E(2)} \, [\, V_{2_E} + c_j \,]} \ > \ \frac{1}{(1 + r_1 + rp)^{M_E(1)} \, V_{1_E}}.$$

Reorganizing the terms (all terms are positive),

$$(1 + r_2 + rp)^{M_E(2)} \, [\, V_{2_E} + c_j \,] \ < \ (1 + r_1 + rp)^{M_E(1)} \, V_{1_E}.$$

Substitute $V_{2_E} = V_{1_E} - PA$ and $M_E(1) = T_E - 1$, $M_E(2) = T_E - 2$:

$$\left(V_{1_E} - PA + c_j \right) (1 + r_2 + rp)^{T_E - 2} \ < \ V_{1_E} \, (1 + r_1 + rp)^{T_E - 1}.$$

Rearranging yields a simple threshold for the net price concession $PA - c_j$:

[8] See Section 1.A for details on the first-order conditions.

$$PA - c_j \ > \ V_{1_E}\left[1 - \frac{(1+r_1+rp)^{T_E-1}}{(1+r_2+rp)^{T_E-2}}\right] \ \equiv \ \Theta_E(r_1, r_2; T_E) \ . \tag{1.3}$$

When (1.3) holds, waiting strictly increases the early-stage return.

For fixed (r_1, r_2), the term Θ_E depends on T_E. A larger T_E magnifies the role of discounting across periods, making the RHS more sensitive to the path of rates. Intuitively, for sufficiently large PA relative to c_j, waiting is more attractive for long-duration projects, consistent with the idea that early-stage is most affected by discount-rate movements and valuation resets.

Using (1.2), early-stage allocations are proportional to return spreads:

$$I_E^* \propto \mathbb{E}[\mu_E(r_1)] - \mu_L, \qquad I_{E2}^* \propto \mathbb{E}[\mu_E(r_2)] - \mu_L.$$

We derive two empirical predictions and take them to the data:

Prediction 1 (Delay: wait-and-see). If $PA - c_j \ > \ \Theta_E(r_1, r_2; T_E)$, then $\mu_E(r_2) > \mu_E(r_1)$. For GP types with low c_j and sufficiently large PA, this implies $I_{E2}^* \geq I_E^*$ and a *delay* of early-stage deployment to $t = 2$. In the cross-section, we expect longer times between early-stage rounds after tightening.

Prediction 2 (Firms' Stage Reallocation). If $PA - c_j \leq \Theta_E(r_1, r_2; T_E)$ (i.e., waiting is not worthwhile) and the rate increase reduces $\mu_E(r_1)$ relative to μ_L sufficiently, then

$$\Delta_E(r) \equiv \mathbb{E}[\mu_E(r)] - \mu_L \quad \text{falls,}$$

and

$$\frac{\partial I_E^*}{\partial r_1} = \frac{1}{\gamma \sigma_E^2}\left(\frac{\partial \mathbb{E}[\mu_E(r_1)]}{\partial r_1} - \frac{\partial \mu_L(r_1)}{\partial r_1}\right) < 0,$$

because $|\partial \mu_E / \partial r| > |\partial \mu_L / \partial r|$ by duration. Hence capital reassigns toward late-stage:

$$I_L^* + I_{L2}^* = K - I_E^* - I_{E2}^* \quad \text{rises as } r \text{ increases.}$$

This yields a cross-sectional reallocation from early- to late-stage after tightening.

Overall, Delay (Prediction 1) reduces contemporaneous deal counts and volumes; Reallocation (Prediction 2) increases late-stage activity but does not fully offset the early-stage contraction. With heterogeneous c_j across GPs and limited PA in practice, the offsets are incomplete, implying a net negative average effect of contractionary shocks on total VC deals, consistent with the empirical results.

1.4 Data

We use three main data sources: Preqin, FRED, and the monetary policy shocks from Bauer & Swanson (2023). The first data set provides deal-level information on VC investments, the second contains country-level macroeconomic variables, and the third offers information on monetary policy shocks. Our baseline sample covers the years 1990 to 2023 and includes 30,805 unique firms. Table 1.1 shows that the total number of deals in our sample is 103,155, with a total volume of USD 868,114.3 million. The average number of deals per year is 882, and the average deal volume is USD 15.1 million. Using Preqin data offers two main advantages. First, we have the exact date of each deal, which allows us to construct a long panel for studying the effect of monetary policy shocks. Second, Preqin provides comprehensive deal-level information, including the stage of VC financing, the deal volume, and the industry of the firm receiving the investment. Figure 1.2 shows the quarterly trend in the number of VC deals per industry over time.

1.4.1 Macroeconomic Variables

When estimating local projections using firm-level data, we control for macroeconomic variables to avoid confounding the effects of monetary policy with the underlying macro conditions that prompted it. We collect the following macroeconomic variables from FRED: CPI (Consumer Price Index), Employment

Ratio (Employment-Population Ratio), GDP Growth (Change in Real Gross Domestic Product). We also use the Excess bond premium of Gilchrist & Zakrajšek (2012).

1.4.2 Measuring Monetary Policy Shocks

Given that monetary policy is endogenous to macroeconomic conditions (e.g., the central bank raises interest rates in response to inflation), we use high-frequency identification to extract monetary policy shocks. It is crucial to use "pure" monetary policy shocks to identify the effects of monetary policy on VC investment because these shocks represent unexpected or exogenous changes in monetary policy, untainted by current or anticipated economic conditions.

The high-frequency approach measures the surprise element of monetary policy by evaluating high-frequency interest rate changes around monetary policy announcements. In particular, the surprise component is constructed by price changes of Federal funds rate futures contracts in the 30-minute window around FOMC announcements. The identifying assumption is that all public information is already incorporated into the prices at the beginning of the narrow window and therefore contains no other news that affect interest rate expectations.

However, as recent studies have shown, this methodology might capture the "information effect" of monetary policy, which could bias estimates of monetary policy transmission Nakamura & Steinsson (2018). For example, an unexpected monetary easing might lead to pessimism among the market participants about economic fundamentals, and these negative changes in beliefs about fundamentals might immediately manifest themselves in the price changes of Federal funds rate future contracts.

To circumvent this concern, in our main analysis we use the monetary policy shocks from Bauer & Swanson (2023). The MPS data are based on the responses of the first four quarterly Eurodollar futures contracts. Following Ottonello & Winberry (2020) and others, we aggregate these high-frequency shocks to a quarterly frequency in order to merge them with our deal-level data and macro controls (see Subsection 1.4.1). Table 1.2 provides the summary statistics and Figure 1.3 plots the monetary policy shocks over time.

Compared to conventional MPS measures, Bauer & Swanson (2023) improve the relevance of monetary policy surprises by substantially expanding the set of monetary policy announcements. In addition to Federal Open Market Committee (FOMC) announcements, they include press conferences, speeches, and testimonies by the Federal Reserve chair. They also address concerns about the exogeneity of the shocks by removing the component of the monetary policy surprises that is correlated with economic and financial data (nonfarm payrolls surprise, employment growth, yield curve slope, S&P 500, commodity prices, and treasury skewness).

1.5 Empirical Strategy and Main Results

This section presents our empirical strategy and main findings. First, we examine how monetary policy shocks affect VC investment. Second, we study how contractionary shocks alter VC terms. Third, we investigate the consequences of a tighter monetary policy environment for the failure rate of early-stage firms. Fourth, we analyze the dynamics of VC investment over time. Fifth, we assess the impact of these shocks at the GP level. Finally, we evaluate how monetary policy affects VC fund fundraising activity.

1.5.1 VC investment Response

1.5.1.1
Average Response

We study the responses of deal-level VC investment to monetary policy shocks by estimating the following specification:

$$\Delta y_{i,t} = \alpha_i + \beta_1 \, MPS_t + \gamma_1^{'} \, X_{t-1} + \epsilon_{i,t}, \tag{1.4}$$

where $y_{i,t}$ is the outcome variable (the log of deal financing size and the log of the number of VC investments) at t quarters after the MPS at time t. MPS_t refers to the monetary policy shocks from Bauer & Swanson (2023), aggregated to the quarterly level. X_{t-1} is a vector of controls that includes the lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. α_i represents firm fixed effects, and standard errors are robust and clustered at the firm level.

Table 1.4 summarizes our findings. In response to a 25 bps contractionary shock, the volume of VC deals decreases by 5.4%. The results for the number of investments (deals) are consistent with our previous findings regarding deal financing size. Specifically, in response to the same contractionary shock, the number of investments decreases by 2.33%.

We next estimate the effect of monetary policy shocks across industries. To do so, we follow Preqin's classification and divide our sample into ten categories: Business Services (BS), Consumer Discretionary (CD), Energy & Utilities (EU), Financial & Insurance Services (FI), Healthcare (H), Industrials (I), Information Technology (IT), Raw Materials & Natural Resources (MR), Real Estate (RE), and Telecoms & Media (TM). Tables 1.5 and 1.6 present the results for deal financing size and the number of investments, respectively. A contractionary monetary policy shock has a strong negative effect across all industries. We also document that financial sectors are among the industries most sensitive to monetary policy shocks. One possible explanation for this result is that these sectors are typically interest-rate sensitive. For instance, in the financial sector, fintech firms can face reduced credit demand and higher funding costs in response to higher rates, thereby affecting their profitability.

1.5.1.2 Financing Stage

Our previous findings show that contractionary monetary policy shocks significantly reduce VC investment. In this section, we examine which financing stage experiences the most pronounced decrease in VC investment. Following Preqin's classification, we define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C- series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). We drop "secondary stock purchase" and "unspecified round" as these are not classified as financing stages. We then estimate Equation (1.4) for each of these five stages.

Table 1.7 presents the results for deal financing size. In response to a contractionary shock, the decrease in VC investment is significant only for firms in the earlier stages of financing. Specifically, a 25 bps contractionary shock leads to a 11.58% decline (relative to the mean) in VC investment for firms in the seed stage. Table 1.8 shows that results using the number of investments as the outcome variable align with the previous findings for deal financing size: the drop in the number of VC investments is most pronounced for firms in the seed stage.

Examining both deal volume and the number of investments provides complementary insights into how monetary tightening affects VC activity. Deal volume captures the aggregate amount of capital deployed, while the number of investments reflects how many distinct financing deals are made. A decline in volume can result from smaller amounts invested per deal, fewer deals, or both, whereas a decline in the number of investments indicates that fewer startups are receiving funding regardless of the amount invested in each deal. In our case, both measures fall, indicating a broad-based contraction in VC activity in which fewer firms are funded and the average investment per deal also declines. This reinforces the interpretation that contractionary shocks affect not only the scale but also the distribution of VC financing across firms.

Table 1.7 also indicates a comparable reallocation of investment toward middle-stage firms, as the coefficient for the effect of MPS in this group is significant. A 25 bps contractionary shock leads to a 10.55% increase (relative to the mean) in VC investment for middle-stage firms. In the VC industry, capital is raised through closed-end vehicles, where funds operate based on capital commitments secured during the fundraising process. These commitments represent guaranteed available capital, which fund managers (GPs) can call upon when they identify attractive investment opportunities. Importantly, LPs cannot withdraw their commitments during the fund's life, meaning the capital remains available for deployment regardless of macroeconomic shocks. Therefore, given that VC investment declines for early-stage startups, it is expected that investment would increase for later stages as part of this reallocation.

Overall, this reallocation reflects the sensitivity of valuations to interest rate changes. When the discount rate rises, the present value of distant cash flows declines more sharply, making early-stage firms—whose revenues may only materialize 5, 10, or more years ahead—less attractive. Later-stage firms, with nearer-term and more predictable cash flows, experience smaller valuation losses, face lower financing costs, and thus become relatively more appealing during periods of monetary tightening. Understanding this shift is important because it highlights that monetary policy not only affects the level of VC investment but also its allocation across the firm life cycle, with implications for the types of innovations that are funded and the long-term growth potential of the startup ecosystem.

1.5.1.3
Number of Quarters Since First Investment

While our baseline results show that early-stage firms are more exposed to contractionary monetary policy shocks than later-stage firms, the early-stage segment itself is heterogeneous. For example, a startup that received its first VC investment last quarter is likely at a very different stage of development than a firm that raised its first round three years ago but has not yet progressed to a later-stage classification. The former may still be building its core product and have minimal operating history, while the latter may have a more established team, customer base, and track record of performance, giving investors greater confidence in its prospects.

Startups that have recently received their first VC investment may be especially vulnerable. These firms typically have the longest cash flow durations and, crucially, investors possess limited information about their prospects, as this is their first financing round.[9] They also tend to require substantial follow-on funding before reaching profitability or liquidity. To test this idea, we restrict the sample to early-stage firms and construct a variable $qsfi_{i,t}$, which represents the number of quarters since the firm's first investment. Each quarter, we sort firms into terciles based on $qsfi_{i,t}$ and estimate Equation (1.4) separately for each subsample.

Tables 1.9 and 1.10 report the results. For both outcome variables—deal financing size and number of investments—the decline is much stronger for firms in the bottom tercile of $qsfi_{i,t}$, indicating that contractionary monetary policy shocks reduce VC investment most significantly for early-stage firms that have only recently received their first investment. For instance, in response to a 25 bps contractionary shock, the volume of VC investments decreases by 37% for firms in the bottom $qsfi_{i,t}$ group, compared to a 2.4% decline relative to the mean for firms in the top $qsfi_{i,t}$ group.

[9] There is evidence that asymmetric learning can affect venture capital decisions (see Hochberg *et al.* (2014)).

These results refine our earlier findings: within the early-stage segment, the most-recently funded startups face the sharpest cuts in investment, consistent with a reallocation of capital toward portfolio companies with shorter paths to liquidity and lower information frictions.

1.5.2 VC Terms

A key feature of venture capital financing is its staged nature: firms typically raise multiple rounds over time to fund development milestones, manage dilution, and align incentives. In this section, we build on our earlier evidence of capital reallocation from early- to later-stage firms to examine whether contractionary monetary policy shocks also affect the timing of follow-on investments. Specifically, we test whether these shocks lengthen the time between rounds (inter-round duration) for early-stage firms—reflecting their lower prioritization when capital is scarce—while shortening the duration for later-stage firms, which are closer to exit and more likely to receive accelerated funding.

To explore this, we construct a measure of the time (in quarters) between non-zero funding rounds for each firm in our sample. We restrict the sample to firms with at least two financing rounds and compute, for each subsequent round, the number of quarters that have elapsed since the previous one. We then regress this inter-round duration on contractionary monetary policy shocks using Equation (1.4).

Table 1.11 reports the results. We find that contractionary monetary policy shocks significantly increase the time between funding rounds, indicating that tighter financial conditions delay follow-on investments. This effect is particularly pronounced for seed and early-stage firms, which are typically riskier, more opaque, and further from generating cash flows. In contrast, inter-round duration shortens for mid-stage firms following contractionary shocks, consistent with a reallocation of capital toward ventures that are closer to exit or profitability.

These findings align with our two theoretical predictions: (i) a reallocation of investment from early- to mid-stage deals, and (ii) a lengthening of funding gaps for early-stage firms.

The evidence suggests that VCs become more cautious in committing additional capital to young firms during periods of monetary tightening—whether due to increased perceived risk, reduced capital availability, or the need to prioritize more mature, less uncertain investments. Longer funding gaps have important implications: they can raise the risk of firm failure, force startups to scale back operations, or push them into less favorable financing arrangements. These results highlight how monetary policy can shape the trajectory of startups not only at the initial entry stage but also throughout their development path.

1.5.3 Failure of Early-Stage Firms

The evidence so far shows that contractionary monetary policy shocks lengthen the time between funding rounds for early-stage firms and reallocate capital toward later-stage deals. We extend our analysis to examine the consequences of a tighter monetary environment on the failure rate of early-stage firms. By studying firm failure, we quantify the destructive side of capital reallocation and show that monetary policy, beyond influencing the flow of investment, also shapes the survival of young companies.

To do this, we first construct the variable "Firm Failure", an indicator that equals one if a firm no longer appears in the venture capital deal database for at least one year after its last observed funding round. We then restrict our analysis to firms for which the very first funding round was a seed or an early-stage round and estimate the impact of monetary policy shocks on "Firm Failure" using Equation (1.4). Although we cannot fully guarantee that a disappearance from the database corresponds to an actual firm failure, given the characteristics of the VC industry, such cases are highly likely to represent firm failures.[10]

Table 1.12 shows the results. Column (1) indicates that, for early-stage firms, a 25 bps contractionary monetary policy shock increases the probability of firm failure one year later by approximately 4.05 percentage points, an increase of 30.7% relative to the baseline probability of failure. The second column disaggregates this effect by separating positive (contractionary) and negative (expansionary) shocks, revealing an important asymmetry. The coefficient on the positive shock interaction term is positive and highly significant, confirming that a contractionary shock substantially increases the probability of failure. In particular, a 25 bps contractionary shock increases the probability of firm failure one year later by approximately 6.3 percentage points. In contrast, the coefficient on the negative shock interaction term is insignificant and near zero, suggesting that expansionary monetary policy does not have a statistically significant effect on reducing firm exit for this sample. This asymmetry highlights the destructive role of monetary tightening in the venture ecosystem, demonstrating that these firms are much more sensitive to negative shocks than they are responsive to positive ones.

1.5.4 Dynamics of VC Investment

To study the dynamic responses of deal-level VC investment to monetary policy shocks, we apply the local projection technique of Jordà (2005) in a panel setting. To the best of our knowledge, our paper is the first to do so for deal-level VC investment. We obtain average impulse response functions (IRFs) across all firms, which illustrate the dynamic response of VC investment. Specifically, we estimate the following specification:

$$y_{i,t+h} - y_{i,t-1} = \beta_1^h \, MPS_t + \gamma_1^{h'} \, X_{t-1} + \alpha_i + \epsilon_{i,t}, \tag{1.5}$$

[10]Another example is mergers. M&A activity is extremely relevant in this market, and we do not have a clear documentation in our database regarding mergers. Mergers would not be failures, reducing the impact of our extensive margin analysis.

where $y_{i,t+h}$ is the outcome variable (the log of deal financing size and the log of the number of VC investments) at h quarters after the MPS at time t. MPS_t refers to the monetary policy shocks from Bauer & Swanson (2023), aggregated to the quarterly level. X_{t-1} is a vector of controls that includes the lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. α_i represents firm fixed effects, and standard errors are robust and clustered at the firm level.

Figure 1.4 presents the results. We find that the negative response of VC investment to contractionary monetary policy shocks dissipates after six quarters, indicating a recovery of flows to baseline levels. However, the cumulative local projection responses remain below zero for several quarters, suggesting that the initial shortfall is not fully offset by subsequent increases. This evidence implies that, although flows normalize, the overall volume of VC investment may still experience a lasting reduction. The nature of the response also indicates that monetary policy influences the timing of capital deployment in private innovation markets, which can ultimately affect startup performance and long-term growth.

1.5.5 The Impact on GPs

In addition to the amount and timing of VC investment, monetary policy may also influence the structure of investment through changes in syndication. In VC markets, the number of GPs involved in a deal reflects how investors share risk, pool capital, and combine expertise. Building on our reallocation results, we examine whether contractionary monetary policy shocks alter syndication patterns across stages. For early-stage deals, reduced attractiveness under tighter conditions could lead to fewer GPs participating, as investors avoid committing jointly to riskier ventures. In contrast, later-stage deals—perceived as safer and closer to exit—may attract more GPs as investors reallocate capital toward them. Analyzing the number of GPs involved in deals provides another dimension to our study, revealing how monetary policy affects not only the volume and timing of VC investment but also its organizational structure.

To do this, we collect the number of GPs in each VC deal from our sample and estimate Equation (1.4) using the log of the number of GPs in a deal as the outcome variable. Figure 1.5 displays the results. In response to a 25 bps contractionary shock, the number of GPs in a VC deal decreases on average by approximately 2.8%. Consistent with our VC investment results, Table 1.13 shows that contractionary shocks significantly reduce the number of GPs in a deal at the seed financing stage, while increasing it in later-stage deals, indicating a significant change in the organizational structure of investments. We also examine the impact of monetary policy across industries. Table 1.14 presents the findings. The reduction in the number of GPs is concentrated in business services, healthcare, and information technology. One possible explanation is that these industries require substantial capital investment and have longer exit horizons, making them riskier when monetary conditions are tighter.

1.5.6 Fundraising Activity

We have focused our analysis at the deal level since our primary goal is to investigate the impact of monetary policy on VC investment. However, monetary policy also affects the fundraising activity of VC funds, which is a key channel shaping future investment capacity. Fundraising represents the supply side of the VC market: it determines the pool of capital that funds can allocate to new deals. When monetary conditions tighten, institutional investors may reduce their commitments to new funds, either because of higher discount rates, portfolio rebalancing toward safer assets, or limited liquidity. This could translate into longer fundraising cycles and fewer new fund launches.

To investigate the impact of monetary policy on fundraising activity, we use fund-level data from Preqin, which report the fund name along with its fundraising launch and close dates. We focus on two outcome variables: (i) the Fundraising Period, defined as the number of days between the launch and close dates, and (ii) the number of funds launched in a given period.[11] We then estimate the effect of monetary policy on these outcomes, controlling for macroeconomic variables as specified in Equation (1.4).

Table 1.15 presents the results. Indeed, monetary policy significantly affect fundraising duration. The coefficient implies that a 25 bps contractionary monetary policy shock raises fundraising duration by about 20%. Given an average fundraising period of 465 days, this corresponds to a delay of roughly 93 days (approximately three months). This extension slows the typical fundraising cycle from 15 months to about 18 months, representing a sizable economic effect in an industry where timely capital commitments are critical. Table 1.15 also shows that the number of funds launched decreases by 36% in response to the same contractionary shock, underscoring the impact of monetary policy on fundraising activity.

1.6 Additional Results and Robustness

This section reports additional results and several robustness tests. We verify that the results are robust to alternative measures of monetary policy shocks, different definitions of financing stages, asymmetric effects, and the inclusion of industry-time fixed effects.

1.6.1 Measures of Monetary Policy Shocks

We confirm that our results are robust to alternative monetary policy shocks. A key consideration in studying the impact of monetary policy on venture capital (VC) investment is which shocks best capture the relevant transmission channel. VC is a long-horizon, equity-financed activity whose valuations depend primar-

[11]Our analysis covers the universe of VC funds included in the Preqin database.

ily on discount rates and investor risk appetite rather than on contemporaneous borrowing costs. For this reason, we also test monetary policy shocks that move long-term interest rates. Specifically, we employ changes in the 10-year treasury rate around the FOMC as in Li (2025). These shocks alter expectations about the future path of policy and term premia, thereby shifting long-run discount rates that are most relevant for VC valuations and investment flows. Tables 1.16 and 1.17 show that our results remain qualitatively similar. Overall, these findings indicate that our results are robust to the choice of shock construction.

1.6.2 Financing Stages

One concern with our baseline results is that, when defining financing stages, we include Exit (merger, PIPE, pre-IPO) and Venture Debt (VD). These stages may respond to different risk-return trade-offs and might not be comparable to other financing stages, such as seed and early-stage deals, which could introduce bias into our estimates. To address this, we construct a sample of deals that excludes both Exit and Venture Debt stages and re-estimate our baseline specification. Figure 1.6 shows our results are virtually unchanged.

1.6.3 Asymmetric Effects

We investigate potential differences between negative and positive monetary policy shocks on VC investment by estimating the coefficient of MPS conditional on its sign. Specifically, we create the variables POS and NEG, which are dummy variables equal to one when the monetary policy shock is greater than zero and less than zero, respectively. Table 1.18 presents two main results. First, both contractionary and expansionary shocks affect VC investment. Second, and most importantly, contractionary shocks have stronger effects. This result is intuitive: tighter policy raises discount rates and lowers valuations, especially for early-stage firms with long cash flow durations, making new deals less attractive relative to holding dry powder for future opportunities. In contrast, when policy is more accommodative, the ability to accelerate investment is inherently limited by the availability of suitable deals, the time required to evaluate startups thoroughly, and the staged nature of VC financing, which prevents a rapid scaling-up of capital deployment.

1.6.4 Industry-Time Fixed Effects

In our main specification, we do not include industry–time fixed effects, as these would absorb all variation in MPS_t that is common to firms within the same industry and time period, effectively subsuming its coefficient. However, industry–time fixed effects are valuable for controlling for unobserved shocks that vary across both industries and time—such as industry-specific demand changes, technology shocks, or regulatory events—that could otherwise bias our estimates. As a robustness check, we re-estimate our baseline regression including industry–time fixed effects to ensure that our results are not driven by such confounding factors. Specifically, we estimate:

$$\Delta y_{i,t} = \beta_1 \, MPS_t \times Seed_{i,t} + \beta_2 \, Seed_{i,t} + \alpha_i + \theta_{j,t} + \epsilon_{i,t}, \tag{1.6}$$

where Seed is a dummy variable equals to one if the firm is at the seed financing stage (seed, angel, grant), MPS is the monetary policy shock from Bauer & Swanson (2023), and α_i and $\theta_{j,t}$ are firm and industry-time fixed effects, respectively.

Table 1.19 reports our results. Including industry–time fixed effects does not alter our main finding that seed-stage firms are more sensitive to monetary policy shocks. In particular, the interaction term between Seed and MPS_t remains highly statistically significant, indicating that the stronger negative response of seed-stage investment to contractionary shocks is not driven by unobserved industry-specific shocks occurring at the same time. This robustness check reinforces our interpretation that the early-stage effect reflects a systematic reallocation away from the riskiest and longest-duration investments when monetary policy tightens.

1.7 Conclusion

This paper provides comprehensive evidence on how monetary policy shocks affect venture capital (VC) markets. We use high-frequency identification and detailed deal-level data to show that contractionary monetary policy reduces both the number of startups receiving funding and the amount invested per deal, with the largest effects concentrated in the seed stage. At the same time, investment in more mature firms increases, consistent with a reallocation of capital toward ventures with shorter durations and more predictable returns.

We document that contractionary shocks lengthen the time between funding rounds for early-stage firms, shorten it for later-stage firms, and reduce syndication in seed-stage deals, indicating that monetary policy affects not only the scale of VC investment but also its pace and organizational structure. We also establish that tighter monetary policy affects the fundraising activity of VC funds, prolonging fundraising cycles and reducing the number of new fund launches, which constrains future investment capacity. Together, these results highlight a new channel through which monetary policy can shape innovation and growth—by restricting access to early-stage venture capital, where firms are most dependent on external equity and where the risk of funding delays is most consequential.

Our findings suggest that the effects of monetary tightening extend well beyond traditional debt markets, influencing the allocation of risk capital in ways that may have lasting implications for the composition of innovation and the long-run growth potential of the economy. Future research could explore whether similar dynamics arise in other segments of private equity, how persistent these reallocation effects are, the extent to which they alter the types of technologies and business models that ultimately reach the market, and how monetary policy affects different types of general partners during the fundraising process.

Figures and Tables

Table 1.1: Summary Statistics

		Per Quarter		
	Total	Average	Median	Std.Dev
Number of Deals	103,155	882	640	809
Deal Volume (U$MM)	868,114.3	15.1	9.7	335.9

This table presents the summary statistics for the number of deals and deal volume (in USD millions). Column (1) reports the total number of deals and the total deal volume for our full sample. Columns (2), (3), and (4) report the average, median, and standard deviation of these variables per quarter. Our sample period spans 1990 to 2023. Source: Preqin.

Table 1.2: Summary Statistics: Monetary Policy Shocks

	Obs	Mean	Std.Dev	Min	Max
MPS	141	0.000	0.068	-0.248	0.192

This table reports the summary statistics of the monetary policy shocks from Bauer & Swanson (2023) aggregated to the quarterly frequency.

Table 1.3: VC Funds and Investment Strategies

Investment Strategy	Number of Funds	Percent (%)
Panel A: Baseline		
Single Stage	3,967	36.20
Two Stage	3,006	27.43
Three or More Stages	3,985	36.37
Total	10,958	100
Panel B: Disaggregated Analysis		
Single Stage	3,967	36.20
Seed and ES	1,820	16.61
Seed and MS	109	0.99
Seed and LS	72	0.66
ES and MS	345	3.15
ES and LS	204	1.86
MS and LS	456	4.16
Three or More Stages	3,985	36.37
Total	10,958	100

Panel A of this table presents the investment strategies of VC funds. Single Stage, Two Stage, and Three or More Stages indicate the number of funds that invest in one, two, and three or more financing stages, respectively. Panel B disaggregates the two-stage strategy, showing the explicit financing stages in which those funds invest. We follow Preqin's classification and define the following financing stages: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C- series L, add on, growth capital/expansion). Our sample period spans 1990 to 2023. Source: Preqin.

Table 1.4: Effect of Monetary Policy on VC deals

	Deal Financing Size	Number of Investments
MPS	-0.219***	-0.0934***
	(0.0382)	(0.0115)
Observations	325,683	325,683
R-squared	0.22	0.21
Macro Controls	Yes	Yes
Firm FE	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variables are Deal Financing Size and Number of Investments. Deal Financing size is the log change in the volume size of the VC investment and Number of Investments is the log change in the number of VC deals. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.5: Industry Analysis and the Effect of Monetary Policy on VC deals: Deal Financing Size

	BS	CD	EU	FI	H	I	IT	MR	RE	TE
MPS	-0.436**	-0.399**	-0.00449	-0.759***	-0.164**	-0.507*	-0.194***	0.0504	-0.972	-0.172
	(0.177)	(0.201)	(0.383)	(0.285)	(0.0818)	(0.262)	(0.0552)	(0.288)	(1.318)	(0.223)
Observations	13,897	17,977	5,905	9,044	76,745	8,568	154,445	4,734	695	11,293
R-squared	0.21	0.24	0.22	0.25	0.21	0.24	0.22	0.22	0.37	0.21
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variable is Deal Financing Size. Deal Financing size is the log change in the volume size of the VC investment. We estimate Equation (1.4) for each industry. We follow Preqin's classification and divide our sample into ten categories: Business Services (BS), Consumer Discretionary (CD), Energy & Utilities (EU), Financial & Insurance Services (FI), Healthcare (H), Industrials (I), Information Technology (IT), Raw Materials & Natural Resources (MR), Real Estate (RE), and Telecoms & Media (TM). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.6: Industry Analysis and the Effect of Monetary Policy on VC deals: Number of Investments

	BS	CD	EU	FI	H	I	IT	MR	RE	TE
MPS	-0.171***	-0.171***	-0.0332	-0.189**	-0.0900***	-0.222***	-0.0822***	0.0808	-0.427	-0.0799
	(0.0541)	(0.0630)	(0.0960)	(0.0781)	(0.0240)	(0.0808)	(0.0171)	(0.103)	(0.290)	(0.0612)
Observations	13,897	17,977	5,905	9,044	76,745	8,568	154,445	4,734	695	11,293
R-squared	0.20	0.22	0.22	0.23	0.21	0.22	0.21	0.21	0.31	0.21
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variables is the Number of Investments. Number of Investments is the log change in the number of VC deals. We estimate Equation (1.4) for each industry. We follow Preqin's classification and divide our sample into ten categories: Business Services (BS), Consumer Discretionary (CD), Energy & Utilities (EU), Financial & Insurance Services (FI), Healthcare (H), Industrials (I), Information Technology (IT), Raw Materials & Natural Resources (MR), Real Estate (RE), and Telecoms & Media (TM). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses.
*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.7: Effect of Monetary Policy on VC deals: Deal Financing Size

	Seed	ES	MS	LS	Exit	VD
MPS	-0.131**	-0.0680	0.230**	-0.0286	0.562	-0.205
	(0.0592)	(0.0617)	(0.105)	(0.117)	(0.398)	(0.207)
Observations	42,723	59,255	48,469	54,178	2,753	17,485
R-squared	0.23	0.22	0.27	0.25	0.23	0.28
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variable is Deal Financing Size. Deal Financing size is the log change in the volume size of the VC investment. We estimate Equation (1.4) for each financing stage. We follow Preqin's classification and define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C- series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.8: Effect of Monetary Policy on VC deals: Number of Investments

	Seed	ES	MS	LS	Exit	VD
MPS	-0.134***	-0.0337	0.0283	-0.0208	0.0755	-0.0522
	(0.0357)	(0.0225)	(0.0293)	(0.0277)	(0.0986)	(0.0711)
Observations	42,723	59,255	48,469	54,178	2,753	17,485
R-squared	0.21	0.23	0.27	0.25	0.24	0.28
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variables is the Number of Investments. Number of Investments is the log change in the number of VC deals. We estimate Equation (1.4) for each financing stage. We follow Preqin's classification and define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C- series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.9: Effect of Monetary Policy on VC deals: Deal Financing Size

	Top qsfi	Mid qsfi	Bottom qsfi
MPS	-0.0962*	-0.336***	-1.505***
	(0.0583)	(0.0917)	(0.178)
Observations	36,301	36,155	14,173
R-squared	0.25	0.30	0.30
Macro Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variable is Deal Financing Size. Deal Financing size is the log change in the volume size of the VC investment. We sort firms into terciles each quarter based on their $qsfi_{i,t}$ (see section 1.5) and estimate Equation (1.4) for each of these three subsamples. Top qsfi, Mid qsfi and Bottom qsfi are the group of firms in the top, medium and bottom tercile of the $qsfi_{i,t}$ distribution. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.10: Effect of Monetary Policy on VC deals: Number of Investments

	Top qsfi	Mid qsfi	Bottom qsfi
MPS	-0.0471*	-0.181***	-0.780***
	(0.0243)	(0.0361)	(0.0755)
Observations	36,301	36,155	14,173
R-squared	0.25	0.31	0.33
Macro Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

This table reports the coefficients from the regression of monetary policy shocks on VC investment. The dependent variables is the Number of Investments. Number of Investments is the log change in the number of VC deals. We sort firms into terciles each quarter based on their $qsfi_{i,t}$ (see section 1.5) and estimate Equation (1.4) for each of these three subsamples. Top qsfi, Mid qsfi and Bottom qsfi are the group of firms in the top, medium and bottom tercile of the $qsfi_{i,t}$ distribution. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.11: Effect of Monetary Policy on Inter-Round Duration

	Seed	ES	MS	LS	Exit	VD
MPS	0.474***	0.0358	-0.208***	0.000711	-0.283	0.164
	(0.0683)	(0.0440)	(0.0632)	(0.0649)	(0.224)	(0.135)
Observations	42,995	59,285	47,853	53,171	2,819	18,124
R-squared	0.11	0.11	0.07	0.04	0.06	0.10
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the coefficients of the regression of monetary policy shocks on Inter-round duration. The dependent variable is Inter-round duration, defined as the number of quarters since a firm's most recent non-zero venture capital funding round. We estimate Equation (1.4) for each financing stage. We follow Preqin's classification and define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C- series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of Inter-round duration to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.12: Effect of Monetary Policy on the Failure of Early-Stage Firms

	Firm Failure	Firm Failure
MPS	0.1617***	
	(0.0147)	
MPS $\times$ pos		0.2542***
		(0.0226)
MPS $\times$ neg		-0.0147
		(0.0306)
Observations	135,956	135,956
R-squared	0.1341	0.1343
Macro Controls	Yes	Yes
Firm FE	Yes	Yes

This table reports the coefficients of the regression of monetary policy shocks on firm failure. The dependent variable is Firm Failure, defined as an indicator that equals one if a firm no longer appears in the venture capital deal database for at least one year after its last observed funding round. We estimate Equation (1.4) restricting our analysis to firms for which the very first funding round was a seed or an early-stage round following the classification in Section 1.5. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of firm failure to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.13: Effect of Monetary Policy on the Number of GPs

	Seed	ES	MS	LS	Exit	VD
MPS	-0.159***	-0.0400	0.0837	-0.0393	0.288*	0.0416
	(0.0545)	(0.0391)	(0.0591)	(0.0635)	(0.173)	(0.0877)
Observations	40,648	58,263	48,114	53,924	2,582	16,777
R-squared	0.23	0.24	0.26	0.24	0.24	0.28
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the coefficients of the regression of monetary policy shocks on the number of GPs. The dependent variable is the Number of GPs. Number of GPs is the log change in the number of GPs. We estimate Equation (1.4) for each financing stage. We follow Preqin's classification and define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C- series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of the number of GPs to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.14: Industry Analysis and the Effect of Monetary Policy on the Number of GPs

	BS	CD	EU	FI	H	I	IT	MR	RE	TE
MPS	-0.205*	-0.218	0.123	-0.213	-0.104**	-0.116	-0.117***	-0.167	-0.700	-0.0252
	(0.113)	(0.133)	(0.217)	(0.158)	(0.0519)	(0.164)	(0.0361)	(0.235)	(0.611)	(0.136)
Observations	10,234	13,183	4,067	7,096	55,032	5,952	118,434	3,032	465	8,323
R-squared	0.20	0.23	0.22	0.24	0.21	0.22	0.21	0.20	0.32	0.21
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the coefficients of the regression of monetary policy shocks on the number of GPs. The dependent variables is the Number of GPs. Number of GPs is the log change in the number of GPs. We estimate Equation (1.4) for each industry. We follow Preqin's classification and divide our sample into ten categories: Business Services (BS), Consumer Discretionary (CD), Energy & Utilities (EU), Financial & Insurance Services (FI), Healthcare (H), Industrials (I), Information Technology (IT), Raw Materials & Natural Resources (MR), Real Estate (RE), and Telecoms & Media (TM). MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of the number of GPs to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.15: Effect of Monetary Policy on Fundraising Activity

	Fundraising Period	Number of Funds Launched
MPS	0.8066**	-1.474**
	(0.3786)	(0.577)
Observations	3,825	91
R-squared	0.032	0.97
Macro Controls	Yes	Yes
Year FE	Yes	Yes

This table reports the estimated coefficients of monetary policy policy shocks on fundraising activity. The dependent variables are Fundraising Period and Number of Funds Launched. Fundraising Period is the log of the number of days between the launch and close dates, and Number of Funds Launched is the log of the number of funds launched. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this table shows the response of fundraising activity to a 100 bps contractionary monetary policy shock. All regressions control for employment ratio, CPI, GDP growth, and the Excess Bond Premium. Standard errors are heteroskedasticity robust and clustered at the time level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.16: Effect of Monetary Policy on VC deals: Deal Financing Size and Changes in 10-year Treasury as Monetary Policy Shocks

	Seed	ES	MS	LS	Exit	VD
MPS	-0.110*	-0.0803	0.494***	0.123	0.156	-0.413*
	(0.0584)	(0.0701)	(0.117)	(0.133)	(0.553)	(0.213)
Observations	42,723	59,255	48,469	54,178	2,753	17,485
R-squared	0.23	0.22	0.27	0.25	0.23	0.28
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the coefficients from the regression of monetary policy shocks on VC investment. The dependent variable is Deal Financing Size. Deal Financing size is the log change in the volume size of the VC investment. We estimate Equation (1.4) for each financing stage. We follow Preqin's classification and define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C-series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). MPS is the change in 10-year treasury rate around FOMC meetings. Therefore, this table shows the response of VC investment to a 100 bps contractionary forward guidance shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.17: Effect of Monetary Policy on VC deals: Number of Investments and Changes in 10-year Treasury as Monetary Policy Shocks

	Seed	ES	MS	LS	Exit	VD
MPS	-0.0803**	-0.0274	0.0972***	0.0237	-0.0271	-0.203**
	(0.0408)	(0.0262)	(0.0337)	(0.0331)	(0.137)	(0.0831)
Observations	42,723	59,255	48,469	54,178	2,753	17,485
R-squared	0.21	0.23	0.27	0.25	0.24	0.28
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the coefficients from the regression of monetary policy shocks on VC investment. The dependent variables is the Number of Investments. Number of Investments is the log change in the number of VC deals. We estimate Equation (1.4) for each financing stage. We follow Preqin's classification and define the following financing stages, along with a debt stage: Seed (seed, angel, grant), Early-stage-ES (series A), Middle Stage-MS (series B), Late-stage-LS (series C-series L, add on, growth capital/expansion), Exit (merger, PIPE, pre-IPO), and Venture Debt (VD). MPS is the change in 10-year treasury rate around FOMC meetings. Therefore, this table shows the response of VC investment to a 100 bps contractionary forward guidance shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.18: Effect of Monetary Policy on VC deals: Asymmetric Effects

	Deal Financing Size	Number of Investments
MPS $\times$ *POS*	-0.3705***	-0.1559***
	(0.0676)	(0.0212)
MPS $\times$ *NEG*	0.2608***	0.1322***
	(0.0979)	(0.0288)
Observations	325,683	325,683
R-squared	0.22	0.21
Macro Controls	Yes	Yes
Firm FE	Yes	Yes

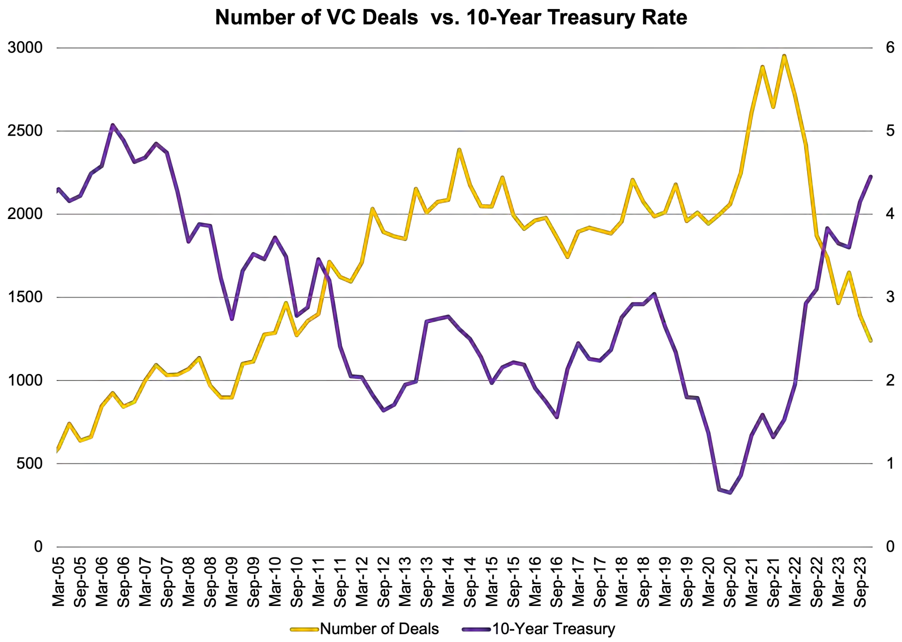
This table reports the estimated coefficients of monetary policy policy shocks on VC investment. The dependent variables are Deal Financing Size and Number of Investments. Deal Financing size is the log change in the volume size of the VC investment and Number of Investments is the log change in the number of VC deals. MPS is the monetary policy shock as in Bauer & Swanson (2023). POS is a dummy equals to one if MPS is greater than zero. NEG is a dummy equals to one if MPS is less than zero. Therefore, this table shows the response of VC investment to a 100 bps contractionary (MPS $\times$ POS) and expansionary (MPS $\times$ NEG) monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 1.19: Effect of Monetary Policy on VC deals: Time Fixed Effects

	Deal Financing Size	Number of Investments
MPS × *Seed*	-0.595***	-0.106***
	(0.104)	(0.0379)
Observations	259,104	259,104
R-squared	0.22	0.23
Firm FE	Yes	Yes
Industry x Time FE	Yes	Yes

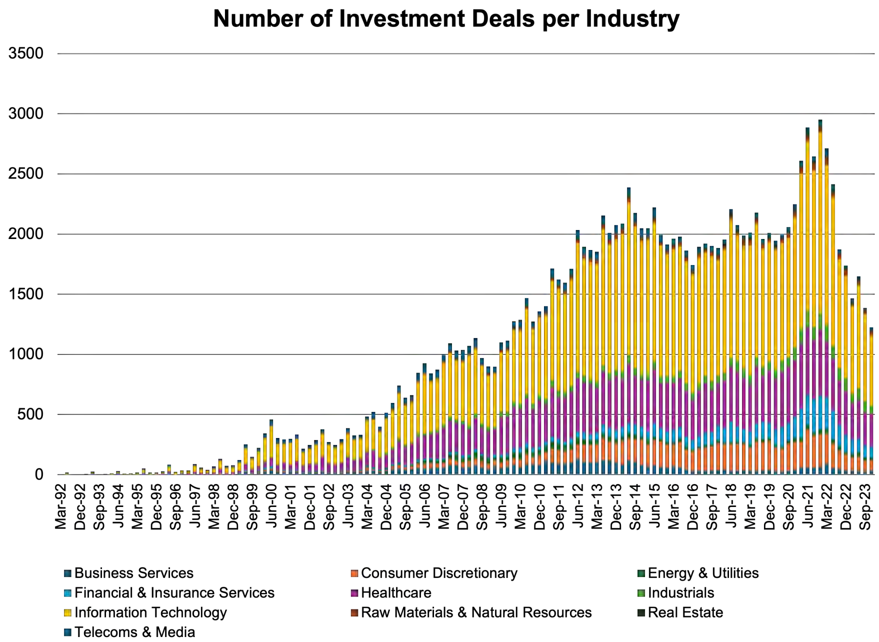
This table reports the coefficients from the regression of monetary policy shocks on VC investment. The dependent variables are Deal Financing Size and Number of Investments. Deal Financing size is the log change in the volume size of the VC investment. Number of investments is the log change in the number of VC deals. MPS is the monetary policy shock as in Bauer & Swanson (2023). Seed is a dummy variable equals to one if the firm is at the seed financing stage (seed, angel, grant). Therefore, this table shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for one lag of the outcome variable. Standard errors are heteroskedasticity robust and clustered at the firm level. We report the respective standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Figure 1.1: Number of Investments Deals vs. 10-Year Treasury Rate



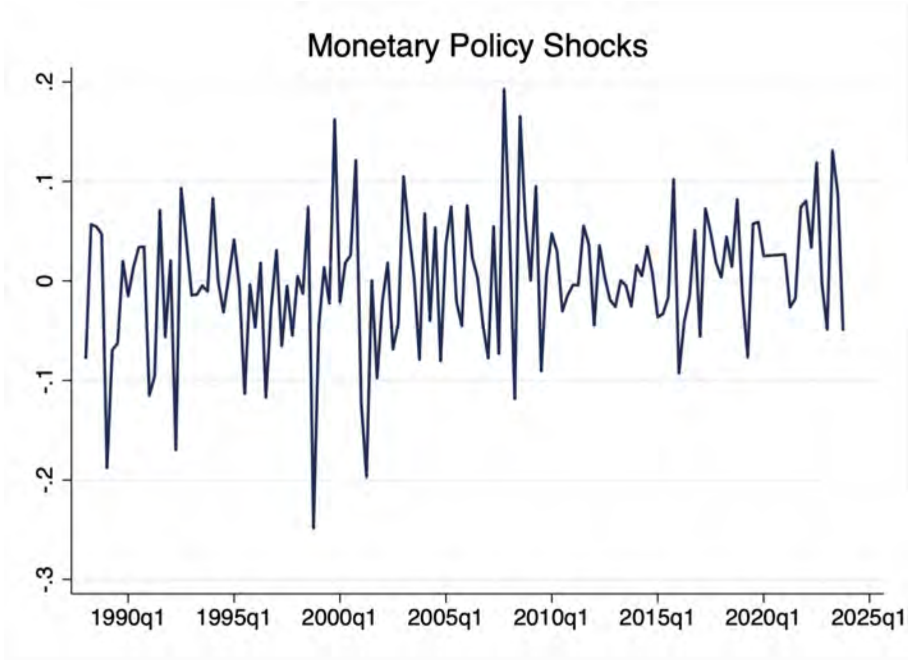
This figure shows the number of VC deals at the quarterly frequency and the market Yield on U.S. Treasury Securities at 10-Year Constant Maturity, Quoted on an Investment Basis [DGS10]. Sources: FRED and Preqin.

Figure 1.2: Number of Investments Deals per Industry



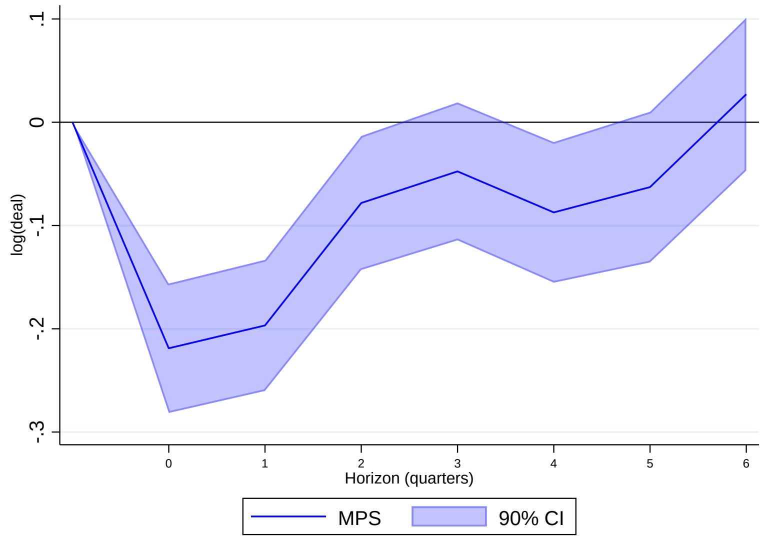
This figure shows the number of VC deals per industry at the quarterly frequency. We follow Preqin's classification and divide our sample into ten categories: Business Services (BS), Consumer Discretionary (CD), Energy & Utilities (EU), Financial & Insurance Services (FI), Healthcare (H), Industrials (I), Information Technology (IT), Raw Materials & Natural Resources (MR), Real Estate (RE), and Telecoms & Media (TM). Source: Preqin.

Figure 1.3: Monetary Policy Shocks

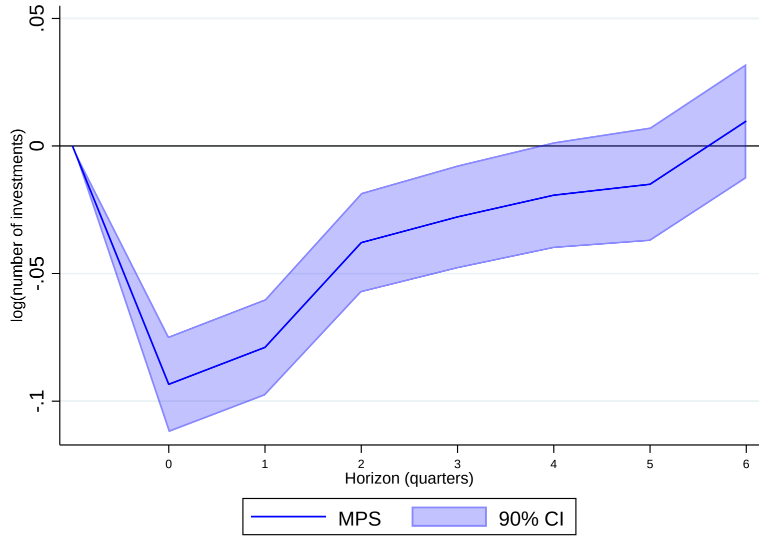


This figure plots quarterly aggregates of the monetary policy shocks from Bauer & Swanson (2023) for 1990–2025.

Figure 1.4: The Effect of Monetary Policy on VC Investment



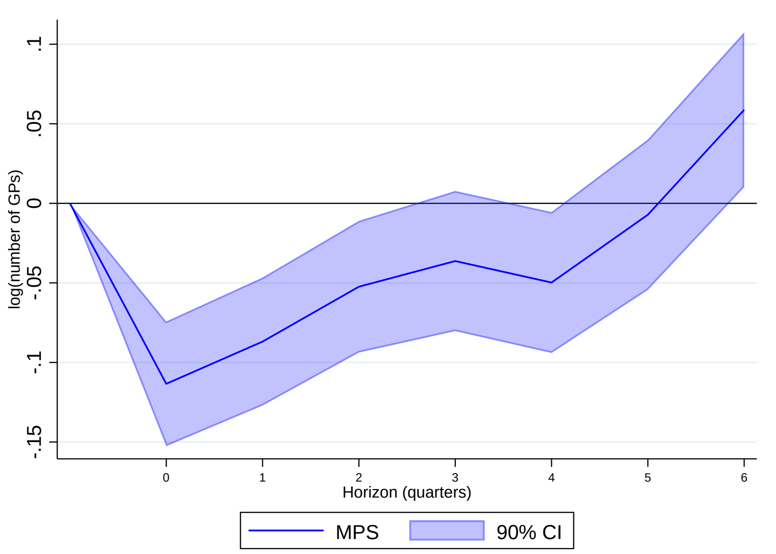
Deal Financing Size



Number of Investments

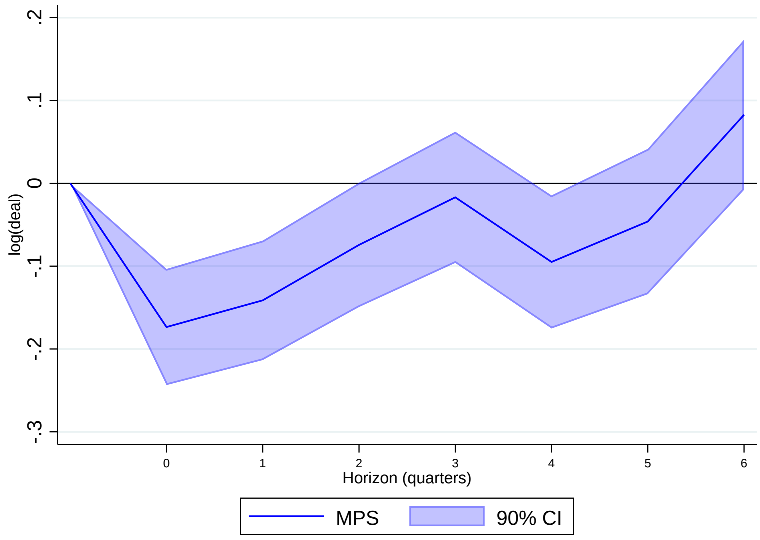
This figure shows the effect of monetary policy on VC investment. The dependent variables are Deal Financing Size (Panel A) and Number of Investments (Panel B). Deal Financing size is the log change in the volume size of the VC investment and Number of Investments is the log change in the number of VC deals. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this figure shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. The dashed line represents 90% confidence intervals. Standard errors are heteroskedasticity robust and clustered at the firm level.

Figure 1.5: The Effect of Monetary Policy on the Number of GPs

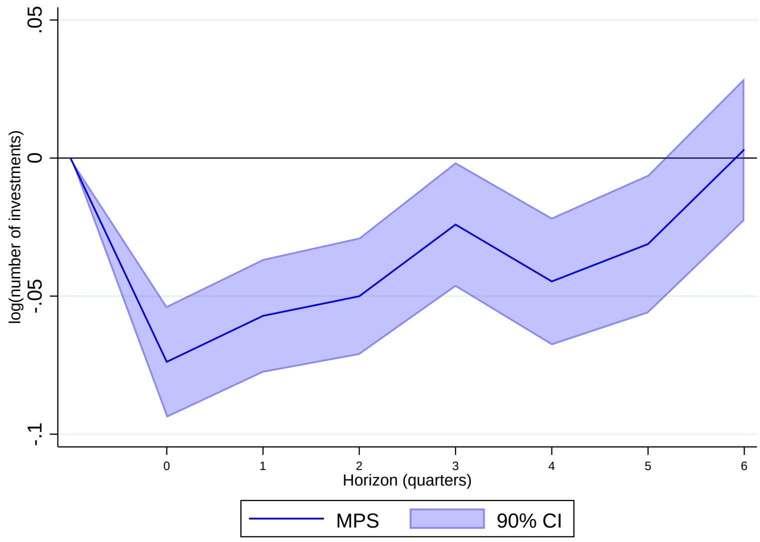


This figure shows the effect of monetary policy on the number of GPs. The dependent variable is Number of GPs. Number of GPs is the log change in the number of GPs. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this figure shows the response of the number of GPs in VC deals to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. The dashed line represents 90% confidence intervals. Standard errors are heteroskedasticity robust and clustered at the firm level.

Figure 1.6: The Effect of Monetary Policy on VC Investment: No Exit and Venture Debt Stages



Deal Financing Size



Number of Investments

This figure shows the effect of monetary policy on VC investment. The dependent variables are Deal Financing Size (Panel A) and Number of Investments (Panel B). Deal Financing size is the log change in the volume size of the VC investment and Number of Investments is the log change in the number of VC deals. MPS is the monetary policy shock as in Bauer & Swanson (2023). Therefore, this figure shows the response of VC investment to a 100 bps contractionary monetary policy shock. All regressions control for lagged employment ratio, CPI, GDP growth, and the Excess Bond Premium. We also include one lag of the shock and the outcome variable. We exclude Exit and Venture Debt stages. The dashed line represents 90% confidence intervals. Standard errors are heteroskedasticity robust and clustered at the firm level.

1.A Setup, Lagrangian, and First-Order Conditions

The VC maximizes mean–variance utility subject to a capital constraint:

$$\max_{I_E, I_L, I_{E2}, I_{L2}} \; I_E\,\mathbb{E}[\mu_E(r_1)] - \frac{\gamma}{2}\,I_E^2\sigma_E^2 + I_L\,\mu_L(r_1) + I_{E2}\,\mathbb{E}[\mu_E(r_2)] - \frac{\gamma}{2}\,I_{E2}^2\sigma_E^2 + I_{L2}\,\mu_L(r_2)$$
$$\text{s.t.} \;\; I_E + I_L + I_{E2} + I_{L2} \le K, \qquad I_{\cdot} \ge 0,$$

where $\gamma > 0$ and $\sigma_E^2 > 0$. Late-stage returns are deterministic for simplicity and tractability. Let $\lambda \ge 0$ be the multiplier on the capital constraint. The Lagrangian and the FOCs are:

$$\mathcal{L} = I_E\,\mathbb{E}[\mu_E(r_1)] - \frac{\gamma}{2}\,I_E^2\sigma_E^2 + I_L\,\mu_L(r_1) + I_{E2}\,\mathbb{E}[\mu_E(r_2)] - \frac{\gamma}{2}\,I_{E2}^2\sigma_E^2 + I_{L2}\,\mu_L(r_2) + \lambda\big(K - I_E - I_L - I_{E2} - I_{L2}\big).$$

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial I_E} &= \mathbb{E}[\mu_E(r_1)] - \gamma I_E\sigma_E^2 - \lambda = 0,\\
\frac{\partial\mathcal{L}}{\partial I_{E2}} &= \mathbb{E}[\mu_E(r_2)] - \gamma I_{E2}\sigma_E^2 - \lambda = 0,\\
\frac{\partial\mathcal{L}}{\partial I_L} &= \mu_L(r_1) - \lambda = 0,\\
\frac{\partial\mathcal{L}}{\partial I_{L2}} &= \mu_L(r_2) - \lambda = 0,
\end{aligned}$$

together with complementary slackness $\lambda\,(K - I_E - I_L - I_{E2} - I_{L2}) = 0$.

From the late-stage FOCs we obtain $\lambda = \mu_L(r_1) = \mu_L(r_2)$ if both I_L and I_{L2} are strictly positive. With deterministic late-stage returns and identical pricing of late-stage risk across periods, we can take $\mu_L(r_1) = \mu_L(r_2) \equiv \mu_L$. Substituting into the early-stage FOCs gives the optimal early-stage allocations:

$$I_E^* = \frac{\mathbb{E}[\mu_E(r_1)] - \mu_L}{\gamma\sigma_E^2}, \qquad I_{E2}^* = \frac{\mathbb{E}[\mu_E(r_2)] - \mu_L}{\gamma\sigma_E^2}.$$

Given K, the residual goes to late-stage:

$$I_L^* + I_{L2}^* = K - I_E^* - I_{E2}^*.$$

2
Cherry-Picking Excellence: A Data-Driven Approach to Choosing the Best Funds and GPs in PE

Abstract. This paper employs high-dimensional econometric models to predict whether a private equity (PE) fund will achieve performance success, using only information available to prospective investors at the time of fundraising. In a novel approach, we incorporate identifiers that help isolate general partners (GPs) with a higher likelihood of launching top-performing funds. For venture capital (VC) funds, our models achieve an average accuracy of approximately 75% in predicting funds with abnormal returns. These results suggest that data-driven approaches can improve investment decisions, reduce due diligence efforts, and decrease reliance on subjective judgment.

Keywords: Machine Learning, Private Equity, Venture Capital, Performance, Limited Partners.

2.1 Introduction

From 2020 to mid-2023, private equity assets under management (AUM) increased from $6.1 trillion to approximately $8.2 trillion, reflecting an annual growth rate of around 8%–10%, driven by a rise in investors' portfolio allocations (McKinsey&Company, 2024). The private equity market has two main players: limited partners (LPs), the investors in this market (endowments, pension funds, institutional investors), and general partners (GPs), the fund managers who seek deals and allocate the raised capital. Private equity funds are closed-end funds that after the fundraising period at their launch operates with committed capital, doing capital calls step by step as investment opportunities arise. As private equity gains more relevance, LPs are spending more time and effort screening for the best general partners GPs[1]. Da Rin & Phalippou (2017) emphasizes the critical role of due diligence in private equity (PE), while David Swensen (Swensen, 2009), former CIO of the Yale Endowment, notes, "*Successful private equity investing requires identifying and affiliating with superior partners*".

In this paper, we aim to improve investors' decision-making by utilizing various econometric tools, with a focus on machine learning (ML) techniques, to predict whether a PE fund will achieve performance success, using only the information available to prospective investors at the time of fundraising. To enhance predictive power, we incorporate novel explanatory variables that reveal key drivers of fund performance, offering deeper insights into the factors that shape PE returns. These findings have the potential to increase returns, reduce due diligence time, and decrease reliance on human judgment.

With the growing availability of detailed datasets on GPs and funds, we can incorporate more comprehensive information in performance analysis. However,

[1] A recent survey by Private Equity International reports that 69% of LPs dedicate most of their time to due diligence in search of top GPs - https://www.privateequityinternational.com/lps-put-emphasis-on-due-diligence-as-they-get-selective/

traditional discrete choice econometric models struggle when the number of variables exceeds the sample size. To overcome this, we use ML methods capable of handling high-dimensional data. In this context, we evaluate six models to identify which one best predicts the performance success of PE funds. These models include traditional approaches like the Linear Probability Model (LPM), Logit, and Probit, as well as ML techniques such as Lasso, Ridge, and Random Forest (RF). This comparative analysis helps us identify the most robust method for forecasting private equity outcomes.

We find that ML techniques generally outperform classical econometric discrete choice models. For ML models, we observe that predictive accuracy is largely driven by VC funds, in some simulations reaching an average accuracy of approximately 75 percent in predicting funds with abnormal returns based on our performance metrics. The most influential predictor that emerges from our models is the GP identifier, a variable that can be traced to each individual GP, highlighting how crucial GP selection is in VC investing and how our approach can help optimize LPs' efforts in identifying high quality GPs.

These results are aligned with theories documenting the existence of performance persistence in VC (Hochberg *et al.*, 2014; Maurin *et al.*, 2023), where empirical work has shown that this persistence can be identified by predicting a fund's future performance using a simple metric, the past performance of funds managed by the same GP (Kaplan & Schoar, 2005; Harris *et al.*, 2023). However, predicting future fund performance at the time of fundraising depending solely on this metric presents challenges, since historical performance may be unavailable and interim results may be unreliable. In particular, the measurement error in the estimated NAV (Net Asset Value) can distort the accuracy of persistence-based forecasts. To address these issues, our approach is different and relies on more information that would be available to LPs at the time of fundraising. We examine a broader set of fundraising-period predictors of abnormal returns, an exercise made possible by high-dimensional machine learning methods.

We measure whether a fund will have performance success or not, using two key metrics: net internal rate of return (Net IRR) and implied Public Market Equivalent (implied PME)[2]. Successful (or Abnormal) performance is defined as funds with a implied PME greater than 1 or a Net IRR falling within the top tercile, which serve as our dependent variables in the analysis.

For our explanatory variables, we capture six critical aspects from PE literature that are linked to performance: GP identification, fund characteristics, PE industry, market environment, macroeconomic conditions, and PE theory. Including GP identification is essential[3], as Hochberg *et al.* (2007) demonstrate that certain GPs consistently outperform their peers. In a novel approach, we include specific identifiers for each GP, enabling us to pinpoint those with stronger explanatory power for fund performance. Ewens & Rhodes-Kropf (2015) demonstrates that VC partners have a greater impact on the positive results of funds than institutional aspects of VC firms (the GPs). We can therefore consider that our model, in a way, leads companies that bring together the best professionals in their partnership to consistently be selected in the ML models used here. Fund-specific factors, particularly fund size, show a concave relationship with performance (Kaplan & Schoar, 2005), and sectors such as technology and healthcare are known for superior returns. Additionally, papers by Kaplan & Schoar (2005); Harris *et al.* (2014); Gompers *et al.* (2016); Robinson & Sensoy (2016); Ljungqvist *et al.* (2020); Gompers *et al.* (1998); Gompers & Lerner (2000) emphasize the importance of PE industry, market environment, and macro conditions, with variables like GDP growth and interest rates providing crucial

[2]The Public Market Equivalent (PME) is a metric introduced by Kaplan & Schoar (2005)) to evaluate private equity performance relative to public market benchmarks. It compares the present value of cash flows from a PE fund with the value of investing those same cash flows in a public index. In our analysis, we employ an implied PME, and the methodology underlying its construction is presented in Section 2.

[3]The GP identification variables are not used in traditional discrete choice econometric models because the number of variables exceeds the sample size. Only the machine learning methods applied here can handle this maximization problem.

insights into PE returns. Lastly, PE theory variables relate to how GPs organize funding (Lerner & Schoar, 2004; Maurin *et al.*, 2023), where the characteristics of LPs play a crucial role in securing capital across different market states, allowing GPs to capture attractive investment opportunities.

We also separate our sample into different groups that have underlying differences that might impact performance. To account for differences in risk-return profiles and investment strategies, we separate BO and VC funds. BO funds typically target mature companies with stable cash flows, while VC funds focus on early-stage firms with higher risks and greater potential upsides (Gompers & Lerner, 1999; Kaplan & Stromberg, 2009). Additionally, we divide our sample into first-time and sequential funds, aligning with the well-known phenomenon in PE called *fund performance persistence*—where prior fund performance predicts future returns for follow-on funds from the same GP (Kaplan & Schoar, 2005; Harris *et al.*, 2023), this persistence, explained by various theories (Lerner & Schoar, 2004; Hochberg *et al.*, 2014; Maurin *et al.*, 2023), highlights the importance of analyzing sequential funds in comparison to first-time funds to better understand performance. A first-time fund is the first fund in a sequence of funds that follows a defined strategy. An experienced GP may have several first-time funds, and it is certain that the first fund launched by a GP will be a first-time fund. A sequence fund is the sequential fund of this strategy; for example, a GP launches a fund called **Healthcare techs VC**, and launches 3 subsequent versions of this fund, raising funds every 2 years. **Healthcare techs VC II**, **Healthcare techs VC III**, and **Healthcare techs VC IV** will be sequence funds.

We distinguish our paper from the prior literature in several key ways. First, we are the first to use a more complete data available at the time of fundraising to predict future fund success, in contrast to studies that rely solely, for example, on past performance information from the same GP (Kaplan & Schoar, 2005; Harris *et al.*, 2023). Such information is often unavailable or dependent on interim performance proxies, which may be noisy or imprecise. Second, we introduce a GP identifier into the regressions, allowing us to pinpoint specific GPs that are more likely to succeed, an insight that can improve the screening processes of LPs. Additionally, our analysis separates Venture Capital (VC) and Buyout (BO) funds, as well as first-time and follow-on funds. VC funds are dedicated to investing in startups, companies with disruptive technology in their sector, ranging from building a new app, providing a new service, creating a new market, developing a new chemical molecule (for silicon batteries, for example), or building a new patent. We have countless examples of established companies today that have received VC investment, such as Facebook, Google, Instagram, and Uber. The companies invested in by these funds are developing a product, creating demand, and often do not yet have revenue, or have very low revenue. Buyout funds, on the other hand, are dedicated to investing in privately held companies that are already established, have a history with a product and revenue, and may need capital to grow or to carry out a turnaround in their segment. While Kruglikov & Forthun (2022) also employ machine learning, they group all fund types together. In our view, this distinction is important, as numerous empirical and theoretical studies highlight significant differences between these fund types, making it more appropriate to compare assets with similar characteristics. Furthermore, Morales & Tiozzo (2019) analyzes the probability that a fund's portfolio company will go public, inferring superior returns (or success) for those funds. In contrast, we use direct fund and GP information to predict performance without relying on indirect success measures such as IPOs. Additionally, Fernández Tamayo *et al.* (2023) employs textual analysis of fund prospectuses to capture performance, while we use other financial data and identification strategies.

The dataset for this study is compiled from multiple sources based on the type of information required. PE data is gathered from Preqin©, a specialized platform for alternative investments, while public market and macroeconomic data are sourced from FRED (Federal Reserve Economic Data) and other specialized providers. The sample period covers the years 2000 to 2017, consisting of 1,267 BO funds and 868 VC funds, totaling 2,135 unique funds.

Our results from applying the six models across different sample group combinations[4], are derived using cross-validation. We train each model with 75% of the total data and test its predictive capacity on the remaining 25%, repeating this process 100 times per model and sample separation. To ensure robustness, we randomly select the training and testing samples with replacement. Additionally, to mitigate survivorship bias—since Preqin's dataset is self-reported and includes over 70% of funds with a PME greater than one—we ensure a balanced distribution by structuring the training and testing datasets to include 50% of funds with PME values above and below one. We apply a similar approach to the IRR threshold, creating a more realistic environment for analysis.

In analyzing the results from our models, we find that ML techniques generally outperformed classical econometric discrete choice models, particularly for VC funds. For ML models, we observe that the predictive accuracy is largely driven by sequential VC funds, achieving an average accuracy of approximately 75% in predicting funds with abnormal returns based on our two performance metrics (Net IRR and Implied PME). In contrast, for first-time VC funds, the accuracy drops to around 60.0%. For buyout funds, the models perform less effectively, especially with first-time funds, where the results are nearly equivalent to random chance. When considering additional metrics such as precision, recall, specificity, and the F2 score, the same pattern emerges, with VC funds—particularly those in sequence—demonstrating stronger model performance across the board.

[4]We categorize our sample into six distinct groups based on fund type—Buyout or Venture Capital—and their classification as either first-time, sequence, or all funds: Buyout all, Buyout first-time funds, Buyout sequence funds, VC all, VC first-time funds, and VC sequence funds.

These prediction results align with Fernández Tamayo *et al.* (2023), where they correctly classify 75% of the outperformer funds in the top tercile using prospectus information. However, textual analysis relies heavily on the quality and completeness of prospectus documents, which are not always as readily available as the data used in this paper. Over time, GPs may standardize their approach in response to these findings, reducing the predictability and effectiveness of such analyses. In contrast, our investigation strategy uses structured, objective data, enabling more accurate and consistent predictions across various fund categories.

ML models have significant potential for uncovering insights in finance, but their "black box" nature often limits transparency. However, by employing variable selection techniques, we successfully identify key financial indicators driving our model predictions. Our analysis shows that variables related to GPs, macroeconomic factors, and fund information consistently hold more relevance across both VC and BO funds, regardless of the ML model or performance threshold. Among these, GPs stand out as the most influential predictor. This highlights the critical role of a thorough GP screening process in determining fund performance. Using ML tools, investors can make data-driven decisions and efficiently identify high-potential GPs.

We then examine the top 10 most influential variables across all simulations and observe the dominance of GP-related variables, particularly those linked to specific GP ID numbers. Additionally, macroeconomic variables such as CPI (Consumer Price Index for All Urban Consumers), federal funds rate, housing starts, and new private housing permits frequently emerge as key drivers. Interestingly, these variables are often interconnected: federal funds rate is typically adjusted in response to inflation (CPI), which in turn influences housing-related indicators like housing starts and new private housing permits.

Analyzing the beta direction for the CPI, our most relevant macroeconomic variable, reveals a negative correlation, indicating that higher inflation during the fundraising process is related to lower fund performance. Inflation has the potential to increase operational costs, complicating efforts to improve efficiency—an essential objective in PE (Kaplan & Stromberg, 2009). Additionally, higher inflation often triggers rising interest rates, slowing economic activity and raising the cost of capital, which may depress future valuations. If GPs plan to exit in such conditions, performance might suffer. While these are just a few reasons for lower performance during investments in inflationary periods, the findings highlight the critical role of market timing in PE. This aligns with literature on timing strategies (Robinson & Sensoy, 2016), with inflation emerging as a key factor for analysis.

In addition, for BO funds, "median first-day return" frequently appears as a key predictive variable across several models. This metric, often used to assess the initial performance of public offerings, seems to serve as a proxy for positive market sentiment (Ritter, 1984), indicating favorable conditions in the broader financial environment. In our Lasso and Ridge models, its positive coefficients indicate higher returns for BO funds during fundraising periods marked by positive market sentiment. This aligns with Ljungqvist *et al.* (2020), who shows that buyout funds tend to earn higher returns when investment opportunities improve, competition for deal flow eases, and credit market conditions loosen. These findings highlight the critical role of timing and sentiment in BO investments, emphasizing the need to carefully evaluate external market conditions.

Our contributions are threefold. First, we apply machine learning tools to predict performance alpha by using hard data available to investors at the time of fundraising, separating VC and BO funds, including a further division into first-time and sequence funds. This unique approach reflects the distinct underlying characteristics of these funds, which impact performance, and our results demonstrate that this separation enhances predictive power. Second, we include GP identifiers in our ML models, offering a new strategy for investors to identify specific GPs. Finally, we address the effects of unbalanced samples, which can introduce survivorship bias typical in PE databases.

The paper is organized as follows. Section II presents the dataset used in the analysis, Section III details the methodology, Sections IV and V address our results, and VI concludes the paper.

2.2 Data

The dataset was compiled from diverse sources, depending on the specific type of information required. For the PE data, we used Preqin©, a specialized platform for alternative investments, while public market and macroeconomic data were sourced from FRED and other specialized providers. The sample period spans from 2000 to 2017.

Specifically for PE, Preqin was used to collect information on funds, performance, LPs, and GPs, referencing data from 2022. Funds post-2017 were excluded to allow a five-year maturation period for performance data, aligning with the approach of Harris *et al.* (2023). From table 2.1, the sample includes 1,267 BO funds and 868 VC funds, totaling 2,135 unique funds. All VC subdivisions available in Preqin were included, while no subdivisions existed for BO funds in Preqin's database. In Table 2.2, we present the global regions and industries in which the funds operate. Most funds focus on North America and Europe, consistent with the higher deal values in these regions, which naturally attract more sophisticated investors (Bain&Company, 2024). Additionally, for BO funds, diversified industries dominate, while for VC funds, healthcare and IT emerge as the top two targeted sectors.

We use two measures of fund performance: Net IRR (net internal rate of return) and implied PME (public market equivalent). PME, as defined by (Kaplan & Schoar, 2005), is the ratio of the sum of discounted fund distributions to the discounted capital invested (with the S&P 500's total return as the discount rate). A PME greater than one indicates that the fund outperformed the S&P 500. However, Preqin does not report PME for all funds. To address this, we followed the approach of Sensoy *et al.* (2014) to derive an implied PME. Essentially, this strategy involves using the regression coefficients from Harris *et al.* (2014) to calculate PMEs from Net IRR, hence the term "implied PME".

Additionally, the performance metrics used in our models were not the original variables. Instead, we modeled these metrics as binary variables: invest or not invest. For implied PME, this adjustment is straightforward since an implied PME above 1 represents a good investment, while below 1 indicates otherwise. For Net IRR, we classified a fund as 1 only if it was in the top 33% of the entire sample, and as 0 otherwise.

The sample of funds was also separated into first-time funds and sequential funds. First-time funds are those classified by Preqin as 1 in their *fund number series* classification. In other words, these are the first-time funds of a specific series by a given GP, but not necessarily the GP's first overall fund. GPs might shift their focus to different regions or industries, positioning themselves in new ventures and taking on unfamiliar risks. This lack of prior experience can result in limited information for both GPs and LPs. We aim to capture this initial exposure through the *fund number series* classification provided by Preqin.

For sequential funds, this refers to subsequent funds launched by the same GP. To accurately separate the sequential funds, we followed the methodology described by Harris *et al.* (2023). Our initial step involved organizing the database by fund name, GP, vintage year, and sequence number. To ensure consistency, we excluded funds that did not align with their subdivision class or primary region of operation, such as those shifting focus from Europe to Asia. We also excluded funds with different characteristics, like annex or side funds. After sorting, we reviewed the sequence numbers to identify only adjacent pairs for analysis.

To examine aspects of private equity theory and empirical findings, we incorporate variables grounded in network theory. Following the methodology of Abreu Neto & Saito (2024), who analyze LP-GP interactions, they find that persistent LPs—also referred to as 1st Quartile LPs—those with strong, consistent reinvestments with GPs, serve as reliable indicators of GP quality/performance. This aligns with the theories proposed by Lerner & Schoar (2004); Hochberg *et al.* (2014); Maurin *et al.* (2023), which highlight a consistent match between successful GPs and specific LPs through follow-on funds, resulting in superior returns for these LPs. Additionally, we apply centrality measures at the GP level as a proxy for GP demand by LPs, akin to the oversubscription phenomenon identified by Kaplan & Schoar (2005), which is associated with higher returns. Similarly, Abreu Neto & Saito (2024) find that GPs with high centrality in a bipartite network of LPs and GPs tend to outperform their peers.

In addition to PE data, our analysis incorporated macroeconomic and public market information. For macroeconomic data, we utilized FRED and the Baker, Bloom, and Davis website, though the latter was used exclusively for a specific Uncertainty Index. Public market data was gathered from various sources, including Eikon, FRED, the Kenneth French website, and Jay R. Ritter's IPO Data website. Table 2.3 provides a detailed overview of the sources and corresponding website links for each variable.

2.3 Method

The central idea of this article is to use various econometric tools, including new machine learning techniques capable of handling big data, to predict a fund's potential success during the fundraising process. We aim to demonstrate, ex ante to the fund launch, the likelihood of a fund achieving abnormal performance based on different characteristic groups. Additionally, we seek to identify the main factors that influence PE returns. Abnormal performance is measured in two ways: PME above 1 or IRR above the first-tercile threshold (our dependent variables). To predict these binary outcomes, we employ discrete choice models.

The set of predictors (our independent variables or factors) is detailed in Table 2.3, grouped into six categories: GPs, fund information, PE theory and centrality measures, PE industry, macroeconomic factors, and market conditions/environment. These categories include 55, 58, and 55 covariates for VC first funds, VC sequence funds, and the full VC sample, respectively, and 54, 57, and 54 covariates for BO first funds, BO sequence funds, and the full BO sample, respectively. In VC, there is an additional strategy variable provided by Preqin, which is absent in BO data. Notably, categorical variables are transformed into dummy variables in the econometric models, significantly increasing the number of variables on the right-hand side of the equations. Due to the small sample size and the cross-validation approach used for population stratification, GP variables were excluded from traditional discrete choice econometric models, as the number of variables exceeded the sample size. Only the machine learning methods applied here are capable of handling this maximization issue to estimate the beta coefficients.

We conduct a horse race between six models to determine which has the best accuracy in predicting the binary events mentioned above. These models include traditional approaches like the Linear Probability Model (LPM), Logit, and Probit, as well as machine learning techniques such as Lasso, Ridge, and Random Forest (RF).

There are distinct idiosyncrasies between BO and VC funds, prompting us to conduct separate experiments for each fund type. BO funds typically target mature companies with stable cash flows, while VC funds focus on early-stage firms with higher risks and greater potential upsides (Gompers & Lerner, 1999; Kaplan & Stromberg, 2009). Additionally, we divide the experiment into three categories: first-time funds, sequential funds, and all funds within each type. The rationale is that during the fundraising process for sequential funds, more information is available, and in a market characterized by information asymmetry, this should make prediction easier—which is indeed the case. Moreover, performance persistence, a well-known phenomenon in PE where prior fund performance predicts future returns for follow-on funds from the same GP (Kaplan & Schoar, 2005; Harris *et al.*, 2023), further justifies separating the analysis for first-time and sequential funds. This persistence is explained by several theories (Lerner & Schoar, 2004; Hochberg *et al.*, 2014; Maurin *et al.*, 2023).

For each group of funds, we implement cross-validation by training the model on 75% of the total data and testing its predictive capacity on the remaining 25%. This process is repeated 100 times for each model and each experiment, and we collect the average results. The training and testing samples are selected randomly, with replacement after each iteration of the 100 times.

There is ongoing discussion about the presence of survivorship bias in PE market data, as it is self-reported and privately held. Indeed, over 70% of the funds in our data show PME greater than one. To create a more realistic environment and mitigate potential survivorship bias, we ensure that both the training and testing samples contain an equal 50/50 split of funds with PME greater than and below one, as well as for the IRR threshold.

In summary, we categorize the funds into six classes (Buyout full, Buyout first funds, Buyout sequence funds, VC full, VC first funds, VC sequence funds) and aim to predict two binary events (implied PME and IRR threshold). We compare the performance of six models (LPM, Logit, Probit, Lasso, Ridge, and Random Forest) using bagging with 100 cross-validation iterations for each experiment, totaling 7,200 prediction exercises. Figure 2.27 below illustrates our methodology.

2.3.1 Models

We use six models to predict binary events (0 for below threshold and 1 for above threshold). The first model assumes a simpler relationship between covariates and the dependent variable, with a linear marginal effect. Probit and

Logit models introduce non-linear functions, aiming for a better fit in binary outcomes. Ridge and Lasso models apply regularization, shrinking coefficients to reduce overfitting. Lastly, the Random Forest (RF) model captures complex, non-linear relationships using an ensemble of decision trees in an unsupervised learning framework.

2.3.1.1
LPM - Linear Probability Model

LPM applies linear regression to binary outcomes, estimating the probability that the dependent variable equals 1. It assumes a constant marginal effect of covariates on the event's probability, providing a straightforward approach but with potential limitations, such as predicted probabilities outside the [0,1] range.

$$P(y = 1|x) = \beta_0 + \sum_{i}^{N} \beta_i x_i \tag{2.1}$$

2.3.1.2
Logit

The Logit model captures non-linear relationships by applying a logistic function, ensuring predicted probabilities remain between 0 and 1.

$$P(y = 1|x) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} \tag{2.2}$$

2.3.1.3
Probit

Similar to Logit, the Probit model uses a normal cumulative distribution function (CDF) to model binary outcomes.

$$\Phi(x'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz \tag{2.3}$$

2.3.1.4
Random Forest (RF)

The Random Forest model builds multiple regression trees and using the principles of Bagging together, where each tree is trained on a random subset of the data. The model generates predictions by averaging the outputs of these trees, enhancing accuracy and reducing overfitting. The regression tree can be described as follows:

$$f(x) = E[P(y = 1|x)] = \sum_{m=1}^{M} W_m I(x \in R_m) = \sum_{m=1}^{M} W_m \phi(x; V_m) \tag{2.4}$$

R_m represents the tree regions m, W_m the average answer of P(y=1 | x) to this region, and V_m is the variable choice to be divided and the threshold. While x is the set of predictors, comprising approximately 55 variables across five groups, as detailed in Table 2.3.

A detailed explanation of how the algorithm determines the number of regions M and selects variables from each bootstrap sample is provided in Medeiros *et al.* (2021), Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. This can be found in the appendix of the Journal of Business and Economic Statistics, Section B8.

2.3.1.5
Lasso

Lasso introduces L1 regularization, shrinking coefficients and selecting variables by setting irrelevant ones to zero (Bühlmann & Van De Geer, 2011). This model is useful in high-dimensional settings to reduce overfitting and improve interpretability.

$$\hat{\mu}(\lambda), \hat{\beta}(\lambda) = argmin_{\mu,\beta}(\frac{1}{n}\sum_{i=1}^{n} \rho_{\mu,\beta}(X_i, Y_i) + \lambda\|\beta\|_1) \tag{2.5}$$

Being $Y_i|X_i = x\ Bernoulli(\pi(x)), \rho(f,y) = ln(1+exp(-(2y-1)f) = ln(1+exp(-yf))\ y = 2y-1 \in -1,1, ln(\frac{\pi(x)}{1-\pi(x)}) = \pi + \sum_{j=1}^{p} \beta_j x^{(j)} = f_{\mu,\beta}(x)$.

2.3.1.6
Ridge

Ridge regression applies L2 regularization, shrinking coefficients without eliminating variables, making it effective in managing multicollinearity (Bühlmann & Van De Geer, 2011).

$$\hat{\mu}(\lambda), \hat{\beta}(\lambda) = argmin_{\mu,\beta}(\frac{1}{n}\sum_{i=1}^{n} \rho_{\mu,\beta}(X_i, Y_i) + \lambda\beta^2) \tag{2.6}$$

Being $Y_i|X_i = x\ Bernoulli(\pi(x)), \rho(f,y) = ln(1+exp(-(2y-1)f) = ln(1+exp(-yf))\ y = 2y-1 \in -1,1, ln(\frac{\pi(x)}{1-\pi(x)}) = \pi + \sum_{j=1}^{p} \beta_j x^{(j)} = f_{\mu,\beta}(x)$.

2.4 Results

To evaluate the results of our prediction models, we use three main metrics: out-of-sample accuracy, ROC curves and confusion matrices. Our confusion matrices, are the mean results from 100 iterations, tested out-of-sample. These matrices summarizes how well the model's predictions align with actual outcomes by categorizing the results into four groups: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). From this breakdown, we can derive key performance metrics that help assess the accuracy and effectiveness of each model (Powers, 2020).

We present our results by first separating Buyout and Venture Capital funds. Within each type of fund, we further divide the analysis into 1st-time, sequence and all funds. The funds have special features that justify this separation, and we also saw an assertiveness improvement of the models. For each case, the results show that ML techniques consistently outperform classical discrete choice econometric models in terms of accuracy—the ability to correctly predict ones and zeros, indicating whether funds exceed or fall below the proposed threshold.

2.4.1 VC results

Panel A of Table 2.4 presents the accuracy results for VC funds. The first noticeable pattern is that ML techniques consistently outperform traditional discrete choice models across all cases. For first-time funds using the PME measure, the results from traditional models are similar than random chance. The Lasso model emerges as the top performer in all experiments.

When analyzing the confusion matrix in table 2.5, the results show a high degree of consistency, with the models being better at predicting 1s for the PME metric and 0s for the IRR threshold metric. Both Type I and Type II errors[5] are not much distant, which is a promising outcome if these models are intended for use as decision-making tools by LPs at the investment stage. Achieving 75% accuracy in predicting a fund's relative success during the fundraising process would be highly beneficial in helping LPs decide whether to invest.

When we analyze the ROC curves in the set of 6 figures in 2.4 and 2.8, the same pattern is observed, corroborating the methodological tool used and the specific characteristics of this market. Market—asymmetric information, skill, and revealed type—is evidenced by the greater predictive capacity for success in sequence funds. First Funds are much more difficult to predict the performance of, and the shape of the ROC curves demonstrate this. Even so, ML models still manage to bring a more satisfactory result. When we look at the ROC curves at the IRR threshold, the curves of the 3 ML models are visibly much better.

The Lasso model remains the best performer in almost all cases, with Ridge having the best AUC at the IRR threshold for all funds, and PME Threshold for first funds.

It's also important to note that the population of VC funds is smaller than BO funds in all our experiments. However, despite the smaller sample size, the prediction results for the VC funds are significantly stronger in all cases.

The superior performance of ML models, particularly in predicting the success of VC funds, is evident. Despite potential complexity—such as having more independent variables than sample size due to the inclusion of GP dummies (as discussed in section 3)—the models used in this study remain interpretable. In the next section, we explore the most significant variables for predicting performance, providing further theoretical insights by dissecting the results.

[5]A Type I Error or False Positive (FP) in our confusion matrix occurs when the model incorrectly predicts a positive outcome (e.g., abnormal fund performance) when the actual outcome is negative (e.g., the fund underperforms). On the other hand, a Type II Error or False Negative (FN) happens when the model incorrectly predicts a negative outcome (e.g., underperformance) when the actual outcome is positive (e.g., the fund outperforms).

2.4.2 Buyout results

Panel B of Table 2.4 presents the models' accuracy in predicting, out of sample, whether a BO fund will achieve abnormal performance. It can be observed that the models perform poorly, especially with first-time funds, with results nearly equivalent to random guessing. Specifically, for the PME measure, the task is to identify which funds underperformed (PME less than one), despite over 90% of funds having a PME greater than one.

The results improve slightly for sequence funds, with Lasso and Ridge achieving 66.1% and 68.4% accuracy respectively. This aligns with the theoretical expectation that, over time, some information becomes more accessible in these markets characterized by asymmetric information.

In table 2.6, which provides the confusion matrices, the results are relatively balanced. The Ridge model, using the PME threshold, performs better than the others in predicting both ones and zeros, supporting its higher accuracy.

In figure 2.12 and 2.16, we can also observe that the Buyout market is a more competitive market where skill and return persistence are not as latent as in the VC market, and this is reflected in lower models' accuracy. Predicting success in first funds is a more difficult task, and the models are generally similar, with the ROC curves proving unstable, even with traditional econometric models having a higher AUC at the IRR threshold. The ROC curve metric maintains the same conclusion; for the other samples, ML models win, with Random Forest having the highest AUC at the IRR Threshold for Sequence Funds, Ridge for PME threshold in Sequence Funds and All Funds, and Lasso for IRR threshold for All Funds.

Overall, the BO results, particularly with the Lasso and Ridge model for sequence funds, provide useful insights for LP decision-makers at the time of investment. However, the more promising results are found in the VC funds, which will be discussed in the next section.

2.5 Results - Variable Selection

Machine learning models have significant potential for uncovering insights in finance, but their "black box" nature often limits transparency. In this section, however, by employing variable selection techniques, we were able to identify key financial indicators that drive our model predictions, thereby shedding light on the complex workings of machine learning models in financial applications.

We divided the variable selection analysis into two parts. Initially, in section 2.5.1, we perform a summary evaluation where it is possible to compare all models in the same way, using shapley values. In this evaluation, we look at the 3 ML models, separating Buyout, and Venture Capital in just one sample and one threshold (funds full, and IRR threshold), and we group the variables. In this way, we can arrive at the main conclusions, in a simplified and directly comparative way, that predict high performance.

In section 2.5.2, we look at the models with their specific characteristics, detailing all the samples, with a detailed drilldown in order to understand all the predictive nuances of performance.

2.5.1 Shapley Values - Group Variables

Shapley Values is a concept derived from Cooperative Game Theory (Shapley (1953)), a way of distributing the total payout among the players of a coalition based on each player's marginal contribution to all possible subsets of the coalition. It measures the importance , the weight of each individual to the final outcome of a "game". When applied to Machine Learning (ML), the "game" is a model's prediction (our case the models Lasso, Ridge, and Random Forest) for a single instance of data (our prediction that a fund will achieve or not achieve top performance), and the "players" are the variables (features) used to make that forecast.

We are using the shapley values with the intention of bringing common interpretability from ML techniques used here. Decomposing the features prediction contribution, with the values of each variable representing the average marginal contribution of that variable to the prediction, considering all possible orders of inclusion of variables in the model. [6]

Evaluating the importance of variables using Shapley Values allows for a direct and powerful comparison between different Machine Learning models. Regardless of whether we are using either a simple linear model (such as Lasso or Ridge) or a tree-based model (such as Random Forest), the Shapley value serves as a unified and model-agnostic metric. This means that, while the Lasso model may have an importance calculated based on coefficients and the Random Forest based on impurity reduction (Gini importance), the Shapley value standardizes this measure. Therefore, it is possible to say that the variable has the same magnitude of importance for prediction, even if one model is linear and the other non-linear, facilitating the comparison of the influence of variables through different algorithmic approaches.

Figure 2.20 shows a set of 3 graphs with Shapley values from the ML models for the IRR threshold in the sample with all funds, and Figure 2.24 shows a set of 3 graphs with Shapley values for the equivalent buyout market. The idea of this analysis is to provide a general summary of the variables that are most relevant to predicting high performance, and the main conclusion lies in the skill of the managers. The private market is a market of asymmetric information, and the choice of manager is the main characteristic for predicting high performance. Nuances are observed, and we see that the buyout market is a more established and competitive market, and the manager's characteristics are less important in this case, with macroeconomic factors and the market environment at the time of fundraising being relevant in 31.8% (in the Lasso model) to predict performance. Choos-

[6] The use of Shapley values as feature importance applied to Machine Learning is found in the seminal work of Lundberg & Lee (2017), that Shap Values can be approximated by solving a weighted least squares optimization problem.

ing a favorable macroeconomic environment and a positive atmosphere, that is, having good timing to market, positively reflects whether buyout funds will have high performance. However, for the venture capital market, macroeconomic conditions and the business environment are less important for predicting high performance, with the GP's ability being the main characteristic in predicting high performance. Not surprisingly, the models perform better with sequence funds.

In the next section, we will evaluate, within the characteristics of each model, the primarily selected variables and the potential economic insights in predicting high performance at the time of fundraising, with different samples and different thresholds.

2.5.2 Coeficients detailed

In Tables 2.7 and 2.8, we present the main descriptive statistics for each category of variables used in our ML models. In Lasso and Ridge models, regressions capture the strength and direction of the linear relationship between variables and outcomes through beta coefficients (β), while RF evaluates variable importance based on its contribution to predictive accuracy in non-linear interactions.

To identify key drivers of performance across different modeling approaches, we aim to compare the absolute values of the beta coefficients for Lasso and Ridge, alongside the variable importance scores from RF, as detailed in Tables 2.9 and 2.10. Additionally, we refer to the outputs from ML models as "feature significance measures" because they indicate the relative importance or contribution of each feature (or variable) in making predictions.

Our results reveal that the categories of GPs and Macro factors are consistently more relevant than others for both VC and BO across different ML models and performance thresholds, with GPs being by far the most influential. This highlights the importance of a thorough screening process when selecting GPs, as they emerge as a key predictive variable for fund performance.

When examining the top 10 most influential variables across all simulations, we find that GPs, macro consistently emerge as the primary groups. This pattern holds true for both BO and VC funds. The dominance of GP-related variables, particularly those identified by GP ID numbers, is evident. Additionally, CPIAUCSL frequently emerges as a key macroeconomic variable, showing a negative beta direction. This shows the critical impact of inflation when investing in VC and BO funds. Interestingly, for BO funds, "median first day return" typically appears as a key predictive variable in several models. This variable, along with other PE industry variables identified in RF models, likely serves as a proxy for positive market sentiment, highlighting the importance of timing when participating in BO funds, particularly for first-time funds.

2.5.2.1
VC - Categories/Group Analysis - feature significance measures

In Table 2.9, we present key statistics for the absolute beta coefficients from Ridge and Lasso, along with variable importance scores from Random Forest (RF), applied to VC funds. These models are used to predict PME (Public Market Equivalent) and Net IRR (Internal Rate of Return) thresholds across different fund samples, including the Full Sample, First-Time Funds, and Sequence. For each sample, the variables used in the predictions are categorized into different groups: Fund Information, GPs, Macro, Market Conditions, PE Industry, and PE Theory & Centrality Measures.

When assessing the main categories, we find that the top three are GPs, Macro, and Fund Information. However, across various metrics—such as the sum of absolute values, averages, medians, and maximums— it is clear that GPs consistently have the most significant impact on both PME and Net IRR models. The GPs category demonstrates the highest sum of absolute values and strong averages, suggesting a considerable and consistent influence on performance metrics across various samples, including first-time and sequence funds. For example, in the full sample, the GPs group shows sum values of 117.47 (Ridge) and 133.46 (Ridge) for PME and Net IRR models, respectively, which far exceeds those of other groups. This indicates that the performance of GPs is crucial across different fund samples, making it the primary focus for decision-making.

The Macro category also plays a important role, especially when considering the maximum values observed. In certain cases, such as with sequence funds, Macro factors achieve a maximum value of 7.23 (PME Ridge) and 7.38 (Net IRR Ridge), indicating that while they may not always have the highest average impact, they can also exert influence under specific conditions/simulations. This variability suggests that macroeconomic variables should be closely monitored, particularly in environments with significant economic fluctuations. The consistency in the direction of action between PME and Net IRR models for Macro factors further strengthens this conclusion.

In summary, GPs are the most influential group in both PME and Net IRR models, across both first-time and sequence funds, with high sums and averages indicating significant impact on performance. Macro factors also play a important role, particularly under certain conditions, as suggested by their maximum values. Fund Information is slightly less influential but still valuable, particularly for its consistency across models. Other groups, such as Market Conditions, PE Industry, and PE Theory & Centrality Measures, are less relevant due to their lower sums and averages. Strategic focus should be on GPs and Macro especially for first-time funds.

2.5.2.2
Top 10 most influential variables in VC funds

Table 2.13 shows that the top 10 variables with the highest feature significance measures in VC funds from our ML models. The analysis reveals that GP-related and Macro variables are the primary groups influencing performance across different models.

The dominance of GP related variables appears prominently in all models, performance threshold and fund sample levels. In the Ridge and Lasso models, the most significant variables often exhibit a negative direction, indicating potential challenges that specific GPs or their decisions might impose on fund performance.

In addition to GP-related variables, macroeconomic indicators such as CPIAUCSL and RPI in the Ridge and Lasso models highlight the sensitivity of private equity performance to broader economic conditions. The negative direction of their betas suggests that higher inflation correlates with lower fund performance, indicating that VC funds may struggle to maintain returns in inflationary environments, particularly at the time the fund begins its operations.

The RF model highlights FEDFUNDS (Federal Funds Rate) along with HOUST (Housing Starts) and PERMIT (New Private Housing Permits) as key drivers of private equity performance. Meanwhile, the Lasso and Ridge models emphasize CPIAUCSL (Consumer Price Index). These variables are interconnected: FEDFUNDS is often adjusted in response to changes in inflation (CPIAUCSL), and this, in turn, affects housing-related indicators like HOUST and PERMIT.

The inclusion of HOUST and PERMIT in the RF model suggests that the model captures broader economic conditions, where interest rates influence both inflation and the housing market. This reinforces the idea that macroeconomic factors—particularly those linked to inflation, monetary policy, and housing market conditions—are critical in shaping both relative (PME) and absolute (Net IRR) returns in private equity. The RF model's focus on these variables complements the direct emphasis on inflation seen in the Lasso and Ridge models, providing a more comprehensive view of how economic conditions impact fund performance.

2.5.2.3
BO - Categories/Group Analysis - feature significance measures

In our analysis of BO fund performance (table 2.10), similar to VC funds, we also found that the key categories are GPs and Macro factors.

GPs consistently emerge as the most influential category across all samples and thresholds. For example, in the Full Sample under the PME threshold, the Ridge model shows a sum of absolute betas of 81.08 with an average of 0.18, median of 0.13, and a maximum value of 0.85. This strong influence is mirrored under the Net IRR threshold, where the Ridge model again highlights the GPs' dominance with a sum of 107.90 and an average of 0.24. These findings suggest that GP characteristics are crucial across different performance measures and stages of fund development, indicating that PE firms should prioritize enhancing the capabilities and track records of their GPs.

Macro factors show varying levels of influence depending on the sample and threshold. In the Sequence Sample under the PME threshold, the Ridge model highlights a significant impact with a sum of 4.88 and an average of 0.24. This impact is slightly reduced under the Net IRR threshold but remains notable, suggesting that macroeconomic conditions are particularly relevant for sequential funds. The variability in the influence of Macro factors indicates that while they are important, their relevance may fluctuate depending on the specific context of the fund, requiring PE firms to stay attuned to broader economic trends.

When comparing the thresholds (PME and Net IRR), both generally indicate the same direction of action, particularly highlighting the dominance of GPs. However, the PME threshold appears to provide slightly more consistent results across different samples, suggesting it might be a more reliable performance measure in this context. Additionally, between the First Fund and Sequence Fund samples, there is a clearer direction and more robust conclusions drawn from the Sequence Sample, likely due to the accumulated experience and data available from multiple fund cycles.

2.5.2.4
Top 10 most influential variables in BO funds

The table 2.14 reveals that GP-related and Macro variables are also, just like for VC funds, important categories influencing performance across different models. Specially for our linear models Ridge and Lasso macro variable CPIAUCSL was the most important.

However, different from VCs funds, in the linear models ridge and lasso the group market conditions and environment delivers interesting predictive power - being represented by "median first day return". This happens independent of the performance threshold, and fund sample.

The "median first day return" is a key predictive variable in Lasso and Ridge models as it effectively be capturing overall market sentiment and investor confidence during the time a fund is launched. This variable serves as a proxy for market conditions and timing (Ritter, 1984), and its positive direction shows that higher "median first day return" provides better performance prediction across thresholds and fund samples.

When analyzing the RF model, in the first-time funds sample, many of the variables identified by the RF model can be seen as capturing sentiments or conditions similar to those indicated by the "median first day return." These variables reflect investor confidence, market conditions, and sectoral or regional attractiveness, which are all factors that could drive both successful first-day returns in IPOs and the performance of first-time funds. While these variables may not be identical to the median first-day return, they can serve as proxies or are correlated with similar market sentiments, making them relevant in assessing the performance of first-time funds.

On the other hand, the lack of a similar pattern in sequential funds compared to first-time funds likely reflects the different nature and drivers of these funds. Sequence funds might benefit from an established GP track record and investor base, which shifts the focus away from short-term market sentiment (captured by the "median first day return") towards broader macroeconomic conditions and long-term strategic factors. The RF model, therefore, emphasizes variables that reflect these more stable and long-term influences, explaining why "median first day return" and similar sentiment-based variables are less prominent in the sequence funds sample.

2.6 Conclusion

We demonstrate that recent advances in machine learning methods, combined with the availability of new, rich datasets, make it possible to improve PE fund performance forecasts. Models such as Ridge, Lasso, and Random Forest generate more accurate predictions than traditional discrete choice econometric methods. These results leverages the value of ML and big data in private equity forecasting.

Our analysis shows that predictions for VC funds yield better outcome projections compared to BO funds, where Lasso models performed better than others. When we separate the sample into first-time and sequence funds, we find that the improved accuracy primarily comes from sequence funds rather than first-time funds.

Examining the variables selected by the ML models during the fundraising process, we identify inflation—measured through the CPI index—as a key driver of performance. Its negative coefficients indicate that inflationary periods challenge GPs to avoid overpaying and their portfolio companies in operating effectively to deliver superior performance. However, this finding promotes a deeper investigation into the reason behind the negative correlation between inflation and performance and creates a promising area for future research. Timing investments around inflation is crucial for generating alpha.

Although we apply ML techniques, the significance of GPs, as demonstrated by several other empirical studies on PE performance, remains consistent with findings using traditional econometric methods. In other words, this offers an alternative way to capture the fixed effects of GPs in a big data environment. Also, by identifying individual GPs throughout the simulations, we can pinpoint the top-performing GPs and show that this group plays a critical role in improving PE prediction accuracy.

2.7 Figures and Tables

Table 2.1: Summary Statistics

The table provides descriptive statistics for a sample of 1,267 venture capital and 868 buyout funds with vintage years between 2000 and 2017, categorized by the type of funds used (full sample, first-time funds, and sequential funds). Net IRR performance data is sourced from Preqin, while the Implied PME was calculated using regression coefficients reported by Harris *et al.* (2014), following a similar approach to Sensoy *et al.* (2014). Vintage year refers to the year the fund begins its operations. Size indicates the fund's volume in terms of invested capital. The overall fund sequence number represents the sequential fund reference number for a specific GP.

Fund Characteristics	Full Sample				First-time Funds				Sequence Funds			
	N	Mean	Median	SD	N	Mean	Median	SD	N	Mean	Median	SD
Buyouts funds												
Net IRR (%)	1267	17,24	15,80	14,34	177	17,72	16,00	14,49	841	17,57	16,00	13,37
Implied PME	1267	1,44	1,40	0,38	177	1,46	1,42	0,40	841	1,45	1,41	0,36
Vintage Year	1267	2009	2008	5	177	2006	2006	4	841	2010	2010	5
Size (millions of dollars)	1267	1.515,9	602,0	2.487,7	177	435,8	265,0	559,5	841	1.951,8	803,2	2.876,6
Overall fund sequence number	1267	5	3	6	177	2	1	4	841	6	4	6
Venture Capital funds												
Net IRR (%)	868	12,18	9,65	31,82	170	11,77	10,15	18,11	479	11,37	9,94	17,09
Implied PME	868	1,37	1,28	1,12	170	1,35	1,31	0,64	479	1,34	1,29	0,60
Vintage Year	868	2008	2007	5	170	2007	2007	5	479	2008	2008	5
Size (millions of dollars)	868	301,3	206,0	358,4	170	134,5	89,5	167,4	479	371,5	275,4	387,3
Overall fund sequence number	868	5	4	6	170	3	1	4	479	5	4	4

Table 2.2: Funds Characteristics

The table presents characteristics of a sample of 1,267 venture capital and 868 buyout funds raised between 2000 and 2017, focusing on their primary investment regions and core industries. The statistics are provided for the full sample, as well as for first-time and sequential funds. The region and core industry classifications were provided by Preqin.

Funds Region Focus and Core Industries	Buyout Funds			Venture Capital Funds		
	Full Sample	First-time Fund Sample	Sequence Fund Sample	Full Sample	First-time Fund Sample	Sequence Fund Sample
Funds Primary Region Focus						
Africa	5	1	1	5	2	2
Americas	25	4	13	8	5	2
Asia	65	11	39	81	28	31
Australasia	37	6	22	9	7	1
Europe	383	52	248	110	35	49
Middle East & Israel	12	3	8	27	5	14
North America	740	100	510	623	87	377
Diversified Multi-Regional	-	-	-	5	1	3
Total	*1.267*	*177*	*841*	*868*	*170*	*479*
Funds Core Industries						
Business Services	18	2	15	5	2	3
Business Services, Consumer Discretionary	1	1	-	-	-	-
Business Services, Information Technology	2	-	1	-	-	-
Consumer Discretionary	79	14	51	14	5	7
Consumer Discretionary, Financial & Insurance Services	1	-	1	-	-	-
Consumer Discretionary, Raw Materials & Natural Resources	1	-	-	-	-	-
Diversified	924	120	626	179	38	82
Energy & Utilities	12	1	7	26	5	14
Energy & Utilities, Real Estate	2	-	2	-	-	-
Financial & Insurance Services	14	5	6	9	3	3
Healthcare	32	7	19	197	44	121
Healthcare, Information Technology	6	-	4	65	5	41
Industrials	73	13	42	1	1	
Information Technology	54	6	35	247	50	136
Information Technology, Telecoms & Media	6	4	2	66	4	40
Others	5	1	2	13	2	6
Raw Materials & Natural Resources	3	-	2	5	1	3
Real Estate	9	1	8	-	-	-
Telecoms & Media	25	2	18	41	10	23
Total	*1.267*	*177*	*841*	*868*	*170*	*479*

Table 2.3: Explanatory Variable description and other information

This table presents all 54 explanatory variables utilized in our prediction models. We have categorized each variable into six distinct groups based on the original exposure: GPs, Fund Information, PE Theory & Centrality Measures, PE Industry, Macro and Market Conditions/Market Environment. (1) Only for VC funds, as Preqin does not detail the strategy for BO funds. The VC strategies are: Early Stage: Seed, Early Stage: Start-up, Early Stage, Venture (General), Expansion / Late Stage. (2) We separate the regions as: North America, Europe, Asia, Australasia, Middle East & Israel, Americas. (3) Funds core industries are given/classified by Preqin. (4) For every vintage year, we found eigenvector centrality measure for GPs using a past 5yr rolling window. (5) For each fund, we calculated the average eigenvector centrality of all LPs invested in that fund. (6) For each fund, we determined the proportion of LPs classified in the 1st quartile (with the highest centrality measures) relative to the total number of LPs in the fund. (7) Time variable, where we have funds vintage year minus 1980 (base year). (8) https://www.policyuncertainty.com/ (9) https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (10) https://site.warrington.ufl.edu/ritter/ipo-data/

Explanatory Variable	Group Classification	Type of Variables	Source
GPs Firm Identification (ID)	GPs	Categorical	Preqin
Fund Strategy [1]	Fund information	Categorical	Preqin
Fund Primary Region of Focus [2]	Fund information	Categorical	Preqin
Fund Core Industries [3]	Fund information	Categorical	Preqin
GPs Headquarter Location/Country	Fund information	Categorical	Preqin
Fund Size (U$MM)	Fund information	Numerical	Preqin
Fund Number Overall in GP	Fund information	Numerical	Preqin
Eigenvector ex-ante Centrality measure GP [4]	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Average LPs Eigenvector Centrality per fund [5]	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Proportion of 1Q LPs (ex-ante) per fund [6]	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Quantity of LPs per fund	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Number of LPs reinvested per fund	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Number of 1Q LPs reinvested per fund	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Fund Size (U$MM) previous fund from same GP	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Industry Dry Powder	PE Industry	Numerical	Preqin
Industry Fund Raising Volume	PE Industry	Numerical	Preqin
Industry Fund Raising Number	PE Industry	Numerical	Preqin
Industry Year [7]	PE Industry	Numerical	Preqin
University Michigan - Consumer Sentiment (UMCSENTx)	Macro	Numerical	FRED
Uncertainty Index (Uncertain)	Macro	Numerical	Baker, Bloom and Davis website [8]
Housing Starts: Total New Privately Owned (HOUST)	Macro	Numerical	FRED
New Private Housing Permits (PERMIT)	Macro	Numerical	FRED
Efective Federal Funds Rate (FEDFUNDS)	Macro	Numerical	FRED
1-Year Treasury Const. Minus Fed Funds (T1YFFM)	Macro	Numerical	FRED
5-Year Treasury Const. Minus Fed Funds (T5YFFM)	Macro	Numerical	FRED
10-Year Treasury Const. Minus Fed Funds (T10YFFM)	Macro	Numerical	FRED
Moody's Seasoned Aaa Corp. Bond Minus Federal Funds Rate (AAAFFM)	Macro	Numerical	FRED
Moody's Seasoned Baa Corp. Bond Minus Federal Funds Rate (BAAFFM)	Macro	Numerical	FRED
Help-Wanted Index for United States (HWI)	Macro	Numerical	FRED
Ratio of Help Wanted/No. Unemployed (HWI Ratio)	Macro	Numerical	FRED
Unemployment rate (UNRATE)	Macro	Numerical	FRED
Initial Jobless Claims (CLAIMSx)	Macro	Numerical	FRED
Industrial Production (INDPRO)	Macro	Numerical	FRED
Retail Price Index (RPI)	Macro	Numerical	FRED
Capacity Utilization: Manufacturing (CUMFNS)	Macro	Numerical	FRED
Crude Oil, spliced WTI and Cushing Price (OILPRICEx)	Macro	Numerical	FRED
Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)	Macro	Numerical	FRED
US GDP (last year)	Macro	Numerical	FRED
S&P 500 returns	Market Conditions/Environment	Numerical	FRED
S&P Dividend Yield	Market Conditions/Environment	Numerical	FRED
S&P Price-Earnings Ratio	Market Conditions/Environment	Numerical	FRED
Cleveland Fed's VIX (VIXCLSx)	Market Conditions/Environment	Numerical	FRED
Risk Premium: Market Return Minus Risk-Free Rate (Mkt_RF)	Market Conditions/Environment	Numerical	Kenneth R. French website [9]
Nasdaq returns	Market Conditions/Environment	Numerical	Eikon
Number of IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter
Average First day Return Proceeds Weighted	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Median First day Return	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Money left on the Table	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Aggregate proceeds	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Venture Capital (VC) IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Buyout IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Tech IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Venture Capital (VC) backed proceeds	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Tech proceeds	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]

Table 2.4: Performance Prediction Accuracy - VC and BO funds

The table shows the accuracy of the models for performance prediction. Panel A presents the results for VC funds, considering different sample types (all funds, first-time funds, and sequence funds) and performance thresholds (PME and Net IRR). Panel B shows the results for BO funds. Each result represents the average of 100 cross-validation procedures. The model parameters are trained using 75% of the total available sample, selected randomly while maintaining the original ratio of ones (funds with PME greater than 1 or above the IRR threshold) and zeros (funds with PME below 1 or below the IRR threshold). The sample is forced into a 50/50 split between ones and zeros, meaning part of the data is disregarded. After training, the model is tested out-of-sample on the remaining 25%, also structured with a 50/50 split between ones and zeros.

Panel A: Venture Capital Funds		RF	Lasso	Ridge	LPM	Probit	Logit
All Funds	PME Threshold	0,694	0,711	0,688	0,691	0,683	0,675
	IRR Threshold	0,707	0,730	0,713	0,726	0,712	0,712
First-time Funds	PME Threshold	0,541	0,600	0,601	0,509	0,515	0,518
	IRR Threshold	0,630	0,616	0,615	0,524	0,511	0,507
Sequence Funds	PME Threshold	0,717	0,747	0,705	0,705	0,685	0,683
	IRR Threshold	0,713	0,746	0,710	0,725	0,709	0,707
Panel B: Buyout Funds							
All Funds	PME Threshold	0,649	0,669	0,676	0,642	0,638	0,634
	IRR Threshold	0,626	0,649	0,618	0,636	0,633	0,632
First-time Funds	PME Threshold	0,626	0,632	0,613	0,547	0,547	0,547
	IRR Threshold	0,506	0,526	0,525	0,505	0,510	0,506
Sequence Funds	PME Threshold	0,648	0,661	0,641	0,601	0,580	0,581
	IRR Threshold	0,636	0,642	0,620	0,631	0,630	0,628

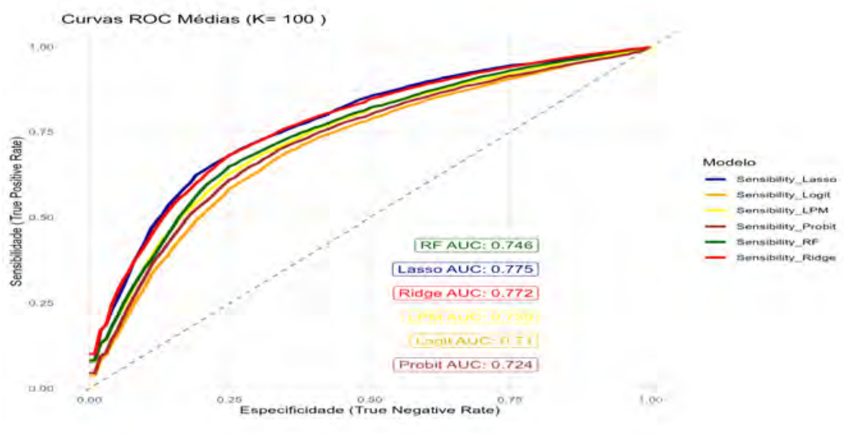
Table 2.5: Confusion Matrices - Venture Capital Funds

The table presents the average results of each confusion matrix from the 100 cross-validation procedures explained for Venture Capital funds across different performance thresholds.Each matrix presents the true positive (TP - bottom right corner), true negative (TN - top left corner), false positive (FP - top right corner), and false negative (FN - bottom left corner) rates. Panels A, B, and C consider different sets of funds included in the analysis. Each panel presents the results based on varying fund types, such as all funds, first-time funds, or sequence funds. Details on accuracy, precision, recall, specificity, and F-score can be found in Powers (2020).

Venture Capital Funds *Panel A: All Funds*	RF		Lasso		Ridge	
PME Threshold	19,0	12,4	18,8	12,5	13,2	18,2
	10,6	33,0	9,1	34,5	5,2	38,5
Accuracy	69,4%		71,1%		68,8%	
Precision	72,8%		73,4%		67,9%	
Recall	75,7%		79,1%		88,1%	
Specificity	60,6%		60,1%		42,0%	
F-score	74,2%		76,1%		76,7%	
IRR Threshold	34,1	11,7	34,7	11,2	39,5	6,4
	11,4	21,8	10,1	23,0	16,3	16,9
Accuracy	70,7%		73,0%		71,3%	
Precision	64,9%		67,3%		72,5%	
Recall	65,7%		69,4%		50,9%	
Specificity	74,4%		75,6%		86,0%	
F-score	65,3%		68,3%		59,8%	

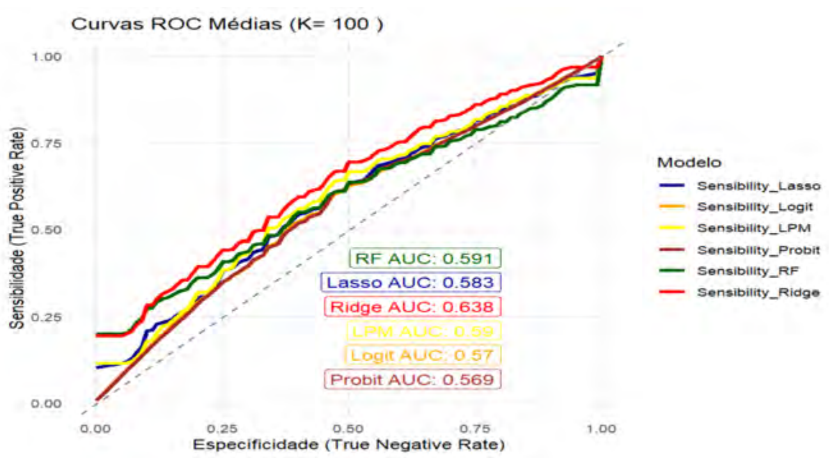
Panel B: First Funds	RF		Lasso		Ridge	
PME Threshold	2,2	3,2	1,7	3,6	1,5	3,9
	2,7	4,9	1,6	6,1	1,3	6,3
Accuracy	54,1%		60,0%		60,1%	
Precision	60,1%		62,7%		62,0%	
Recall	64,1%		79,3%		82,5%	
Specificity	39,9%		32,1%		28,2%	
F-score	62,1%		70,0%		70,8%	
IRR Threshold	6,7	2,6	7,3	2,0	7,6	1,7
	3,3	3,4	4,1	2,6	4,5	2,2
Accuracy	63,0%		61,6%		61,5%	
Precision	56,5%		56,1%		56,7%	
Recall	51,0%		38,1%		32,9%	
Specificity	71,7%		78,6%		82,0%	
F-score	53,6%		45,4%		41,6%	

Panel C: Sequence Funds	RF		Lasso		Ridge	
PME Threshold	16,9	10,2	17,6	9,5	10,7	16,3
	8,4	30,5	7,2	31,7	3,2	35,8
Accuracy	71,7%		74,7%		70,5%	
Precision	74,9%		77,0%		68,7%	
Recall	78,3%		81,4%		91,9%	
Specificity	62,3%		65,0%		39,7%	
F-score	76,6%		79,1%		78,6%	
IRR Threshold	30,7	11,3	31,2	10,7	35,6	6,3
	9,4	20,7	7,5	22,5	14,6	15,5
Accuracy	71,3%		74,6%		71,0%	
Precision	64,7%		67,7%		70,9%	
Recall	68,8%		74,9%		51,5%	
Specificity	73,2%		74,4%		84,9%	
F-score	66,7%		71,1%		59,7%	



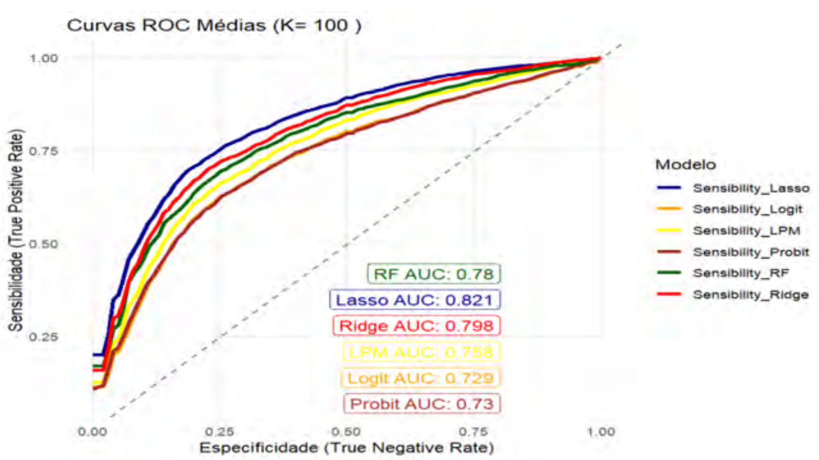
2.1(a):
b

Figure 2.1: ROC Curve - PME Threshold - All Funds



2.2(a):
b

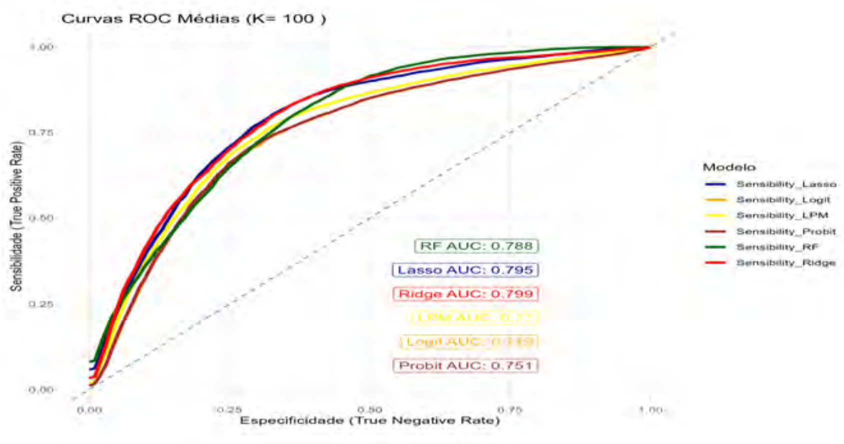
Figure 2.2: ROC Curve - PME Threshold - First Funds



2.3(a):
b

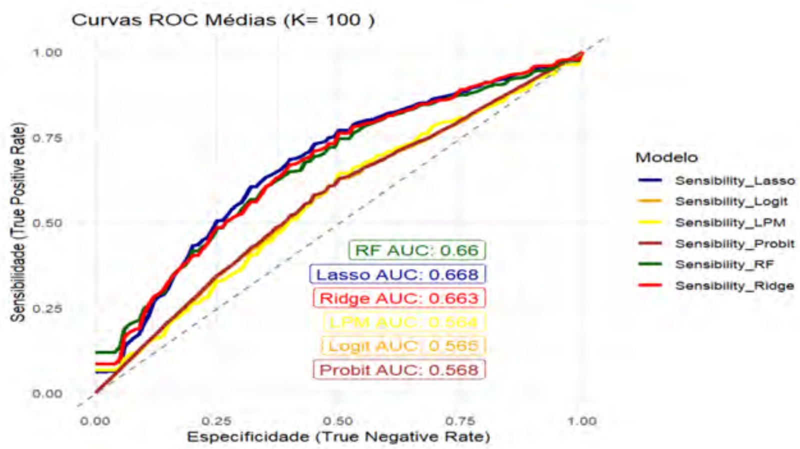
Figure 2.3: ROC Curve - PME Threshold - Sequence Funds

Figure 2.4: ROC Curves - VC Funds (PME Threshold)



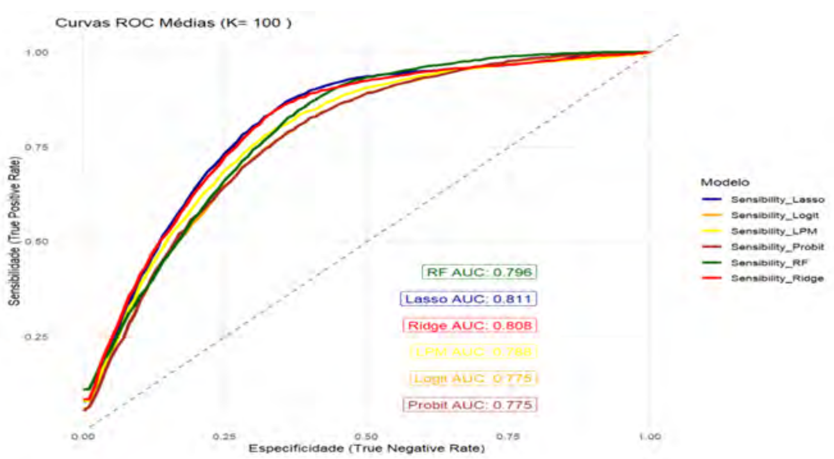
2.5(a):
b

Figure 2.5: ROC Curve - IRR Threshold - All Funds



2.6(a):
b

Figure 2.6: ROC Curve - IRR Threshold - First Funds



2.7(a):
b

Figure 2.7: ROC Curve - IRR Threshold - Sequence Funds

Figure 2.8: ROC Curves - VC Funds (IRR Threshold)

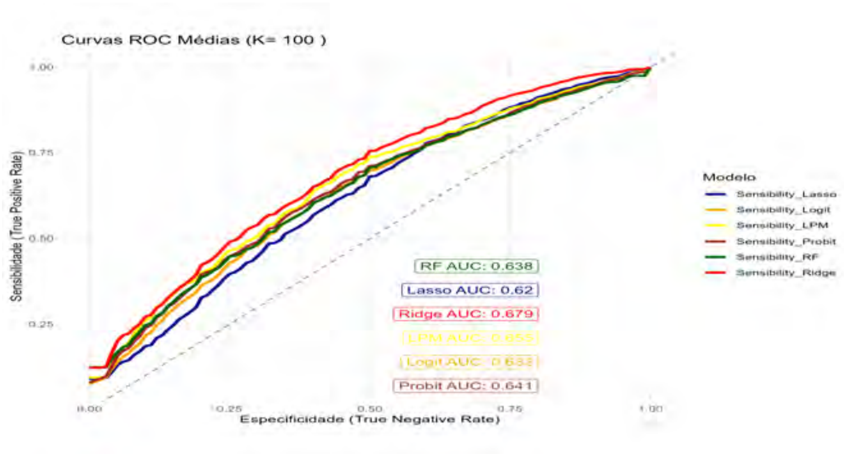
Table 2.6: Confusion Matrices - Buyout Funds

The table presents the average results of each confusion matrix from the 100 cross-validation procedures explained for Buyout funds. Details on accuracy, precision, recall, specificity, and F-score score can be found in Powers (2020).

Buyout Funds *Panel A: All Funds*	RF		Lasso		Ridge	
PME Threshold	5,5	11,6	2,7	14,4	2,6	14,5
	5,2	25,7	1,5	29,4	1,1	29,8
Accuracy	64,9%		66,9%		67,6%	
Precision	68,8%		67,2%		67,3%	
Recall	83,1%		95,2%		96,6%	
Specificity	32,0%		15,8%		15,1%	
F-score	75,3%		78,8%		79,3%	
IRR Threshold	86,8	35,8	95,5	27,1	96,5	26,1
	45,0	48,4	48,7	44,7	56,4	37,0
Accuracy	62,6%		64,9%		61,8%	
Precision	57,5%		62,3%		58,7%	
Recall	51,8%		47,9%		39,7%	
Specificity	70,8%		77,9%		78,7%	
F-score	54,5%		54,2%		47,3%	

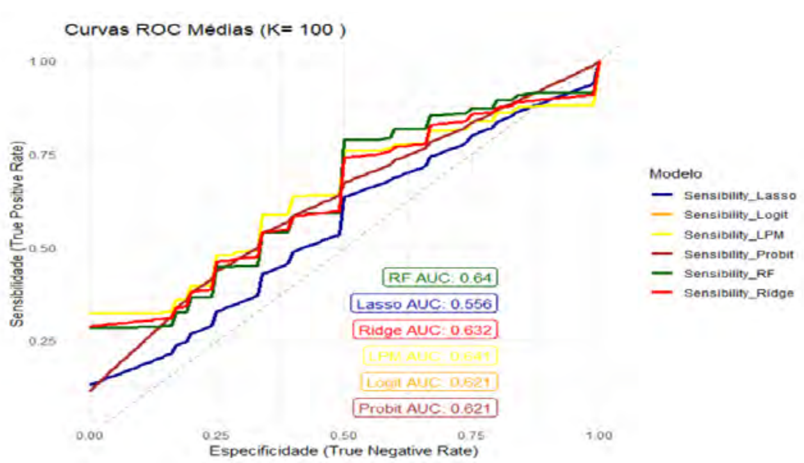
Panel B: First Funds	RF		Lasso		Ridge	
PME Threshold	0,8	1,9	0,5	2,2	0,4	2,4
	1,1	4,2	0,7	4,5	0,7	4,5
Accuracy	62,6%		63,2%		61,3%	
Precision	68,4%		66,8%		65,6%	
Recall	79,9%		86,6%		86,0%	
Specificity	29,6%		19,3%		14,3%	
F-score	73,7%		75,4%		74,4%	
IRR Threshold	11,1	6,4	13,9	3,2	14,4	2,9
	9,0	4,7	11,4	2,4	11,8	1,8
Accuracy	50,6%		52,6%		52,5%	
Precision	42,3%		42,9%		39,0%	
Recall	34,4%		17,6%		13,5%	
Specificity	63,3%		81,0%		83,3%	
F-score	37,9%		25,0%		20,1%	

Panel C: Sequence Funds	RF		Lasso		Ridge	
PME Threshold	2,6	7,0	1,4	8,3	1,3	8,3
	2,8	15,5	1,2	17,2	0,5	17,8
Accuracy	64,8%		66,1%		68,4%	
Precision	68,8%		67,4%		68,2%	
Recall	84,6%		93,5%		97,1%	
Specificity	27,3%		14,1%		13,9%	
F-score	75,9%		78,3%		80,1%	
IRR Threshold	57,6	25,5	61,7	21,4	63,6	19,5
	28,3	36,5	31,6	33,2	36,8	28,1
Accuracy	63,6%		64,2%		62,0%	
Precision	58,9%		60,8%		59,0%	
Recall	56,3%		51,2%		43,2%	
Specificity	69,3%		74,2%		76,5%	
F-score	57,6%		55,6%		50,0%	



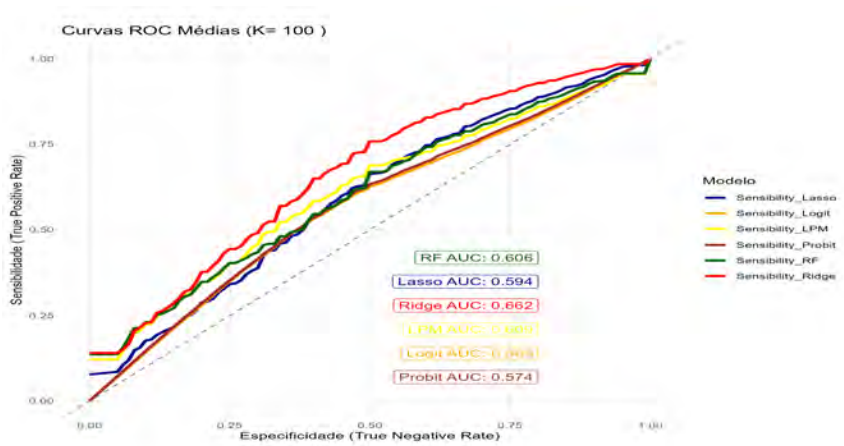
2.9(a):
b

Figure 2.9: ROC Curve - PME Threshold - All Funds



2.10(a):
b

Figure 2.10: ROC Curve - PME Threshold - First Funds



2.11(a):
b

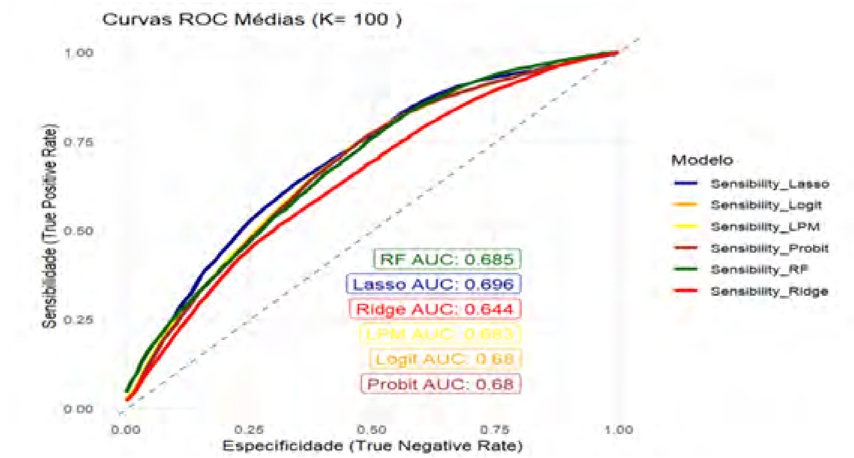
Figure 2.11: ROC Curve - PME Threshold - Sequence Funds

Figure 2.12: ROC Curves - Buyout Funds (PME Threshold)

Table 2.7: Summary of Feature Importance for Ridge, Lasso, and RF in VC Funds

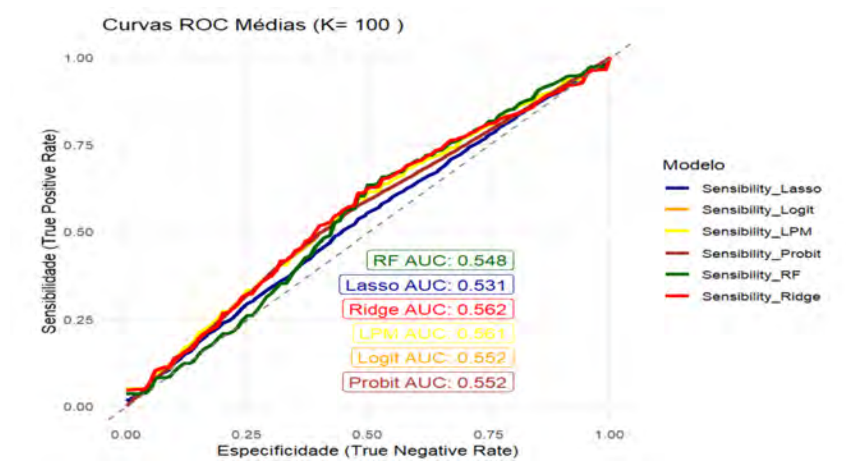
The table presents the summary statistics of feature significance measures, including beta coefficients for Ridge and Lasso, and variable importance scores for RF, applied to VC funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

Betas Categories/Groups	Ridge						Lasso						RF					
	N	Avg	Med	Stdev	Min	Max	N	Avg	Med	Stdev	Min	Max	N	Avg	Med	Stdev	Min	Max
PME Threshold Performance																		
All funds sample																		
Fund information	35	0,01	- 0,00	0,17	- 0,35	0,49	35	- 0,01	-	0,14	- 0,67	0,31	6	0,00	0,00	0,00	0,00	0,00
GPs	373	- 0,02	0,09	0,36	- 0,94	0,63	373	0,01	-	0,12	- 0,83	0,38	1	0,08	0,08	-	0,08	0,08
Macro	20	- 0,39	0,00	1,26	- 5,51	0,32	20	- 0,04	-	0,14	- 0,59	0,05	20	0,01	0,00	0,00	0,00	0,01
Market Conditions - Market Environment	16	0,02	- 0,00	0,13	- 0,33	0,28	16	0,01	-	0,02	- 0,00	0,08	16	0,00	0,00	0,00	0,00	0,02
PE Industry	4	0,01	0,00	0,01	0,00	0,03	4	0,04	0,00	0,09	-	0,17	4	0,01	0,00	0,01	0,00	0,02
PE Theory & Centrality Measures	4	0,10	0,04	0,21	- 0,07	0,40	4	0,13	0,00	0,26	- 0,01	0,52	4	0,00	0,00	0,00	0,00	0,00
Total	*452*	*- 0,03*	*0,00*	*0,43*	*- 5,51*	*0,63*	*452*	*0,00*	-	*0,12*	*- 0,83*	*0,52*	*51*	*0,01*	*0,00*	*0,01*	*0,00*	*0,08*
First time funds sample																		
Fund information	35	- 0,01	- 0,02	0,10	- 0,21	0,13	35	0,00	-	0,03	- 0,04	0,09	6	0,02	0,02	0,01	0,01	0,03
GPs	153	0,00	0,07	0,14	- 0,31	0,20	153	0,00	-	0,02	- 0,10	0,10	1	0,36	0,36	-	0,36	0,36
Macro	20	- 0,12	- 0,00	0,44	- 1,94	0,17	20	- 0,07	-	0,24	- 1,05	0,08	20	0,03	0,03	0,01	0,02	0,07
Market Conditions - Market Environment	16	0,02	0,00	0,10	- 0,15	0,24	16	0,02	-	0,07	- 0,01	0,29	16	0,02	0,01	0,01	0,01	0,04
PE Industry	4	0,00	0,00	0,00	- 0,00	0,01	4	0,00	0,00	0,00	- 0,00	0,01	4	0,03	0,02	0,01	0,02	0,04
PE Theory & Centrality Measures	4	- 0,09	- 0,02	0,19	- 0,37	0,04	4	- 0,00	-	0,00	- 0,01	0,00	4	0,01	0,01	0,00	0,01	0,02
Total	*232*	*- 0,01*	*0,04*	*0,18*	*- 1,94*	*0,24*	*232*	*- 0,00*	-	*0,08*	*- 1,05*	*0,29*	*51*	*0,03*	*0,02*	*0,05*	*0,01*	*0,36*
Sequence funds sample																		
Fund information	33	0,00	- 0,01	0,18	- 0,59	0,33	33	- 0,03	-	0,21	- 1,19	0,12	6	0,01	0,01	0,00	0,00	0,01
GPs	259	- 0,00	0,12	0,37	- 1,06	0,67	259	0,01	-	0,19	- 1,41	0,55	1	0,13	0,13	-	0,13	0,13
Macro	20	- 0,58	- 0,00	1,71	- 7,23	0,36	20	- 0,14	- 0,00	0,52	- 2,28	0,01	20	0,01	0,01	0,01	0,01	0,03
Market Conditions - Market Environment	16	0,03	- 0,00	0,19	- 0,51	0,32	16	0,01	-	0,04	- 0,05	0,13	16	0,01	0,01	0,01	0,00	0,05
PE Industry	4	0,01	0,00	0,02	0,00	0,03	4	0,05	0,00	0,11	-	0,22	4	0,02	0,01	0,02	0,00	0,04
PE Theory & Centrality Measures	7	0,05	0,00	0,23	- 0,12	0,55	7	0,16	-	0,43	- 0,01	1,14	7	0,01	0,01	0,00	0,01	0,01
Total	*339*	*- 0,03*	*0,08*	*0,54*	*- 7,23*	*0,67*	*339*	*0,00*	-	*0,23*	*- 2,28*	*1,14*	*54*	*0,01*	*0,01*	*0,02*	*0,00*	*0,13*
Net IRR Threshold Performance																		
All funds sample																		
Fund information	35	0,02	0,03	0,19	- 0,63	0,47	35	- 0,03	0,00	0,23	- 1,32	0,13	6	0,00	0,00	0,00	0,00	0,00
GPs	373	- 0,03	- 0,14	0,42	- 0,96	1,39	373	- 0,00	-	0,26	- 0,80	1,74	1	0,06	0,06	.	0,06	0,06
Macro	20	- 0,31	- 0,00	1,19	- 5,29	0,29	20	- 0,04	-	0,11	- 0,49	0,00	20	0,01	0,00	0,00	0,00	0,02
Market Conditions - Market Environment	16	- 0,00	- 0,00	0,17	- 0,59	0,19	16	- 0,02	-	0,10	- 0,40	0,01	16	0,00	0,00	0,00	0,00	0,02
PE Industry	4	0,01	0,00	0,02	- 0,00	0,03	4	0,05	0,00	0,10	- 0,00	0,19	4	0,01	0,01	0,01	0,00	0,02
PE Theory & Centrality Measures	4	0,06	0,06	0,08	- 0,02	0,14	4	0,01	0,00	0,02	- 0,00	0,03	4	0,00	0,00	0,00	0,00	0,00
Total	*452*	*- 0,04*	*- 0,09*	*0,46*	*- 5,29*	*1,39*	*452*	*- 0,01*	-	*0,25*	*- 1,32*	*1,74*	*51*	*0,01*	*0,00*	*0,01*	*0,00*	*0,06*
First time funds sample																		
Fund information	35	- 0,00	- 0,01	0,10	- 0,22	0,30	35	0,00	-	0,02	- 0,05	0,10	6	0,01	0,01	0,00	0,01	0,02
GPs	153	0,00	- 0,09	0,17	- 0,23	0,41	153	0,00	-	0,02	- 0,05	0,10	1	0,30	0,30	.	0,30	0,30
Macro	20	- 0,07	- 0,00	0,44	- 1,78	0,69	20	- 0,01	-	0,09	- 0,37	0,10	20	0,02	0,02	0,01	0,01	0,06
Market Conditions - Market Environment	16	0,01	- 0,00	0,11	- 0,25	0,22	16	0,00	-	0,01	- 0,03	0,03	16	0,01	0,01	0,01	0,00	0,04
PE Industry	4	0,00	0,00	0,01	- 0,00	0,01	4	0,01	-	0,02	-	0,04	4	0,02	0,02	0,02	0,01	0,06
PE Theory & Centrality Measures	4	0,17	0,10	0,23	- 0,01	0,50	4	0,03	0,02	0,03	- 0,00	0,07	4	0,01	0,01	0,00	0,00	0,01
Total	*232*	*- 0,00*	*- 0,06*	*0,20*	*- 1,78*	*0,69*	*232*	*0,00*	-	*0,03*	*- 0,37*	*0,10*	*51*	*0,02*	*0,01*	*0,04*	*0,00*	*0,30*
Sequence funds sample																		
Fund information	33	0,07	0,09	0,24	- 0,83	0,46	33	- 0,06	-	0,28	- 1,55	0,10	6	0,00	0,00	0,00	0,00	0,01
GPs	259	- 0,03	- 0,14	0,36	- 0,87	1,17	259	- 0,01	-	0,16	- 0,68	0,95	1	0,12	0,12	.	0,12	0,12
Macro	20	- 0,53	- 0,00	1,70	- 7,38	0,12	20	- 0,20	-	0,69	- 2,97	0,00	20	0,01	0,01	0,01	0,01	0,04
Market Conditions - Market Environment	16	- 0,03	- 0,00	0,27	- 1,01	0,29	16	- 0,01	-	0,12	- 0,40	0,23	16	0,01	0,01	0,01	0,00	0,04
PE Industry	4	0,01	0,00	0,02	- 0,00	0,03	4	0,04	0,00	0,09	- 0,00	0,18	4	0,02	0,02	0,02	0,01	0,04
PE Theory & Centrality Measures	7	0,05	0,00	0,09	- 0,05	0,15	7	0,01	0,00	0,02	- 0,00	0,06	7	0,01	0,01	0,00	0,00	0,01
Total	*339*	*- 0,04*	*- 0,07*	*0,54*	*- 7,38*	*1,17*	*339*	*- 0,03*	-	*0,24*	*- 2,97*	*0,95*	*54*	*0,01*	*0,01*	*0,02*	*0,00*	*0,12*



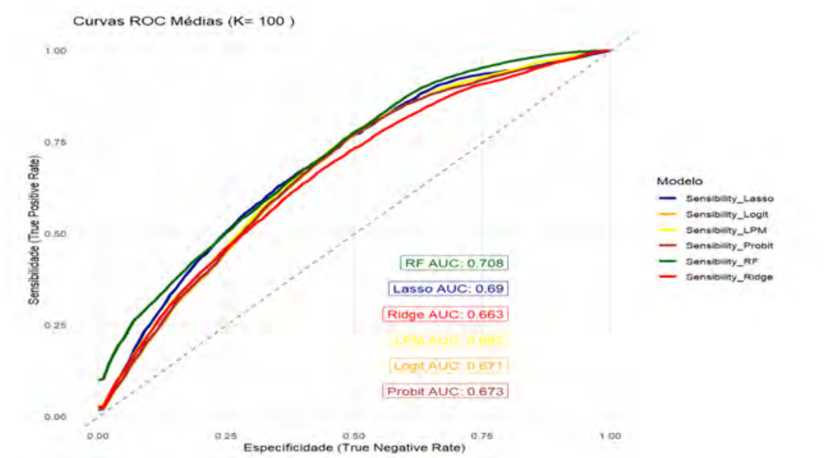
2.13(a):
b

Figure 2.13: ROC Curve - IRR Threshold - All Funds



2.14(a):
b

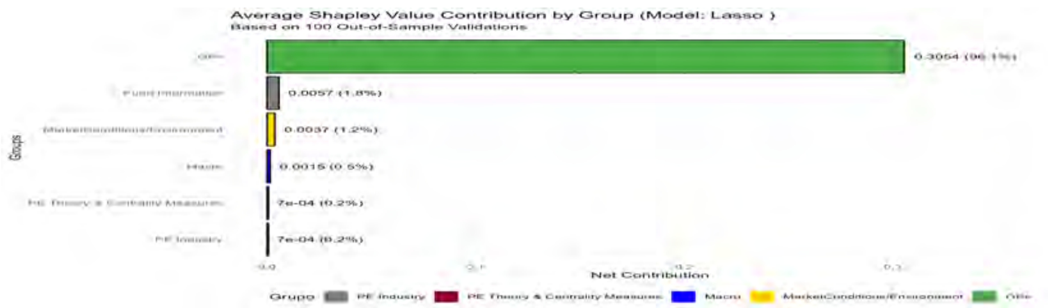
Figure 2.14: ROC Curve - IRR Threshold - First Funds



2.15(a):
b

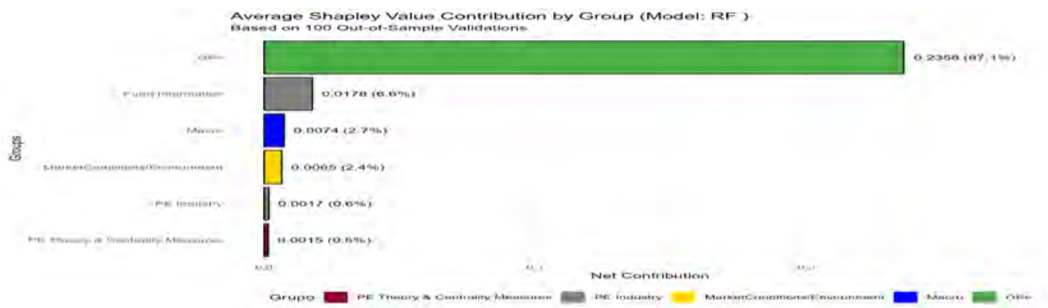
Figure 2.15: ROC Curve - IRR Threshold - Sequence Funds

Figure 2.16: ROC Curves - Buyout Funds (IRR Threshold)



2.17(a):
b

Figure 2.17: Shapley - VC - All Funds - IRR Threshold - Lasso

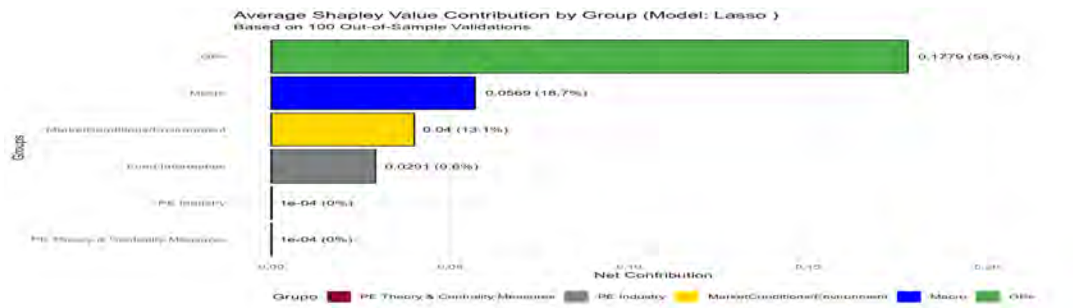


2.18(a):
b

Figure 2.18: Shapley - VC - All Funds - IRR Threshold - Random Forest

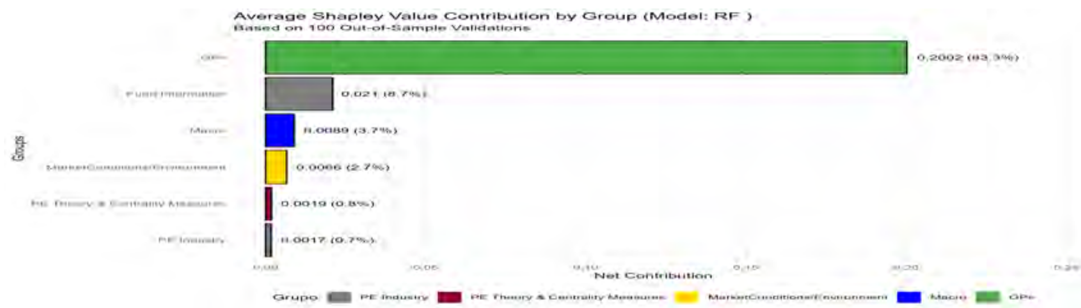
2.19(a):
b

Figure 2.19: Shapley - VC - All Funds - IRR Threshold - Ridge
Figure 2.20: Shapley Values - VC



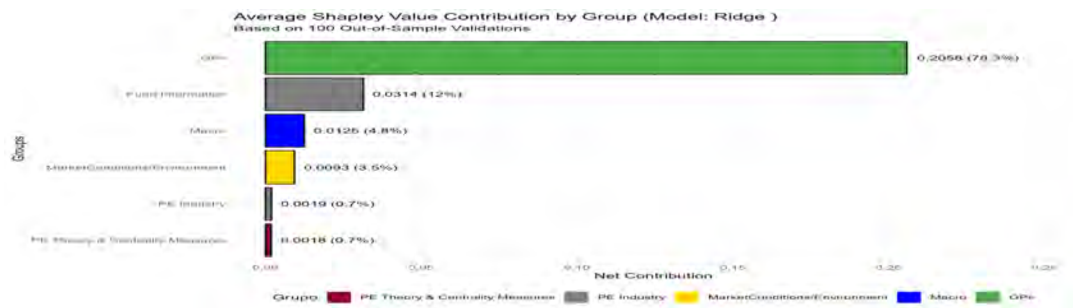
2.21(a):
b

Figure 2.21: Shapley - Buyout - All Funds - IRR Threshold - Lasso



2.22(a):
b

Figure 2.22: Shapley - Buyout - All Funds - IRR Threshold - Random Forest



2.23(a):
b

Figure 2.23: Shapley - Buyout - All Funds - IRR Threshold - Ridge
Figure 2.24: Shapley Values - Buyout

Table 2.8: Summary of Feature Importance for Ridge, Lasso, and RF in BO Funds

The table presents the summary statistics of feature significance measures, including beta coefficients for Ridge and Lasso, and variable importance scores for RF, applied to BO funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

Betas Categories/Groups	Ridge						Lasso						RF					
	N	Avg	p50	Stdev	Min	Max	N	Avg	p50	Stdev	Min	Max	N	Avg	p50	Stdev	Min	Max
PME Threshold Performance																		
All funds sample																		
Fund information	35	- 0,00	0,01	0,24	- 0,53	0,51	35	- 0,02	0,00	0,22	- 0,65	0,59	5	0,01	0,01	0,01	0,01	0,02
GPs	452	0,00	0,09	0,24	- 0,85	0,34	452	- 0,00	0,01	0,06	- 0,71	0,13	1	0,19	0,19	.	0,19	0,19
Macro	20	- 0,06	0,00	0,59	- 2,39	0,88	20	- 0,05	-	0,35	- 1,53	0,20	20	0,02	0,02	0,00	0,02	0,02
Market Conditions - Market Environment	16	0,11	0,00	0,40	- 0,07	1,59	16	0,22	-	0,86	- 0,03	3,42	16	0,02	0,01	0,00	0,01	0,03
PE Industry	4	0,00	- 0,00	0,01	- 0,00	0,01	4	0,00	- 0,00	0,01	- 0,00	0,01	4	0,02	0,02	0,00	0,01	0,02
PE Theory & Centrality Measures	4	0,14	0,16	0,10	0,00	0,23	4	0,11	0,07	0,13	0,00	0,28	4	0,02	0,02	0,00	0,01	0,02
Total	*531*	*0,00*	*0,08*	*0,27*	*- 2,39*	*1,59*	*531*	*0,00*	*0,01*	*0,18*	*- 1,53*	*3,42*	*50*	*0,02*	*0,02*	*0,03*	*0,01*	*0,19*
First time funds sample																		
Fund information	28	- 0,00	0,00	0,05	- 0,10	0,08	28	0,00	-	0,03	- 0,08	0,13	5	0,04	0,04	0,02	0,01	0,06
GPs	171	0,00	0,01	0,03	- 0,15	0,03	171	0,00	-	0,01	- 0,11	0,04	1	1,00	1,00	.	1,00	1,00
Macro	20	0,07	0,00	0,24	- 0,07	1,07	20	0,13	-	0,54	- 0,03	2,41	20	0,04	0,03	0,02	0,02	0,11
Market Conditions - Market Environment	16	0,04	0,00	0,20	- 0,06	0,80	16	0,17	0,00	0,69	- 0,00	2,76	16	0,05	0,04	0,04	0,01	0,15
PE Industry	4	0,00	- 0,00	0,00	- 0,00	0,00	4	- 0,00	- 0,00	0,00	- 0,00	0,00	4	0,06	0,06	0,02	0,04	0,09
PE Theory & Centrality Measures	4	- 0,37	0,02	0,82	- 1,59	0,08	4	- 0,23	0,01	0,54	- 1,04	0,08	4	0,03	0,04	0,02	-	0,04
Total	*243*	*0,00*	*0,01*	*0,14*	*- 1,59*	*1,07*	*243*	*0,02*	-	*0,25*	*- 1,04*	*2,76*	*50*	*0,06*	*0,04*	*0,14*	-	*1,00*
Sequence funds sample																		
Fund information	33	- 0,02	0,00	0,16	- 0,35	0,31	33	- 0,00	-	0,07	- 0,26	0,19	5	0,02	0,02	0,01	0,01	0,03
GPs	373	0,00	0,04	0,13	- 0,45	0,16	373	- 0,00	-	0,02	- 0,13	0,04	1	0,32	0,32	.	0,32	0,32
Macro	20	- 0,12	0,00	0,62	- 2,16	0,81	20	- 0,04	-	0,18	- 0,68	0,18	20	0,03	0,02	0,01	0,02	0,04
Market Conditions - Market Environment	16	0,06	0,00	0,19	- 0,04	0,76	16	0,04	0,00	0,16	- 0,01	0,64	16	0,02	0,02	0,01	0,01	0,05
PE Industry	4	0,00	0,00	0,00	- 0,00	0,01	4	0,00	0,00	0,00	- 0,00	0,01	4	0,02	0,02	0,01	0,02	0,03
PE Theory & Centrality Measures	7	0,05	0,00	0,10	- 0,07	0,18	7	0,02	0,00	0,05	- 0,04	0,12	7	0,03	0,03	0,00	0,02	0,03
Total	*453*	*- 0,00*	*0,03*	*0,18*	*- 2,16*	*0,81*	*453*	*- 0,00*	-	*0,05*	*- 0,68*	*0,64*	*53*	*0,03*	*0,02*	*0,04*	*0,01*	*0,32*
Net IRR Threshold Performance																		
All funds sample																		
Fund information	35	0,01	- 0,00	0,19	- 0,38	0,41	35	- 0,03	-	0,23	- 0,96	0,47	5	0,00	0,00	0,00	0,00	0,00
GPs	452	- 0,02	- 0,01	0,27	- 0,48	0,57	452	0,01	- 0,00	0,20	- 0,53	0,96	1	0,02	0,02	.	0,02	0,02
Macro	20	- 0,10	0,00	0,63	- 2,72	0,54	20	0,07	-	0,50	- 0,97	1,92	20	0,00	0,00	0,00	0,00	0,00
Market Conditions - Market Environment	16	0,06	- 0,00	0,29	- 0,08	1,16	16	0,19	-	0,81	- 0,13	3,22	16	0,00	0,00	0,00	0,00	0,00
PE Industry	4	0,00	0,00	0,00	- 0,00	0,01	4	0,00	-	0,00	- 0,00	0,00	4	0,00	0,00	0,00	0,00	0,00
PE Theory & Centrality Measures	4	- 0,05	- 0,06	0,05	- 0,10	- 0,00	4	- 0,04	- 0,04	0,04	- 0,07	- 0,00	4	0,00	0,00	0,00	0,00	0,00
Total	*531*	*- 0,02*	*- 0,00*	*0,29*	*- 2,72*	*1,16*	*531*	*0,01*	*- 0,00*	*0,26*	*- 0,97*	*3,22*	*50*	*0,00*	*0,00*	*0,00*	*0,00*	*0,02*
First time funds sample																		
Fund information	28	- 0,00	- 0,00	0,07	- 0,13	0,12	28	0,00	-	0,05	- 0,13	0,20	5	0,02	0,02	0,01	0,01	0,03
GPs	171	- 0,00	- 0,05	0,09	- 0,13	0,17	171	0,00	-	0,01	- 0,02	0,06	1	0,69	0,69	.	0,69	0,69
Macro	20	0,02	0,00	0,14	- 0,22	0,57	20	0,02	-	0,13	- 0,15	0,56	20	0,01	0,01	0,01	0,00	0,03
Market Conditions - Market Environment	16	0,01	0,00	0,06	- 0,05	0,23	16	0,01	-	0,04	- 0,00	0,14	16	0,01	0,01	0,01	0,00	0,02
PE Industry	4	- 0,00	- 0,00	0,00	- 0,00	- 0,00	4	- 0,00	- 0,00	0,00	- 0,00	-	4	0,04	0,04	0,00	0,04	0,04
PE Theory & Centrality Measures	4	- 0,03	- 0,04	0,03	- 0,06	- 0,00	4	- 0,01	- 0,01	0,01	- 0,03	-	4	0,03	0,03	0,01	0,01	0,04
Total	*243*	*0,00*	*- 0,03*	*0,09*	*- 0,22*	*0,57*	*243*	*0,00*	-	*0,04*	*- 0,15*	*0,56*	*50*	*0,03*	*0,01*	*0,10*	*0,00*	*0,69*
Sequence funds sample																		
Fund information	33	0,00	-	0,20	- 0,45	0,39	33	- 0,03	-	0,25	- 1,16	0,34	5	0,00	0,00	0,00	0,00	0,00
GPs	373	- 0,01	- 0,06	0,26	- 0,56	0,61	373	- 0,00	- 0,00	0,15	- 0,52	0,80	1	0,04	0,04	.	0,04	0,04
Macro	20	- 0,15	0,00	0,74	- 3,27	0,29	20	0,01	-	0,23	- 0,66	0,73	20	0,00	0,00	0,00	0,00	0,01
Market Conditions - Market Environment	16	0,09	- 0,00	0,41	- 0,12	1,61	16	0,23	-	0,98	- 0,19	3,89	15	0,00	0,00	0,00	0,00	0,00
PE Industry	4	0,00	0,00	0,01	0,00	0,01	4	0,01	0,00	0,01	-	0,03	4	0,00	0,00	0,00	0,00	0,00
PE Theory & Centrality Measures	7	- 0,01	- 0,01	0,02	- 0,05	0,01	7	- 0,01	- 0,00	0,01	- 0,02	-	7	0,00	0,00	0,00	0,00	0,00
Total	*453*	*- 0,01*	*- 0,00*	*0,30*	*- 3,27*	*1,61*	*453*	*0,00*	-	*0,24*	*- 1,16*	*3,89*	*52*	*0,00*	*0,00*	*0,00*	*0,00*	*0,04*

Table 2.9: Summary of Absolute Betas and Variable Importance for Ridge, Lasso, and RF in VC Funds

The table provides key statistics of the absolute betas for Ridge and Lasso, along with variable importance scores for Random Forest, applied to VC funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

	Ridge						Lasso						Randon Forest (RF)					
Betas Categories/Groups	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.
PME Threshold Performance																		
All funds Sample																		
Fund information	35	3,87	0,111	0,060	0,49	0,00004	35	1,89	0,054	0,007	0,67	-	6	0,01	0,002	0,002	0,004	0,001
GPs	373	117,47	0,315	0,279	0,94	-	373	21,57	0,058	0,004	0,83	-	1	0,08	0,076	0,076	0,076	0,076
Macro	20	8,56	0,428	0,010	5,51	0,00012	20	0,92	0,046	0,000	0,59	-	20	0,12	0,006	0,004	0,015	0,003
Market Conditions - Market Environment	16	1,08	0,068	0,002	0,33	0,00000	16	0,10	0,006	0,000	0,08	-	16	0,05	0,003	0,002	0,018	0,001
PE Industry	4	0,03	0,007	0,001	0,03	0,00017	4	0,17	0,043	0,000	0,17	-	4	0,03	0,008	0,004	0,020	0,002
PE Theory & Centrality Measures	4	0,54	0,135	0,069	0,40	0,00032	4	0,53	0,132	0,006	0,52	-	4	0,01	0,003	0,003	0,004	0,002
Total	452	131,55	0,291	0,260	5,51	-	452	25,18	0,056	0,003	0,83	-	51	0,30	0,006	0,003	0,076	0,001
First Fund Sample																		
Fund information	35	2,72	0,078	0,063	0,21	0,00003	35	0,42	0,012	0,001	0,09	-	6	0,12	0,020	0,019	0,033	0,007
GPs	153	18,33	0,120	0,092	0,31	0,00705	153	1,03	0,007	-	0,10	-	1	0,36	0,359	0,359	0,359	0,359
Macro	20	2,97	0,149	0,015	1,94	0,00006	20	1,55	0,077	0,001	1,05	-	20	0,68	0,034	0,033	0,066	0,018
Market Conditions - Market Environment	16	0,76	0,048	0,001	0,24	0,00000	16	0,39	0,025	0,000	0,29	-	16	0,27	0,017	0,013	0,039	0,009
PE Industry	4	0,01	0,002	0,001	0,01	0,00002	4	0,01	0,002	0,000	0,01	-	4	0,10	0,025	0,021	0,042	0,017
PE Theory & Centrality Measures	4	0,45	0,113	0,037	0,37	0,00817	4	0,01	0,002	0,001	0,01	-	4	0,04	0,010	0,009	0,015	0,007
Total	232	25,24	0,109	0,084	1,94	0,00000	232	3,40	0,015	0,000	1,05	-	51	1,57	0,031	0,021	0,359	0,007
Sequence funds Sample																		
Fund information	33	4,20	0,127	0,090	0,59	-	33	1,9	0,057	0,006	1,19	-	6	0,03	0,006	0,005	0,012	0,001
GPs	259	79,66	0,308	0,266	1,06	-	259	20,5	0,079	0,005	1,41	-	1	0,13	0,126	0,126	0,126	0,126
Macro	20	12,47	0,624	0,008	7,23	0,00017	20	2,8	0,142	0,000	2,28	-	20	0,26	0,013	0,009	0,030	0,005
Market Conditions - Market Environment	16	1,52	0,095	0,004	0,51	0,00000	16	0,2	0,015	0,000	0,13	-	16	0,14	0,009	0,006	0,047	0,003
PE Industry	4	0,04	0,009	0,001	0,03	0,00022	4	0,22	0,054	0,000	0,22	-	4	0,08	0,019	0,014	0,044	0,005
PE Theory & Centrality Measures	7	0,94	0,135	0,084	0,55	0,00012	7	1,17	0,167	0,003	1,14	-	7	0,06	0,008	0,008	0,011	0,005
Total	339	98,83	0,292	0,238	7,23	-	339	26,84	0,079	0,004	2,28	-	54	0,69	0,013	0,008	0,126	0,001
Net IRR Threshold Performance																		
All funds Sample																		
Fund information	35	4,49	0,128	0,096	0,63	0,00002	35	2,12	0,061	0,007	1,32	-	6	0,01	0,002	0,002	0,003	0,001
GPs	373	133,46	0,358	0,328	1,39	-	373	43,94	0,118	0,001	1,74	-	1	0,06	0,055	0,055	0,055	0,055
Macro	20	7,12	0,356	0,015	5,29	0,00022	20	0,79	0,040	0,000	0,49	-	20	0,11	0,006	0,004	0,017	0,003
Market Conditions - Market Environment	16	1,19	0,074	0,002	0,59	0,00000	16	0,41	0,025	0,000	0,40	-	16	0,05	0,003	0,002	0,016	0,001
PE Industry	4	0,03	0,009	0,001	0,03	0,00023	4	0,19	0,048	0,000	0,19	-	4	0,03	0,008	0,008	0,017	0,002
PE Theory & Centrality Measures	4	0,29	0,073	0,074	0,14	0,00456	4	0,04	0,011	0,007	0,03	0,00	4	0,01	0,002	0,002	0,003	0,002
Total	452	146,6	0,324	0,284	5,29	-	452	47,5	0,105	0,001	1,74	-	51	0,27	0,005	0,003	0,055	0,001
First Fund Sample																		
Fund information	35	2,65	0,076	0,057	0,30	0,00009	35	0,37	0,011	0,001	0,10	-	6	0,09	0,014	0,014	0,018	0,009
GPs	153	24,08	0,157	0,135	0,41	0,06308	153	1,50	0,010	-	0,10	-	1	0,30	0,302	0,302	0,302	0,302
Macro	20	3,36	0,168	0,011	1,78	0,00012	20	0,64	0,032	0,000	0,37	-	20	0,49	0,024	0,021	0,055	0,011
Market Conditions - Market Environment	16	0,91	0,057	0,002	0,25	0,00000	16	0,08	0,005	0,000	0,03	-	16	0,19	0,012	0,008	0,042	0,005
PE Industry	4	0,01	0,003	0,001	0,01	0,00006	4	0,04	0,010	-	0,04	-	4	0,10	0,024	0,016	0,056	0,009
PE Theory & Centrality Measures	4	0,71	0,177	0,100	0,50	0,00648	4	0,11	0,028	0,021	0,07	-	4	0,02	0,006	0,005	0,008	0,005
Total	232	31,7	0,137	0,117	1,78	0,00000	232	2,7	0,012	-	0,37	-	51	1,2	0,023	0,014	0,302	0,005
Sequence funds Sample																		
Fund information	33	5,84	0,177	0,119	0,83	-	33	2,58	0,078	0,008	1,55	-	6	0,03	0,005	0,005	0,009	0,002
GPs	259	79,32	0,306	0,290	1,17	-	259	17,58	0,068	0,000	0,95	-	1	0,12	0,120	0,120	0,120	0,120
Macro	20	11,17	0,559	0,015	7,38	0,00027	20	4,03	0,201	0,000	2,97	-	20	0,27	0,013	0,010	0,040	0,006
Market Conditions - Market Environment	16	1,68	0,105	0,002	1,01	0,00000	16	0,63	0,040	-	0,40	-	16	0,13	0,008	0,006	0,042	0,004
PE Industry	4	0,04	0,009	0,002	0,03	0,00023	4	0,18	0,044	0,000	0,18	-	4	0,08	0,021	0,019	0,041	0,007
PE Theory & Centrality Measures	7	0,51	0,074	0,053	0,15	0,00011	7	0,10	0,015	0,005	0,06	0,00	7	0,04	0,006	0,006	0,007	0,003
Total	339	98,6	0,291	0,226	7,38	-	339	25,1	0,074	0,000	2,97	-	54	0,7	0,012	0,007	0,120	0,002

Table 2.10: Summary of Absolute Betas and Variable Importance for Ridge, Lasso, and RF in BO Funds

The table provides key statistics of the absolute betas for Ridge and Lasso, along with variable importance scores for Random Forest, applied to BO funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

Betas Categories/Groups	Ridge						Lasso						Randon Forest (RF)					
	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.
PME Threshold Performance																		
All funds Sample																		
Fund information	35	5,87	0,17	0,10	0,53	-	35	3,65	0,10	0,01	0,65	-	5	0,07	0,01	0,01	0,02	0,007
GPs	452	81,08	0,18	0,13	0,85	-	452	14,22	0,03	0,02	0,71	-	1	0,19	0,19	0,19	0,19	0,193
Macro	20	4,39	0,22	0,03	2,39	0,00000	20	2,13	0,11	0,00	1,53	-	20	0,39	0,02	0,02	0,02	0,016
Market Conditions - Market Environment	16	1,94	0,12	0,00	1,59	0,00000	16	3,52	0,22	0,00	3,42	-	16	0,25	0,02	0,01	0,03	0,010
PE Industry	4	0,01	0,00	0,00	0,01	0,00002	4	0,01	0,00	0,00	0,01	0,00	4	0,07	0,02	0,02	0,02	0,014
PE Theory & Centrality Measures	4	0,56	0,14	0,16	0,23	0,00278	4	0,42	0,11	0,07	0,28	0,00	4	0,07	0,02	0,02	0,02	0,013
Total	531	93,86	0,18	0,12	2,39	-	531	23,95	0,05	0,01	3,42	-	50	1,04	0,02	0,02	0,19	0,007
First Fund Sample																		
Fund information	28	1,06	0,04	0,04	0,10	-	28	0,42	0,01	0,00	0,13	-	5	0,21	0,04	0,04	0,06	0,013
GPs	171	3,26	0,02	0,01	0,15	-	171	0,82	0,00	-	0,11	-	1	1,00	1,00	1,00	1,00	1,000
Macro	20	1,51	0,08	0,00	1,07	0,00001	20	2,65	0,13	0,00	2,41	-	20	0,71	0,04	0,03	0,11	0,016
Market Conditions - Market Environment	16	0,96	0,06	0,00	0,80	0,00000	16	2,78	0,17	0,00	2,76	-	16	0,81	0,05	0,04	0,15	0,010
PE Industry	4	0,00	0,00	0,00	0,00	0,00011	4	0,00	0,00	0,00	0,00	-	4	0,25	0,06	0,06	0,09	0,037
PE Theory & Centrality Measures	4	1,71	0,43	0,06	1,59	0,00293	4	1,15	0,29	0,05	1,04	0,00	4	0,11	0,03	0,04	0,04	-
Total	243	8,49	0,03	0,01	1,59	-	243	7,82	0,03	-	2,76	-	50	3,08	0,06	0,04	1,00	-
Sequence funds Sample																		
Fund information	33	3,87	0,12	0,08	0,35	-	33	1,13	0,03	0,00	0,26	-	5	0,10	0,02	0,02	0,03	0,013
GPs	373	30,27	0,08	0,05	0,45	-	373	2,72	0,01	0,00	0,13	-	1	0,32	0,32	0,32	0,32	0,318
Macro	20	4,88	0,24	0,02	2,16	0,00001	20	1,34	0,07	0,00	0,68	-	20	0,50	0,03	0,02	0,04	0,017
Market Conditions - Market Environment	16	1,10	0,07	0,00	0,76	0,00000	16	0,72	0,05	0,00	0,64	-	16	0,30	0,02	0,02	0,05	0,011
PE Industry	4	0,01	0,00	0,00	0,01	0,00009	4	0,01	0,00	0,00	0,01	0,00	4	0,09	0,02	0,02	0,03	0,015
PE Theory & Centrality Measures	7	0,58	0,08	0,07	0,18	0,00003	7	0,25	0,04	0,02	0,12	0,00	7	0,18	0,03	0,03	0,03	0,020
Total	453	40,72	0,09	0,05	2,16	-	453	6,17	0,01	0,00	0,68	-	53	1,48	0,03	0,02	0,32	0,011
Net IRR Threshold Performance																		
All funds Sample																		
Fund information	35	4,88	0,14	0,12	0,41	-	35	3,36	0,10	0,01	0,96	-	5	0,01	0,00	0,00	0,00	0,001
GPs	452	107,90	0,24	0,23	0,57	-	452	52,36	0,12	0,05	0,96	-	1	0,02	0,02	0,02	0,02	0,021
Macro	20	4,04	0,20	0,03	2,72	0,00003	20	3,33	0,17	0,00	1,92	-	20	0,05	0,00	0,00	0,00	0,002
Market Conditions - Market Environment	16	1,39	0,09	0,00	1,16	0,00000	16	3,41	0,21	0,00	3,22	-	16	0,03	0,00	0,00	0,00	0,001
PE Industry	4	0,01	0,00	0,00	0,01	0,00002	4	0,01	0,00	0,00	0,00	-	4	0,01	0,00	0,00	0,00	0,002
PE Theory & Centrality Measures	4	0,21	0,05	0,06	0,10	0,00109	4	0,15	0,04	0,04	0,07	0,00	4	0,01	0,00	0,00	0,00	0,002
Total	531	118,43	0,22	0,21	2,72	-	531	62,62	0,12	0,04	3,22	-	50	0,12	0,00	0,00	0,02	0,001
First Fund Sample																		
Fund information	28	1,37	0,05	0,03	0,13	-	28	0,54	0,02	0,00	0,20	-	5	0,08	0,02	0,02	0,03	0,006
GPs	171	14,24	0,08	0,08	0,17	-	171	0,93	0,01	0,00	0,06	-	1	0,69	0,69	0,69	0,69	0,693
Macro	20	1,01	0,05	0,00	0,57	0,00000	20	0,73	0,04	-	0,56	-	20	0,27	0,01	0,01	0,03	0,005
Market Conditions - Market Environment	16	0,35	0,02	0,00	0,23	0,00000	16	0,14	0,01	-	0,14	-	16	0,17	0,01	0,01	0,02	0,001
PE Industry	4	0,00	0,00	0,00	0,00	0,00012	4	0,00	0,00	0,00	0,00	-	4	0,17	0,04	0,04	0,04	0,041
PE Theory & Centrality Measures	4	0,13	0,03	0,04	0,06	0,00049	4	0,05	0,01	0,01	0,03	-	4	0,10	0,03	0,03	0,04	0,011
Total	243	17,09	0,07	0,07	0,57	-	243	2,40	0,01	0,00	0,56	-	50	1,49	0,03	0,01	0,69	0,001
Sequence funds Sample																		
Fund information	33	4,77	0,14	0,11	0,45	-	33	3,46	0,10	0,01	1,16	-	5	0,01	0,00	0,00	0,00	0,001
GPs	373	85,73	0,23	0,23	0,61	-	373	28,30	0,08	0,02	0,80	-	1	0,04	0,04	0,04	0,04	0,036
Macro	20	4,32	0,22	0,03	3,27	0,00004	20	1,88	0,09	0,00	0,73	-	20	0,09	0,00	0,00	0,01	0,003
Market Conditions - Market Environment	16	1,93	0,12	0,00	1,61	0,00000	16	4,13	0,26	0,00	3,89	-	15	0,04	0,00	0,00	0,00	0,002
PE Industry	4	0,01	0,00	0,00	0,01	0,00000	4	0,03	0,01	0,00	0,03	-	4	0,01	0,00	0,00	0,00	0,003
PE Theory & Centrality Measures	7	0,12	0,02	0,01	0,05	0,00000	7	0,04	0,01	0,00	0,02	-	7	0,03	0,00	0,00	0,00	0,003
Total	453	96,88	0,21	0,20	3,27	-	453	37,84	0,08	0,02	3,89	-	52	0,22	0,00	0,00	0,04	0,001

Table 2.11: Category Importance for Ridge, Lasso, and RF in VC Funds

The table presents the proportion of absolute betas for Ridge and Lasso, along with the proportion of variable importance scores for Random Forest, applied to VC funds across each category/group. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds.

PME Threshold				**IRR Treshold**			
All Funds	**Ridge**	**Lasso**	**RF**	**All Funds**	**Ridge**	**Lasso**	**RF**
GPs	89,3%	85,6%	25,1%	GPs	91,0%	92,5%	20,3%
Fund information	2,9%	7,5%	4,5%	Fund information	3,1%	4,5%	4,9%
PE Theory & Centrality Measures	0,4%	2,1%	4,4%	PE Theory & Centrality Measures	0,2%	0,1%	3,4%
PE Industry	0,0%	0,7%	10,2%	PE Industry	0,0%	0,4%	12,5%
Macro	6,5%	3,6%	38,2%	Macro	4,9%	1,7%	40,8%
Market Conditions - Market Environment	0,8%	0,4%	17,7%	Market Conditions - Market Environment	0,8%	0,9%	18,1%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
1st Funds	**Ridge**	**Lasso**	**RF**	**1st Funds**	**Ridge**	**Lasso**	**RF**
GPs	72,6%	30,1%	22,8%	GPs	75,9%	54,6%	25,4%
Fund information	10,8%	12,3%	7,6%	Fund information	8,4%	13,5%	7,2%
PE Theory & Centrality Measures	1,8%	0,2%	2,6%	PE Theory & Centrality Measures	2,2%	4,1%	2,0%
PE Industry	0,0%	0,2%	6,4%	PE Industry	0,0%	1,5%	8,2%
Macro	11,8%	45,5%	43,1%	Macro	10,6%	23,5%	41,0%
Market Conditions - Market Environment	3,0%	11,6%	17,5%	Market Conditions - Market Environment	2,9%	3,0%	16,2%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
Sequence	**Ridge**	**Lasso**	**RF**	**Sequence**	**Ridge**	**Lasso**	**RF**
GPs	80,6%	76,3%	18,2%	GPs	80,5%	70,0%	17,9%
Fund information	4,3%	7,0%	4,9%	Fund information	5,9%	10,3%	4,4%
PE Theory & Centrality Measures	1,0%	4,3%	8,1%	PE Theory & Centrality Measures	0,5%	0,4%	5,9%
PE Industry	0,0%	0,8%	10,9%	PE Industry	0,0%	0,7%	12,6%
Macro	12,6%	10,6%	37,1%	Macro	11,3%	16,0%	39,9%
Market Conditions - Market Environment	1,5%	0,9%	20,8%	Market Conditions - Market Environment	1,7%	2,5%	19,3%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*

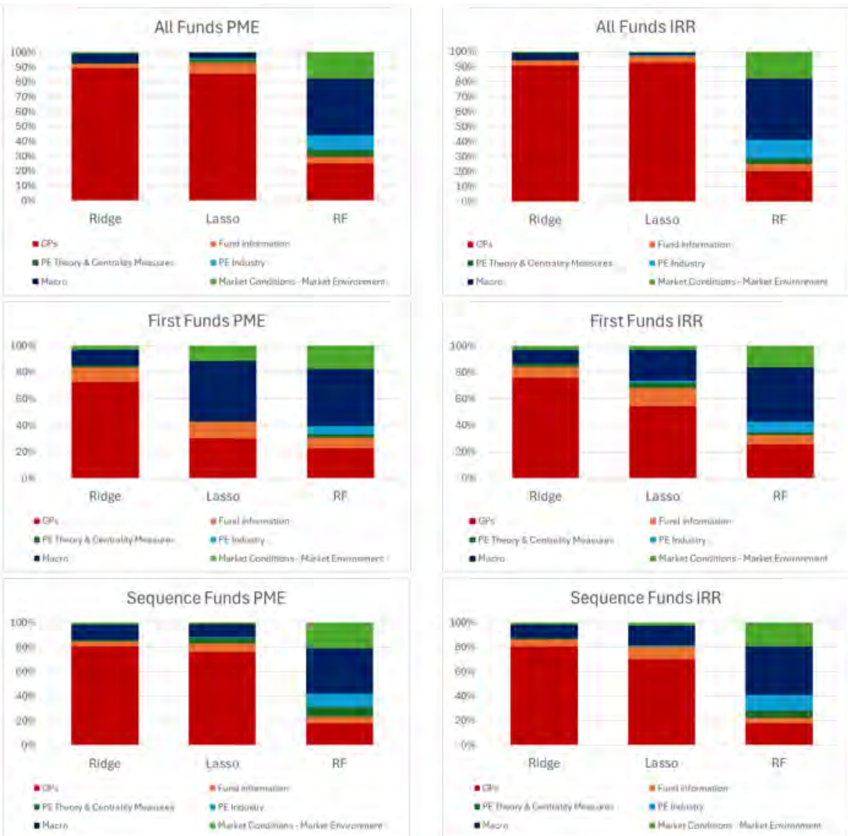


Figure 2.25: Proportion of Category Importance for Ridge, Lasso, and RF in VC Funds

Table 2.12: Category Importance for Ridge, Lasso, and RF in BO Funds

The table presents the proportion of absolute betas for Ridge and Lasso, along with the proportion of variable importance scores for Random Forest, applied to BO funds across each category/group. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds.

PME Threshold				**IRR Treshold**			
All Funds	**Ridge**	**Lasso**	**RF**	**All funds**	**Ridge**	**Lasso**	**RF**
GPs	86,4%	59,4%	18,5%	GPs	91,1%	83,6%	17,2%
Fund information	6,3%	15,2%	6,3%	Fund information	4,1%	5,4%	6,8%
PE Theory & Centrality Measures	0,6%	1,8%	7,0%	PE Theory & Centrality Measures	0,2%	0,2%	7,4%
PE Industry	0,0%	0,1%	6,7%	PE Industry	0,0%	0,0%	7,1%
Macro	4,7%	8,9%	37,2%	Macro	3,4%	5,3%	40,0%
Market Conditions - Market Environment	2,1%	14,7%	24,3%	Market Conditions - Market Environment	1,2%	5,4%	21,6%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
1st Funds	**Ridge**	**Lasso**	**RF**	**1st Funds**	**Ridge**	**Lasso**	**RF**
GPs	38,3%	10,5%	32,4%	GPs	83,3%	38,9%	46,6%
Fund information	12,5%	5,4%	6,8%	Fund information	8,0%	22,4%	5,5%
PE Theory & Centrality Measures	20,1%	14,7%	3,7%	PE Theory & Centrality Measures	0,8%	2,1%	6,8%
PE Industry	0,0%	0,0%	8,1%	PE Industry	0,0%	0,0%	11,4%
Macro	17,7%	33,9%	23,0%	Macro	5,9%	30,6%	18,5%
Market Conditions - Market Environment	11,3%	35,5%	26,1%	Market Conditions - Market Environment	2,0%	6,0%	11,2%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
Sequence	**Ridge**	**Lasso**	**RF**	**Sequence**	**Ridge**	**Lasso**	**RF**
GPs	74,3%	44,2%	21,4%	GPs	88,5%	74,8%	16,1%
Fund information	9,5%	18,3%	6,5%	Fund information	4,9%	9,1%	6,0%
PE Theory & Centrality Measures	1,4%	4,0%	12,1%	PE Theory & Centrality Measures	0,1%	0,1%	11,8%
PE Industry	0,0%	0,1%	6,1%	PE Industry	0,0%	0,1%	6,5%
Macro	12,0%	21,7%	33,7%	Macro	4,5%	5,0%	39,9%
Market Conditions - Market Environment	2,7%	11,7%	20,3%	Market Conditions - Market Environment	2,0%	10,9%	19,8%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*

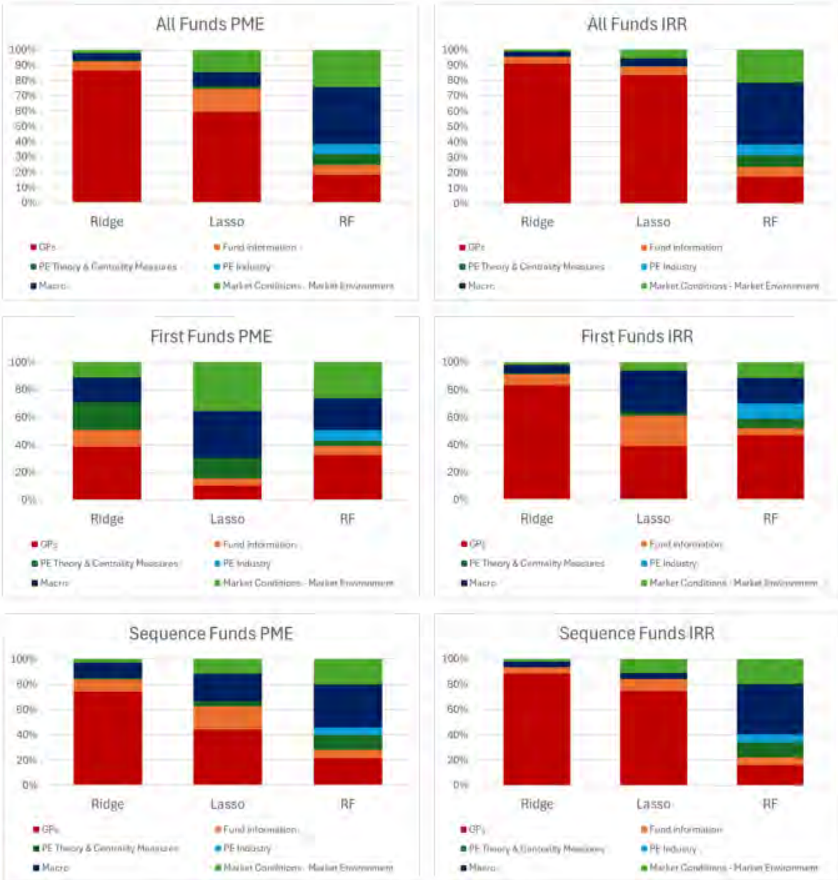


Figure 2.26: Proportion of Category Importance for Ridge, Lasso, and RF in BO Funds

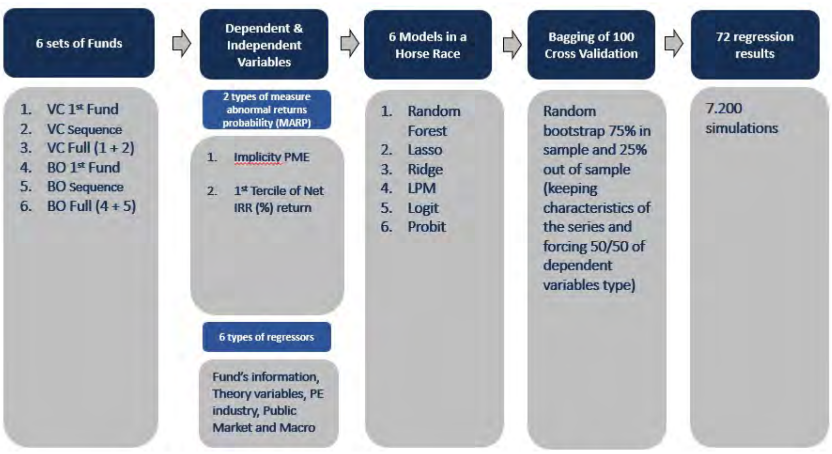


Figure 2.27: Methodology scheme.

Table 2.13: Top 10 variables - Venture Capital funds

The table presents the top 10 variables with the highest absolute betas, considering all ML models (Lasso, Ridge, and RF), performance thresholds (PME and Net IRR), and samples of VC funds (full sample, first-time, and sequential funds). For each variable, we identified its group classification as outlined in Table 2.3, the beta direction (positive or negative), and the absolute beta value. Note that for RF models, beta direction is not applicable, so we have marked it as "NA" in the respective column. The variables are ranked from the largest (#1) to the 10th largest absolute beta value. In the variables column, you can match the names with Table 2.3 for more details. Variables associated with a number represent the ID of a specific GP in our dataset. Due to confidentiality, the names of the GPs are not disclosed.

PME Threshold Performance

Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Full sample				*Full sample*				*Full sample*			
Macro	CPIAUCSL	-	5,511	GPs	102485	-	0,827	GPs	Firm(GP) ID	NA	0,076
Macro	RPI	-	1,336	Fund information	Energy & Utilities	-	0,667	PE Industry	Industry years	NA	0,020
GPs	2290	-	0,939	Macro	CPIAUCSL	-	0,590	Market Cond. & Environ.	S&P Dividend Yield	NA	0,018
Macro	US GDP (last year)	-	0,921	GPs	53913	-	0,583	Macro	FEDFUNDS	NA	0,015
GPs	102485	-	0,817	GPs	72331	-	0,544	Macro	HOUST	NA	0,014
GPs	53913	-	0,816	PE Theory & Cent. Meas.	Eigenv ex-ante GP	+	0,516	Macro	PERMIT	NA	0,012
GPs	49310	-	0,809	GPs	178323	-	0,469	Macro	OILPRICEx	NA	0,010
GPs	72331	-	0,800	GPs	2290	-	0,423	Macro	INDPRO	NA	0,007
GPs	4561	-	0,800	GPs	6712	+	0,383	Macro	UNRATE	NA	0,007
GPs	62315	-	0,788	GPs	247	+	0,372	PE Industry	Fund Raising Number	NA	0,006
First time fund sample				*First time fund sample*				*First time fund sample*			
Macro	CPIAUCSL	-	1,940	Macro	CPIAUCSL	-	1,049	GPs	Firm(GP) ID	NA	0,359
PE Theory & Cent. Meas.	Eigenv ex-ante GP	-	0,371	Market Cond. & Environ.	Mkt_RF	+	0,286	Macro	CLAIMSx	NA	0,066
Macro	RPI	-	0,361	Macro	CLAIMSx	-	0,258	Macro	FEDFUNDS	NA	0,056
GPs	62315	-	0,309	GPs	3798	+	0,100	Macro	CPIAUCSL	NA	0,046
GPs	18446	-	0,304	GPs	62315	-	0,099	PE Industry	Industry years	NA	0,042
GPs	102485	-	0,287	Macro	HWIURATIO	-	0,099	Market Cond. & Environ.	S&P Dividend Yield	NA	0,039
GPs	838	-	0,271	Market Cond. & Environ.	S&P 500 returns	+	0,098	Macro	Uncertain	NA	0,039
GPs	768	-	0,270	Fund information	Asia	+	0,094	Market Cond. & Environ.	Mkt_RF	NA	0,036
GPs	2616	-	0,265	Macro	UNRATE	+	0,080	Macro	CUMFNS	NA	0,036
GPs	7809	-	0,250	Fund information	LOCATION_Asia	+	0,073	Macro	UNRATE	NA	0,036
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	7,234	Macro	CPIAUCSL	-	2,282	GPs	Firm(GP) ID	NA	0,126
Macro	RPI	-	2,263	GPs	178323	-	1,406	Market Cond. & Environ.	S&P Dividend Yield	NA	0,047
Macro	US GDP (last year)	-	2,135	Fund information	Energy & Utilities	-	1,189	PE Industry	Industry years	NA	0,044
GPs	178323	-	1,059	PE Theory & Cent. Meas.	Eigenv ex-ante GP	+	1,144	Macro	HOUST	NA	0,030
GPs	9398	-	0,971	GPs	14527	-	1,076	Macro	PERMIT	NA	0,028
GPs	14527	-	0,953	GPs	238	-	0,695	Macro	FEDFUNDS	NA	0,028
GPs	9443	-	0,881	GPs	5462	+	0,549	Macro	OILPRICEx	NA	0,025
GPs	1697	-	0,837	GPs	6479	-	0,530	Macro	INDPRO	NA	0,020
GPs	7473	-	0,799	GPs	9398	-	0,518	PE Industry	Fund Raising Number	NA	0,019
GPs	1986	-	0,768	GPs	553	-	0,507	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,016

Net IRR Threshold Performance

Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Full sample				*Full sample*				*Full sample*			
Macro	CPIAUCSL	-	5,289	GPs	2280	+	1,737	GPs	Firm(GP) ID	NA	0,055
GPs	2280	+	1,391	Fund information	Energy & Utilities	-	1,317	Macro	HOUST	NA	0,017
GPs	1390	+	1,243	GPs	1390	+	1,251	PE Industry	Industry years	NA	0,017
GPs	8595	+	1,173	GPs	140	+	1,170	Macro	PERMIT	NA	0,016
GPs	48370	+	0,990	GPs	48370	+	1,085	Market Cond. & Environ.	S&P Dividend Yield	NA	0,016
GPs	7379	+	0,971	GPs	381	+	1,055	Macro	FEDFUNDS	NA	0,012
GPs	9403	-	0,957	GPs	8846	+	1,034	PE Industry	Fund Raising Number	NA	0,011
GPs	381	+	0,950	GPs	2214	+	0,977	Macro	CUMFNS	NA	0,006
GPs	11944	+	0,925	GPs	11944	+	0,957	Macro	Uncertain	NA	0,005
GPs	2214	+	0,917	GPs	7379	+	0,918	Macro	OILPRICEx	NA	0,005
First time fund sample				*First time fund sample*				*First time fund sample*			
Macro	CPIAUCSL	-	1,779	Macro	CPIAUCSL	-	0,369	GPs	Firm(GP) ID	NA	0,302
Macro	RPI	+	0,689	Macro	RPI	+	0,102	PE Industry	Industry years	NA	0,056
PE Theory & Cent. Meas.	Eigenv ex-ante GP	+	0,501	Fund information	Early Stage	+	0,099	Macro	PERMIT	NA	0,055
GPs	1390	+	0,408	GPs	9225	+	0,099	Macro	HOUST	NA	0,054
GPs	140	+	0,407	GPs	1390	+	0,091	Market Cond. & Environ.	S&P Dividend Yield	NA	0,042
GPs	3795	+	0,353	GPs	7379	+	0,089	Macro	FEDFUNDS	NA	0,042
GPs	7379	+	0,351	GPs	65	+	0,087	Macro	CUMFNS	NA	0,035
GPs	11944	+	0,342	GPs	140	+	0,073	Macro	CPIAUCSL	NA	0,030
GPs	65	+	0,338	GPs	11944	+	0,072	Macro	Uncertain	NA	0,029
GPs	8710	+	0,328	GPs	15991	+	0,072	Market Cond. & Environ.	S&P 500 returns	NA	0,025
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	7,376	Macro	CPIAUCSL	-	2,967	GPs	Firm(GP) ID	NA	0,120
Macro	RPI	-	2,287	Fund information	Energy & Utilities	-	1,551	Market Cond. & Environ.	S&P Dividend Yield	NA	0,042
GPs	8595	+	1,168	Macro	RPI	-	1,037	PE Industry	Industry years	NA	0,041
Market Cond. & Environ.	Med. 1st day Return	-	1,012	GPs	8846	+	0,946	Macro	PERMIT	NA	0,040
GPs	247	+	0,909	GPs	8595	+	0,822	Macro	HOUST	NA	0,040
Macro	US GDP (last year)	-	0,907	GPs	140	+	0,727	Macro	FEDFUNDS	NA	0,029
GPs	664	+	0,896	GPs	247	+	0,700	PE Industry	Fund Raising Number	NA	0,027
GPs	2017	-	0,865	GPs	2017	-	0,677	Macro	Uncertain	NA	0,014
Fund information	Energy & Utilities	-	0,831	GPs	1449	+	0,533	Macro	OILPRICEx	NA	0,012
GPs	140	+	0,741	GPs	8792	+	0,524	Macro	CPIAUCSL	NA	0,012

Table 2.14: Top 10 variables - Buyout funds

The table presents the top 10 variables with the highest absolute betas, considering all ML models (Lasso, Ridge, and RF), performance thresholds (PME and Net IRR), and samples of BO funds (full sample, first-time, and sequential funds). For each variable, we identified its group classification as outlined in Table 2.3, the beta direction (positive or negative), and the absolute beta value. Note that for RF models, beta direction is not applicable, so we have marked it as "NA" in the respective column. The variables are ranked from the largest (#1) to the 10th largest absolute beta value. In the variables column, you can match the names with Table 2.3 for more details. Variables associated with a number represent the ID of a specific GP in our dataset. Due to confidentiality, the names of the GPs are not disclosed.

PME Threshold Performance

Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Full sample				*Full sample*				*Full sample*			
Macro	CPIAUCSL	-	2,386	Market Cond. & Environ.	Median First day Return	+	3,423	GPs	Firm(GP) ID	NA	0,193
Market Cond. & Environ.	Median First day Return	+	1,595	Macro	CPIAUCSL	-	1,529	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,029
Macro	US GDP (last year)	+	0,883	GPs	1390	-	0,708	Macro	T5YFFM	NA	0,025
GPs	1390	-	0,852	Fund information	Africa	-	0,655	Macro	AAAFFM	NA	0,022
GPs	43451	-	0,728	Fund information	Americas	-	0,639	Macro	HWIURATIO	NA	0,022
GPs	581	-	0,720	Fund information	Australasia	+	0,588	Macro	T10YFFM	NA	0,022
GPs	12067	-	0,693	Fund information	Information Technology	+	0,524	Macro	OILPRICEx	NA	0,021
GPs	429	-	0,658	GPs	8837	-	0,341	Fund information	Fund Size (U$MM)	NA	0,021
GPs	782	-	0,651	GPs	11936	-	0,314	PE Theory & Cent. Meas.	Quant_all_lps	NA	0,021
GPs	6629	-	0,647	Fund information	Latin America & Caribbean	-	0,290	Macro	CUMFNS	NA	0,021
First time fund sample				*First time fund sample*				*First time fund sample*			
PE Theory & Cent. Meas.	Eigenv_ex_ante_GP	-	1,592	Market Cond. & Environ.	Median First day Return	+	2,765	GPs	Firm(GP) ID	NA	1,000
Macro	RPI	+	1,072	Macro	RPI	+	2,410	Market Cond. & Environ.	Median First day Return	NA	0,153
Market Cond. & Environ.	Median First day Return	+	0,799	Fund information	LOCATION_Europe	+	0,130	Market Cond. & Environ.	Tech_proceeds	NA	0,112
GPs	1390	-	0,149	Macro	HWIURATIO	+	0,089	Macro	RPI	NA	0,107
GPs	6839	-	0,143	PE Theory & Cent. Meas.	Proportion 1Q LPs ex ante	+	0,084	Market Cond. & Environ.	VC backed proceeds	NA	0,100
GPs	895	-	0,136	Fund information	Europe	+	0,071	Market Cond. & Environ.	VC IPOs	NA	0,089
GPs	25072	-	0,127	Macro	HWI	+	0,059	PE Industry	Fund Raising Volume	NA	0,088
GPs	9389	-	0,123	Macro	CLAIMSx	+	0,040	PE Industry	Dry Powder	NA	0,069
Macro	US GDP (last year)	+	0,121	GPs	46046	+	0,038	Fund information	GP Headquarter Location	NA	0,064
GPs	8734	-	0,110	Fund information	Australasia	+	0,027	Market Cond. & Environ.	Aggregate proceeds	NA	0,057
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	2,159	Macro	CPIAUCSL	-	0,675	GPs	Firm(GP) ID	NA	0,318
Macro	RPI	-	1,491	Market Cond. & Environ.	Median First day Return	+	0,639	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,048
Macro	US GDP (last year)	+	0,812	Macro	RPI	-	0,341	Macro	T5YFFM	NA	0,041
Market Cond. & Environ.	Median First day Return	+	0,756	Fund information	Americas	-	0,262	Macro	T10YFFM	NA	0,038
GPs	7550	-	0,449	Fund information	Australasia	+	0,192	Macro	FEDFUNDS	NA	0,033
GPs	581	-	0,445	Macro	US GDP (last year)	+	0,183	PE Industry	Industry years	NA	0,033
GPs	8935	-	0,445	GPs	11936	-	0,132	Macro	AAAFFM	NA	0,032
GPs	11936	-	0,439	Fund information	Information Technology	+	0,124	Fund information	Fund Size (U$MM)	NA	0,031
GPs	429	-	0,432	PE Theory & Cent. Meas.	Proportion 1Q LPs ex ante	+	0,119	Macro	OILPRICEx	NA	0,031
GPs	386	-	0,409	GPs	8837	-	0,113	Macro	BAAFFM	NA	0,031

Net IRR Threshold Performance

Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Macro	CPIAUCSL	-	2,724	Market Cond. & Environ.	Median First day Return	+	3,221	GPs	Firm(GP) ID	NA	0,021
Market Cond. & Environ.	Median First day Return	+	1,164	Macro	HWIURATIO	+	1,919	Macro	HWIURATIO	NA	0,003
GPs	6674	+	0,569	Macro	CPIAUCSL	-	0,971	Macro	CUMFNS	NA	0,003
GPs	129	+	0,568	Fund information	Americas	-	0,963	Macro	OILPRICEx	NA	0,003
GPs	8723	+	0,566	GPs	438	+	0,956	Fund information	Fund Size (U$MM)	NA	0,003
GPs	939	+	0,562	GPs	4936	+	0,829	Macro	AAAFFM	NA	0,003
GPs	438	+	0,559	GPs	2317	+	0,820	Macro	BAAFFM	NA	0,003
Macro	RPI	+	0,544	GPs	14457	+	0,790	Macro	HOUST	NA	0,003
GPs	21861	+	0,535	GPs	6674	+	0,714	PE Theory & Cent. Meas.	Eigen fund exante 5yr	NA	0,003
GPs	2317	+	0,529	GPs	21861	+	0,702	Macro	PERMIT	NA	0,003
First time fund sample				*First time fund sample*				*First time fund sample*			
Macro	RPI	+	0,567	Macro	RPI	+	0,559	GPs	Firm(GP) ID	NA	0,693
Market Cond. & Environ.	Median First day Return	+	0,234	Fund information	Healthcare	+	0,196	PE Industry	Dry Powder	NA	0,043
Macro	CPIAUCSL	-	0,219	Macro	CPIAUCSL	-	0,153	PE Industry	Industry years	NA	0,043
GPs	12952	+	0,175	Market Cond. & Environ.	Median First day Return	+	0,141	PE Industry	Fund Raising Number	NA	0,042
GPs	6955	+	0,168	Fund information	Australasia	-	0,127	PE Industry	Fund Raising Volume	NA	0,041
GPs	14457	+	0,164	GPs	10177	+	0,061	PE Theory & Cent. Meas.	Eigenv_ex_ante_GP	NA	0,038
GPs	7995	+	0,156	GPs	46046	+	0,058	Macro	OILPRICEx	NA	0,033
GPs	14055	+	0,151	Fund information	IT, Telecoms & Media	-	0,055	PE Theory & Cent. Meas.	Eigen fund exante 5yr	NA	0,032
GPs	21861	+	0,148	GPs	12952	+	0,052	Fund information	Fund Number Overall in GP	NA	0,025
GPs	14052	+	0,147	Fund information	Industrials	+	0,041	Macro	AAAFFM	NA	0,025
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	3,271	Market Cond. & Environ.	Median First day Return	+	3,891	GPs	Firm(GP) ID	NA	0,036
Market Cond. & Environ.	Median First day Return	+	1,611	Fund information	Americas	-	1,161	Macro	CUMFNS	NA	0,006
GPs	438	+	0,605	GPs	438	+	0,797	Macro	HWIURATIO	NA	0,006
GPs	548	+	0,605	Macro	HWIURATIO	+	0,732	Macro	Uncertain	NA	0,005
GPs	2317	+	0,601	GPs	6674	+	0,677	Macro	HOUST	NA	0,005
GPs	129	+	0,599	Macro	CPIAUCSL	-	0,664	Macro	AAAFFM	NA	0,005
GPs	7168	+	0,570	GPs	548	+	0,634	Fund information	Fund Size (U$MM)	NA	0,005
GPs	455	-	0,564	Fund information	LOCATION_Asia	-	0,626	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,005
GPs	718	+	0,561	GPs	4936	+	0,624	Macro	T10YFFM	NA	0,005
GPs	6674	+	0,545	GPs	2317	+	0,593	Macro	BAAFFM	NA	0,005

3
Breakeven inflation as the level of disbelief in monetary policy

Abstract. Machine learning techniques and algorithms that interpret natural language are used to observe the impact of newspapers on the difference in inflation expectations (breakeven inflation observed through the yield curve versus the Focus survey). The Lasso model is used to observe the main news stories that predict the market and interpret the Brazilian economy over the last decade. Despite the useful comprehension observed, the Brazilian yield curve market is extremely efficient with main information already incorporated in prices, newspaper news provides little additional information when predicting the future direction of the curves.

Keywords: breakeven inflation; monetary policy; machine learning; LLM - Large language models.

3.1 Introduction

Brazil has two similar economic indicators that which are generally used as indicators for future inflation, but, consistently, with different values over time: the inflation expectation given by the breakeven inflation (formed by the difference between the nominal and real yield curves) and the inflation expectation measured by the Central Bank's Focus survey. Indeed, the construction of the indicators is quite different; one is based on government bonds and derivative instruments traded on the market, and the other on data collected from participating institutions in the Focus survey. The act of buying and selling financial assets generates a different pecuniary commitment than simply answering a survey. However, the Focus survey has on average 140 active participants (an apparently representative sample of the Brazilian financial market) with an incentive mechanism to answer correctly; the Central Bank rewards the institutions that most accurately predict the results.

The breakeven inflation rate is usually higher than the inflation expectation given by the Focus survey. In 3.1 and 3.2 tables, we can observe this evolution in the recent period, year by year. The data is daily; the annual reference is the average observed in the year. For example, in 2015, the inflation expectation for one year ahead is 2016, and for four years ahead it is 2019, with the accumulated four years from 2015 to 2019 represented annually.

Table 3.1: **Inflation Expectations during the year - Average values**

	1 Year ahead		4 Years ahead		4 Years cumulative	
	Breakeven Inflation	Focus Survey	Breakeven Inflation	Focus Survey	Breakeven Inflation	Focus Survey
2016	5.72%	5.15%	5.52%	4.50%	5.44%	4.69%
2017	4.07%	4.30%	5.10%	4.12%	4.76%	4.23%
2018	4.23%	4.13%	5.28%	3.89%	4.91%	3.99%
2019	4.02%	3.88%	4.10%	3.63%	4.12%	3.72%
2020	3.43%	3.29%	4.18%	3.37%	3.97%	3.38%
2021	4.79%	3.93%	5.29%	3.11%	5.16%	3.37%
2022	6.31%	4.52%	6.39%	3.00%	6.29%	3.46%
2023	4.95%	3.97%	6.21%	3.65%	5.69%	3.73%

This table presents the average inflation expectation from the difference sources, Focus survey and the difference from nominal and real B3 Yield curves. Our sample period spans may/2016 to dec/2023. Source: Bacen and B3.

Table 3.2: **Breakeven Inflation rate premium during the year - Average values**

	1 Year ahead	4 Years ahead	4 Years cumulative
2016	0.57%	1.01%	0.75%
2017	-0.22%	0.98%	0.53%
2018	0.10%	1.39%	0.92%
2019	0.14%	0.47%	0.40%
2020	0.14%	0.81%	0.59%
2021	0.86%	2.17%	1.79%
2022	1.79%	3.38%	2.83%
2023	0.98%	2.56%	1.95%

This table presents the difference of inflation expectation values presented in 3.1. Source: Bacen and B3.

Consistently, the breakeven inflation expects higher inflation. Furthermore, the short-term breakeven inflation premium is lower than the long-term breakeven inflation premium. Do respondents in the Focus survey say one thing, but in practice, when trading assets in the market, expect a different value? Does the Focus survey sample not provide a true representation of the financial market? Are investors shortsighted and leaving money on the table? We are calling these indicators as expectations, but is the expectations theory valid in Brazil? There are numerous papers invalidating expectations theory in the yield curve. If realized inflation follows the Focus survey, which consistently presents values below the breakeven inflation rate, investors can achieve higher returns by investing in nominal interest rate instruments. In practice, and in the recent past, this has not been the case. The table 3.3 shows the average returns realized by each type of instrument.[1] Realized inflation has consistently exceeded market expectations, and even with a premium observed in the breakeven inflation rate, nominal bonds investments have received lower returns.

The market's demand for a premium demonstrates a disbelief that monetary policy will converge inflation to the target. The recent realized past has been

[1]This table was published in newspaper article in Valor Econômico on 20/03/2024 - Juros e expectativas de inflação - Márcio Garcia e Eduardo Marinho

Table 3.3: **Returns expected and realized**

	Yield expected ex-ante		**Yield realized ex-post (annual return)**		
	5 years		**5 years**		
Year	**IPCA +**	**CDI**	**IPCA +**	**Fixed**	**CDI**
2008	7.2%	12.2%	13.2%	**14.0%**	10.0%
2009	8.3%	10.2%	**14.7%**	12.3%	9.5%
2010	6.7%	9.7%	**13.7%**	12.1%	10.0%
2011	6.4%	11.7%	**14.0%**	12.0%	10.6%
2012	6.5%	8.6%	**13.9%**	9.6%	11.1%
2013	5.8%	7.9%	**12.4%**	10.8%	11.1%
2014	5.9%	10.7%	12.0%	**12.2%**	10.5%
2015	6.8%	13.2%	11.7%	**13.9%**	9.0%
2016	6.3%	14.1%	10.9%	**12.8%**	6.8%
2017	5.0%	10.4%	**10.9%**	10.1%	5.9%
2018	4.7%	6.5%	**10.8%**	10.0%	7.0%

This table presents the average return expected for 5 years on real Bonds, Fixed (right side expected and realized) and the average short term rates CDI on left. Against the average return achieved keeping the instruments for 5 years on right. Source: Bacen and B3.

confirmed that this disbelief is not in vain.

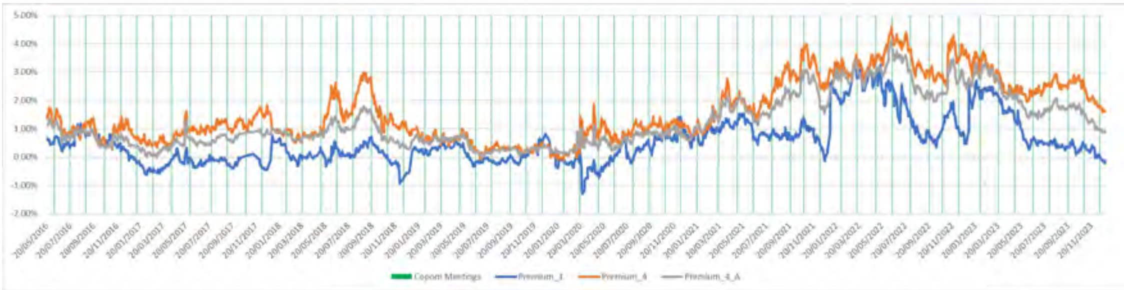
In figure 3.1, we have the behavior of these premiums in conjunction with Central Bank meetings. We have interesting events such as the March 2020 meeting. After this meeting, the short-term premium reversed, and the long-term premium increased. The drop in nominal rates, referenced by the Selic rate, was not accompanied by a drop in the real interest rate, either in the short or long term. One can think of lower inflation due to a supply shock from the pandemic in the short term, and a long-term inflationary rebound due to the low interest rates practiced and the excess of programmed fiscal stimulus. This demonstration provides a greater signal of a faster adaptation speed from the market in relation to expectations disclosed in the Focus survey.

In figure 3.2 we can observe the premium rising during stressful events such as Joesley Day (17/05/2017) and Dilma's impeachment [2], events that seem

[2]Three important dates: 15/10/2015, 02/12/2015 and 12/05/2016, the complaint for crime of responsibility offered by retired public prosecutor Hélio Bicudo and lawyers Miguel Reale Júnior and Janaina Paschoal; the acceptance by the president of the Chamber of Deputies, and the ended with the Senate approval, respectively.

to be much more linked to interest rates and risk premiums themselves than to inflationary expectations, but, the premiuns increased after these events.

Figure 3.1: **Breakeven inflation premium and Copom Meetings**



This figure shows the behavior of breakeven inflation premiums and the copom meetings

Figure 3.2: **Breakeven inflation premium and Market events**

This figure shows the behavior of breakeven inflation premiums and stress market events

The study of inflation and its impacts is a central topic in economics, finance, economic development, and other sub fields. In Brazil the attention is greater because the hyperinflation history that struck the country during the 80s and 90s. The premium movements on this daily frequency are the market's response to news. The market's reaction to the monetary and fiscal policy environment. The higher the premium, the greater the agents' disbelief in macroeconomic policies. Beyond that, understanding the premium movements in advance provides trading opportunities on yield curves. Is it possible to predict the behavior of the breakeven inflation premium? The data is presented daily; do newspaper news influence its movement? In the fourth chapter of this paper, the Fama-Bliss (1987) textbook test does not reject the hypothesis that expectations theory is valid for Brazilian yield curve. Thus, Machine Learning techniques were used in three ways:

- **Natural language processing**: an extensive database of newspaper articles from Valor Econômico, Folha de São Paulo, and Estado de São Paulo were classified by topic and sentiment. News was transformed into indicators of the daily environment.
- **Forecasting:** with news data, Lasso, Ridge, and Random Forest models were used in an attempt to predict the direction of the premium (whether the premium will rise or fall on the day in question). Also was attempted to forecast the premium itself (against the best AR model chosen via the Akaike method).
- **Interpretation:** Using the shrinkage property from Lasso (the winner model) the selected coeficients were analysed for each year, newspaper and object of study.

The main conclusion is that the Brazilian yield curve market proves to be extremely efficient; daily newspaper news adds little value when trying to predict the direction of premiums. The out-of-sample RMSE is beaten by the Lasso model only in the long-term premium (4 years ahead); AR models perform better for short-term premiums and the cumulative premium. In general, the market granger causes newspaper news, not the other way around. This is another demonstration of market efficiency. Events that occurs throughout a day ”t” is newspaper material for ”t+1”, but in general, it is already incorporated into the price at ”t”, demonstrating the difficulty of news in adding value for trading (morning news of t+1 to the market day of t+1, again, prices on t+1 have already contained the information of events that occurred in ”t”). The selection of coefficients demonstrated as expected, with news about inflation and fiscal policy being dominant in adding predictive value to the Lasso model.

This paper is organized as follows, the chapter 2 a brief literature review with 3 main areas of yield curve and expectations, methodology decomposition of yield curve and brazilian studies of breakeven inflation rate. The chapter 3 we described the data used (yield curve and the breakeven inflation rate data, the focus survey and the feature engineering of news, the chapter 4 we did a test of expectations theory to brazilian data following the classical paper from Fama & Bliss (1987), the chapter 5 our empirical strategy to forecast the movements of breakeven inflation premium, chapter 6 we tried to understand which news moved the curve using Lasso shrinkage property, and finally the chapter 7 concludes the paper.

3.2
Literature Review

This work relates to several areas of economic literature in Macroeconomics and Finance. Seminal studies such as those by Fama & Bliss (1987); Campbell & Shiller (1991) seek to understand whether prices negotiated on the yield curve can be interpreted as rational expectations of agents. That is, good proxies for future spot rates added to a constant liquidity premium. In the American case, it is common to reject this hypothesis, but not in the Brazilian case, as observed in the analysis of chapter 4. Recent literature seeks to understand the yield curve not exactly as rational expectations, but as heterogeneous subjective beliefs, analyzing the yield curve and survey data, as in Molavi *et al.* (2025), and Xiong & Yan (2009); Ehling *et al.* (2018); Buraschi & Whelan (2022) that modeled beliefs heterogeneity by assuming that investors "agree to disagree" about economic fundamentals, such as future interest rates or inflation.

Since Nelson & Siegel (1987); Litterman & Scheinkman (1991) we have had more methodological works in the literature, with the aim of deconstructing the shape of the yield curve through the factors of level, slope and curvature, trying to shape a curve capable of filling the maximum number of vertices (maturities) that go beyond the securities traded in the market. The curves published by B3 and Anbima are constructed using the Svensson (1995) model, following this research chain.

Going beyond the analysis of the nominal yield curve, we find in Bernanke (2007) that it is crucial for the stabilization of monetary policy that central banks appropriately manage the inflationary expectations of economic agents. The indicators analyzed in this article, inflation expectations via the Focus survey and the breakeven inflation rate, are of great importance. There is a literature that deconstructs the difference between the nominal and real yield curves in breakeven inflation rate, liquidity premium, and inflation risk premium in D'Amico *et al.* (2018), Abrahams *et al.* (2016) and Pflueger & Viceira (2011).

For the Brazilian case, we have several articles that seek to understand the nuances of expectations regarding these objects. Vicente & Guillen (2013) show lower predictive capacity in horizons of up to 18 months, in the breakeven inflation rate; however, the way the breakeven inflation rate is extracted makes a difference in the accuracy of the forecasts, as demonstrated by Val & Araujo (2019) and Araujo *et al.* (2025), using government bonds, derivatives, but mainly taking into account in the short term the seasonality of inflation and its 15-day lag in the short term embedded in government bonds. In Doi *et al.* (2017), the authors study the disagreement in inflation forecasts with the risk premium found in the breakeven inflation rate, using a VAR model. They find a relationship between the dispersion of forecasts and the disagreement shocks affecting the breakeven inflation risk premium at the level, but not in the curvature or slope. In fact, the results justify a higher premium demanded by investors in fixed-rate instruments when there is greater uncertainty regarding future inflation. Val & Araujo (2019) demonstrate the details in calculating the short-term breakeven inflation rate given the 15-day lag embedded in government bonds and the seasonality observed in the months. When correcting for these factors, the authors conclude that the proposed breakeven inflation measure of forecast: reacts more quickly to unexpected price shocks; is more accurate than the naive breakeven inflation that does not consider the indexation lag and seasonality; is statistically equal to or higher than measures based on market survey (Focus Top 5 surveys) and the breakeven inflation extracted from the futures market (through DAP contracts). The authors updated the study in Araujo *et al.* (2025), finding that the breakeven inflation rate formed by derivative instruments is more accurate than that calculated through government bonds to forecast short-term inflation. For longer terms, the pure expectations hypothesis for breakeven inflation is not corroborated, aligning with previous research and with our present study. The results suggest that investors demand higher compensation to hold fixed-income securities with longer maturities, the disbelief of convergence in monetary policy in the future. Costa Filho (2025) have studied trend inflation and its stochastic volatility. His aim is to understand, in the Brazilian case, whether expectations are well anchored, because there is a debate in Brazil that inflation expectations from professional forecasters are over reliant on the views of the financial sector, which may behave opportunistically. Their main conclusion is that stochastic volatility is mildly correlated with inflation disagreement, with no direct relationship was observed between periods of inflationary uncertainty (when inflation disagreement rises), and stochastic volatility.

This work differs from the others because it is the first to use newspaper articles to analyze the breakeven inflation rate premium, and our intention is not to measure the quality of these expectations, or if the breakeven inflation rate performs well than focus survey. Our exercise is to understand how the expectations move, what factors govern expectations and premiums, and whether we are able to predict the behavior of agents' expectations in advance.

3.3 Data

3.3.1 Breakeven inflation rate

Before specifically discussing the main subject of this article, the breakeven inflation rate premium, it is necessary to explain the breakeven inflation rate itself. The breakeven inflation rate is obtained by the difference observed between the nominal yield curve and the real yield curve. The breakeven inflation rate is a price of inflation, because is derived from market prices. It is a sum of agents expectations and a term premium. It can be used to predict the realized inflation that is measured in the economy's prices, day by day. From a theoretical standpoint, in the next chapter, we will demonstrate that, unlike what is normally observed in the American yield curve, the theory of expectations holds true for Brazil.

Yield curves present spot interest rates for different future maturities, a snapshot of the economy at any given moment, but not its actual realization. For example, on October 31, 2025, the spot rate of the nominal curve published by ANBIMA for one year was 14.0011%, and the spot rate for the real curve was 9.4578%. The consensus of market expectations (with a premium embedded together that is not possible to disentangle) for one-year inflation (the period between October 31, 2025, and October 31, 2026) is $\pi^e_{BEIR} = \frac{1+14.0011\%}{1+9.4578\%} - 1 = 4.1507\%$. One argument supporting the theory of pure expectations implicit in the nominal and real yield curves is the possibility of arbitrage. In the Brazilian financial market, for a large portion of the observed points on the curves, there are instruments available to "arbitrage" the value of 4.1507%. If an investor believes in a higher or lower expected inflation rate for the period between October 31, 2025, and October 31, 2026, she can trade "against the market". If an investor's expected inflation is higher than 4.1507%, the investor can short some Notional amount in the nominal curve futures and go long in the real curve futures with the same Notional amount. The daily market adjustments from the both future positions will provide the profit from the operation:

$$Notional * (1 + 9.4578\%) * (1 + \pi_{0\to 1}) - Notional * (1 + 14.0011\%) = Profit$$
$$Profit \geq 0 \, for\, all \; \pi_{0\to 1} \geq 4.1507\%$$

One additional point is that the market and economists are usually interested in inflation in closed calendar periods. We are in 2025, and the expected inflation of interest is the inflation for 2026, 2027, 2028, and 2029 (for the next 4 years). Which match the set of questions and answers in the Central Bank's Focus survey for annual inflation. Our interest will be in short-term inflation expectations for the 1 year ahead (for instance, in our current moment, 2026), in long-term inflation expectations, 4 years ahead (in our current moment, 2029), and in accumulated inflation over 4 years (in our current period from 2026 to 2029).

These data are not directly observed in the spot rates of the yield curve, is observed only in the month of December. The breakeven inflation rate for the next year, for the next 4 years, and for the 4 years cumulative, is usually obtained through the forward rates implied in the yield curves. Which will obey the same no-arbitrage principle; therefore, it's the market consensus regarding the realization of future inflation expressed in prices.

For example, if we are collecting data from the yield curves during the month of June, the breakeven inflation rate for the following year will be given by the formula below:

$$\pi_{BEIR2026} = \frac{1+i_{2026}}{1+r_{2026}} - 1$$
$$1 + i_{(2026=126\to 378)} = \frac{(1+i_{0\to 378})^{\left(\frac{378}{252}\right)}}{(1+i_{0\to 126})^{\left(\frac{126}{252}\right)}}$$

The same vertex will be used for real rates.

For the **Nominal interest rate data** the theory says that government bonds establish the curve (from various maturities of LTNs & NTN-Fs). However, other instruments are used, because it contains more vertices, for instance: Fixed x DI swaps, and DI Futures. The data of this paper is the DI Futures. [3] For the **Real interest rate curve** the theory says that it is formed by the various maturities of NTN-Bs, real government bonds. Other instruments are also used and contain more vertices. The B3 DAP- DI x IPCA coupon is the data used in this paper. [4] Daily data from August 2016 to December 2023 were used, with 14 vertices:

- 1 Month
- 3Months
- 6Months
- 9Months
- 1Year
- 2Years
- 3Years
- 4 Years
- 5 Years
- 6 Years
- 7 Years
- 8 Years
- 9 Years
- 10 Years

Based on this data, a new time series was constructed, showing the breakeven inflation for the 1 year ahead, the breakeven inflation for 4 years ahead, and the cumulative breakeven inflation for the next four years. For each month of data collected, the set of vertices of the spot rates to construct the forward rates is different. Following this spirit on the example above from June data, 36 different

[3]Thanks to Carlos Viana, Marco Bonomo and Marcos Mendes, who provided the data.
[4]Thanks to Insper doctoral student Matheus Patrocinio who provided the data

formulas were used. One for each month for each of the three time frames chosen, creating the new data series of breakeven inflation rate. In this way, the data series becomes comparable to the Central Bank's expectations system provided by the Focus survey.

3.3.2
Focus survey and the Premiuns

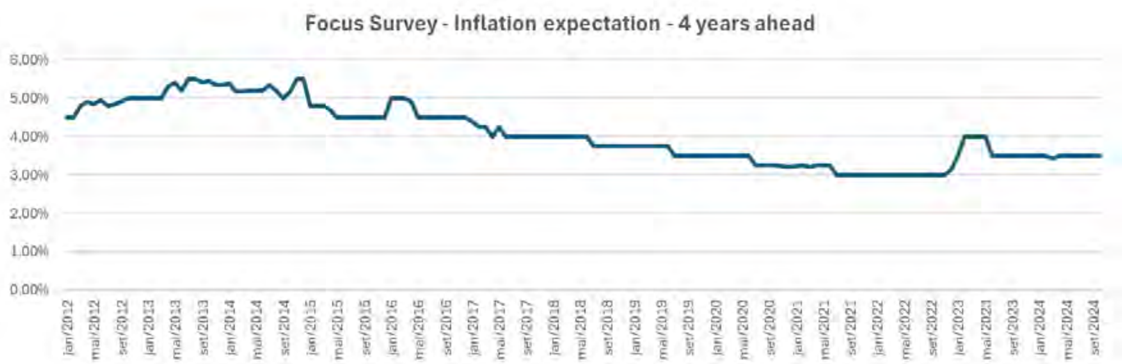
The nominal and real interest rate curves are used to obtain the data series on inflation expectations given by breakeven inflation. The Central Bank has available the historical data on inflation expectations collected by the Focus survey over time. The Focus survey exits since 1999; the daily median data of inflation expectations for 1 year ahead, the 4 years ahead, and the 4 years cumulative were matched with the yield curve database timeseries described above. With the data compared, it is possible to compare the expectations implicit in the yield curves and the expectations disclosed by agents in the Focus survey for these different maturities. Consistently, the expectation contained in the breakeven inflation is higher than the inflation expectation collected by the Focus survey; there is a premium in the inflation expectation measured by the breakeven inflation.

$$
\begin{aligned}
p_{1y} &= \pi^e_{BEIR\,1y} - \pi^e_{focus\,1y} \\
\mathrm{p}_{4y} &= \pi^e_{BEIR\,4y} - \pi^e_{focus\,4y} \\
\mathrm{p}_{4ac} &= \pi^e_{BEIR\,4ac} - \pi^e_{focus\,4ac}
\end{aligned}
$$

3.3.3
Why is the level of disbelief in monetary policy?

The figure 3.1 represents the average of long-term inflation expectations (4 years ahead) recorded by the Focus survey since the beginning of the series. We can observe that expectations have been falling over recent times. Indeed, these expectations are equal to the inflation targets of the Brazilian Central Bank during the recent years. The fall reflects an institutional change in the target being gradually reduced over time. In other words, economists, when answering the Focus survey, believe in the convergence of inflation to the target in the long term, or at least, they pretend to believe it. The same cannot be observed in breakeven inflation, looking to the figure 3.2 the market has a disbelief in the convergence of inflation to the target in the long term. However, according to the liquidity premium hypothesis, the market is actually charging a premium. Put another way, if the market charges a premium, it's the same as saying it doesn't believe in convergence. The higher the premium, the greater the disbelief. This data is collected daily and shaped by the market, which makes it interesting for us to assess how much news influences this premium.

Figure 3.1: **Focus Survey - Long term inflation expectation**



This figure shows the evolution of long term inflation expectation median collected by Focus Survey

3.3.4
News - Transforming textual data in information - Feature engineering

This study used news articles from three major Brazilian newspapers: Valor Econômico, Folha de São Paulo, and Estado de São Paulo. [5]; The number of news

[5]We are grateful for Marcelo Medeiros e Leonardo Ladalardo who provided the database , contained in Ladalardo & Medeiros (2022).

Figure 3.2: **Focus Survey and Breakeven Inflation - Long term inflation expectation**

This figure shows the evolution of long term inflation expectation median collected by Focus Survey and the inflation expectation observed in breakeven inflation

articles per year is detailed in Table 3.1 The daily breakeven inflation premium data based on the yield curve and the focus survey were matched with the daily news data. The database of the paper is from 2016 to 2021.

Table 3.1: **News database**

Year ↓ | Newspaper →	**Valor Econômico**	**Estado de São Paulo**	**Folha de São Paulo**
2012	65,545	0	0
2013	51,057	0	0
2014	45,509	24,168	30,628
2015	45,048	24,180	24,048
2016	47,010	24,192	22,073
2017	46,195	24,204	18,479
2018	47,001	24,216	25,179
2019	46,059	24,228	16,377
2020	60,773	24,240	15,287
2021	22,985	10,105	4,797
Total	**477,182**	**179,533**	**156,868**

This table presents the news database used in the article.

The set of daily news were transformed into news counts according to the 6 themes below, combined with 3 sentiments (positive, negative, and neutral), totalizing 18 variables. In addition to these variables, the counts of overall sentiment of the day's newspaper news was added, totalizing 21 variables at the end.

To ensure comparability between days, given that the number of daily news items changes from one day to the next, the quantities were transformed into proportions.

Themes:

- Atividade Econômica (Economic Activity)
- Economia Internacional (International Economy)
- Política Fiscal e Contas Públicas (Fiscal Policy and Public Accounts)
- Inflação e política monetária (inflation and monetary policy)
- Politica (Politics)
- Outro assunto (Other topics)

For textual analysis data and its transformation, 3 natural language processing models available in Hugging Face community were used, 1 for sentiment analysis, and 2 for them classification:

- **The FinBERT-PT-BR**[6] is a pre-trained Natural Language Processing (NLP) model specifically developed to analyze sentiment in Brazilian Portuguese financial texts. The model's development occurred in two main stages: language modeling and sentiment modeling. In the first phase, the language model was trained using a vast corpus of over 1.4 million texts of financial news in Portuguese. Thanks to this initial training, it was possible to create a sentiment classifier that achieved satisfactory convergence, even when utilizing a relatively small number of labeled texts (just 500). However, the algorithm used in this paper sliced larger texts and evaluated each part of the texts, creating a weighting for each part of the larger texts. Thus, we evaluated sentiment using all labels, even when texts exceeded the number of labels supported by the model. The FinBERT-PT-BR has a detailed comparative analysis with other models, the comparison results showed that it outperformed current state-of-the-art models. Among its applications, it was demonstrated that FinBERT-PT-BR is highly useful

[6] Santos *et al.* (2023)

for building sentiment indices, formulating investment strategies, and analyzing macroeconomic data, such as inflation.

- For theme classification, first we used a dictionary to translate[7] the news to english and after a theme classification model was used. The theme classification model is **deberta-v3-base-zeroshot-v1** [8], this model is a pre-trained Natural Language Processing (NLP) tool built upon the efficient DeBERTa-v3 (Base-sized) architecture. Its primary distinction is that it has been optimized to work as a universal text classifier. The core purpose of this model is to perform zero-shot classification, which grants it the ability to classify a piece of text into categories it has never explicitly encountered during its training, relying solely on the provided category descriptions. This capability is achieved by reframing every classification task into a problem of Natural Language Inference (NLI). In this process, the model takes the text to be classified (the premise) and the category label (the hypothesis) as input. It then determines the degree to which the hypothesis is "true" (entailment) or "not true" (not entailment) given the context of the text. To classify a text against multiple categories (e.g., "sports," "politics," "finance"), the model tests each category as a separate hypothesis, and the label that yields the highest "entailment" score becomes the final prediction.

  The model was fine-tuned on a massive and diverse mix of multiple classification tasks and established NLI datasets, such as MNLI(Multi-Genre Natural Language Inference) and ANLI (Adversarial Natural Language Inference), all converted into this universal NLI format.

3.4 Do the expectations theory is valid for Brazil?

Throughout the text, we are assuming expectations. However, is the expectations theory valid for the term structure of Brazilian interest rates? One

[7] The tradutor used was: model="Helsinki-NLP/opus-mt-ROMANCE-en"
[8] Laurer *et al.* (2023)

way to represent agents' expectations is given by the difference between forward rates and the spot rate observed today ($f_t^{N \to N+1} - y_t^1$). Is this difference observed today a good predictor of the spot rate differential that will be realized at t+N ($y_{t+N}^1 - y_t^1$)?

We will use the regression 3.1 below inspired by Fama & Bliss (1987) to test the validity of expectations theory for the Brazilian case. The **1 year** period from the regression was generalized to **P periods**. We will test the coefficient "b" for different vertices available on the yield curve. For "b" close to 1, the forward-spot premium is a good predictor of the future realization of spot rates.

$$y_{t+N}^P - y_t^P = a + b(f_t^{N \to N+P} - y_t^P) + \epsilon_{t+N} \tag{3.1}$$

Being:

y_{t+N}^P: Spot rate of term P observed in t+N

y_t^P**:** Spot rate of term P observed in t

$f_t^{N \to N+P}$**:** Forward rate of term P (from N till N + P) observed in t.

The forward rates were calculated as follows:

$$f_t^{N \to N+P} = \left(Ln\left(1 + y_t^{N+P}\right) * \frac{P+N}{12} - Ln\left(1 + y_t^N\right)\frac{N}{12}\right) * \frac{12}{P} \tag{3.2}$$

First, the table 3.1 below shows a feasibility analysis for carrying out the experiment. That is, which combinations of N and P are feasible in the term structure of Brazilian interest rates to perform the generalized Fama-Bliss regression. The feasibility analysis involves checking if, at time "t", the rates y_t^{P+N} and y_t^N, necessary to calculate the forward rate, are available.

In table 3.2, we can observe that the result of Fama-Bliss generalized regressions is pervasive. The results for coefficients "b" are satisfactory for almost every pair N X P combinations. Forward rates observed at time t represents expectations of the spot rates to be realized at t + N, especially for longer terms.

In general, this paper is a forecasting paper; our objective is to predict future movements of the breakeven inflation premium. We could proceed agnostically, attempting to predict the behavior of this indicator, regardless of what it represents. However, the test in this chapter allows us to state that the behavior of Brazilian yield curves can be supported by expectations theory, unlike what occurs in tests of the American yield curve, for example. Thus, our work is an attempt to predict changes in the behavior of market forecasts, understand which drives market expectations.

Figure 3.1: **Feasibility Table N X P**

	P = Spot term (months)																									
N =	1	2	3	4	5	6	7	8	9	10	11	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180
1	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
2	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
3	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
4	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
5	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
6	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
7	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
8	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
9	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
10	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
11	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
12	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible
18	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
24	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible
30	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
36	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible
42	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
48	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible
54	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
60	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Not Feasible	Not Feasible	Not Feasible	Not Feasible	Not Feasible
Qtt of possible regressions	11	10	9	8	7	6	5	4	3	2	1	5	5	5	5	5	5	5	5	5	5	4	3	2	1	0

This table presents the possible combinations to test the Fama-Bliss regression

Figure 3.2: **Fama-Bliss generalized regressions - May/2003 to Dec/2023**

			N = Time (Expectations)															
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**24**	**36**	**48**	**60**
Interest term (Months) = P	1	a	-0.003***	-0.005***	-0.005***	-0.005***	-0.004***	-0.002*	-0.002*	0.003.	-0.001	0	0.006*					
		b	0.496***	0.985***	1.047***	1.18***	1.239***	0.978***	1.218***	1.036***	0.639***	1.31***	0.602***					
		R^2	0.265	0.626	0.734	0.755	0.74	0.568	0.674	0.517	0.305	0.57	0.23					
		Df	241	240	239	238	237	127	126	125	124	123	122					
	6	a	-0.001***	-0.002**	-0.002***	-0.001	-0.001	-0.001										
		b	0.777***	1.023***	1.315***	1.119***	1.178***	1.309***										
		R^2	0.251	0.408	0.478	0.416	0.423	0.449										
		Df	132	131	239	129	128	236										
	12	a												0.003	0.007**	0.011***	0.027***	0.027***
		b												1.033***	1.231***	1.189***	1.491***	1.28***
		R^2												0.303	0.473	0.473	0.625	0.532
		Df												230	218	206	194	182

This table presents the results from different combinations of N X P tested by the Fama-Bliss generalized regression

3.5
Empirical Approach

This section presents the empirical strategy and the results found. The work consists primarily of forecasting, supported by machine learning techniques. Newspaper reports, transformed into data, are used to try to predict the daily behavior of the breakeven inflation premium. In sections 5.1, 5.2, and 5.4, we attempt to predict the direction of the premium day by day (1 if premiums rise, and 0 if premiums fall). In section 5.3, we attempt to predict the premium itself, and in section 5.5 we perform a Granger causality test between the news indicators and the premiums (and vice versa).

3.5.1
Discrete choice models and the market direction

In this section, we attempt to predict the direction of the premium using discrete choice models. Three machine learning techniques were used: Lasso, Ridge, and Random Forest, following the logic of the simplified formula below.

$$E[\mathbb{1}_{(p_t^j - p_{t-1}^j) >= 0}] = f(\beta' X_t^{is}) \tag{3.3}$$

Being:

p^j: Premiuns of j = 1 year ahead, 4 years ahead, and 4 cumulative years.

f(): Lasso, Ridge or Random forest methods for discrete choice models[9]

i: The newspaper set of news (Valor Econômico, Estado de São Paulo, Folha de São Paulo

s: The set of variables Xs (21 or 81 subjects combined with sentiment)

[9]The Random Forest function is not linear and does not have coefficients. We ask the reader for their comprehension regarding the simplification adopted. Details with due precision regarding these functions are in the second paper of this dissertation, chapter 3, methodology.

t: Represents the time series and the size of the windows (considering rolling or extending)

The right side consists of two set of news variables derived from newspaper articles (one set with 21 news variables, and another with 81 news variables that allows combination of two subjects[10], with newspapers tested separately). The models are trained in an in-sample window and tested out-of-sample with the morning news of the day. We tested methodological variations for each model: in-sample windows with 126, 252, 315, 378, 441 and 504 working days, extended windows, and rolling windows. For each 3 premiums (1 year ahead, 4 years ahead, 4 year cumulative), were tested 3 models (Lasso, Ridge, Random Forest), 3 (newspapers) * 7 (training window sizes) * 2 types of in sample windows (rolling or extending windows), totalizing 756 methods tested.

Given the multiplicity of models, we report the average results of the top 20 best methodologies in Table 3.1

Table 3.1: **Out-of-sample Accuracy**

Direction Forecast (Raise or Fall)			
Object	**Out-of-sample Accuracy**	**Winner Model**	**Winner Newspaper**
1 year ahead premium	0.528	Random Forest	Estado de São Paulo
4 years ahead premium	0.507	Random Forest	Folha de São Paulo
4 years cumulative premium	0.520	Random Forest	Valor Econômico

This table presents the average out-of-sample accuracy for the best 20 methodologies.

The random forest model shows a clear characteristic of overfitting; the in-sample accuracy is 100%. In general, for the 3 premiuns, the models are practically equivalent to flipping a coin.

3.5.2 Tomorrow news and the market direction

The unsatisfactory results made us wonder if the news was already outdated by morning. What if we could travel to the future, see tomorrow's news,

[10]For instance - Inflation and International Economy together is a exclusive variable, on bigger set, on the smaller set the Inflation would count 1 and international Economy also count 1.

and try to predict the market today? In this section, we conducted an experiment using the following day's news to try to predict the market direction, following the formula below. The results are in 3.2.

$$E[\mathbb{1}_{(p_t^j - p_{t-1}^j)>=0}] = f(\beta' X_{t+1}^{is}) \tag{3.4}$$

Table 3.2: **Out-of-sample Accuracy - Tomorrow News**

Direction Forecast - Tomorrow news (Raise or Fall)			
Object	**Out-of-sample Accuracy**	**Winner Model**	**Winner Newspaper**
1 year ahead premium	0.526	Random Forest	Estado de São Paulo
4 years ahead premium	**0.518**	**Lasso**	Folha de São Paulo
4 years cumulative premium	0.519	Random Forest	Valor Econômico

This table presents the average out-of-sample accuracy for the best 20 methodologies, using the next day news as explanatory variables.

Average out-of-sample accuracy increases for the 4-year-ahead premium. And the winning models are from the Folha de São Paulo newspaper. Folha's news seems to be a reflection of the market and not a shaper of the market. However, the results improvement are very subtle; news adds little to predicting the market direction. Machine learning and feature engineering analysis provides indicators of subject matter and sentiment, not a perfect foresight of what the market would be like the following day.

3.5.3 Forecasting the premium

In this session, we attempted to predict the premium itself and evaluated the predictive capacity of each model using out-of-sample RMSE. Minor changes were made. We added an extended AR model, where the selected lag was determined using the akaike selection criterion. The model is extended because we added a dummy variable for the occurrence of the Copom meeting the day before. The AR model was chosen because the premium shows stationarity over time. For the machine learning models, in addition to the news variables, the lags of the three premiums and the dummy variable of the Bacen meeting eve were added.

$$\hat{p}_t^j = f(\beta' X_t^{is} + \alpha' p_{t-h}^j + D_{copom}) \tag{3.5}$$

Being:

p^j: Premiuns of 1 year ahead, 4 years ahead, and 4 cumulative years.

f(): Lasso, Ridge (Gaussian models), Random forest and the best AR selected by Akaike method

i: The newspaper set of news (Valor Econômico, Estado de São Paulo, Folha de São Paulo

s: The set of variables Xs (21 or 81 subjects combined with sentiment)

t: Represents the time series and the size of the windows (considering rolling or extending)

h: Equal 1 for Lasso, Ridge and Random Forest model, lag operator selected by akaike model when f() is an AR model.

The same methodological variations regarding window size, rolling or extended windows, use of newspapers separately, news of the day or of the future were adopted, totalizing 252 models per premium. The main results are reported in Table 3.3 and in Table 3.4.

Table 3.3: **Out-of-sample RMSE**

Object	Premium Forecast (Out-of-sample RMSE/Mean)	Winner Model	Winner Newspaper
1 year ahead premium	41.99%	AR / Lasso	Draw
4 years ahead premium	11.82%	Lasso	Valor Econômico
4 years cumulative premium	9.748%	AR / Lasso	Estadão / Valor

This table presents the average out-of-sample RMSE divided by the mean from 20 the best methodologies

Table 3.4: **Out-of-sample RMSE - Tomorrow news**

Premium Forecast - Tomorrow news			
Object	**(Out-of-sample RMSE/Mean)**	**Winner Model**	**Winner Newspaper**
1 year ahead premium	41.96%	AR / Lasso	Draw
4 years ahead premium	11.88%	Lasso	Estado de São Paulo
4 years cumulative premium	9.753%	AR / Lasso	Draw

This table presents the average out-of-sample RMSE divided by the mean from 20 the best methodologies, using tomorrows news: $\hat{p_t^j} = f(\beta' X_{t+1}^{is} + \alpha' p_{t-h}^j + D_{copom})$

There is no clear pattern regarding the window size and the newspaper with the most informational content when we look at the results. We noticed that for the 1-year and 4-year forward premiums, the winning models are the AR and Lasso models, and the winning model for the 4-year ahead premium is the Lasso model. The Lasso model, combined with the variables constructed from newspaper news, manages to bring a lower RMSE when compared to a traditional AR model for the long-term inflation premium. We can conclude that the long-term inflation premium, that is, the level of market disbelief in the convergence of inflation to the target by monetary policy, is influenced by the current economic atmosphere portrayed in the news. However, since news occurs with high frequency, the contribution observed by the morning news in newspapers is very subtle. In general, as we will observe in the section 5.5, the information brought in the morning by newspapers is already contained in the prices.

3.5.4 The market direction with Gaussian models

In this section, we tried to predict the direction of the premium, but using Gaussian models from the last section. We predicted the value of each premium and compared the result with the previous day, following the metodology of equations 3.6, 3.7 and 3.8 below.

$$\hat{p_t^j} = f(\beta' X_t^{is} + \alpha' p_{t-h}^j + D_{copom}) \tag{3.6}$$

The exercise with tomorrow news:

$$\hat{p}_t^j = f(\beta' X_{t+1}^{is} + \alpha' p_{t-h}^j + D_{copom}) \tag{3.7}$$

$$Premium\hat{D}irection = \mathbb{1}\{(\hat{p}_t^j - p_{t-1}^j) \geq 0\} \tag{3.8}$$

The results are shown in Table 3.1, the best out-of-sample result was found using this approach. However, the accuracy remains below 60%, which does not provide a relevant motivation to use the method for trading purposes.

Figure 3.1: **Market direction with Gaussian models**

Direction Forecast (Raise or Fall) - Gaussian Models			
Object	**Out-of-sample Accuracy**	**Winner Model**	**Winner Newspaper**
1 year ahead premium	0.507	AR / Lasso	Estadão / Folha
4 years ahead premium	0.543	Ridge	Valor Econômico
4 years cumulative premium	0.5281	AR / Lasso	Estadão / Valor

Direction Forecast - Tomorrow news (Raise or Fall) - Gaussian Models			
Object	**Out-of-sample Accuracy**	**Winner Model**	**Winner Newspaper**
1 year ahead premium	0.506	AR / Lasso	Estadão / Folha
4 years ahead premium	0.541	Ridge	Valor / Estadão
4 years cumulative premium	0.5275	AR / Lasso	Draw

This table presents the average out-of-sample accuracy for the best 20 methodologies for each line. First step we predicted the premium following 3.6 or 3.7, second step we verified if it raised or declined following 3.8, and calculated out-of-sample accuracy.

3.5.5 Granger Causality

In order to evaluate the "chicken and egg" problem, the variables representing newspaper articles against the different premiuns were assessed using Granger causality testing. The test was performed in both directions, evaluating whether the market Granger causes the news, or whether the news Granger causes the market, during the entire data window.

In general, the market Granger causes newspaper articles, not the other way around, explaining the low effectiveness of the models, also a demonstration of high efficiency in brazillian yield curve market.

On table 3.2 is possible to see the results. The highlight is the inflation-related news in Folha de São Paulo, which provides informative content for all premiuns, fiscal news for Valor Econômico, and overall positive sentiment for Valor and Estadão. These are topics directly related to the economic variables in question, which have informational value in shaping the expectations of economic agents.

Figure 3.2: **Granger Causality Test - QTT Greater than 95% of confidence**

Valor Econônico	**Market to News**	**Ambiguous**	**News to Market**	**Subjects (News to Market)**
Premium_1	11	8	1	Fiscal.Negative
Premium_4	7	11	2	International Economics.Negative / Overall sentiment_Positive
Premium_4_A	8	11	1	Politics.Neutral
Total Valor	26	30	4	
Estado de São Paulo	**Market to News**	**Ambiguous**	**News to Market**	**Subjects (News to Market)**
Premium_1	7	13	0	
Premium_4	4	13	3	Economic Activity.Neutral /International Economics.Neutral / Overall sentiment_Positive
Premium_4_A	9	8	3	Economic Activity.Positive / Politics.Neutral / Overall sentiment_Positive
Total Estado de São Paulo	20	34	6	
Folha de São Paulo	**Market to News**	**Ambiguous**	**News to Market**	**Subjects (News to Market)**
Premium_1	6	13	1	Inflation.Negative
Premium_4	7	11	2	Inflation.Negative / Politics.Negative
Premium_4_A	6	12	2	Inflation.Negative / Other topics.Neutral
Total Folha de São de Paulo	19	36	5	

This table presents the quantity of news variables that granger causes the market and other way around.

3.6
Variable interpretation - News selected

The Lasso model was the winner, providing the lowest out-of-sample RMSE predictions. Despite its subtlety, the daily morning news contains informational content when predicting the breakeven inflation premium. The Lasso model has the shrinkage (coefficient "selection") feature, allowing the interpretation of results. We can understand which news variables were selected, that is, which news variables contribute most to predicting the premium outcome.

In section 5, various out-of-sample methods of forecasting were tested for the entire sample. In this section, the model specification follows the formula 3.9 below:

$$\hat{p_t^j} = f(\beta' X_t + \alpha' p_{t-1}^j + D_{copom}) \tag{3.9}$$

Being:

p^j: Premiuns of 1 year ahead, 4 years ahead.

f(): Lasso function

X: Now the explanatory variables belongs the 3 newspaper together (Valor Econômico, Estado de São Paulo, Folha de São Paulo) with 21 subjects combined with the sentiments for each one, totalizing 63 variables

t: Represents the time series for each year of news database (2016, 2017, 2018, 2019, 2020, 2021)

Instead of testing the newspapers separately, we tested all newspapers in the same environment, tripling the number of variables. By this way, we let the data speak for itself and can observe if the model automatically selects any specific newspaper most than the others. We also circumscribed the test window to a closed year, so we can observe the selection of variables in different periods and understand which news items are most representative in understanding the breakeven inflation premium at different moments of the recent Brazilian economy. The desire were to observe the top five coeficients selected by the lasso model from each year, to the 2 premiums (1 year ahead and 4 years ahead).

Initially, the general results are presented. Table 3.1 lists the themes (variables) that appear in the top five of the variables selected by Lasso for each premium. Market expectations are mainly influenced by issues correlated with inflation, with the short-term premium having more noise effects. Table 3.2 shows that Valor Econômico is a newspaper with more informational content in shaping market expectations, and negative news has a greater influence on market expectations.

More details following figures 3.1 and 3.2 show an exercise with the news in the morning, the same day of market close. This assessment reveals the main newspapers news that help predict the 1-year and 4-year ahead premiums on that day, **the news that shapes the market**. Fiscal and inflation topics dominate the selection, but observe the 2020 year is interesting, Covid pandemic put another topics on top five helping to form the expectations.

The figures 3.3 and 3.4 show the same exercise with the following day's news, an attempt to evaluate market efficiency and the dynamics of Brazilian newspapers. What information becomes news the next day? Indeed, **this exercise is the market that shapes the news**.

Table 3.1: **News topics - most selected by Lasso**

Number of times being in the top five		
Subjects	**Premium 1 y**	**Premium 4 y**
Fiscal	11	14
Inflation	10	15
Other.topics	10	9
Politics	9	10
International.Economy	8	5
Economic.activity	5	6
Overall.Sentiment	5	1

This table presents the number of times that each news topic was selected as top five coefficient by the Lasso model

Table 3.2: Newspapers and sentiments most selected by Lasso

Number of times being in the top five		
Newspaper	**Premium 1 y**	**Premium 4 y**
Valor Econômico	22	36
Estado de São Paulo	29	16
Folha de São Paulo	7	8
Sentiment	**Premium 1 y**	**Premium 4 y**
Negative	29	31
Positive	10	18
Neutral	19	11

This table presents the number of times that each newspaper and news sentiment was selected as top five coefficient by the Lasso model

Figure 3.1: **News (today) to Market - 1 Year ahead**

2016	2017	2018	2019	2020	2021
Valor_Fiscal-Negative	Estadao_Other.topics-Negative	Valor_Fiscal-Negative	Valor_International.Economy-Neutral	Estadao_Other.topics-Negative	Valor_Politics-Negative
Valor_Other.topics-Neutral	Estadao_Other.topics-Neutral	Folha_Fiscal-Negative	Estadao_Inflation-Neutral	Valor_Fiscal-Neutral	Valor_Economic.activity-Positive
Estadao_International.Economy-Negative	Estadao_Overall.Sentiment-Negative	Estadao_Inflation-Positive	Estadao_Inflation-Negative	Estadao_Fiscal-Neutral	Estadao_Politics-Neutral
Estadao_Economic.activity-Neutral	Estadao_Other.topics-Positive	Folha_Politics-Negative	Folha_International.Economy-Positive	Valor_Economic.activity-Positive	Estadao_Economic.activity-Neutral
Estadao_Other.topics-Negative	Estadao_Politics-Positive	Valor_Inflation-Neutral	Valor_Politics-Negative	Estadao_Overall.Sentiment-Negative	Valor_Overall.Sentiment-Positive

This figure presents the top five variables selected by the Lasso model, for each year regression

Figure 3.2: **News (today) to Market - 4 Year ahead**

2016	2017	2018	2019	2020	2021
Folha_Inflation-Negative	Estadao_Fiscal-Negative	Folha_Inflation-Negative	Valor_Fiscal-Negative	Estadao_Inflation-Negative	Valor_Politics-Negative
Valor_Other.topics-Negative	Estadao_Other.topics-Negative	Valor_Economic.activity-Positive	Valor_Economic.activity-Neutral	Valor_Fiscal-Neutral	Valor_Other.topics-Negative
Valor_Inflation-Neutral	Valor_Politics-Neutral	Valor_Fiscal-Negative	Estadao_Inflation-Positive	Valor_International.Economy-Positive	Valor_Politics-Positive
Valor_Inflation-Positive	Valor_Inflation-Positive	Valor_Other.topics-Positive	Estadao_International.Economy-Neutral	Valor_Fiscal-Negative	Estadao_Politics-Negative
Folha_Politics-Negative	Valor_Economic.activity-Positive	Valor_Politics-Positive	Folha_Inflation-Negative	Valor_Other.topics-Negative	Valor_Fiscal-Positive

This figure presents the top five variables selected by the Lasso model, for each year regression

Figure 3.3: **Market to News (tomorrow) - 1 Year ahead**

2016	2017	2018	2019	2020	2021
Valor_Fiscal-Negative	Estadao_Other.topics-Neutral	Valor_Fiscal-Negative	Estadao_Inflation-Negative	Estadao_Other.topics-Negative	Valor_Politics-Negative
Valor_Other.topics-Neutral	Estadao_Overall.Sentiment-Negative	Folha_Fiscal-Negative	Valor_Politics-Negative	Estadao_Fiscal-Neutral	Estadao_Politics-Negative
Valor_Politics-Negative	Estadao_Other.topics-Negative	Valor_Inflation-Neutral	Valor_International.Economy-Neutral	Estadao_Overall.Sentiment-Negative	Estadao_Fiscal-Neutral
Estadao_International.Economy-Negative	0	Estadao_Inflation-Positive	Estadao_Inflation-Neutral	Valor_Fiscal-Neutral	Folha_International.Economy-Negative
Estadao_International.Economy-Neutral	0	Folha_Inflation-Negative	Folha_International.Economy-Positive	Valor_Economic.activity-Positive	Valor_Inflation-Negative

This figure presents the top five variables selected by the Lasso model, for each year regression

Figure 3.4: **Market to News (tomorrow) - 4 Year ahead**

2016	2017	2018	2019	2020	2021
Folha_Inflation-Negative	Estadao_Fiscal-Negative	Folha_Inflation-Negative	Valor_Fiscal-Negative	Estadao_Inflation-Negative	Valor_Politics-Negative
Folha_Fiscal-Negative	Estadao_Other.topics-Negative	Valor_Economic.activity-Positive	Valor_Economic.activity-Neutral	Valor_Fiscal-Neutral	Estadao_Politics-Negative
Valor_Inflation-Neutral	Valor_Politics-Neutral	Valor_Fiscal-Negative	Estadao_International.Economy-Neutral	Valor_Fiscal-Negative	Estadao_Fiscal-Neutral
Valor_Inflation-Positive	Valor_Economic.activity-Positive	Valor_Other.topics-Positive	Estadao_Other.topics-Negative	Valor_International.Economy-Positive	Valor_Overall.Sentiment-Negative
Folha_Politics-Negative	Estadao_International.Economy-Positive	Estadao_Fiscal-Negative	Estadao_Inflation-Positive	Valor_Other.topics-Negative	Valor_Inflation-Positive

This figure presents the top five variables selected by the Lasso model, for each year regression

3.7 Conclusion

This paper presents a new approach to evaluating inflation expectations for the Brazilian economy.

The main conclusion concerns the efficiency of the Brazilian yield curve market; information is quickly priced. Little added value to use morning newspaper articles to predict the market's direction for the day.

The market assesses current topics of economic relevance, as news with more informational content for predicting the breakeven inflation premium changes over time, but keeping the main focus with economic substance paying attention on Fiscal and inflation news.

The forecast of long term breakeven premium (4 years ahead) can be improved using Machine Learning techniques, such as Lasso models combined with Natural Language processment, however it is very difficulty to trade with this information.

Understanding institutional changes in fiscal policy, taxation rules in bond issuance, and market issuances at all can help explain the persistent premium observed in breakeven inflation, a fruitful area for future research.

Bibliography

Abrahams, Michael, Adrian, Tobias, Crump, Richard K, & Moench, Emanuel. 2016. Decomposing real and nominal yield curves. *Journal of Monetary Economics*, **84**, 182–200.

Abreu Neto, Jose Carlos, & Saito, Richard. 2024. Unlocking the Power of Relationships: Limited Partner Networks and Performance in Private Equity. *Available at SSRN 4754637*.

Almeida, Heitor, Johnson, Timothy C, Oliveira, Sebastiao, & Zhou, Yucheng. 2025. The Equity Constraint Channel of Monetary Policy. *Available at SSRN 5031788*.

Araujo, Gustavo Silva, Ándres Bergallo, José Ignacio, & Val, Flávio de Freitas. 2025. Do Inflation-Linked Bonds Predict Future Inflation? a reassessment using novel methodologies and instruments. *Working Paper Series*, aug.

Bain&Company. 2024. Global Private Equity Report 2024. *Global Private Equity Report 2024*, **1**(1).

Bauer, Michael D, & Swanson, Eric T. 2023. A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, **37**(1), 87–155.

Bernanke, Ben S. 2007. *Inflation Expectations and Inflation Forecasting*. Speech delivered at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute.

Bernanke, Ben S, Gertler, Mark, & Gilchrist, Simon. 1999. The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, **1**, 1341–1393.

Beyhaghi, Mehdi, Frank, Murray Z, McLemore, Ping, & Sanati, Ali. 2024. Monetary Policy and Corporate Investment: The Equity Financing Channel. *Available at SSRN 4793393*.

Boyer, Brian H, Nadauld, Taylor D, Vorkink, Keith P, & Weisbach, Michael S. 2023. Discount-Rate Risk in Private Equity: Evidence from Secondary Market Transactions. *The Journal of Finance*, **78**(2), 835–885.

Bühlmann, Peter, & Van De Geer, Sara. 2011. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Buraschi, Andrea, & Whelan, Paul. 2022. Speculation, sentiment, and interest rates. *Management Science*, **68**(3), 2308–2329.

Caggese, Andrea, & Pérez-Orive, Ander. 2022. How stimulative are low real interest rates for intangible capital? *European Economic Review*, **142**, 103987.

Campbell, John Y., & Shiller, Robert J. 1991. Yield Spreads and Interest Rate Movements: A Bird's Eye View. *The Review of Economic Studies*, **58**(3), 495–514.

Campbell, John Y, & Viceira, Luis M. 2001. Who should buy long-term bonds? *American Economic Review*, **91**(1), 99–127.

Cloyne, James, Ferreira, Clodomiro, Froemel, Maren, & Surico, Paolo. 2023. Monetary policy, corporate finance, and investment. *Journal of the European Economic Association*, **21**(6), 2586–2634.

Costa Filho, Adonias. 2025. The role of breakevens in trend inflation in Brazil. *MONETÁRIA - Centro de Estudos em Macroeconomia e Finanças*, **45**(jun).

Da Rin, Marco, & Phalippou, Ludovic. 2017. The importance of size in private equity: Evidence from a survey of limited partners. *Journal of Financial Intermediation*, **31**, 64–76.

D'Amico, Stefania, Kim, Don H, & Wei, Min. 2018. Tips from TIPS: The informational content of Treasury Inflation-Protected Security prices. *Journal of Financial and Quantitative Analysis*, **53**(1), 24–57.

Doi, Jonas, Fernandes, Marcelo, & Nunes, Clemens V. A. 2017. Disagreement in Inflation Forecasts and Inflation Risk Premia in Brazil. *Brazilian Review of Econometrics*, **37**(1), 45–59.

Döttling, Robin, & Ratnovski, Lev. 2023. Monetary policy and intangible investment. *Journal of Monetary Economics*, **134**, 53–72.

Ehling, Paul, Gallmeyer, Michael, Heyerdahl-Larsen, Christian, & Illeditsch, Philipp. 2018. Disagreement about inflation and the yield curve. *Journal of Financial Economics*, **127**(3), 459–484.

Ewens, Michael, & Rhodes-Kropf, Matthew. 2015. Is a VC Partnership Greater than the Sum of Its Partners? *The Journal of Finance*, **70**(3), 1081–1113.

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

Fama, Eugene F., & Bliss, Robert R. 1987. The Information in Long-Maturity Forward Rates. *The American Economic Review*, **77**(4), 680–692.

Fernández Tamayo, Borja, Braun, Reiner, Lopez-de Silanes, Florencio, Phalippou, Ludovic, & Sigrist, Natalia. 2023. Limited Partners versus Unlimited Machines; Artificial Intelligence and the Performance of Private Equity Funds. *Artificial Intelligence and the Performance of Private Equity Funds (December 15, 2023)*.

Gertler, Mark, & Gilchrist, Simon. 1994. Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics*, **109**(2), 309–340.

Gilchrist, Simon, & Zakrajšek, Egon. 2012. Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, **102**(4), 1692–1720.

Gompers, Paul, & Lerner, Josh. 1999. *What drives venture capital fundraising?*

Gompers, Paul, & Lerner, Josh. 2000. Money chasing deals? The impact of fund inflows on private equity valuations. *Journal of financial economics*, **55**(2), 281–325.

Gompers, Paul, Lerner, Josh, Blair, Margaret M., & Hellmann, Thomas. 1998. What Drives Venture Capital Fundraising? *Brookings Papers on Economic Activity. Microeconomics*, **1998**, 149–204.

Gompers, Paul, Kaplan, Steven, & Mukharlyamov, Vladimir. 2016. What do private equity firms say they do? *Journal of Financial Economics*, **121**(3), 449–476.

Gompers, Paul, Gornall, Will, Kaplan, Steven, & Strebulaev, A. 2020. How do venture capitalists make decisions? *Journal of Financial Economics*, **135**(1), 169–190.

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

Harris, Robert S, Jenkinson, Tim, Kaplan, Steven N, & Stucke, Ruediger. 2023. Has persistence persisted in private equity? Evidence from buyout and venture capital funds. *Journal of Corporate Finance*, 102361.

Hochberg, Yael V, Ljungqvist, Alexander, & Lu, Yang. 2007. Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, **62**(1), 251–301.

Hochberg, Yael V, Ljungqvist, Alexander, & Vissing-Jørgensen, Annette. 2014. Informational holdup and performance persistence in venture capital. *The Review of Financial Studies*, **27**(1), 102–152.

Holmstrom, Bengt, & Tirole, Jean. 1997. Financial intermediation, loanable funds, and the real sector. *the Quarterly Journal of economics*, **112**(3), 663–691.

Jordà, Òscar. 2005. Estimation and inference of impulse responses by local projections. *American economic review*, **95**(1), 161–182.

Jungherr, Joachim, Meier, Matthias, Reinelt, Timo, & Schott, Immo. 2024. Corporate debt maturity matters for monetary policy. *International Finance Discussion Paper*.

Kaplan, Steven, & Schoar, Antoinette. 2005. Private Equity Performance: Returns, persistence, and capital flows. *The Journal of Finance*, **60**(4), 1791–1823.

Kaplan, Steven N, & Stromberg, Per. 2009. Leveraged buyouts and private equity. *Journal of economic perspectives*, **23**(1), 121–46.

Kruglikov, Nikita, & Forthun, Andreas. 2022. *Predicting Private Equity Fund Performance with Machine Learning*. M.Phil. thesis, Norwegian School of Economics.

Ladalardo, Lucas, & Medeiros, Marcelo. 2022. *From Micro to Macro: Essays in Textual Analysis*.

Laurer, Moritz, Van Atteveldt, Wouter, Casas, Andreu, & Welbers, Kasper. 2023. Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis*, June, 1–33.

Lerner, Josh, & Schoar, Antoinette. 2004. The Illiquidity Puzzle: theory and evidence from private equity. *Journal of Financial Economics*, **72**(1), 3–40.

Li, Ziang. 2025. Long rates, life insurers, and credit spreads. *Life Insurers, and Credit Spreads (April 13, 2025)*.

Litterman, Robert, & Scheinkman, José. 1991. Common factors affecting bond returns. *Journal of Fixed Income*, **1**(1), 54–61.

Ljungqvist, Alexander, Richardson, Matthew, & Wolfenzon, Daniel. 2020. The investment behavior of buyout funds: Theory and evidence. *Financial Management*, **49**(1), 3–32.

Lundberg, Scott M, & Lee, Su-In. 2017. A unified approach to interpreting model predictions. *Pages 4765–4774 of: Advances in neural information processing systems (NIPS 2017)*. Curran Associates, Inc.

Ma, Yueran, & Zimmermann, Kaspar. 2023. *Monetary policy and innovation*. Tech. rept. National Bureau of Economic Research.

Maurin, Vincent, Robinson, David T, & Strömberg, Per. 2023. A theory of liquidity in private equity. *Management science*, **69**(10), 5740–5771.

McKinsey&Company. 2024. Private markets in a slower era. *McKinsey Global Private Markets Review 2024*, **1**(1).

Medeiros, Marcelo C, Vasconcelos, Gabriel FR, Veiga, Álvaro, & Zilberman, Eduardo. 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, **39**(1), 98–119.

Molavi, Pooya, Tahbaz-Salehi, Alireza, & Vedolin, Andrea. 2025 (aug). *Expectations and the Term Structure of Interest Rates*. Working Paper. Northwestern University, Kellogg School of Management.

Morales, Marisa Hillary, & Tiozzo, Vittorio. 2019. *Machine Learning and Survival Prediction Models for Decision Support in the Private Equity Market*. Ph.D. thesis, Politecnico di Torino.

Nakamura, Emi, & Steinsson, Jón. 2018. High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, **133**(3), 1283–1330.

Nanda, Ramana, & Rhodes-Kropf, Matthew. 2013. Investment cycles and startup innovation. *Journal of financial economics*, **110**(2), 403–418.

Nelson, Charles R., & Siegel, Andrew F. 1987. Parsimonious modeling of yield curves. *Journal of Business*, **60**(4), 473–489.

Ottonello, Pablo, & Winberry, Thomas. 2020. Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, **88**(6), 2473–2502.

Perez-Orive, Ander, Timmer, Yannick, & van der Ghote, Alejandro. 2024. Monetary Policy Under Multiple Financing Constraints.

Pflueger, Carolin E, & Viceira, Luis M. 2011. Inflation-indexed bonds and the expectations hypothesis. *Annual Review of Financial Economics*, **3**(1), 139–158.

Powers, David MW. 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Ritter, Jay R. 1984. The" hot issue" market of 1980. *Journal of business*, 215–240.

Robinson, David T, & Sensoy, Berk A. 2016. Cyclicality, performance measurement, and cash flow liquidity in private equity. *Journal of Financial Economics*, **122**(3), 521–543.

Santos, Lucas L, Bianchi, Reinaldo AC, & Costa, Anna HR. 2023. FinBERT-PT-BR: Análise de Sentimentos de Textos em Português do Mercado Financeiro. *Pages 144–155 of: Anais do II Brazilian Workshop on Artificial Intelligence in Finance*. SBC.

Sensoy, Berk A, Wang, Yingdi, & Weisbach, Michael S. 2014. Limited partner performance and the maturing of the private equity industry. *Journal of Financial Economics*, **112**(3), 320–343.

Shapley, L. S. 1953. A value for n-person games. *Contributions to the Theory of Games, II*, **28**(5), 307–317.

Svensson, Lars E. O. 1995. Estimating forward interest rates with the Extended Nelson and Siegel method. *Sveriges Riksbank Quarterly Review*, **1995**(3), 13–26.

Swensen, David F. 2009. *Pioneering portfolio management: An unconventional approach to institutional investment, fully revised and updated*. Simon and Schuster.

Val, Flávio de Freitas, & Araujo, Gustavo Silva. 2019. Breakeven Inflation Rate Estimation: an alternative approach considering indexation lag and seasonality. *Working Paper Series - Banco Central do Brasil*.

Vicente, João Victor M., & Guillen, Osvaldo T. C. 2013. Do inflation-linked bonds contain information about future inflation? *Revista Brasileira de Economia*, **67**, 251–260.

Xiong, Wei, & Yan, Hongjun. 2009. Heterogeneous expectations and bond markets. *The Review of Financial Studies*, **23**(4), 1433–1466.