THE UNIVERSITY OF CHICAGO

ESSAYS IN FINANCIAL MARKETS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS
AND
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
EHSAN AZARMSA

CHICAGO, ILLINOIS
JUNE 2022

In memory of Yiran Fan, a truly smart and kind friend.

Contents

List of Figures

List of Tables

Acknowledgments

University of Chicago, Department of Economics and Booth Business School. I am extremely grateful to my Ph.D. committee members Will Cong, Lars P. Hansen, Ralph S.J. Koijen, Pietro Veronesi, and Constantine Yannelis for detailed feedback and suggestions, and to Vera Ye and Zhi Zhang for exceptional research assistance. I also thank Chuck Boyer, Huaizhi Chen (Discussant), George Constantinides, Ramona Dagostino, Douglas Diamond, Mohammad Ghaderi, Niels Gormsen, Zhiguo He, Jian Li, Lisa Liu, Simon Mayer, Stefan Nagel, Scott Nelson, Yosh Nozowa (Discussant), Lubos Pastor, Kelly Posenau, Nagpurnanand Prabhala (Discussant), Peyman Shahidi, Amir Sufi, Quentin Vandeweyer, Michael Weber, and seminar participants at Chicago Booth and the Office of Financial Research PhD Symposium (2021) for helpful comments and discussions. This research was funded in part by the John and Serena Liew Fellowship Fund at the Fama-Miller Center for Research in Finance, University of Chicago Booth School of Business.

Abstract

Chapter 1. In the United States, building infrastructure is primarily the responsibility of municipal governments. However, prior empirical evidence suggests these governments are borrowing-constrained. This paper provides new evidence and theory that link the constraint to the dominance of retail investors in the municipal bond market, who pay less attention to new bond issues than more specialized investors, such as municipal mutual funds. Supporting this hypothesis, I find that the mutual funds disproportionately buy newly issued bonds and gradually resell them to other investors. Furthermore, a 1% inflow to the mutual fund sector increases bond issuance by county governments by 0.2% and reduces the interest rate by 0.2 basis points in the next quarter. To rationalize these observations, I develop a dynamic model featuring end investors who exhibit sluggish portfolio adjustments and invest in bonds directly or indirectly through some attentive mutual funds. By calibrating the model with the empirical estimates, I find that the elasticity of bond demand is at least one order of magnitude smaller in the short run than in the long run, suggesting that the municipal bond market is not resilient against shocks in the short run. This finding supports market interventions by the federal government in times of crisis, especially when they accompany massive outflows from municipal mutual funds.

Chapter 2. After initial investments, relationship financiers typically observe interim information about projects before continuing financing them. Meanwhile, entrepreneurs produce information endogenously and issue securities to incumbent insider and competitive outsider investors. In such persuasion games with differentially informed receivers and contingent transfers, entrepreneurs' endogenous experimentation reduces insiders' information monopoly but impedes relationship formation through an "information production hold-up." Insiders' information production and interim competition mitigate this hold-up, and jointly

explain empirical links between competition and relationship lending. Optimal contracts restore first-best outcomes using convertible securities for insiders and residuals for outsiders. Our findings are robust under various extensions and alternative specifications.

Chapter 3. I examine how differences in the ability to identify profitable investment opportunities contribute to wealth inequality. I analyze a model of financial markets with investors heterogeneously informed about future returns. The unconditional wealth share distribution features a thick right-tail populated by the best-informed investors, explaining the prominent representation of high-profile investors among the super-rich. Wealth inequality increases with the cost of information acquisition and market liquidity. It is non-monotone in public information precision and the size of investments delegated to the best-informed investors.

CHAPTER 1

FINANCING INFRASTRUCTURE WITH INATTENTIVE INVESTORS: THE CASE OF US MUNICIPAL GOVERNMENTS

1.1 Introduction

Education, healthcare, road maintenance and construction, utility, and police protection are just part of the services US households receive daily from their state and local governments. Most of these services require some initial investment in infrastructure. Available estimates indicate that between 2007 and 2016, 72% of investments in the US national infrastructure were financed by the bonds issued by the municipal governments (Cestau, Hollifield, Li, and Schürhoff, 2019). Furthermore, state and local governments are in the frontline of responding to natural disasters and pandemics, which makes their timely access to credit crucial for the recovery of their local economy.

Prior empirical evidence suggests that municipal governments are credit-constrained. This is evidenced by their borrowing and expenditures being highly sensitive to their investors' lending capacity (Dagostino, 2018; Yi, 2020), their credit rating (Adelino, Cunha, and Ferreira, 2017), and the backing of their bond insurers (Agrawal and Kim, 2021). This is despite the fact that they possess statutory taxing power and their historical default rate is low.

I add to this discussion by putting forward a novel theory- supported by new empirical evidence- that sheds light on the nature of the impediment municipal governments face in their borrowing. I argue that the impediment is related to the dominance of retail investors in the municipal bond market. The household sector is the largest direct owner of municipal bonds, holding 44% of the outstanding amount. The corresponding number is 6% for the US Treasuries, and 7% for corporate and foreign bonds (See Figure 1.A.1). Municipal bonds are attractive for retail investors, especially high net-worth ones, since the interest incomes are usually exempt from federal and state taxes. However, it drives unqualified and tax-exempt investors out of the market (e.g., pension funds and foreign investors).[1] Retail investors are typically characterized as buy-and-hold investors that might not monitor the new bond issues as closely as more specialized intermediaries, such as municipal mutual funds. This means that newly issued bonds should be initially purchased by the mutual funds and other more attentive investors, before gradually reselling the bonds to retail and other less attentive investors.

This is what Figure 1.1 displays about the mutual fund ownership of municipal bonds throughout their lifetime. We see that the mutual fund ownership is initially high, and the funds gradually resell the bonds to other investors. This observation is robust with respect to the choice of the bonds' initial maturity, rating, and state of origination (See Figures 1.A.3, 1.A.4, and 1.A.5). Note that the ownership transfer happens at a slow pace, as the reselling takes more than ten years to complete on average. This is in line with the hypothesis that mutual funds monitor new bond issues more closely than an average investor in the market.[2]

This hypothesis implies that the debt capacity of municipal governments should depend on the capital available to mutual funds because they should initially hold a substantial

1. The tax-benefit of municipal bonds pushes up their price. It lowers their expected return compared to other taxable fixed-income securities for institutional or foreign investors who do not benefit from or qualify for the tax exemption. It makes municipal bonds less attractive for those groups of investors.

2. For corporate bonds, data suggest a similar pattern, as displayed by Figure 1.A.6. However, the transition from mutual funds to other investors is faster for long-term bonds (about one year), while it is slower for medium-term and short-term bonds.

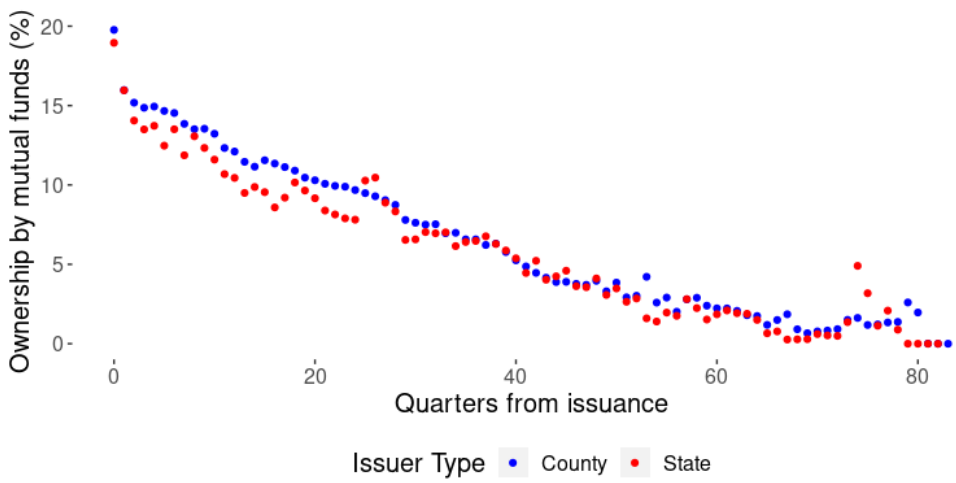


Figure 1.1: Mutual funds' ownership of municipal bonds by quarters after issuance. The figure depicts the average percentage of mutual fund ownership for each quarter after issuance for the period between 2009Q1-2021Q1, and municipal bonds that were issued in 2000 and after. The historical bond issuance data are obtained from Bloomberg for US state governments, and county governments with at least 100K population and their subsidiaries, upon availability. Mutual funds' holding data of municipal bonds are obtained from the Center for Research in Security Prices (CRSP) database.

portion of the issued bonds. To test this hypothesis, I estimate how capital flows in and out of mutual funds impact the timing, size, and interest rate of borrowing by municipal governments.

To obtain the estimates, I collect the historical debt issuance data of 262 largest and most frequently issuing county governments from Bloomberg. I identify their mutual fund owners from CRSP[3] between 2009 and 2019 and examine their trading behavior. The data reveal that the mutual funds respond to inflows and outflows in a predictable manner: In response to inflows, they are more likely to increase their investment in governments whose ownership has been relatively high over the past three years.[4]

3. The Center for Research in Security Prices Database

4. I define the ownership of a fund as its share in the total outstanding debt of the government that is owned by mutual funds. Section 1.3 provides the details.

To gauge the impact of the fund flows on the governments' borrowing behavior, I construct a measure of "flow-induced demand" for each county government in my sample, inspired by the flow-induced trade measure utilized in earlier research (e.g., Lou (2012); Li (2021)). Intuitively, the flow-induced demand is an instrument for the bond demand a municipal government faces from the mutual fund sector. In my empirical analysis, I examine how the flow-induced demand affects the timing, size, and interest rate of the bond issues by the county governments in my sample. In this estimation, the key identification assumption is that the fund flows are not correlated with the county governments' funding needs. The assumption is plausible since the bonds issued by the county governments in my sample, altogether, comprise 5.9% of the mutual funds' assets. The maximum exposure to a single county government is 0.8% on average and less than 2.1% for 90% of the observations. Moreover, the cross-section of the flows cannot be predicted by the previous fund returns and flows. Therefore, it is unlikely that the cross-section of the fund flows is connected to the funding need of the county governments. This is the exclusion restriction that I exploit in my empirical analysis.

I find that in response to a 1% point increase in the flow-induced demand for a county government's bonds, the government borrows 0.224% more in the next quarter, provided a new bond has been issued. The impact on the size of borrowing goes down over the subsequent quarters, reaching 0.088% in the fourth quarter. The evidence that the fund flows can predict the timing of the bond issues is mixed, which could reflect the procedural challenges that municipal governments encounter to receive authorization for new bond issuance. For instance, most local governments need to hold public elections and obtain supermajority approval from their residents to issue new general obligation bonds. However, once the authorization is granted, the issue size is determined in negotiations with an underwriter or an underwriting syndicate in most municipal bond issues (known as negotiated bond sales). Since the underwriters have a limited risk capacity and need to sell the bonds quickly, their perception of the bond demand becomes crucial for the issue size agreed upon.

There are some potential concerns in the estimations that I address in a set of robustness checks. A potential concern is that fund flows are part of individuals' optimal portfolio decisions, so they could be correlated with other economic variables that impact the governments' borrowing decisions. For instance, investment in mutual funds might be affected by changes in the tax rates, which affect municipal governments' revenue. I address this concern by adding state-by-year fixed effects to absorb all state-level shocks at the annual level. Moreover, fluctuations in the wealth and income of the residents of a municipality could impact their investment in the mutual funds, and the tax revenue of the governments serving the municipality. I address this concern by controlling for fluctuations in the local income and house prices. It lowers the coefficient of interest from 0.224% to 0.196%. However, the effect is still significant at the 1% confidence level. Furthermore, the result is robust even after directly controlling for the recent changes in the governments' revenue, expenditure, and liability size.

Another concern is that the empirical observations are connected to the tax treatment of municipal bonds and its impact on investors' behavior. In most states, the interest incomes on municipal bonds issued in that state are exempt from state taxes for the state residents. However, the state residents would need to pay state taxes on the interest incomes earned on municipal bonds issued outside the state. Such differential tax treatments create market segmentation along the states' border lines (Schultz, 2013). This complementary hypothesis has two implications, which I test. First, most mutual funds are only active in a single state to produce tax-free income for their investors (Babina, Jotikasthira, Lundblad, and Ramadorai, 2021). It implies that the aggregate fund flows at the state level should absorb most of the fluctuations in the size of municipal borrowing. However, accounting for the state-level flows reduces the coefficient of interest from 0.224% to 0.180%, and is still significant at the 5% level. Second, the flow-induced demand should be most impactful in states with the highest degree of market segmentation, such as California or New York, where the in-state bond ownership is more concentrated due to their high top marginal tax rates. I test this hypothesis by interacting the flow-induced demand with the amount of tax-saving associated with in-state investing for the investors at the highest tax bracket. The results do not support the hypothesis, and lend further credence to the proposed mechanism in this paper.

Do the fund flows impact the borrowing interest rates as well? I find that the impact is significant, although two orders of magnitudes smaller than its impact on the borrowing quantity. It implies that in response to demand shocks, municipal governments mostly adjust the quantity of their borrowing instead of offering a higher interest rate to attract more investors. Yi (2020) also documents a similar effect on the borrowing quantities and interest rates. She finds that in response to a regulation that curtailed the lending ability of some banks to their municipal borrowers, the most affected municipalities cut their borrowing by 20% more than the least affected ones, while their yield at issuance only increased by ten basis points more than the least affected group. These observations are at odds with standard theory, as it implies that large interest rate movements should accompany large quantity adjustments.

To explain the puzzling observations, I develop a dynamic model of the municipal bond market that captures the essential elements impacting the demand and supply of the bonds. On the supply side, a representative municipal government issues bonds. The municipal government faces a convex cost of bond issuance, capturing the statutory borrowing limitations.

The bonds are purchased by two groups of investors: *Direct investors*, who buy the bonds directly, and *indirect investors*, who buy the shares of some mutual funds that invest in a mix of the municipal bonds and a risk-free asset. Specifically, the mutual funds are modeled as intermediaries that could face mandates in their investments.

The key behavioral assumption of the model is that both direct and indirect investors rebalance their portfolios intermittently. It is consistent with earlier empirical observations (See, e.g., Giglio, Maggiori, Stroebel, and Utkus (2021)), and could be motivated by investor inattention. The intermittent rebalancing creates two frictions that cause the capital flows in and out of the mutual funds to affect the government's borrowing size. First, the direct investors do not immediately adjust their investment to absorb the buying or selling pressure generated by the fund flows. Second, the fund investors (the indirect investors) are slow in adjusting their allocations to "undo" the excess buying or selling induced by the investment mandates the funds face. Overall, both frictions cause the demand elasticity to be low in the short run, which is crucial to explain the large short-term quantity response I find in my empirical analysis.

Next, I calibrate the model to inform discussions about what limits the supply of infrastructure in the US. The quality of infrastructure in many US states is far from ideal: More than 30% of the residents in Alaska and Arkansas do not have access to high-speed Internet.[5] Close to 20% of the bridges in West Virginia and Rhode Island are in poor condition.[6] The water quality is low in many areas.[7] The question is what limits the financing of infrastructure projects in the US. Is the binding element the constitutional limitations that municipal

5. Best and Worst States for Internet Coverage, Prices and Speeds, September 2021

6. Bridge Condition by Highway System, December 2020

7. Millions of Americans drink potentially unsafe tap water, Science 2018

governments face in their borrowing? Or, is accessing credit difficult by these governments?

The calibration results imply that in the long run, the bond supply is substantially less elastic than bond demand. Still, the demand elasticity is quite low in the short run. It is concerning regarding the depth of the market, especially when municipalities need to expand their borrowing to respond to natural disasters or other crisis events. Particularly, I find that the elasticity of bond demand for the county governments in my sample is 11.6 in the short run and 158.4 in the long run. Therefore, the demand elasticity in the short run is one order of magnitude smaller than that in the long run.

Overall, the results suggest that the municipal bond market is not resilient, and its effectiveness depends on the capital available to mutual funds. Therefore, market interventions by the federal government, such as introducing emergency municipal liquidity programs (Li and Momin, 2020; Haughwout, Hyman, and Shachar, 2021), are necessary in times of crisis, especially when mutual funds experience massive outflows. This was the situation in the first few months following the outset of the COVID-19 pandemic. In fact, I find that an outflow of 5% from municipal mutual funds, which is comparable with the outflow amount in March and April 2020, reduces the bond issuance by municipal governments by 10.5 billion dollars (0.25% of the outstanding municipal debt). The estimate explains 46% of the 23 billion dollar reduction in the bond issuance in March and April 2020, compared to January and February 2020.

This paper is at the intersection of three burgeoning research areas: Public finance and municipal bond market, investor inattention, and demand-based asset pricing.

This paper adds to the growing literature of local and state government finances by putting forward novel evidence and theory that shed light on the determinants of municipal governments' borrowing capacity. Adelino, Cunha, and Ferreira (2017) and Dagostino (2018) find that credit shocks to municipal governments significantly impact their expenditure, through which they affect the local economic condition. Yi (2020) investigates the impact of a banking regulation that disrupted access to credit by the municipal governments that relied on bank lending. The previous studies chiefly examined the effect of large credit shocks on municipal borrowing. However, my empirical results indicate that even small credit shocks, such as mutual fund flows, are not promptly absorbed in the municipal bond market. Furthermore, my theoretical analysis demonstrates that the limited borrowing capacity can be explained by the sluggishness of the households' portfolio adjustments, as documented by Giglio, Maggiori, Stroebel, and Utkus (2021).

The model developed in this paper can be employed to analyze how the price and quantity of municipal bonds respond to different types of supply and demand shocks. Thus, it complements the earlier models of the municipal bond market. Myers (2019) examines the portfolio choice problem of pension funds, and how it interacts with the borrowing and default decision of the municipal government supervising the funds. Boyer (2020) presents a model of municipal bond pricing in which the bonds could have an arbitrary level of seniority compared to pension liabilities. My model complements the earlier studies by highlighting the role of mutual fund flows and investor inattention in understanding the demand for municipal bonds.

This paper also contributes to the literature of demand-based asset pricing by investigating the impact of demand shocks on the real economy, and by exploring the role of investors' infrequent rebalancing as a candidate contributing to the low demand elasticity observed for risky assets (Shleifer, 1986; Wurgler and Zhuravskaya, 2002; Koijen and Yogo, 2019; Gabaix and Koijen, 2020). Numerous studies in the literature exploit fund flows to identify demand shocks and examine their impact on asset prices (Coval and Stafford, 2007; Lou, 2012; Peng and Wang, 2019; Li, 2021). I borrow their methodology to study the effect of demand shocks on the quantity of borrowing by municipal governments.[8]

Lastly, I contribute to the literature of investor inattention by putting forward a model in which the intermediary sector creates value by paying attention to the market condition and adjusting its clients' portfolios accordingly. Abel, Eberly, and Panageas (2013) demonstrate that informational costs cause sluggishness in portfolio adjustments, a feature that appears in my model. Chien, Cole, and Lustig (2012, 2016) study a general equilibrium model in which some investors rebalance their portfolios infrequently. Moreover, in a companion paper (Azarmsa, 2021), I study an asset pricing model with heterogeneously attentive investors, and I find that the model can explain the downward term structure of the equity premium, documented by Van Binsbergen, Hueskes, Koijen, and Vrugt (2013); Van Binsbergen and Koijen (2017); Gormsen and Lazarus (2019). In an appendix, Gabaix and Koijen (2020) examine the effect of fund flows on the equilibrium prices in the presence of portfolio inertia, a feature I employ to model investor inattention. Heterogeneity in attention is also featured in Duffie (2010); however, the asset quantity is also endogenized in my model, which drastically alters the asset pricing implications.

The rest of the paper is organized as follows. Section 1.2 describes the data used for this study. Section 1.3 introduces the flow-induced demand measure, and empirically examines its impact on the size, timing, and interest rate of the bond issues for a sample of county governments and their subsidiaries. Section 1.4 presents a dynamic model of the municipal bond market. I use the empirical estimates in Section 1.3 to calibrate the model parameters and estimate the elasticities of supply and demand for municipal bonds. Section 1.5 concludes. Appendix A, B, and C contain the supplementary figures, proofs, and discussions

8. Relatedly, Lou and Wang (2018) study the impact of mutual fund flows on corporate investment.

respectively.

1.2 Data

1.2.1 Debt issuance data

I collected the historical debt issuance data of all state governments and county governments with a population of at least 100,000 (as of 2012) from Bloomberg, along with their subsidiaries, upon availability. For each issued bond, the data contain the 9-digit CUSIP, bond purpose, maturity size, deal size, issuance date, maturity date, refunded date (if applicable), coupon rate, insured status, tax provision, maturity type (indicates whether the bond is callable or not), and S&P credit rating. A desirable feature of Bloomberg data is that it also lists issues from subsidiary agencies that are morally backed by the government's credit.[9]

Overall, the data contain CUSIP-level information for 157,991 bonds issued by 43 state governments and their subsidiaries,[10] and 246,870 bonds issued by 417 county governments and their subsidiaries for the period between 1950 and 2020.

The bond market is generally considered segmented across the state lines since interest incomes on municipal bonds are typically exempt from state taxes for the in-state investors, but not exempt for the out-of-state investors (Schultz, 2013). To reap the tax benefit, the majority of municipal mutual funds, both in numbers and assets under management, are only active in one state (Babina, Jotikasthira, Lundblad, and Ramadorai, 2021). Since any

9. Bloomberg System provides all bonds issued by the "credit family" of an issuer. The credit family of an issuer refers to the issuer's debt as well as the debt of its subsidiaries which it guarantees but does not have further recourse from any parent, direct or indirect, of the reference issuer.

10. For seven states (Indiana, Kansas, Kentucky, North Dakota, Oklahoma, South Dakota, and Wyoming), the issuance data of the state government and its subsidiaries were not available.

inflow to single-state funds should be invested within their state, it is difficult to identify the impact of fund flows on debt issuance decisions for state-level governments; they could merely reflect fluctuations in the overall bond demand of the state's residents, which is affected by numerous state-level and national-level variables, such as tax rates and interest rates.

For identification, I exploit heterogeneities across county governments, the largest sub-state governments, in their exposure to mutual funds. As we see later, mutual funds do not invest equally in all counties of their target states. Thus, the relative flows determine which counties relatively receive more demand. I will expand on the identification strategy in Section 1.3.

Bond issuance by county governments is sparse. The median county government in the sample issued bonds only in 7 out of the 44 quarters (15.9%) between 2009 and 2019. It points to the challenges that municipal governments face in their debt issuance. First, most municipal governments are not allowed to issue long-term bonds to fill their budget deficits when revenue is low, which drastically limits the scope of municipal bonds. Municipalities issue bonds to finance long-term projects, primarily to strengthen their local infrastructure. Second, many municipal governments have to seek voter approval to pledge their full faith and credit for a new bond issue, which are known as general obligation bonds. For instance, local governments in California are required to hold elections and obtain the approval of two-thirds of their residents.[11] However, the authorization rule varies across states. For instance, counties in the state of New York only need the super-majority approval from the members of their county legislature, who are allowed to delegate their authorization power to the county's chief fiscal officer (Myers and Goodfriend, 2009). An alternative for municipalities is to issue revenue bonds, for which they only pledge the revenue from the project being

11. California Constitution Article XVI, sec. 18

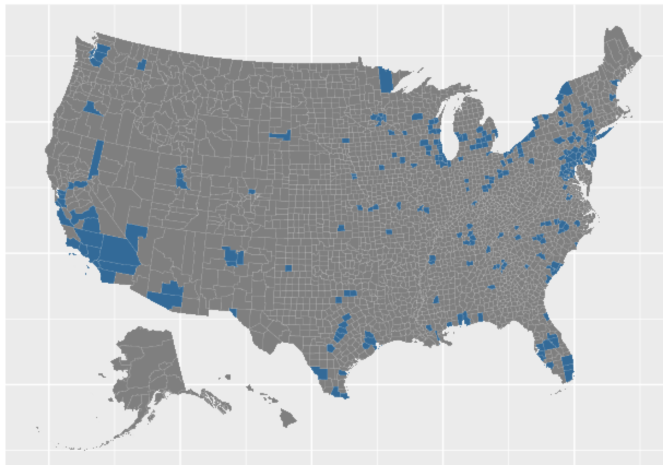
financed, thus the regulatory restrictions are laxer. Third, municipalities need to incur costs to hire financial and legal advisors to prepare the documentation needed for the bond sale, which could be restrictive for smaller municipalities.

Therefore, I only include governments that exhibit a frequent record of bond issuance. Particularly, I exclude governments that have issued in four years or less during my sample period of 2009 to 2019. This leaves me with 262 counties, which I henceforth refer to as the "selected counties." The selected counties are depicted in Figure 1.2a. Figure 1.2b presents their distribution across the states.

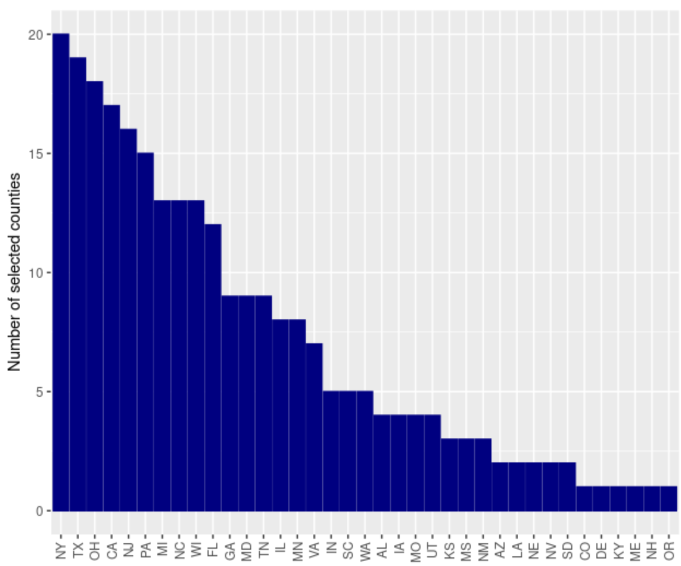
Table 1.1 provides the summary statistics of the issuance data for all and the selected counties. On average, a selected county government issues bonds roughly in one out of four consecutive quarters. The average amount of bonds issued in a quarter is 82.3 million dollars conditional on some issuance takes place, and $23.0(= 27.9\% \times 82.3)$ million dollars, unconditionally. The average size of issuance in a quarter as a fraction of the total outstanding debt is 19.4%. The average (median) borrowing interest rate is 2.4% (2.3%). The majority of the bonds are long-term, as the median bond has 8.4 years to mature at its origination.

In my sample, only 7.3% of the bonds are insured. It reflects the fact that the percentage of insured municipal bonds dropped drastically after the financial crisis, due to the sharp decline in the credit rating of major municipal bond insurers.[12] Lastly, we see that more than 90% of the bonds issued by the selected county governments are exempt from federal taxes.

12. The Fundamentals of Municipal bonds, Sixth Edition, Page 86.



(a) Selected counties (in blue)



(b) Distribution of the selected counties

Figure 1.2: The distribution of the selected counties. Figure (a) displays the selected counties on a US map. Figure (b) presents the distribution of the counties across the states. The selected counties are the ones that i) had a population of at least 100,000 in 2012, ii) their issuance data are available at Bloomberg system, iii) they and their subsidiaries combined have issued debt at least in five out of the eleven years between 2009 and 2019.

	All counties (417) (66364 bonds)					Selected counties (262) (58952 bonds)				
	Mean	SD	10th	50th	90th	Mean	SD	10th	50th	90th
Maturity size ($M)	4.3	18.0	0.2	1.4	7.7	4.6	18.9	0.2	1.5	8.2
Deal size ($M)	57.1	113.8	3.3	23.0	140.7	59.9	11.6	3.4	24.6	150
Overall issuance in a quarter ($M)	77.0	143.5	5.0	30.4	179.4	82.3	148.3	5.0	33.7	194.2
Overall issuance in a quarter (% of outstanding debt)	21.0	30.0	2.8	13.5	43.4	19.4	29.6	2.6	12.6	38.7
Outstanding debt ($M)	426.9	1146.2	19.9	124.6	900.9	564.3	1327.4	37.7	195.0	1414.5
Quarters with issuance (Percentage)	8.9 (20.1%)	6.8 (15.4%)	2 (4.5%)	7 (15.9%)	18 (40.9%)	12.3 (27.9%)	6.3 (14.2%)	6 (13.6%)	11 (25%)	21 (47.7%)
Yield at issuance (%)	2.4	1.2	0.9	2.3	3.9	2.4	1.2	0.9	2.3	3.9
Years to maturity	9.3	6.2	2	8.4	18	9.3	6.2	2	8.4	18
Coupon	3.7	1.2	2	4	5	3.7	1.2	2	4	5
Insured ($Y = 1$)	0.077	0.268	0	0	0	0.073	0.260	0	0	0
Federally taxable ($Y = 1$)	0.099	0.299	0	0	0	0.101	0.301	0	0	1

Table 1.1: Summary statistics of the bond issues by county governments. This table provides the summary statistics for the county governments examined in this study for the period between 2009 and 2019. The historical data of the bond issues are obtained from Bloomberg for counties with at least 100,000 population in 2012. The data on the outstanding debt are obtained from Municipal Atlas database. I select counties with a positive bond issuance in at least five out of eleven years of the study. The summary statistics for the selected counties are presented in the right panel. Maturity size is the par value of a bond CUSIP. Deal size is the total par value of the bonds in an issue.

1.2.2 Municipal bond holding by mutual funds

I obtain the holding data of municipal mutual funds from the Center for Research in Security Prices database (CRSP) for the years between 2000 and 2019. Since the data coverage is less comprehensive before 2009, I primarily work with the quarter-end holding data of the funds between 2009 and 2019.

Table 1.2 provides information about the universe of the mutual funds examined in this study. The table reveals that the funds hold small cash buffers and invest almost entirely in municipal bonds. The average share of cash and municipal bonds in the funds' portfolio are 0.9% and 98.1%, respectively. This observation suggests that mutual funds respond to inflows (outflows) mostly by purchasing new bonds (liquidating some of their holdings). At the end of 2019, the last year of the sample, the total value of municipal bonds held by the mutual funds was 830.9 million dollars, which almost matches the number reported in the US Financial Accounts database (831.0 million dollars). The coverage is more than 94% for the other years of the sample. The only exception is 2009, for which CRSP does not provide the overall values of municipal bond holding. The last column reports the funds' market share. We see that their market share has increased drastically over the years of the sample, indicating their increasing importance in the municipal bond market.

Year	# of funds	TNA ($ Million)		Cash holding	Municipal bond holding			
		Mean	Median	% of all assets	Total ($ Billion)	% of all assets	% of all owned by mutual funds	% of all outstanding
2009	1799	262.7	35.5	Not Avail.	Not Avail.	Not Avail.	Not Avail.	Not Avail.
2010	1804	267.6	36.8	0.4	476.9	98.8	97.3	12.5
2011	1745	291.1	42.1	1.3	498.8	98.3	96.1	12.4
2012	1733	342.8	51.7	0.9	582.8	98.2	95.7	14.1
2013	1769	289.5	40.8	0.9	501.9	98.2	94.7	13.0
2014	1780	325.8	45.3	1.2	566.6	97.9	96.5	14.0
2015	1820	337.3	47.5	1.5	593.4	97.3	98.2	14.6
2016	1822	350.1	50.0	0.8	624.9	98.4	98.9	15.4
2017	1905	365.7	45.9	1.0	678.4	98.0	98.6	16.4
2018	1892	380.0	47.8	0.7	691.8	98.0	99.6	17.3
2019	1841	463.4	62.0	0.6	830.9	98.3	99.9	20.2

Table 1.2: Summary statistics for municipal mutual funds. This table provides information about the size of overall assets, cash holding, and municipal bond holding for municipal mutual funds studied between 2009 and 2019. The data are obtained from the database of the Center for Research in Security Prices (CRSP). The first column reports the number of funds studied in each year. The second and third columns provide the average and median size of the funds. The average cash holding of the funds is reported in the forth column. For 2010 and after, CRSP provides the total value of municipal bond holdings for each fund, which are the input to compute the numbers in the last four columns. In the last two columns, I compare the total value of municipal bonds held by the funds with the overall numbers provided by the financial accounts of the united states database available at the Federal Reserve board of governors website.

The mutual fund holdings dataset also provides the CUSIP of every holding, which enables me to link it to the bond issuance data. Since each share represents a fixed amount of investment at par value, one could infer the total par value of each holding from the corresponding number of shares. Moreover, the data provide the total net asset value (TNA), return, and merger information of each fund at monthly frequency, which I employ to construct a quarterly measure of flow for every fund.

Table 1.3 presents some information about the funds' trading behavior (Panel A), their exposure to the selected counties (Panel B), and their portfolio composition by the issuance category (Panel C). Panel A reveals that the funds, on average, make no adjustments in 88.2% of their positions over a quarter. Furthermore, they liquidate 7.1% of their positions over this period. Therefore, less than 5% of their positions are adjusted to a larger or smaller non-zero investment. It suggests that municipal funds, once they start a new position, they make none to few adjustments before the liquidation.

Panel B shows that the median fund invests at most 0.34% of its assets in a selected county government. The median exposure to the selected county governments is 4.0%. Therefore, most mutual funds have little exposure to the county governments, which makes it unlikely that households use the mutual funds to channel fundings to their county governments. This observation is important for understanding the empirical strategy, which I describe in Section 1.3.

Panel C presents a summary of the portfolio composition of the mutual funds. To obtain the issuer category, I link each CUSIP to its issuer with Muni Atlas dataset, which contains information about the type and sector of each bond's issuer. The dataset provides a match for 230913 out of 278666 bond CUSIPs held by the funds (82.9%). We see that bonds directly issued by state and local governments constitute only 18.5% of the funds' bond holding. The rest includes bonds issued by independent special district governments, such as school districts, and state and local governments' subsidiaries, such as enterprise funds.

1.2.3 Fund flows

Following prior studies (e.g., Lou (2012)), I construct quarterly investment flows to fund f at quarter t based on the formula below, by using the information CRSP provides on the return, total net asset value, and merger status of each fund.

$$Flow_{f,t} = \frac{TNA_{f,t} - TNA^{adj}_{f,t-1}(1 + Ret_{f,t})}{TNA^{adj}_{f,t-1}} \tag{1.1}$$

In (1.1), $TNA_{f,t}$ is the total net asset value of fund f at the end of quarter t. $TNA^{adj}_{f,t}$ accounts for the fund mergers between quarter-ends $t-1$ and t. Since the exact date of the mergers are not available, I assume that the target fund's assets are acquired at the first day after the date of the latest available TNA for the target fund. $Ret_{f,t}$ is the return of fund f between quarter-ends $t-1$ and t.

1.3 Flow-induced demand and government debt issuance

In this section, I examine the impact of mutual fund flows on the borrowing behavior of the selected county governments. To this end, I follow an empirical strategy that exploits biases in the trading behavior of the mutual funds. For instance, when a fund receives outflows, it is more likely to sell from its existing positions, rather than short-selling other bonds, due to short-sale constraints. Therefore, the county governments that are exposed to the fund face a higher selling pressure compared to the unexposed governments. Such biases cause the trading behavior of the funds to be partly predictable, which we can exploit to

	Mean	SD	10th	50th	90th
Panel A: Trading behavior					
Number of holdings at quarter-ends	299.8	458.5	60	164	663
Percentage of no trade (%)	88.2	10.1	78.4	90.6	96.1
Fraction liquidated (%)	7.1	6.6	1.7	5.4	14.3
Fraction of new holdings (%)	8.7	11.9	1.7	7.1	16.9
Panel B: Exposure to the Selected Counties					
Exposure to a selected county government (%)	0.8	1.6	0.0	0.34	2.14
Overall exposure to the selected county governments (%)	5.9	7.1	1.4	4.0	11.7
Panel C: Portfolio Composition					
State governments (%)	7.5	7.6	0.2	5.5	17.2
County governments (%)	4.0	5.8	0.0	2.3	9.3%
City governments (%)	5.8	4.6	1.1	4.7	11.7
By sector					
State and local governments (%)	18.5	11.6	6.5	16.2	33.7
Utility (%)	10.4	6.8	2.6	9.4	19.1
Transportation (%)	9.3	6.2	1.8	8.5	17.6
Health (%)	11.9	6.7	3.0	11.6	20.7
Housing (%)	4.2	4.6	0.0	2.9	10.1
Education (%)	17.2	10.2	6.3	15.0	32.0
Others (%)	28.4	16.6	9.5	26.1	48.7

Table 1.3: Summary of trading behavior and portfolio composition of municipal funds. This table is constructed based on the holding data obtained from CRSP, for the period between 2009 and 2019. Panel A reports some summary statistics about the trading behavior of the mutual funds in the study. Panel B provides some summary statistics on the exposure of the funds to the selected counties. In panel C, each row presents information about the distribution of the percentage of municipal bond investments allocated to each issuer group or category. The bond categorization is obtained from merging the holding data with the Municipal Atlas database.

predict which governments relatively receive more or less demand based on the observable flows. Consequently, I examine how differently the governments respond to the predicted differences in demand.

Section 1.3.2 verifies that the funds respond to inflows and outflows in a predictable manner. In particular, in Section 1.3.1, I introduce a backward-looking measure of a fund's significance for a county government's bond market, which is based on the fund's past market share in the government's outstanding bonds. Building on this finding, I introduce my main demand instrument, "Flow-induced demand," in Section 1.3.3. Moreover, the validity of the demand instrument is discussed. Then, I analyze how the governments respond to flow-induced demand shocks.

1.3.1 A measure of fund significance for a government

For each fund f, county government c, and quarter t, I define significance measure $SIG_{f,c,t}$ as

$$SIG_{f,c,t} = \max_{t-11 \leq t' \leq t} OWN_{fct'} \tag{1.2}$$

,where

$$OWN_{fct} = \frac{\text{Par-value investment of } f \text{ in } c \text{ at } t}{\text{Total par-value investment of mutual funds in } c \text{ at } t} \tag{1.3}$$

In (1.3), $OWN_{f,c,t}$ is the fraction of outstanding debt of county government c held by the mutual funds that is owned by fund f. For instance, $OWN_{f,c,t} = 1$ means that f is the only mutual fund at t that holds the debt of government c, and $OWN_{f,c,t} = 0$ implies that fund f does not hold any bond of government c at t. By construction, $\sum_{f \in \text{Funds}} OWN_{f,c,t} = 1$

for any t and county government c. Building on this ownership variable, $SIG_{f,c,t}$ denotes the largest ownership of fund f over the twelve quarters ending at t. Put differently, funds with large values of $SIG_{f,c,t}$ are the ones that are or were a major investor of government c at some point in the past three years up to t. Table 1.4 presents the summary statistics for OWN and SIG.

	Mean	SD	1th	10th	50th	90th	99th
$\text{OWN}_{f,c,t}$ (%)	2.1	9.7	0.0	0.0	0.0	3.1	51.8
$\text{SIG}_{f,c,t}$ (%)	5.0	15.8	0.0	0.0	0.0	11.7	100.0

Table 1.4: Summary statistics of ownership and significance variables. $\text{OWN}_{f,c,t}$ is the share of fund f at the end of quarter t in the total par-value investments of the mutual funds in my sample in the bonds of county government c. $\text{SIG}_{f,c,t}$, as defined in (1.2), is the maximum value of the ownership variable for the twelve quarters ending at t.

1.3.2 Fund flows and trading behavior

Now, I investigate whether SIG can be used to predict how a fund invests its inflows. To this end, I estimate the coefficients in the panel regression below:

$$\Delta \log Inv^{Par}_{c,f,t} = \beta_0 + \beta_1 Flow_{f,t} + \gamma_2 SIG_{c,f,t-1} + \gamma_3 Flow_{f,t} \times SIG_{c,f,t-1} + \varepsilon_{c,f,t} \quad (1.4)$$

In (1.4), $Inv^{Par}_{c,f,t}$ represents the par-value investment of fund f in the bonds issued by government c at the end of quarter t. I add \$1 to all investments to avoid infinite values for zero positions.[13] The regression results for Specification 1.4, along with for some benchmark

13. Out of 23,916 observations of increase in investment ($\Delta Inv^{Par}_{f,c,t} > 0$), 15,586 (65%) are initiations of a

cases, are provided in Table 1.5.

	$\Delta \log Inv^{Par}_{f,c,t}$					
	Inflow Sample			Outflow Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
$Flow_{f,t}$	-0.002	-0.002^{*}	-0.002	0.524^{***}	0.501^{***}	0.360^{**}
	(0.001)	(0.001)	(0.001)	(0.153)	(0.154)	(0.157)
$\text{OWN}_{f,c,t-1}$		-4.562^{***}			-4.381^{***}	
		(0.140)			(0.121)	
$\text{OWN}_{f,c,t-1} \times Flow_{f,t}$		0.209			5.716^{***}	
		(0.199)			(1.451)	
$\text{SIG}_{f,c,t-1}$			-0.176^{**}			-0.286^{***}
			(0.069)			(0.073)
$\text{SIG}_{f,c,t-1} \times Flow_{f,t}$			0.489^{***}			2.229^{***}
			(0.156)			(0.775)
Observations	144,312	144,312	144,312	151,088	151,088	151,088
Quarter FE	Y	Y	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y	Y	Y
R^2	0.031	0.043	0.032	0.029	0.044	0.032

Table 1.5: This table reports the regression results of mutual fund trading in response to capital inflows and outflows. The dependent variable ($\Delta \log Inv_{f,c,t}$) is the change in the log of par-value investment of fund f in the bonds issued by county government c at the end of quarter t. The investments are aggregated at government level. 1\$ is added to all investments to avoid infinite values for zero positions. $Flow_{f,t}$ is the quarterly flow to fund f between $t-1$ and t, as a fraction of the merger-adjusted total net asset value at the end of quarter $t-1$. $OWN_{f,c,t}$ is the market share of fund f in the total par-value investment of mutual funds in government c at t. $SIG_{f,c,t}$ is the maximum value of $OWN_{c,f,t'}$ in the twelve quarters ending at t.

Table 1.5 informs how mutual funds respond to inflows and outflows. The results of

new position, i.e, $Inv_{f,c,t-1} = 0$. Using the percentage change in positions as the dependent variable would eliminate most of these observations. To verify the robustness of the results, Table 1.A.2 reports the results for different definitions of investment change with respect to a government.

specifications (1) and (2) suggest that inflows do not result in more investment in the current positions. It is in line with the observation in Table 1.3 that the funds hardly expand their current positions; most of the portfolio adjustments by municipal mutual funds are in the form of liquidating an existing position or initiating a new position. In fact, regression results for specifications (2) and (5) suggest that the funds reduce their current positions regardless of the flow sign, just they do so more intensely when they receive outflows.

Specifications (3) and (6) of Table 1.5 demonstrate the role of past positions, captured by SIG, in a fund's trading. By juxtaposing specifications (2) and (3), we see that when the fund flow is sufficiently large, the funds use the inflow to invest in the bonds of governments that they used to have a larger position over the past three years. Table 1.A.1 illustrates this point by adding a dummy of $\mathbb{I}\{SIG_{f,c,t-1} > OWN_{f,c,t-1}\}$ to Specification 1.4.

This trading behavior can arise for multiple reasons. For the outflows, the impact is somehow mechanical. The funds need to liquidate some of their positions to meet the withdrawals. For the inflow sample, the behavior can be explained with the presence of some informational costs. Mutual funds avoid investing in governments with a poor financial condition since they are more likely to be downgraded. Therefore, they need to perform due diligence in their investments, which is costly. The cost is lower for the governments that they have already examined their financials in their previous investments, which could lead to the pattern we see in Table 1.5. Another potential reason is that underwriters contact investors that are potentially interested in an issue. If a fund had repeatedly purchased the bonds of a particular government, it is more likely to be contacted by the underwriter or underwriting syndicate of a new bond issue of that government.

Overall, the main takeaway from the results of specifications (3) and (6) is that SIG at $t-1$ is informative about how mutual funds respond to inflows and outflows at t. It is the basis of utilizing the fund flows to predict the cross-sectional changes in the demand for the bonds of the county governments. The next section details the procedure.

1.3.3 Constructing flow-induced demand

In this section, I examine whether the fund flows have an impact on the timing, size, and the interest rate of the governments' borrowing. To this end, I construct a measure of flow-induced demand (FID):

$$FID_{c,t} = \sum_{f \in \text{Funds}} \text{SIG}_{f,c,t} \times \text{Flow}_{f,t} \times \text{PSF} \tag{1.5}$$

In (1.5), $SIG_{f,c,t}$ is the maximum ownership of fund f in the bond market of government c over the twelve quarters ending at t. $Flow_{f,t}$ is the dollar flow to fund f during quarter t, scaled by its merger-adjusted AUM at the beginning of the quarter. Lastly, PSF is the point coefficient estimate under specifications (3) and (6) in Table 1.5. In fact, it is either 0.489 or 2.229, depending on the sign of flow, and captures how flows to fund f are translated into trades (Lou, 2012). In short, $FID_{c,t}$ captures the dollar demand for the bonds of government c by mutual funds at t, scaled by the total market value of the bonds held by the funds.[14]

14. The following heuristic calculations elucidate the intuition had we defined $SIG = OWN$:

$$\begin{aligned} FID_{c,t} &\simeq \frac{1}{MV_{Funds,c,t-1}} \sum_{f \in \text{Fund}} MV_{f,c,t-1} \times Flow_{f,t-1} \times PSF \\ &= \frac{1}{MV_{Funds,c,t-1}} \sum_{f \in \text{Fund}} \frac{MV_{f,c,t-1}}{AUM_{f,t-1}} \times Flow^{\$}_{f,t-1} \times PSF \end{aligned}$$

In the first line, $MV_{Funds,c,t}$ and $MV_{f,c,t} \simeq MV_{Funds,c,t} \times OWN_{f,c,t}$ are the total market value of government c's bonds held by mutual funds and fund f, respectively. Note that I use the approximation that all funds hold the same composition of the bonds of government c so that they have the same market-to-book value. In the second line, $AUM_{f,t-1}$ is the total value of assets under the management of fund f, and $Flow^{\$}_{f,t}$ is the dollar flow to fund f. When $PSF = 1$, the demand corresponds to the case that each fund scales its

FID can be interpreted as a Bartik instrument that isolates an exogenous component in demand for the bonds. The exclusion restriction for the instrument's validity is that the cross-section of the fund flows is uncorrelated with the funding need of the county governments. It is a plausible assumption for three reasons. First, mutual funds tend to hold diversified portfolios in their target states. It is evidenced by Table 1.3, as it demonstrates that most funds hold large portfolios and do not have large exposure to any specific sector in the market or selected county government. Second, the table reveals that the bonds of the selected county governments only comprise 5.9% of the funds' assets on average. Third, Table 1.A.3 suggests that the investors might not be perfectly strategic in their fund selection, as it shows that investors do not significantly respond to the funds' past returns. Therefore, it is unlikely that a fund's exposure to a specific county government affects the investors' capital allocation to that fund.

Figure 1.3 presents the average flow-induced demand for each year (left-axis), along with the total par value of the bonds issued by the selected counties for those years (right-axis). The figure reveals a comovement between these two variables. We see that the flow-induced demand is negative for most observations. It is due to the asymmetry between responses to inflows and outflows by the mutual funds, as observed in Table 1.5. The impact of outflows is about four times larger than the impact of inflows, which causes the asymmetry we observe in the predicted demand values.

---

portfolio uniformly across its holdings.

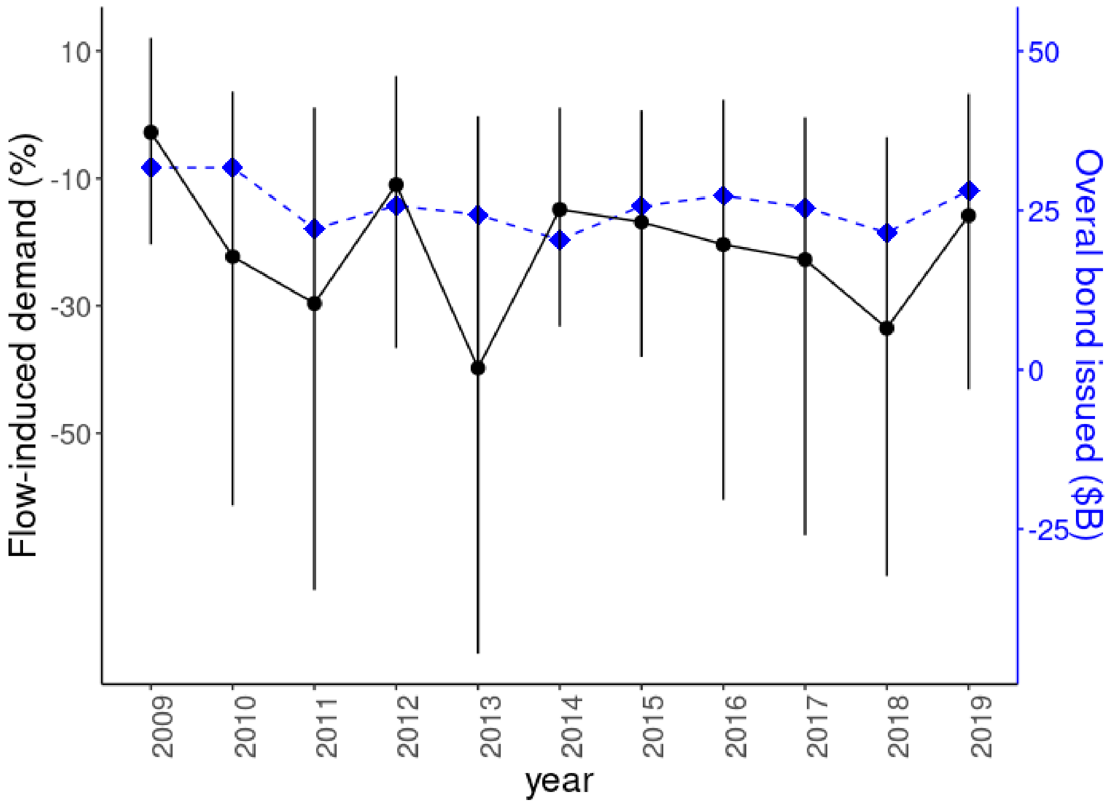


Figure 1.3: This figure presents the average value of FID for each year (left-axis), along with the overall par value of bonds issued by the selected counties (dashed-line, right-axis). Flow-induced demand (FID) is computed based on Equation (1.5) for each county at every quarter. The vertical lines display the range of values between the 90th percentile and 10th percentile of FID for each year.

1.3.4 Flow-induced demand and the timing of borrowing

To examine the impact of flow-induced demand on the timing of debt issuance by the selected county governments, I estimate the coefficients in Equation 1.6. In (1.6), $\text{Issue}_{c,t}$ is a dummy variable that reflects whether government c issued debt at quarter t :

$$Prob(\text{Issue}_{c,t+1}) = \Phi(\alpha_0 + \alpha_1 FID_{c,t} + \alpha_2 X_t) \tag{1.6}$$

The regression results are provided in Table 1.6. In Specifications (1) and (2), which do not include state-by-year fixed effect, flow-induced demand predicts the timing of issuance. Note that $FID_{c,t}$ is constructed based on time-t variables. Therefore, at least in terms of timing, the fund flows lead the governments' bond issuance. However, Specification 3 suggests that the evidence that FID predicts the timing is not strong, as is absorbed by the state-by-year fixed effect. It is, in fact, consistent with the long and costly nature of the process to obtain authorization for a new bond issue, especially for general obligation bonds. The procedural challenges make it difficult for many municipalities to time the market.

	Prob($Issue_{c,t+1}$)		
	(1)	(2)	(3)
$\text{FID}_{c,t}$	0.128***	0.093**	0.026
	(0.037)	(0.037)	(0.037)
Observations	8,661	8,661	8,661
Type	Probit	Probit	Probit
County FE	Y	Y	Y
Season FE	N	Y	Y
Year-State FE	N	N	Y
Log Likelihood	−4,815	−4,753	−4,646

Table 1.6: This table provides the regression results inspecting the impact of flow-induced demand on the probability of bond issuance in the next quarter. The dependent variable is a dummy variable that is one if government c issues bond at quarter $t+1$. A probit model is employed for the estimation. $\text{FID}_{c,t}$ is the flow-induced demand faced by government c at quarter t, computed based on Equation 1.5.

Table 1.7 presents how the flow-induced demand affects the probability of bond issuance in the previous and subsequent quarters. It further confirms that the bond issuance does not lead FID. In fact, the results imply that the opposite is the case. Furthermore, there is no evidence of the governments delaying (expediting) their issuance in response to unfavorable (favorable) demand conditions. If that was the case, there should have been a negative relationship between FID and probability of issuance in the subsequent quarters, which is not what Table 1.7 suggests.

	$Prob(Issue_{c,t+j})$							
	$j=-3$	$j=-2$	$j=-1$	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
$\text{FID}_{c,t}$	−0.010	−0.050	−0.047	0.065*	0.093**	0.103***	0.030	0.028
	(0.038)	(0.038)	(0.038)	(0.037)	(0.037)	(0.038)	(0.038)	(0.038)
Observations	7,973	8,218	8,465	8,714	8,661	8,414	8,169	7,925
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Season FE	Y	Y	Y	Y	Y	Y	Y	Y
Log Likelihood	−4,374	−4,503	−4,640	−4,788	−4,753	−4,623	−4,492	−4,338

Table 1.7: This table provides the results of regressions that examine how the flow-induced demand impacts the probability of bond issuance in previous and subsequent quarters. $Issue_{c,t+j}$ is a dummy variable that is one if county government c issues bonds at quarter $t+j$. $\text{FID}_{c,t}$ is the flow-induced demand faced by government c at quarter t, computed based on Equation 1.5.

1.3.5 Flow-induced demand and the size of borrowing

I estimate the coefficients in Specification 1.7 to examine the impact of flow-induced demand on the size of issuance. Issue $\text{Size}_{c,t}$ is the total par value of debt issued at quarter t by government c. X_t denotes the vector of control variables. Table 1.8 presents the results.

$$\log\left(\text{Issue Size}\right)_{c,t+1} = \beta_0 + \beta_1 FID_{c,t} + \beta_2 X_t + \varepsilon^{size}_{c,t+1} \tag{1.7}$$

The results provide a strong evidence that FID substantially impacts the size of borrowing by the county governments. In fact, it suggests that a 1% inflow to mutual funds causes the borrowing size to increase by 0.224%. It is equivalent to $0.224\% \times 19.4\% = 0.043$ percentage point increase in the size of outstanding debt, where 19.4% is the average ratio between the issue size and outstanding debt and obtained from Table 1.1.

	log(Issue Size$_{c,t+1}$)				
	(1)	(2)	(3)	(4)	(5)
FID$_{c,t}$	0.197***	0.196***	0.224***	0.245***	0.196***
	(0.060)	(0.060)	(0.068)	(0.081)	(0.075)
Observations	2,590	2,590	2,590	2,273	2,119
County FE	Y	Y	Y	Y	Y
Season FE	N	Y	Y	Y	Y
Year-State FE	N	N	Y	Y	Y
Additional Controls	N	N	N	Revenue gr. + lag Expenditure gr. + lag Liability gr. + lag	Income gr. + lag House pr gr. + lag
SE-clustered	State-Year	State-Year	State-Year	State-Year	State-Year
R^2	0.601	0.601	0.643	0.656	0.669

Table 1.8: This table provides the regression results inspecting the impact of flow-induced demand on the size of debt issued by the selected county governments. The dependent variables is the logarithm of total par value of debt issued at quarter t by government c. FID$_{c,t}$ is the flow-induced demand faced by government c at quarter t, computed based on Equation 1.5. Specifications (3)-(5) include county, year-by-state, and season fixed-effects. All standard errors are clustered at state-by-year level. The controls in Specification (4) are revenue growth, expenditure growth, and liability growth of each county government, along with their lag, obtained from the Muni Atlas database. The controls in Specification (5) are the change in the overall gross income reported to IRS and house price growth, along with their lagged values, for the corresponding county.

Specifications (1)-(3) provide the estimation results in the presence of different sets of fixed effects. The coefficient in all specifications is stable at around 0.2%. Note that the state-by-year fixed effect absorbs all state-level shocks at annual frequency, such as changes in the tax rates, or years with natural disasters, potentially associated with a higher-than-normal supply of bonds. Specification (4) tests whether changes in the financial condition can explain the correlation between the flow-induced demand and issue size. Comparing with the baseline results of Specification (2), we see that the regression coefficient of flow-induced demand slightly increases and is significant at 1% level.

Another potential concern is that the demand for the fund shares could be correlated with the demand for the government's bonds. For instance, when the local investors receive positive wealth or credit shocks, the demand for both the bonds and fund shares might increase, following Mian and Sufi (2014); Chodorow-Reich, Nenov, and Simsek (2021). Motivated by this channel, Specification (5) controls for the changes in the local house prices and gross income in the current and previous year. The coefficient of interest decreases slightly, however it is still significant at 1% level.

What is the mechanism through which the fund flows impact the size of bond issuance? In the municipal bond market, around 80% of total nominal value of municipal bonds issued are sold through negotiated sales, in which an underwriter or underwriting syndicate directly negotiates on different terms of the deal, including the issue size. Before finalizing the terms, the underwriting team surveys the market to gauge the demand for the new issue. Thus, their perception of the bond demand is crucial for the issue size since they have a limited risk capacity and need to sell their bonds quickly.

Even if the government auctions the issue to multiple underwriters, a deal type known as "competitive sale," the issuer's financial advisors potentially advise the issuing government on the issue size to help it attract more potential underwriters. Setting a large issue size when the demand is perceived to be weak deters underwriters from participating in the auction.

Table 1.9 examines the persistency of the impact of the fund flows on the governments' issue size. The results are visualized in Figure 1.4. First, the regression results confirm that there is no pre-trend for the impact of the fund flows. In other words, the governments' bond issuance does not seem to predict the fund flows. Moreover, the results suggest that FID impacts the size of issuance even after four quarters. However, the magnitude of the impact is substantially smaller after one quarter. These facts are used in the calibration of the model presented in Section 1.4.

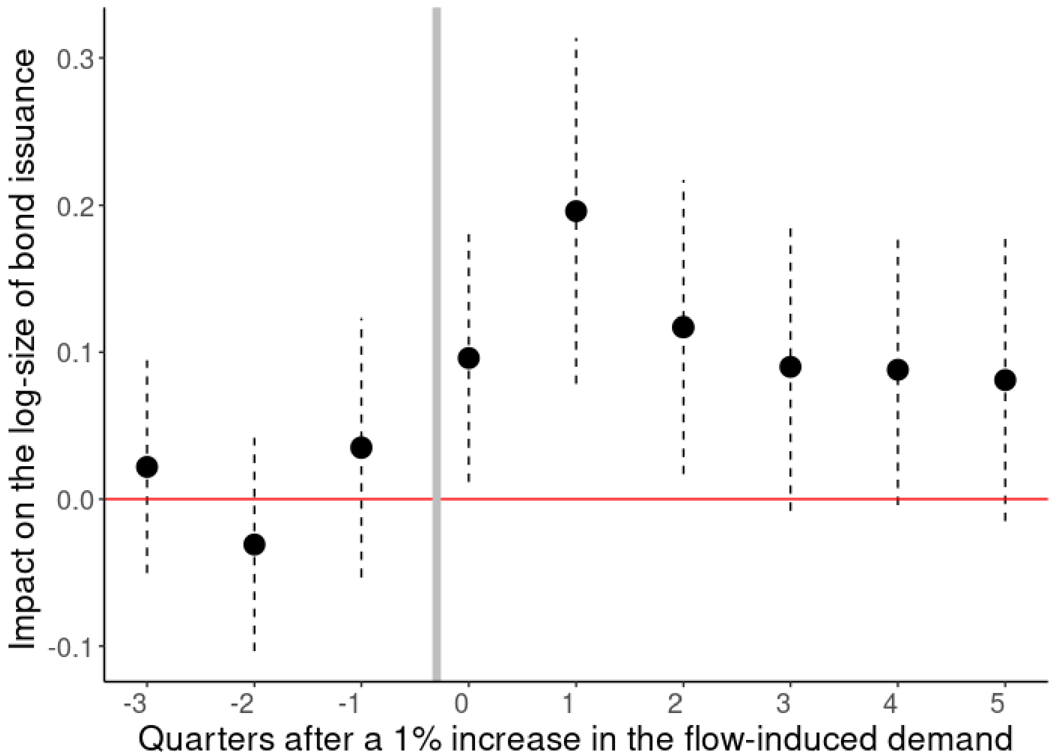


Figure 1.4: This figure displays how a 1% increase in the flow-induced demand impacts the overall size of bond issuance in the subsequent quarters. The points represent the point estimates provided in Table 1.9. Each point estimates is obtained by regressing the log size of bond issuance at quarter $t+j$ ($\log(\text{Issue Size}_{c,t+j})$) on $FID_{c,t}$. Quarter and county fixed effects are included, and standard errors are clustered at state-by-year level. The dashed-lines provide the confidence interval at 95% level. $\text{FID}_{c,t}$ is the flow-induced demand faced by government c at quarter t, computed based on Equation 1.5.

	log($Issue\ Size_{c,t+j}$)								
	$j=-3$	$j=-2$	$j=-1$	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
$\text{FID}_{c,t}$	0.022	−0.031	0.035	0.096**	0.196***	0.117**	0.090*	0.088*	0.081
	(0.037)	(0.037)	(0.045)	(0.043)	(0.060)	(0.051)	(0.050)	(0.047)	(0.049)
Observations	2,381	2,449	2,520	2,610	2,590	2,526	2,446	2,354	2,273
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Season FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
SE-Clustered	State-Year	State-Year	State-Year	State-Year	State-Year	State-Year	State-Year	State-Year	State-Year

Table 1.9: This table provides the regression estimates on the impact of flow-induced demand at quarter t on the size of bond issuance for quarters before and after t. $\text{FID}_{c,t}$ is the flow-induced demand faced by government c at quarter t, computed based on Equation 1.5. All regressions include county and season fixed effects. The standard errors are clustered at state-by-year level.

Does the tax-benefit associated with the in-state investing explain the results?

Interest income on municipal bonds are typically exempt from state and federal taxes as long as the investment is within the state of residency. This fact creates market segmentation along the states' border lines since the tax benefit pushes up the bond prices, makes the investment unattractive for out-of-state investors. Moreover, the majority of funds operate in only one state to deliver tax-free income to their investors (Babina, Jotikasthira, Lundblad, and Ramadorai, 2021). A potential concern is that the empirical results are connected to the tax treatment of municipal bonds. In Table 1.10, I address this concern by testing two direct implications of the tax-induced market segmentation.

First, if the observed impact of the fund flows is related to the tax-induced market segmentation, then the fund flows should be more impactful in states with a higher degree of segmentation. The degree of market segmentation can be captured by the tax-advantage that the investors in a state receive by investing in the bonds of their residency state, rather than investing in the bonds of other states. For instance, investors in California typically do not need to pay state taxes on the interest incomes of bonds issued by California governments. However, they would need to pay state taxes, up to 13.3% of the interest income, if they hold bonds issued outside the state boundary. Since such tax benefits tend to push up the prices, it would make the bonds less appealing for investors outside the state. In the first specification of Table 1.10, I interact this tax-advantage for different states with FID. The results do not suggest that FID has a stronger impact in states with a higher degree of segmentation.

As mentioned earlier, another concern is that many funds face mandates in their investments, which forces them to invest inside their domicile state. In Specification (2) of the table, I add a state-level FID to the regression, which absorbs the demand induced by such mandates. State-level FID, in fact, is a weighted average of the fund flows, multiplied by some scaling factor, where every fund is weighted based on its significance for the municipal bond market of that state. If the market shares are persistent at the state level, then those market shares would be close to the weights used to construct the state-level FID. We see that the impact of the original FID is still significant at 5% level. It suggests that all governments within a state are not perfect substitutes for the funds.

	$\log(\text{Issue Size})_{c,t+1}$	
	(1)	(2)
$\text{FID}_{c,t}$	0.328***	0.180**
	(0.119)	(0.088)
$\text{FID}_{c,t}$ × In-state tax privilege	−0.027	
	(0.022)	
State-level $\text{FID}_{c,t}$		0.020
		(0.019)
Observations	2,542	2,549
Type	OLS	OLS
County FE	Y	Y
Year-State FE	Y	Y
Season FE	Y	Y
SE-clustered	State-Year	State-Year
R^2	0.642	0.643

Table 1.10: This table provides the regression results examining the impact of tax-benefits of in-state investment on the earlier results. State-level FID_t is the state-level flow-induced demand, which is computed similar to (1.5). "In-state tax privilege" for a state is the largest difference between the tax rate applied to the interest income of in-state and out-of-state investment for the residents of each state. $\text{FID}_{c,t}$ is the flow-induced demand faced by government c at quarter t, computed based on Equation 1.5. Both regressions include county fixed effects, season fixed effects, and state-by-year fixed effects. All standard errors are clustered at state-by-year level.

1.3.6 *Flow-induced demand and the borrowing rate*

This section examines the impact of flow-induced demand on municipal governments' borrowing rate. Particularly, the coefficients in the following equation are estimated:

$$\begin{aligned}\text{yield-spread}_{c,t,bond} &= \beta_0 + \beta_1 FID_{c,t} + \beta_2 \text{YTM} + \beta_3 \text{log-size} \\ &\quad + \beta_4 FID_{c,t} \times \text{log-size} + \beta_5 FID_{c,t} \times YTM + \gamma X_{c,t,bond} + \varepsilon_{c,t,bond}\end{aligned} \tag{1.8}$$

In (1.8), yield-spread is the tax-adjusted spread of the bond over a maturity-adjusted treasury at t.[15] YTM denotes the bond's maturity in years, and log-size is the deal size of the bond issue. $X_{c,t,bond}$ represents a vector of bond characteristics (e.g., coupon rate, insured status, maturity type) that I control for. Quarter-rating fixed effect are also included in the regression specification. Table 1.11 presents the results.

The results reveal that the impact of flow-induced demand on the credit spread, while significant, is small and limited to few basis points. The coefficient estimates imply that 100 basis points increase (equivalent to one percentage point) in the flow-induced demand reduced the credit spread by 0.2 basis points. Recall from Section 1.3.5 that the same magnitude of change in the flow-induced demand increase the size of borrowing by 22 bps and outstanding debt by 3.2 basis points. It indicates that the impact of flow-induced demand on the debt

15. To compute the tax-adjusted spread, I use the equation below following Schwert (2017):

$$\text{yield-spread}_{bond_t} = \frac{\text{yield}}{(1-\tau^{State}\mathbb{I}_{State-exempt})(1-\tau^{Fed}\mathbb{I}_{Fed-exempt})} - r_t \tag{1.9}$$

The tax rates are obtained from NBER's TaxSim data set. r_t in (1.9) is the interest rate on a maturity-matched treasury at t.

	Tax-adjusted yield-spread (%)	
	(1)	(2)
$FID_{c,t}$	-0.002^{**}	-0.002^{*}
	(0.001)	(0.001)
log-size	0.026^{***}	0.026^{***}
	(0.004)	(0.004)
YTM	0.137^{***}	0.135^{***}
	(0.001)	(0.001)
log-size × $FID_{c,t}$	0.0001^{**}	0.0001^{**}
	(0.00005)	(0.00005)
YTM × $FID_{c,t}$		-0.00005^{***}
		(0.00001)
Observations	33,705	33,705
County FE	Y	Y
Coupon rate	Y	Y
quarter-rating FE	Y	Y
Insured-status dummy	Y	Y
Maturity-option FE	Y	Y
R^2	0.817	0.817

Table 1.11: This table provides the results of regression analysis pertaining to the impact of flow-induced demand on the credit spread of the bonds issued by the selected counties. The dependent variable, tax-adjusted yield spread, is constructed based on Equation 1.9, which adjusts for the bonds' tax-exemptions. The spread is computed relative to a maturity-matched treasury at the same month of the issuance. $FID_{c,t}$ is the flow-induced demand faced by county government c at quarter t, derived by Equation 1.5. log-size is the natural logarithm of the deal size. YTM denotes the bond years-to-maturity in years. Both regressions include quarter-rating fixed effect, county fixed effect, and maturity type fixed effect, capturing whether the bond is callable or not. Other controls are the bond coupon rate and insured-status.

quantity is one order of magnitude larger than its impact on the debt price.

This borrowing behavior is consistent with the earlier empirical findings. Yi (2020) examined how municipal governments responded to a regulation that reduced the lending ability of some banks to their municipal borrowers. She finds that the borrowing size of the most affected municipal governments decreased by 90% more than the least affected ones, while the relative increase in the interest rate was only 0.1% for the most affected group. Adelino, Cunha, and Ferreira (2017) and Cornaggia, Cornaggia, and Israelsen (2018) study the impact of Moody's recalibration of credit ratings in 2010 on municipal borrowing. Both studies compare the response of upgraded governments with not-upgraded ones. Their estimates indicate that the upgraded governments responded by borrowing 18% more annually. However, their interest rate only dropped by 0.2% more than the not-upgraded group.

Since most previous studies explore the impact of large credit shocks, the new evidence provided in this section contributes to literature by suggesting that even small credit supply shocks are not easily absorbed in the municipal bond market. Furthermore, I use my empirical estimates to inform the parameters of the model presented in Section 1.4 to drive conclusions about the elasticity of supply and demand for municipal bonds.

1.4 A dynamic model of the municipal bond market

Empirical results in Section 1.3 reveal that capital flows to mutual funds substantially impact the borrowing quantity of municipal governments, especially in the short run. This result is at odds with standard theory, which implies an elastic demand for municipal bonds: when mutual funds' demand for bonds increases due to capital inflows, other market participants should promptly absorb the shock by supplying their bonds at the competitive price; thus, the impact of capital flows on the governments' borrowing quantity should be small.

The model presented here explains this puzzling empirical observation by considering two deviations from the frictionless benchmark: First, the households, the end investors of the bonds, exhibit sluggish portfolio behavior, motivated by earlier empirical evidence (See, e.g., Giglio, Maggiori, Stroebel, and Utkus (2021)). In my model, the households can invest in municipal bonds either directly or indirectly, through mutual funds. Since the directly investing households do not respond to price changes immediately, the capital flows in and out of mutual funds substantially impact the market outcomes.

The second friction is that the mutual funds do not invest optimally on behalf of their investors due to the presence of investment mandates (Gabaix and Koijen, 2020). Investment mandates further lower the short-term demand elasticity as they restrict the funds' ability to absorb flow shocks. The households' sluggish portfolio behavior intensifies the problem since the households would have undone the friction had they reoptimized their allocation to the funds frequently.

I use the empirical estimates in Section 1.3.3 to calibrate the model. The calibration results suggest that the bond supply elasticity by the county governments in my sample is 19, which implies that the governments increase their outstanding debt by 1.9% when the interest rate falls by ten basis points. Furthermore, the short-term demand elasticity is at one order of magnitude smaller than the long-term demand elasticity. I conclude the section by discussing in depth how the model can be deployed to gauge the impact of significant market events on the borrowing behavior of municipal governments, such as the massive outflows from municipal mutual funds in March and April 2020.

1.4.1 Setup

Time is discrete, i.e., $t = 0, 1, 2, \ldots$. The economy could represent the municipal bond market in one of the US states since most investors only invest in the bonds issued in their state of residence to maximize their tax savings. Municipal bonds are supplied by a representative municipal government. The government sets the nominal amount of outstanding debt, denoted by Q_t, so that maximizes its objective function:

$$P_t(\frac{Q_t}{W_t^R}) - \frac{\phi}{1+\gamma}(\frac{Q_t}{W_t^R})^{1+\gamma} \tag{1.10}$$

In (1.10), P_t is the unit price of the bonds, and W_t^R is the overall wealth of the residents. The government trades off the bond issuance revenue against a convex cost, capturing the statutory limitations that the government faces in expanding its debt. The convex cost could represent different tiers of municipal debt with varying levels of procedural difficulty for their issuance. For instance, cities in the State of California can issue general obligation bonds with no voting requirement up to 1.5% of the assessed value of the taxable properties in their municipality. However, they can exceed the limit up to 2.5% of that assessed value by obtaining the supermajority approval from their residents. A higher γ corresponds to a more restrictive borrowing regulation, which causes the elasticity of bond supply to be lower. Lastly, $\phi > 0$ is a constant.

The first order condition implies that the bond supply at t only depends on the residents' wealth and price at t:

$$q_t \equiv \frac{Q_t}{W_t^R} = \phi^{-\gamma^{-1}} P_t^{\gamma^{-1}} \tag{1.11}$$

This reduced-form approach in modeling the budget constraint, despite its limitations, has some realistic features, in addition to helping yield tractability. First, financial authorities in municipal governments serve for limited terms, which gives them an incentive to be relatively myopic in their borrowing behavior. In fact, Table 1.7 suggests that the governments do not act strategically in their bond issuance; namely, they do not postpone (expedite) their issuance when the demand is relatively weak (strong). Second, the statutory limitations hardly change over time, which could set a tighter debt limit than the discounted value of the future revenues.

Note that the government effectively optimizes the level of Q_t/W_t^R, the borrowing amount per one unit of the residents' wealth. It accounts for two considerations that impact the government's borrowing decisions: First, the demand for infrastructure is naturally higher when the residents are wealthier. Second, the statutory debt capacity is typically determined as a percentage of the total value of the taxable assets in the municipality, which highly correlates with the overall local wealth. I assume that the funding need of the government does not change over time. In Appendix 1.C.1, I consider a more general model that features a time-varying need for external funding.

Suppose D_0 is the promised coupon payment that each unit of the bond pays at each period, and R^F is the risk-free rate. Since investors compare tax-exempt municipal bonds with other taxable fixed income securities, let $D \equiv \frac{D_0}{1-\tau}$ be the coupon payment for an equivalent taxable bond, where τ is the effective tax-benefit of municipal bonds. The government defaults on a fraction of its bonds with probability δ, independent across the periods. The bond return is $R^D < R^F$ if a default happens. The default risk is the only source of uncertainty in this economy, and once it happens, all investors adjust their portfolios if necessary and the government never defaults again, implying that the remaining bonds become risk-free.

The economy is populated with overlapping generations of atomistic investors. Each investor dies with probability $1 - x \in (0,1)$ at the beginning of each period. The investors' objective is to maximize their expected log-wealth at the time of death. More formally, investor i chooses its portfolio allocations, under the restrictions that are explained later, so that maximizes the following objective function, where $t + \tau^{\text{death}}$ denotes the random time of death.

$$U_t^i = \mathbb{E}_t[\log w^i_{t+\tau^{\text{death}}}] \tag{1.12}$$

There are two groups of investors: "direct investors" who invest in municipal bonds directly and "indirect investors" who invest in the bonds indirectly through some mutual funds. I assume no overlap between these two groups, which is consistent with the institutional feature of the market. In 2013, among the 2.4% US households with a positive investment in municipal bonds, 1.6% had indirect investments and 0.9% directly invested in municipal bonds, leading to a 0.1% overlap (equivalent with 4.2% of the municipal bond investors) (Bergstresser and Cohen, 2016). The difference in the investment approach can be explained by differences in the level of sophistication regarding municipal bond trading or level of trust to municipal fund managers. In the model, I denote the overall wealth of the direct and indirect investors by W_t^D and W_t^{ID}, respectively. By definition, $W_t^R = W_t^D + W_t^{ID}$. Upon death, the investors are replaced by their child who inherits their investment approach (i.e., indirect or direct) and entire wealth.

The key behavioral assumption of the model is that the investors rebalance their portfolio intermittently. In every period, a fraction $1 - \lambda \in (0,1)$ of the direct and indirect investors reoptimize their portfolio, and the rest keep the same portfolio allocation as in the previous period. As such, λ captures the extent of sluggishness in the investors' portfolio allocations. The sluggish portfolio adjustments cause the bond demand to be less elastic in the short-run, which is crucial to explain the large short-term quantity response observed in Figure 1.4.

Let α_t^D and α_t^{ID} be the average wealth fraction that the direct and indirect investors allocate to the municipal bonds and the mutual funds, respectively. Therefore, the portfolio dynamics for these investors can be specified as below:

$$\alpha_t^J = \lambda \alpha_{t-1}^J + (1-\lambda)\alpha_t^{J-Reb} \quad J \in \{D, ID\} \tag{1.13}$$

,where α_t^{J-Reb} denotes the optimal portfolio allocation for the rebalancing investors at period t, cognizant of the fact that they will hold the same portfolio until their next rebalancing time.

The mutual funds invest fraction $\alpha^F(R_t)$ of their assets in municipal bonds, where $R_t \equiv \frac{P_{t+1}+D}{P_t}$ is the one-period return conditional on no default happening at $t+1$. Furthermore, $\alpha^F(R_t)$ is an increasing function with a bounded second derivative. Note that R_t is a sufficient statistic to describe the bond return since the return is binary, and it is either R_t with probability $1-\delta$ or R^D with probability δ. The indirect investors only have access to the risk-free asset other than the mutual funds. Therefore, they effectively allocate fraction $\alpha_t^{ID}\alpha^F(R_t)$ of their wealth to municipal bonds.

Lemma 1 provides the optimal portfolios for the direct and indirect investors that rebalance at t, provided the government does not default by t.

Lemma 1. *Let $\{R_t\}_{t=0}^{\infty}$ be the sequence of the bond return prior to the government's default. Therefore, the optimal allocation of the rebalancing investors at t, α_t^{D-Reb} and α_t^{ID-Reb}, are the solutions to the following maximization problems:*

$$\alpha_t^{D-Reb} = \arg\max_{\alpha} \quad \frac{\delta}{1-\nu} \log\{R^F + \alpha(R^D - R^F)\}$$
$$+(1-\delta)\sum_{s=0}^{\infty} \nu^s \log\{R^F + \alpha(R_{t+s} - R^F)\} \tag{1.14}$$

$$\alpha_t^{ID-Reb} = \arg\max_{\alpha} \quad \delta \sum_{s=0}^{\infty} \nu^s \log\{R^F + \alpha\alpha^F(R_{t+s})(R^D - R^F)\}$$
$$+(1-\delta)\sum_{s=0}^{\infty} \nu^s \log\{R^F + \alpha\alpha^F(R_{t+s})(R_{t+s} - R^F)\} \tag{1.15}$$

,where $\nu = x\lambda(1-\delta)$.

In (1.14) and (1.15), we see that the optimal portfolios depend on all the future returns since the investors are uncertain about their next reoptimization time. The demand sensitivity with respect to the future returns is captured by ν, which is increasing in the inattention parameter λ. In the extreme case that the investors reblanace their portfolio in every period, i.e., $\lambda = 0$, the optimal portfolios only depend on R_t, that is they become myopic in their portfolio decisions. More specifically, the optimal portfolios are $\alpha^*(R_t)$ and $\frac{\alpha^*(R_t)}{\alpha^F(R_t)}$ in this extreme case, where:

$$\alpha^*(R_t) = \arg\max_{\alpha} \quad (1-\delta)\log\{R^F + \alpha(R_t - R^F)\} + \delta \log\{R^F + \alpha(R^D - R^F)\} \tag{1.16}$$

Unlike the investors, the mutual funds are responsive to changes in the bond price. In fact, they can implement the optimal exposure to municipal bonds for their investors when $\alpha^F(R_t) = \alpha^*(R_t)$. In this case, the indirect investors achieve their optimal portfolio allocation by perfectly delegating their investment decisions to the funds, i.e., $\alpha_t^{ID} = 1$. It means that even though the indirect investors never change their allocation to the funds, their exposure to municipal bonds is always optimal. However, when $\alpha^F(R_t) \neq \alpha^*(R_t)$, possibly due to agency frictions (He and Krishnamurthy, 2013) or investment mandates (Gabaix and Koijen, 2020), the funds might over-react or under-react to changes in the bond price, and their investors are slow in adjusting their capital allocation and undoing the misallocation caused by the funds' suboptimal behavior.

In fact, municipal mutual funds typically face mandates in their portfolio allocations. For instance, Vanguard high-yield tax-exempt municipal bond fund has to invest at least 80% of its assets in investment-grade municipal bonds.[16] As reported by Table 1.2, the municipal funds invest their funds almost entirely in municipal bonds and hold little cash buffers. This feature is reflected in my calibration exercise in Section 1.4.3.

Lastly, I let the mutual funds receive investments from sources other than the investors inside the economy. For instance, an investor in Texas might invest in a California fund, i.e., a municipal mutual fund that invests in the bonds issued by California governments, possibly to diversify its bond portfolio nationally. That would increase the investment in California bonds, and reduce the investment in Texas bonds. Let F_t^{Out} be the net amount of such investments.

Figure 1.5 depicts the structure of the bond market. Equation 1.17 provides the market clearing condition:

16. See more information about the fund here.

$$\alpha_t^F(R_t)W_t^F + \alpha_t^D W_t^D = P_t Q_t$$
$$W_t^F = F_t^{Out} + \alpha_t^{ID} W_t^{ID} \tag{1.17}$$

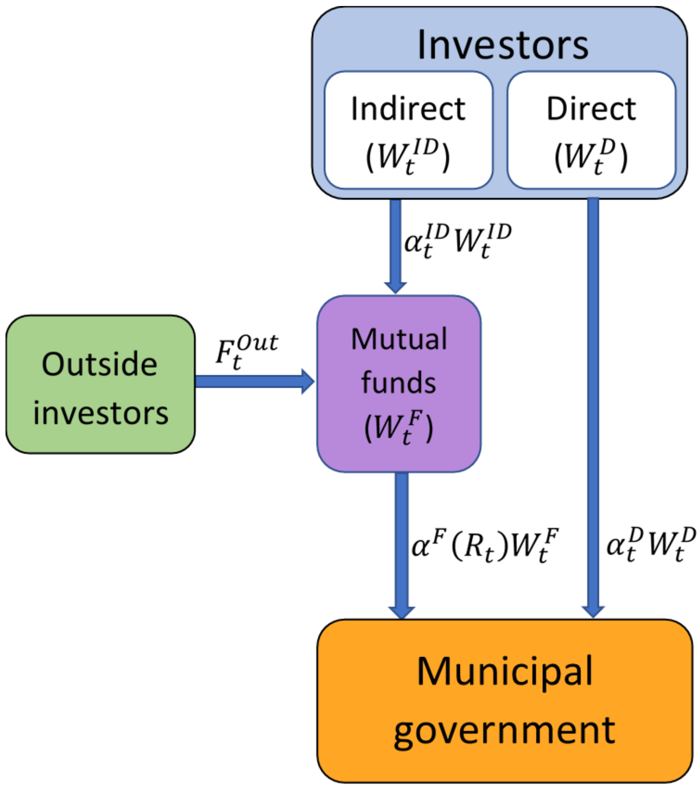


Figure 1.5: Visual representation of the market structure. The values next to each arrow presents the capital flows from one sector to another.

1.4.2 The impact of fund flows on the bond price and quantity

In practice, mutual funds receive three types of inflows (similarly, outflows): The first type is that mutual fund investors might add to their investments in mutual funds. The second type is that some investors who used to invest directly in municipal bonds decide to have some indirect investment through mutual funds. The third type is that some investors living outside the economy (here, state) decide to increase their investment in municipal bonds by allocating more funding to the mutual funds active in that state.

Flows originated from different sources have different implications for the market outcomes. The first flow type has a temporary impact on the market outcomes since the indirect investors eventually reoptimize and fix their collective excess allocations (withdrawals). The second flow type has no first-order impact on the market outcomes; because, on average, both direct and indirect investors allocate the same fraction of their wealth to the bonds, which is their unconditionally optimal portfolio share of municipal bonds. Therefore, only the third type of fund flows has a permanent effect on the equilibrium quantity and prices.

Note that the model considers the mutual funds homogeneous, which is not the case in the data. The capital flows across the funds could also have a temporary or permanent impact on the governments' outstanding debt. For instance, if the familiarity bias in mutual funds' investments is persistent, the capital flows have a lasting effect. Since my empirical strategy mostly exploits such heterogeneities across the funds, I interpret the flows as the third type in my analysis, unless otherwise stated.

In particular, suppose initially (at $t = 0-$) $F_{0-}^{Out} = 0$, namely the bonds are only held by the investors residing in the economy. At $t = 0$, the mutual funds receive inflow $f_0 W_0^{ID}$ from investors outside the economy. In my empirical analysis, I examine the impact of a 1% inflow to the funds, which corresponds to $f_0 = 0.01$.

Section 1.4.2 examines the impact of the inflow on the steady-state price and quantity of the bonds, and reveals that the portfolio sluggishness parameter λ has no impact on the steady-state outcomes. In Section 1.4.2, I log-linearize the model around its steady-state and investigate the dynamic implications of the permanent inflow. I find that the short-term responses indeed depend on λ. In Section 1.4.3, I calibrate the model with the empirical estimates obtained in Section 1.3.

Long-run effects of a permanent inflow

At steady state, both Q_t and $W_t^R \equiv W_t^D + W_t^{ID}$ increase at the same rate. Therefore, their ratio, $q_t \equiv \frac{Q_t}{W_t^R}$, and the bond unit price (P_t) are constant. Let q_{0-} and P_{0-} be those constants before the inflow. Therefore, the return on the bond before the inflow is $R_{0-} = \frac{D+P_{0-}}{P_{0-}}$. Prior to the inflow, the bond return is i.i.d across the periods, implying that the portfolio allocation decision for any number of inactive periods coincides with that for a single period. Therefore,

$$\alpha_{0-}^D = \alpha^*(R_{0-}), \qquad \alpha_{0-}^{ID} = \frac{\alpha^*(R_{0-})}{\alpha^F(R_{0-})}. \tag{1.18}$$

Since $F_{0-}^{Out} = 0$, the market clearing for the bond at $t = 0-$ implies:

$$\alpha^*(R_{0-})W_0^R = Q_{0-}P_{0-} = \phi^{-\gamma^{-1}} {P_{0-}}^{1+\gamma^{-1}} W_0^R \tag{1.19}$$

$$\Rightarrow \alpha^*(\frac{D + P_{0-}}{P_{0-}}) = \phi^{-\gamma^{-1}} {P_{0-}}^{1+\gamma^{-1}}. \tag{1.20}$$

After the inflow at $t = 0$, the total wealth available to mutual funds increases by $f_0 W_0^F$. Particularly, assume $F_t^{Out} = f_0 \alpha_t^{ID} W_t^{ID}$, implying that the outside investors allocate funds proportional to the indirect investors. By combining the market clearing condition with the bond supply equation (1.11), we can obtain the following market clearing condition for $t \geq 0$:

$$(1 + f_0)\alpha^F(R_t)\alpha_t^{ID} W_t^{ID} + \alpha_t^D W_t^D = P_t Q_t = \phi^{-\gamma^{-1}} P_t^{1+\gamma^{-1}} W_t^R \tag{1.21}$$

By repeating the same steps, we can find the new steady-state bond price, which I denote by P_{SS}:

$$\alpha^*(\frac{D + P_{SS}}{P_{SS}}) = \phi^{-\gamma^{-1}} {P_{SS}}^{1+\gamma^{-1}}(1 - S^F f_0) \tag{1.22}$$

,where $S^F \equiv \frac{W_0^{ID}}{W_0^D + W_0^{ID}}$ is the fraction of bonds held by the mutual funds. Likewise, let $S^D \equiv 1 - S^F$ be the market share of the direct investors. Appendix 1.B.2 provides the details of the derivation of (1.22).[17] Equation 1.22 implies that the rebalancing parameter λ does not have a first-order effect on the steady-state outcomes. Note that the measure of bond quantity used here is debt per unit of the residents' wealth. The inflow impacts the overall wealth of the residents as it pushes up the bond prices. However, the government is assumed to fully absorb the fluctuations in the overall wealth by adjusting its outstanding debt.

Suppose η is the demand semi-elasticity with respect to the bond return. That is,

$$\eta \equiv \frac{\partial \log \alpha^*(R_t)}{\partial R_t}|_{R_t = R_{0-}}. \tag{1.23}$$

By dividing both sides of (1.22) on those of (1.20), we get:

17. In the derivation, we see that the inflow changes the wealth of both indirect and direct investors. However, the change in the residents' overall wealth does not impact the steady-state bond price since it is perfectly absorbed by the government, by the construction of the government's objective function 1.10. The wealth ratio between the two groups also has no long-term impact since all investors eventually hold the same portfolio.

$$\frac{\alpha^*(\frac{D+P_{SS}}{P_{SS}})}{\alpha^*(\frac{D+P_{0-}}{P_{0-}})} = (\frac{P_{SS}}{P_{0-}})^{1+\gamma^{-1}}(1 - S^F f_0) \Rightarrow \eta dp(1 - \frac{P_{SS}}{P_{0-}}) = (1+\gamma^{-1})(\frac{P_{SS}}{P_{0-}} - 1) - S^F f_0$$
$$\Rightarrow \frac{P_{ss}}{P_{0-}} = 1 + \frac{S^F}{\eta dp + \gamma^{-1} + 1} f_0 \tag{1.24}$$

The impact of the inflow on the steady-state value of q_t can be found by combining (1.24) with (1.11):

$$\frac{q_{SS}}{q_{0-}} = (\frac{P_{SS}}{P_{0-}})^{-\gamma} \simeq 1 + \frac{\gamma^{-1} S^F}{\eta dp + \gamma^{-1} + 1} f_0 \tag{1.25}$$

Short-run effects of a permanent inflow

In this section, I derive the dynamics induced by the inflow in the market outcomes. In the next section, I use the derivations to calibrate the model with the elasticity estimates obtained in the empirical section.

Let the hatted variables represent the deviations from their corresponding steady-state values. That is, for variable Y_t with steady-state value Y_{SS}, I define $\hat{y}_t = \frac{Y_t - Y_{SS}}{Y_{SS}}$. Appendix 1.B.3 provides the details of the derivations below. The log-linearized version of the bond supply equation (1.11) is

$$\hat{q}_t = \gamma^{-1} \hat{p}_t. \tag{1.26}$$

Likewise, the deviations of the bond return from its steady-state value $R_{SS} \equiv \frac{P_{SS}+D}{P_{SS}}$ can be written as follows

$$
\begin{aligned}
R_t &= R_{SS} + r_t, \\
r_t &= \hat{p}_{t+1} - (1+dp)\hat{p}_t, \qquad dp = \frac{D}{P_{SS}}.
\end{aligned} \tag{1.27}
$$

The share of assets that the funds and direct investors invest in municipal bonds, and the indirect investors allocate to the funds are $\alpha_t^F = \alpha_{SS}^F(1+\hat{\alpha}_t^F)$, $\alpha_t^D = \alpha_{SS}^*(1+\hat{\alpha}_t^D)$, and $\alpha_t^{ID} = \frac{\alpha_{SS}^*}{\alpha_{SS}^F}(1+\hat{\alpha}_t^{ID})$, respectively. Let η^F be the demand semi-elasticity of the funds with respect to the returns:[18]

$$
\eta^F \equiv \frac{\partial \log \alpha^F(R_t)}{\partial R_t}|_{R_t=R_{0-}}. \tag{1.28}
$$

It directly implies that

$$
\hat{\alpha}_t^F = \eta^F r_t. \tag{1.29}
$$

The dynamics of $\hat{\alpha}_t^D$ and $\hat{\alpha}_t^{ID}$ can be obtained by log-linearizing the first-order conditions of (1.14) and (1.15):

$$
\hat{\alpha}_{t+1}^D = \lambda\hat{\alpha}_t^D + (1-\lambda)(1-\nu)\eta\sum_{s=0}^{\infty}\nu^s r_{t+s} \qquad t \geq 0 \tag{1.30}
$$

$$
\hat{\alpha}_{t+1}^{ID} = \lambda\hat{\alpha}_t^D + (1-\lambda)(1-\nu)(\eta-\eta^F)\sum_{s=0}^{\infty}\nu^s r_{t+s} \qquad t \geq 0 \tag{1.31}
$$

The initial condition is determined by the deviation of the initial portfolios from their corresponding new steady-state portfolios:

18. Since $\alpha^F(\cdot)$ has a bounded second derivative, the difference between the demand sensitivity estimated at the initial and final steady states is small, provided the inflow f_0 is small. Thus, using either of these demand sensitivities has no first-order effect in the derivations.

$$
\begin{aligned}
\hat{\alpha}^{D}_{0-} &= \frac{\eta S^F dp}{\eta dp + \gamma^{-1} + 1} f_0 \\
\hat{\alpha}^{ID}_{0-} &= \frac{(\eta - \eta^F) S^F dp}{\eta dp + \gamma^{-1} + 1} f_0
\end{aligned} \tag{1.32}
$$

The market clearing condition 1.17 can be written as below.

$$
S^F \hat{\alpha}^F_t + S^F \hat{\alpha}^{ID}_t + S^D \hat{\alpha}^D_t = \hat{p}_t + \hat{q}_t \tag{1.33}
$$

$\hat{\alpha}^F_t$, $\hat{\alpha}^{ID}_t$, and $\hat{\alpha}^D_t$ represent the deviations from the steady-state portfolio allocation by mutual funds, indirect investors, and direct investors, respectively. Therefore, the left-hand side in (1.33) presents the total deviation in the demand for bonds, divided by the residents' wealth, from its steady-state value. Likewise, the right-hand side presents the deviation of the market value of the government's outstanding debt, as a fraction of the total wealth of the residents, from its steady-state value. Thus, the condition states that the deviations in the demand and supply of the bonds should coincide at each period. By expanding and combining the equations above, we can obtain the following equation that describes the price dynamics

$$
\begin{aligned}
-S^F \eta^f \nu \hat{p}_{t+3} + \Big(\nu(1+\gamma^{-1}) + \lambda\nu S^F \eta^F + S^F \eta^F + M + (1+dp) S^F \eta^F \nu\Big)\hat{p}_{t+2} \\
-\Big((1+\lambda\nu)(1+\gamma^{-1}) + \lambda S^F \eta^F + (1+dp)(\lambda\nu S^F \eta^F + S^F \eta^F + M)\Big)\hat{p}_{t+1} \\
+\Big(\lambda(1+\gamma^{-1}) + (1+dp)\lambda S^F \eta^F\Big)\hat{p}_t = 0.
\end{aligned} \tag{1.34}
$$

,where $M = (1-\lambda)(1-\nu)(\eta - S^F \eta^F)$.

Lemma 2. *The characteristic polynomial corresponding to dynamic process 1.34 has three positive roots, only one of which is smaller than one.*

Lemma 2 states that the dynamic process 1.34 has only one stable root, which I hereafter denote by κ. It means that the solution should have the following form

$$\hat{p}_t = \kappa^t \hat{p}_0. \tag{1.35}$$

To solve for $\hat{p}_0$, we can utilize the market clearing condition at $t = 0$ and the initial conditions provided by (1.32). By so doing, we obtain

$$\hat{p}_0 = \frac{\lambda(\eta - \eta^F S^F)S^F}{(1 + \gamma^{-1} + dp)(1 + \gamma^{-1} + (1 + dp - \kappa)(\frac{M\nu}{1-\nu\kappa} + M + S^F \eta^F))} f_0. \tag{1.36}$$

In the extreme case that the investors rebalance at every period ($\lambda = 0$), or the case that the mutual funds manage the entire wealth of the economy optimally ($S^F = 1, \eta^F = \eta$), $\hat{p}_t$ is zero, implying that the market outcomes reach their steady-state value at $t = 0$. However, Figure 1.4 suggests that the short-term response of the bond quantity is larger than its long-term response, which is not consistent with the implication of those extreme cases. The second S^F in the numerator reflects the fact that the inflow to the market is proportional to the size of the mutual funds' market share.

Figure 1.6 illustrates how the portfolio sluggishness parameter λ and the funds' semi-elasticity of demand η^F impact the magnitude of the short-term response and speed of convergence to the new steady-state. In Panel (a), we see that as the investors rebalance less frequently, the magnitude of short-term response rises, and the convergence to the new steady-state slows down. In the extreme case of $\lambda = 0$, there is no over-shooting and the market immediately adjusts after the inflow.

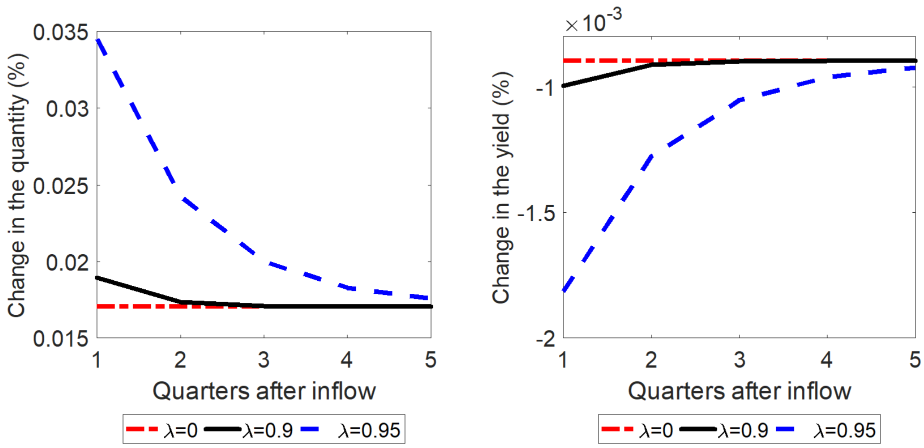
The intuition is that the funds demand more bonds when they receive the inflow. However, the direct investors do not sell their bonds as much as they would were they responsive to the bond price changes. It means that either the government needs to provide more bonds, or the bond price spikes, depending on the elasticity of bond supply. The bond supply should be quite elastic to justify the large quantity movements and small price movements we see in the data. A larger λ implies that the direct investors are slower in responding to the increased bond demand from the mutual fund sector.

Figure 1.6b reveals that the magnitude of the short-term responses decrease with the funds' demand semi-elasticity η^F. Note that a combination of two forces determines the overall impact of the permanent inflow on the bond demand by the mutual fund sector. First, the funds need to buy more bonds due to the inflow. However, the relationship is theoretically less than one-to-one since the funds increase their asset allocation to risk-free securities as they become relatively cheaper compared to the municipal bonds. As η^F increases, a larger portion of the flow induced demand is offset by this substitution effect. When η^F is close to zero, like when the funds face tight mandates, the bond demand by the mutual fund sector increases close to proportional in response to the inflow, which generates a larger spike in the bond price and quantity.

1.4.3 Calibration

In the empirical section, we obtained some estimates on how capital flows to mutual funds impact the borrowing behavior of municipal governments. Particularly, Figure 1.4 revealed that the short-term effect of the fund flows are substantially larger than their long-term effect, and we learned that the impact on the bond quantity is about 16 times larger than the impact on the yield at issuance. With these two empirical observation, I calibrate three

(a) Portfolio Sluggishness (λ)



(b) Mutual funds' demand sensitivity (η^F)

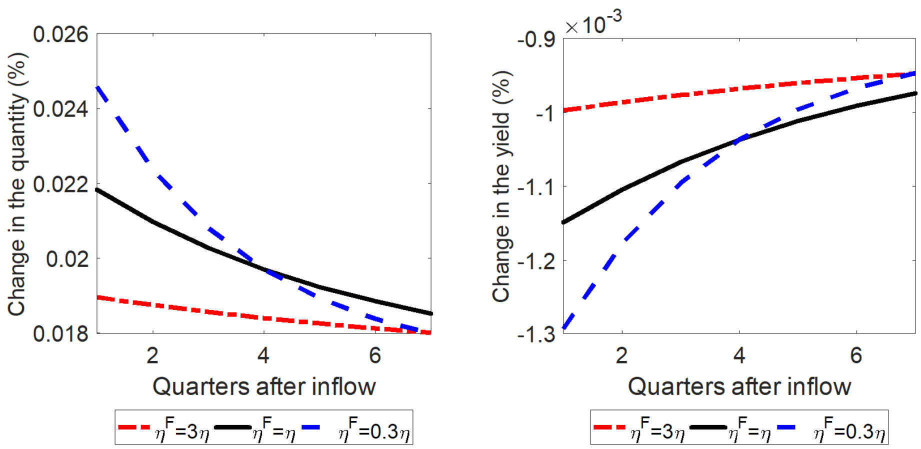


Figure 1.6: The impact of a permanent inflow to the funds on the bond yield and quantity. Panel (a) presents the results for a 1% inflow to mutual funds. Panel (b) illustrates the impact of a permanent 1% inflow to the mutual funds for different values of the funds' demand sensitivity (η^F). The parameter values reported in Table 1.12 are used for the illustrations, unless otherwise is stated.

free parameters of the model: The bond supply elasticity (γ^{-1}), the demand semi-elasticity parameter (η), and the portfolio inertia parameter λ.[19] Table 1.12 provides the calibrated values for the model parameters.

Parameters	Symbol	Value
Mutual funds' market share	S^F	0.16
Dividend-price ratio (Quarterly)	dp	1.68×10^{-2}
Bond supply elasticity by the municipal government	γ^{-1}	19
Portfolio inertia	λ	0.924
Survival rate	x	0.994
Funds' demand elasticity	$\eta^F dp$	0
Default probability	δ	3.75×10^{-5}
Long-run demand elasticity	ηdp	158.4
Short-run demand elasticity	$((1-\lambda)\eta + \lambda\eta^F S^F)dp$	11.6

Table 1.12: Parameter values used for the calibration

.

In Table 1.12, S^F is the average market share of the mutual funds in my sample. The dividend-price ratio is equal to the average tax-adjusted quarterly coupon rate in my bond issuance data. γ^{-1} and η are chosen so that they match the short-term impact of fund flows, estimated in Tables 1.8 and 1.11. The long-run demand elasticity is ηdp since 1% change in the steady-state bond price causes the bond one-period return to move by $dp \times 1\%$ in the opposite direction:

$$R_{SS} - R_{0-} = \frac{D + PSS}{P_{SS}} - \frac{D + P0-}{P_{0-}} = -dp(\frac{P_{SS} - P_{0-}}{P_{0-}}) \tag{1.37}$$

The investors' survival rate (x) is selected so that they survive for 120 quarters in expectation. The default probability is chosen to match the 10-year accumulated default probability of Moody's rated municipal bonds, which is 0.15% (Cestau, Hollifield, Li, and Schürhoff,

19. Note that η captures the risk-tolerance of the investors, thus there is a one-to-one relationship between η and the perceived return at default R^D. Therefore, I directly calibrate η.

2019). I set $\eta^F = 0$ since Table 1.2 suggests that the funds barely change their allocation to municipal bonds due to the tight mandates they face.

Under the calibrated parameters, the short-run demand elasticity is 12.0, which is one order of magnitude smaller than the long-run demand elasticity (171.7). This large gap between short-run and long-run demand elasticity is necessary to explain the overshooting in Figure 1.4, and is generated by the sluggish portfolio adjustments of the investors.

Figure 1.7 illustrates the dynamics implied by the model as a result of a 1% one-time and permanent inflow to mutual funds at $t = 0$ ($f_0 = 0.01$). Under the calibrated parameters, the size of outstanding debt increases by 0.038% in the first quarter and 0.017% in the long-run. These numbers match their empirical counterparts.[20] Note that even though the investors are very slow in their portfolio adjustments, the bond quantity and price converge quite quickly to their new steady state values. The intuition is that the investors who adjust earlier respond more aggressively to benefit from the temporary mispricing.

How differently would the bond quantity and price respond if the demand or supply were more or less elastic? The results are provided in Figure 1.8. In Figure 1.8a, we see that a substantially larger demand elasticity (ten times larger) would have produced quantity and price responses four times smaller than data. It is intuitive since more elastic demand means the market is more effective in absorbing the flows. Likewise, a substantially smaller demand elasticity generates short-term responses three time larger than the empirical estimates. Moreover, it generates dynamic responses quite slower than what data suggest. Figure 1.8b and 1.8c perform the same exercise for the bond supply elasticity and portfolio inertia.

20. To see this, note that Table 1.6 reveals that the size of issuance increases by 0.196% in the first quarter and 0.088% in the forth quarter after a 1% inflow. The numbers are obtained by adjusting (multiplying) them with 19.4%, the average issue size as a fraction of outstanding debt.

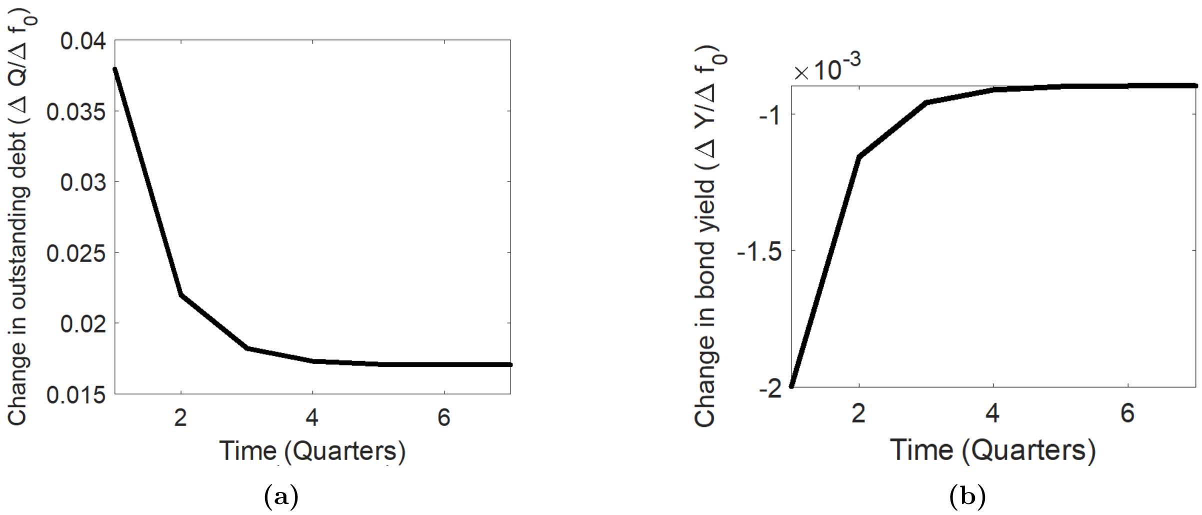


Figure 1.7: Response of the bond quantity and price to a 1% inflow to mutual funds. This figure presents the dynamic response of (a) the bond quantity, and (b) the bond yield, which is inversely related to the bond price. The parameters used to generate the figures are provided in Table 1.12.

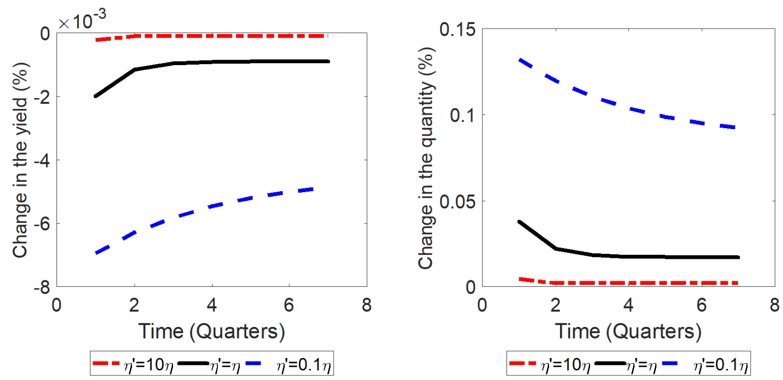
1.4.4 Impact of the massive outflows in March and April 2020

In March and April 2020, municipal mutual funds experienced a massive outflow of about 5% of their assets. We can use the framework developed here to gauge their impact on the borrowing behavior of the municipal governments. Specifically, suppose the funds receive an uninformative outflow of $W_0^F f_0$ from their investors at $t = 0$. It means that at $t = 0$, the allocation to mutual funds by the indirect investors is:

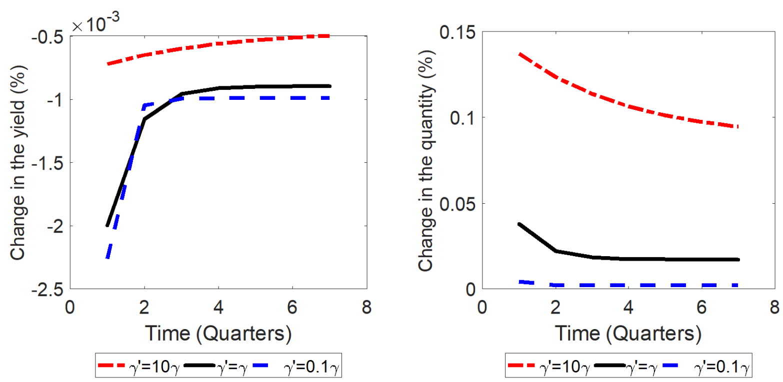
$$\alpha_0^{ID} = \lambda \alpha_{0-}^{ID} + (1-\lambda)\hat{\alpha}_0^{ID-Reb} + f_0 \alpha_{0-}^{ID} \tag{1.38}$$

For simplicity, assume that there is no investment from the investors outside the economy, i.e., $F_t^{Out} = 0$. Since the outflow is temporary, the bond price and quantity should eventually converge to their steady-state values. As a result, $\hat{\alpha}_{0-}^{ID} = \hat{\alpha}_{0-}^{D} = 0$. Therefore, the initial condition in this case is

(a) Changing the demand elasticity (η)



(b) Changing the supply elasticity (γ^{-1})



(c) Changing the investor inattention parameter (λ)

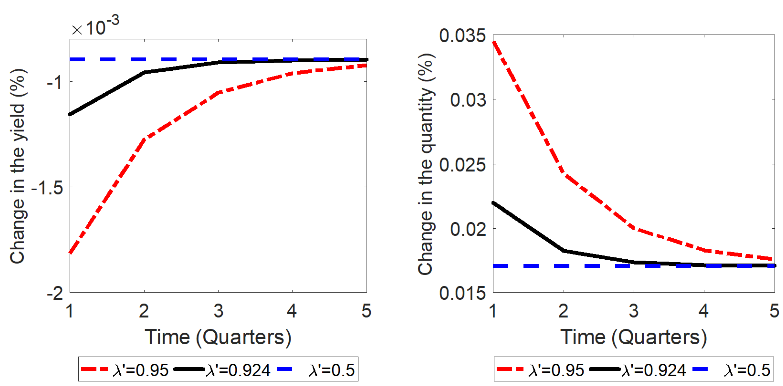


Figure 1.8: The response of bond yield and quantity for various choices of parameters

Figure 1.8, continued: This figure presents the dynamic response of the bond yield and quantity if we were choosing substantially smaller (dashed blue lines) or larger parameters (dashed red lines) (a) for the bond demand elasticity (η), (b) bond supply elasticity (γ^{-1}), (c) household inattention parameter (λ). The black solid lines display the results under the calibration parameters. The parameters used to generate the figures are provided in Table 1.12.

$$S^F \eta^F r_0 + S^F f_0 + M \sum_{s=0}^{\infty} \nu^s r_s = (1 + \gamma^- 1)\hat{p}_0. \tag{1.39}$$

We can use the condition $\hat{p}_t = \kappa^t \hat{p}_0$ to solve for the price dynamics

$$\hat{p}_t = \frac{S^F \kappa^t}{1 + \gamma^{-1} + (1 + dp - \kappa)(\frac{M\nu}{1-\nu\kappa} + M + S^F \eta^F)} f_0. \tag{1.40}$$

Figure 1.9 illustrates the price dynamics for different values of the mutual funds' demand sensitivity (η^F).

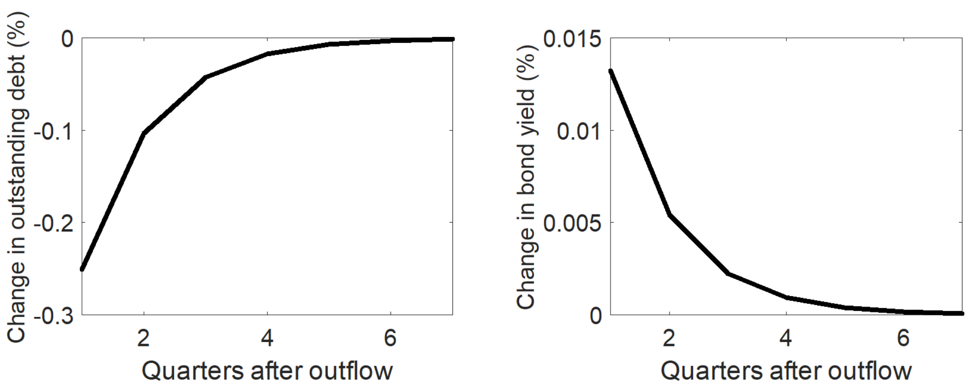


Figure 1.9: This figures displays the response of the bond quantity and yield at issuance to a 5% temporary outflow from mutual funds. The parameters in Table 1.12 are used for this illustration.

We see that the model predicts a decline in the bond issuance by 0.25% of the market size, equivalent to 10.5 billion dollars. It explains 46% of the 23 billion dollar actual decline in the bond issuance in March and April 2020 compared to January and February 2020.[21]

21. The number is obtained from the Securities Industry and Financial Markets Association (SIFMA)

Note that the it is an out-of-sample exercise since the study period ends at 2019. This result suggests that investment frictions faced by mutual funds could drastically affect the bond issuance by municipal governments during economic downturns. Therefore, federal financial assistance is particularly needed when the funds face massive outflows.

1.5 Conclusion

The municipal bond market provides the funding needed for most infrastructure projects in the US. Hence, the well-functioning of this capital market is imperative for the quality of health care services, schools, water, roads, recreational areas, and many other public services that US residents receive daily. The market effectiveness is especially crucial during economic downturns when raising tax rates is a contentious policy choice. The empirical findings in this paper indicate that the municipal bond market is not resilient in the short run, and the market effectiveness depends on the capital available to mutual funds. Therefore, market interventions by the federal government are necessary in times of crisis to ensure that municipal governments have the resources to manage the situation effectively. The federal government's support is even more needed if the crisis accompanies massive outflows from mutual funds. It was the situation in the first few months following the outset of the COVID-19 pandemic.

What makes mutual funds special for policy decisions? First, they control a large portion of the municipal bond market. As of the second quarter of 2021, they own more than 21% of the outstanding municipal debt. Second, this number chiefly reflects the ownership by "municipal" mutual funds, which are the specialized players in the municipal bond market. It is their responsibility to pay attention to market trends and act accordingly on behalf

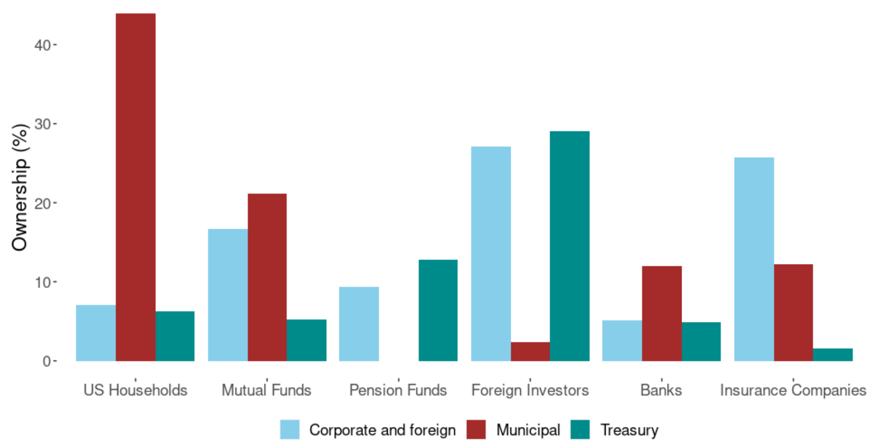
database.

of their clients. As the mutual funds receive more outflows, such attentive intermediaries manage a smaller portion of the market, making the market less resilient as a whole.

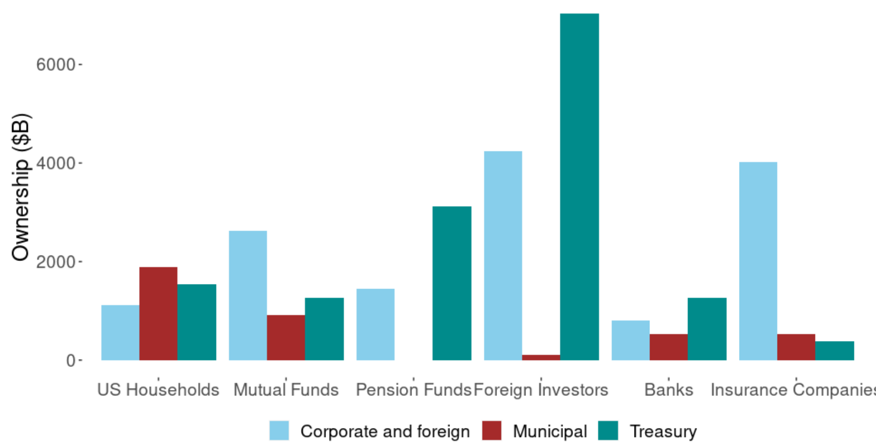
My study also puts forward a model of the municipal bond market that captures the key elements representing its supply and demand side. I find that the model performs quite well in predicting the impact of the massive outflows in March 2020. Thus, the model can be employed to analyze and evaluate different economic policies targeting the US infrastructure. For instance, investors are concerned about how the recently proposed infrastructure bill, which increases the federal support for infrastructure projects by about 500 billion dollars (11.6% of the overall size of the outstanding municipal debt)[22], would affect the issuance of new municipal bonds in the future. The model can provide an answer, as it is calibrated by the recent responses of municipal governments to smaller funding shocks, i.e., mutual fund flows.

22. Senate Passes $1 Trillion Infrastructure Bill, Handing Biden a Bipartisan Win, August 2021

Appendix 1.A: Supplementary Figures and Tables



(a) Ownership by percentage



(b) Ownership by dollar value

Figure 1.A.1: Ownership of different bond categories. Data are obtained from the US Financial Accounts database available at the Federal Reserve Board of Governors webpage. They present the ownership for the end of 2021Q1.

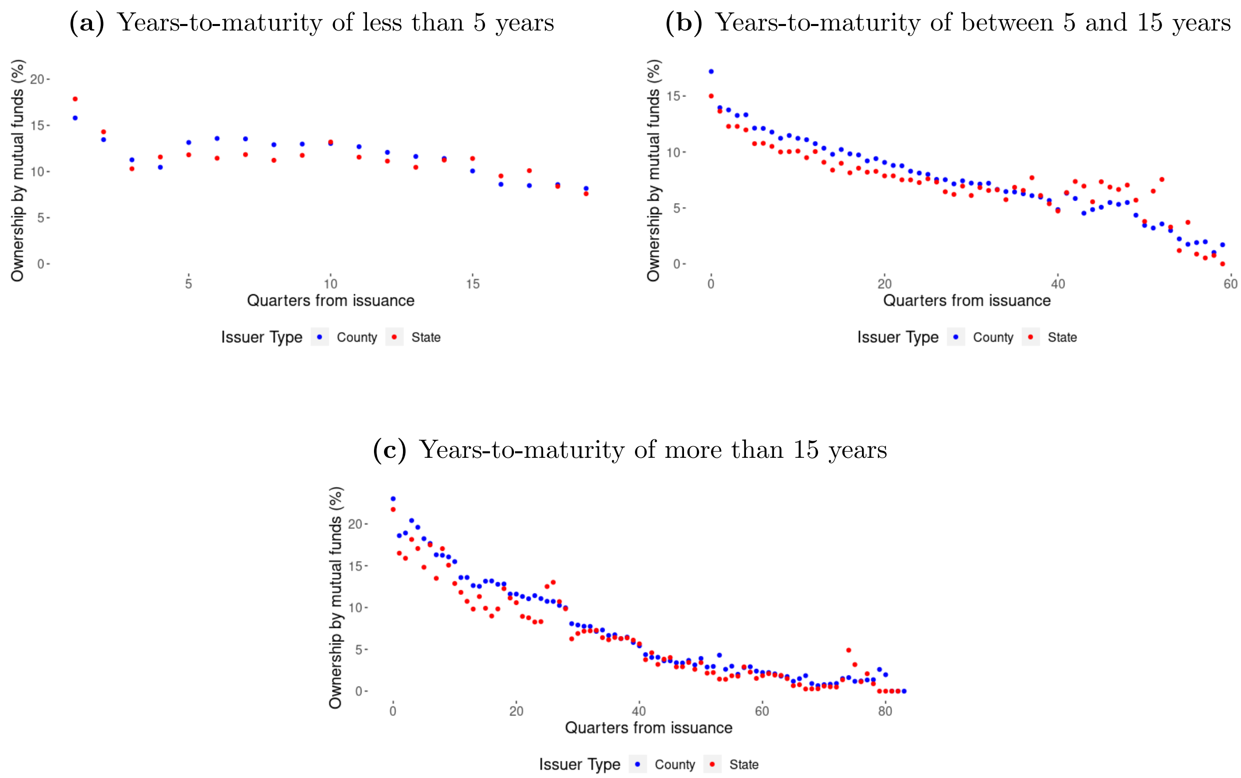
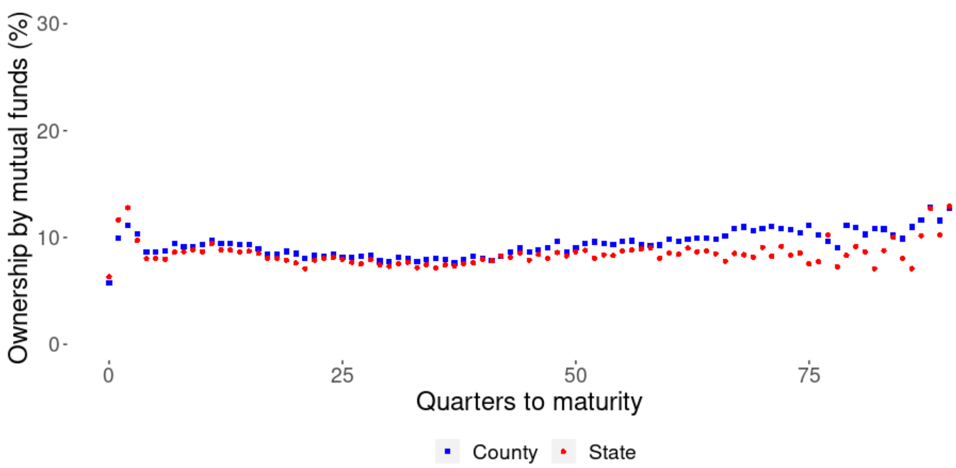


Figure 1.A.2: Mutual funds' municipal bond ownership by quarters after issuance for short-term, medium-term, and long-term bonds. The figure depicts the average mutual fund ownership for each number of quarters after issuance for the period between 2009Q1-2021Q1 and municipal bonds that were issued in 2000 and after. Panel (a) provides the mutual fund ownership for bonds with maturity of less than 5 years at issuance. Likewise, Panels (b) and (c) provide the mutual fund average ownership for medium-term (5 years - 15 years) and long-term bonds (more than 15 years), respectively. The historical bond issuance data are obtained from Bloomberg for US state governments, county governments with at least 100K population, and their subsidiaries, upon availability. Mutual funds' holding data of municipal bonds are obtained from CRSP.

(a) All bonds



(b) Bonds with less than one year after issuance

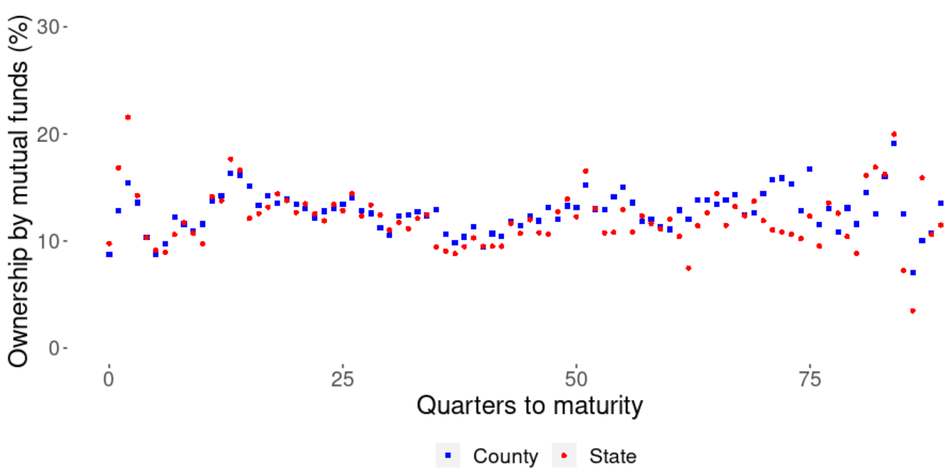


Figure 1.A.3: Mutual funds' municipal bond ownership by quarters to maturity. The figure depicts the average mutual fund ownership based on the quarters to maturity for the period between 2009Q1-2021Q1 and municipal bonds that were issued in 2000 and after. Panel (a) provides the mutual fund ownership for all bonds. Panel (b) provides the mutual fund average ownership for bonds with less than one year after their issuance . The historical bond issuance data are obtained from Bloomberg for US state governments, county governments with at least 100K population, and their subsidiaries, upon availability. Mutual funds' holding data of municipal bonds are obtained from CRSP.

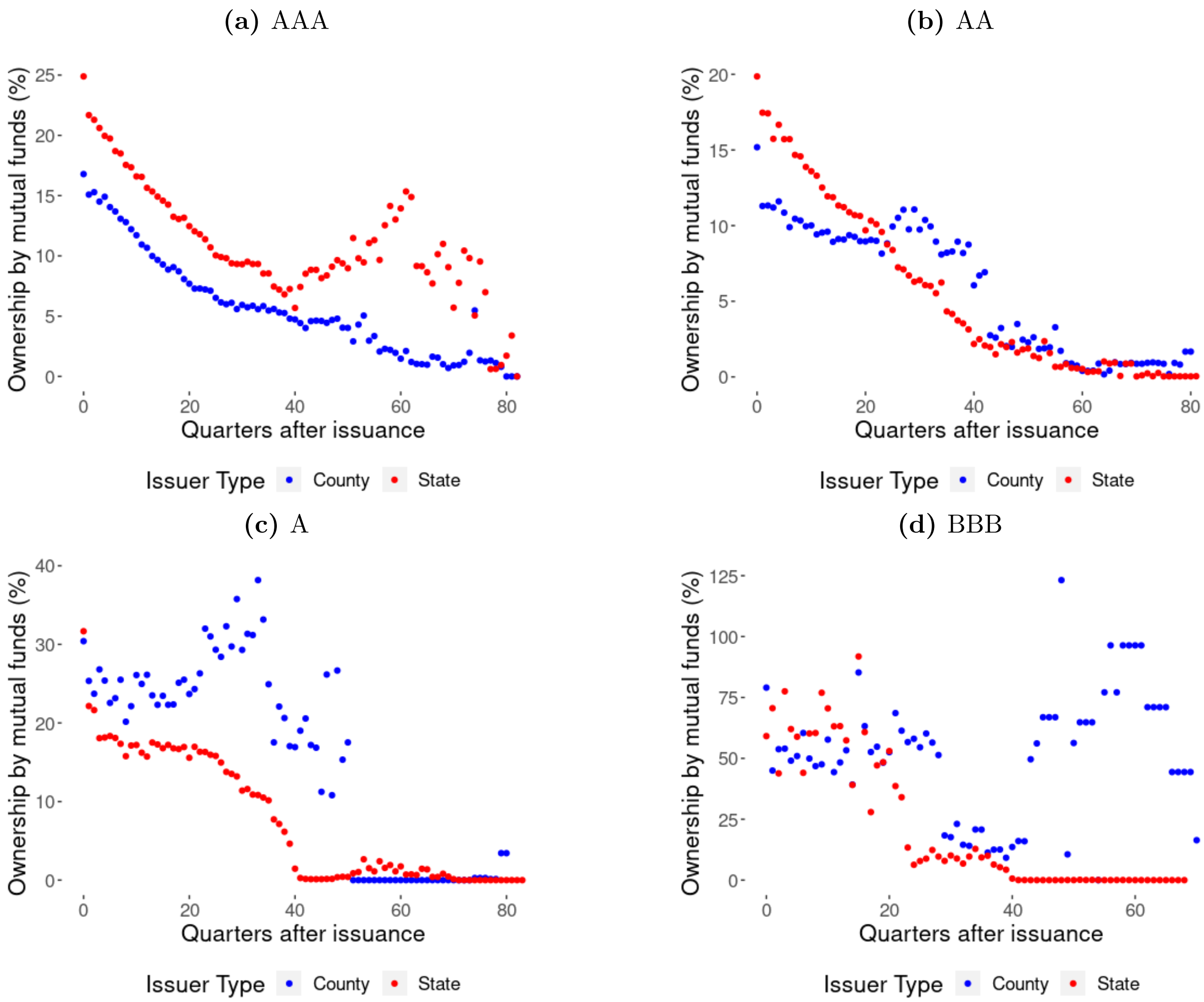


Figure 1.A.4: Mutual funds' municipal bond ownership by quarters after issuance for different bond ratings. The figure depicts the average mutual fund ownership for each number of quarters after issuance for the period between 2009Q1-2021Q1 and municipal bonds that were issued in 2000 and after. The panels provide the ownership dynamics for different investment-grade credit rating categories. The historical bond issuance data are obtained from Bloomberg for US state governments, county governments with at least 100K population, and their subsidiaries, upon availability. Mutual funds' holding data of municipal bonds are obtained from CRSP.

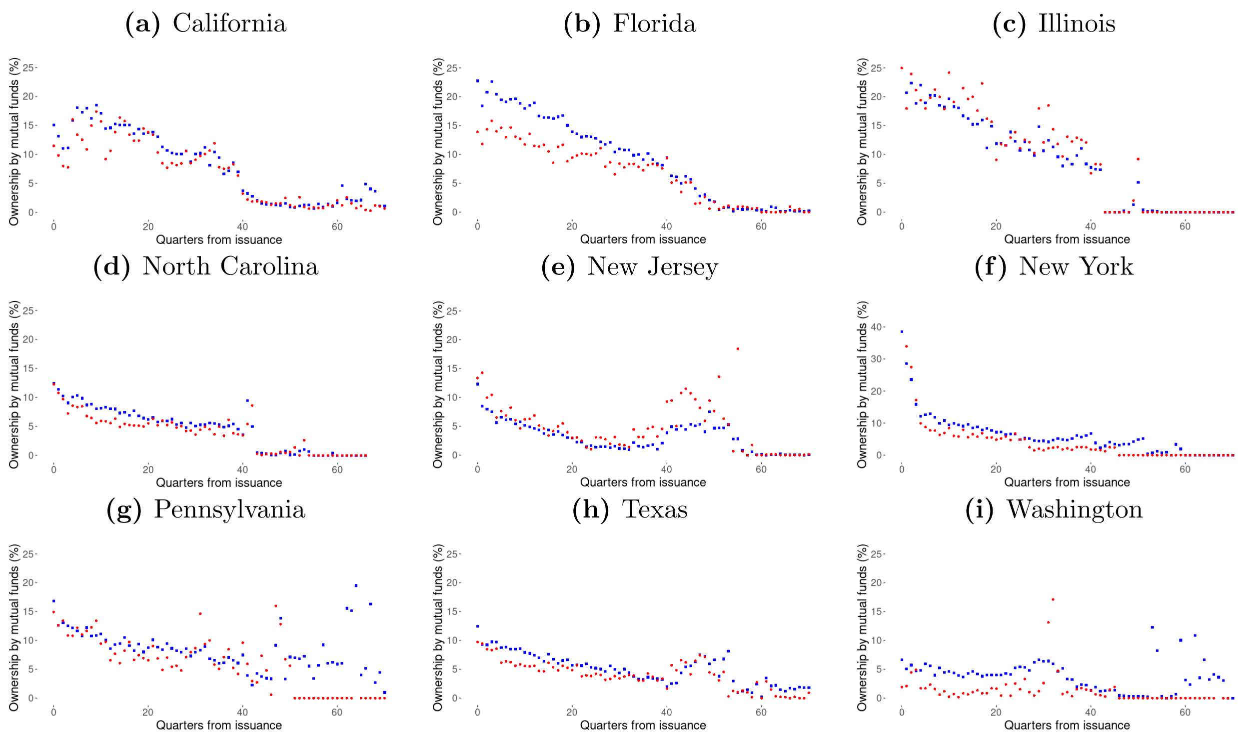


Figure 1.A.5: Mutual funds' municipal bond ownership by quarters after issuance for some US states. The figure depicts the average mutual fund ownership for each number of quarters after issuance for the period between 2009Q1-2021Q1 and municipal bonds that were issued in 2000 and after. The panels provide the ownership dynamics for different states. In each panel, the blue squares represent the mutual fund ownership for bonds issued by county governments and their subsidiaries (Counties with a population of at least 100K). Red circles represent the ownership for the bonds issued by the state government and its subsidiaries. The historical bond issuance data are obtained from Bloomberg. Mutual funds' holding data of municipal bonds are obtained from CRSP.

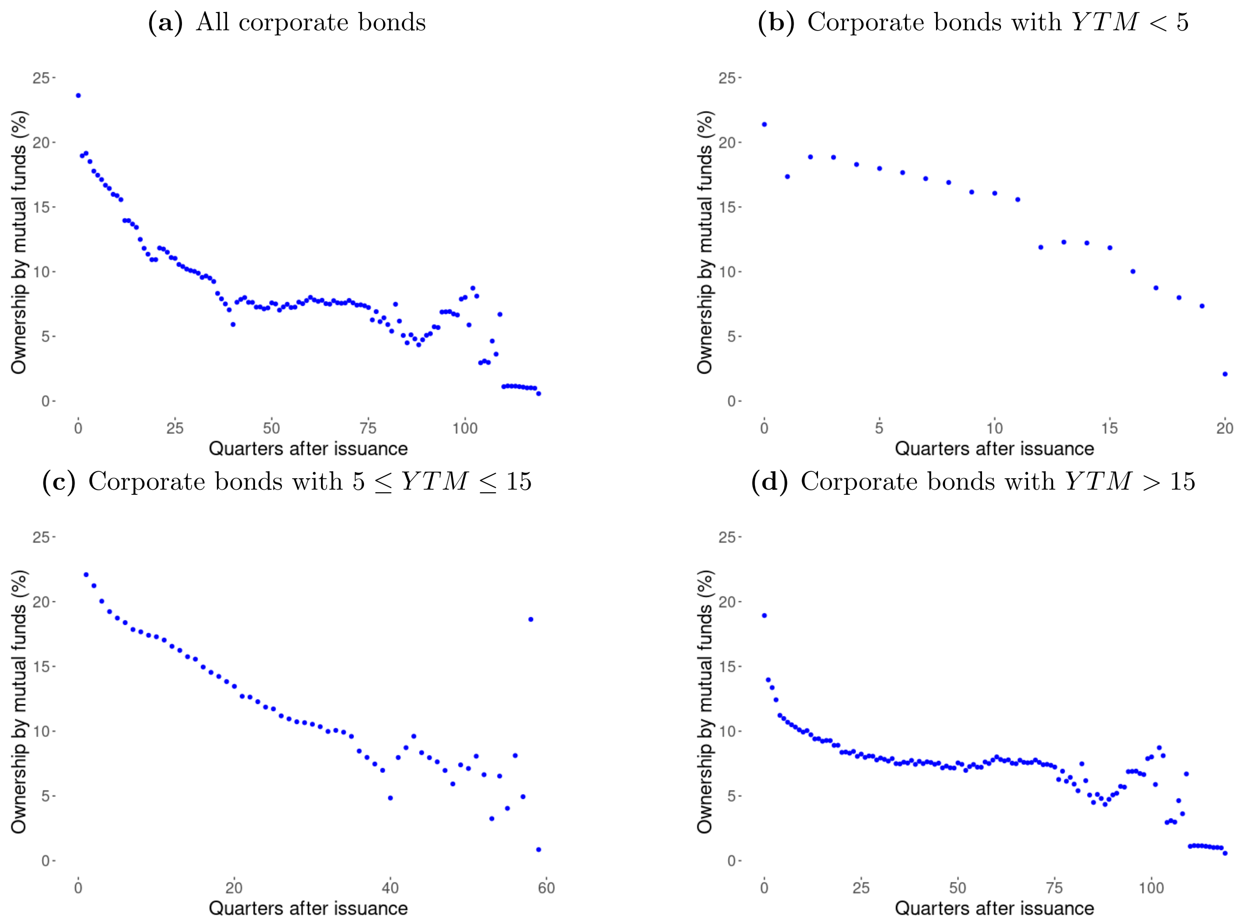


Figure 1.A.6: Mutual funds' ownership of Corporate bonds by quarters after issuance. The figure depicts the average mutual fund ownership for corporate bonds issued in 2000 and after for the period between 2009Q1-2021Q1. The holding data are obtained from CRSP, and covers more than 90% of the mutual funds' holding of corporate bonds, as reported by the US financial accounts. The panels provide the ownership dynamics for bonds with different years-to-maturity at origination. The historical bond issuance data are obtained from Mergent Fixed Income Securities Database (FISD).

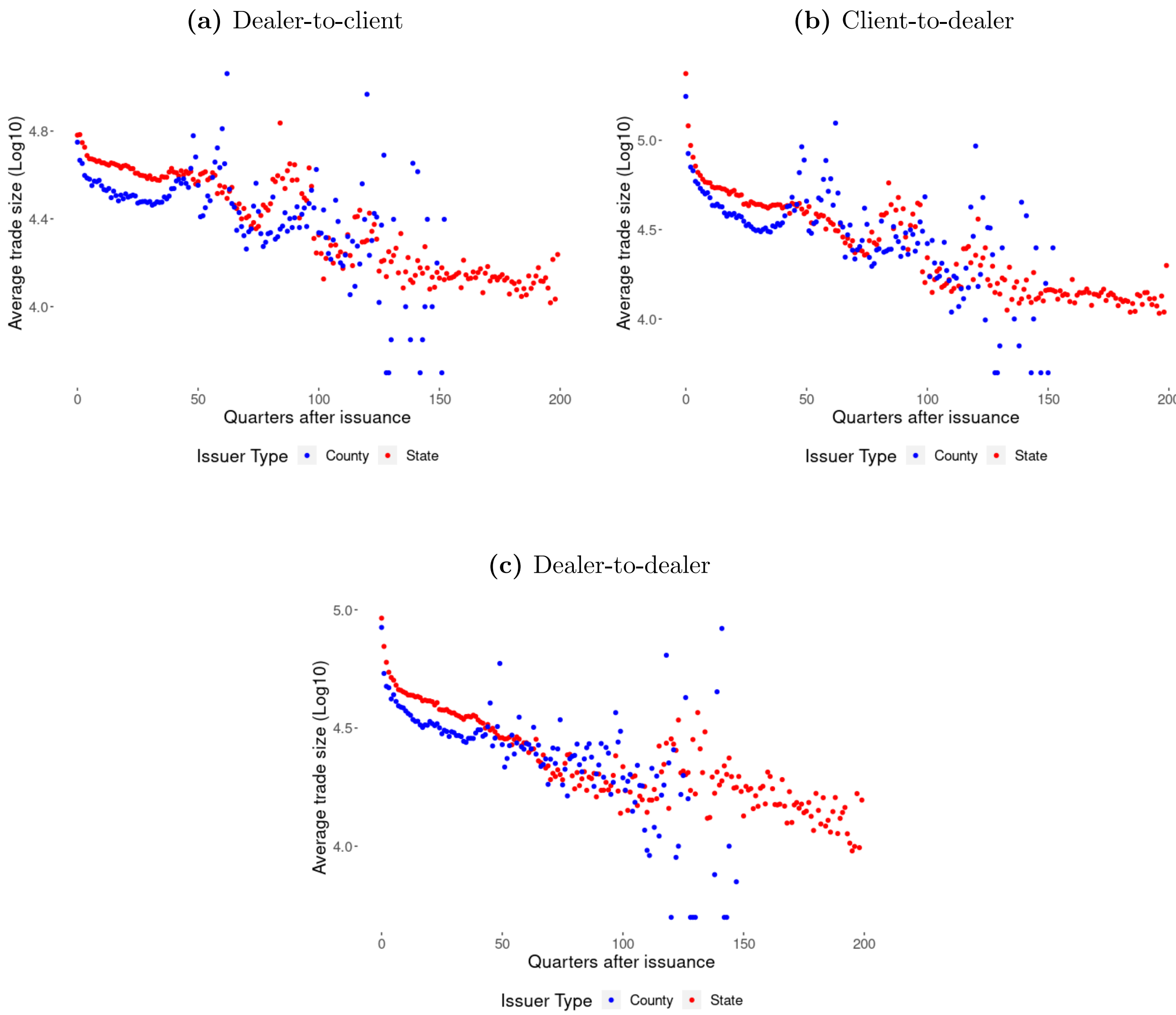


Figure 1.A.7: Average trade size by the number of quarters after issuance and type of transaction. This figure displays the average log-size (logarithm in base 10) of a) dealer-to-client, b) client-to-dealer, c) dealer-to-dealer transactions of municipal bonds. In a dealer-to-client transaction, the client (an institutional or retail investor) purchases bonds from a dealer. Client-to-dealer and dealer-to-dealer transactions are defined likewise. The transaction data are obtained from MSRB database. The blue dots represents the transactions of the bonds of the selected county governments. The red dots represents the average trade size for bonds issued by state governments.

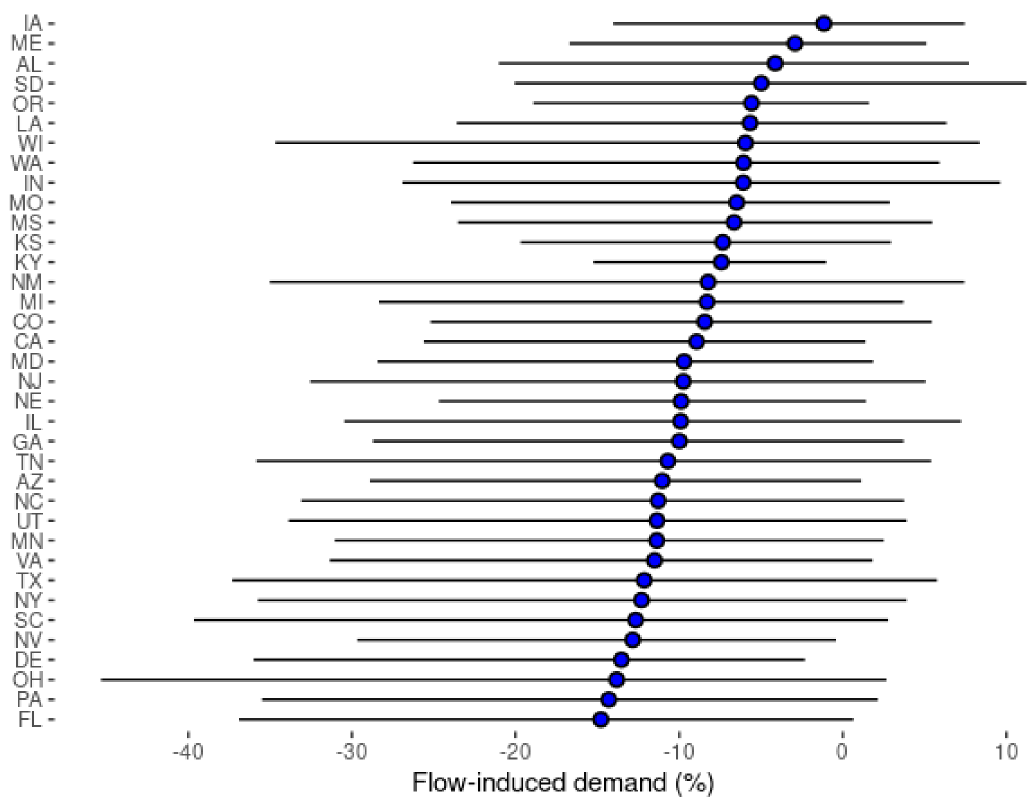


Figure 1.A.8: Average and distribution of flow-induced demand for the selected counties of each state. Flow-induced demand is computed based on Equation 1.5 for each county at every quarter between 2009 and 2019. The filled circle presents the average value, and the horizontal line presents the range between the 10th percentile and 90th percentile observations.

	$\Delta Inv^{Par}_{f,c,t}$	
	(Inflow Sample)	(Outflow Sample)
$Flow_{f,t}$	-0.002	0.002
	(0.001)	(0.190)
$\text{SIG}_{f,c,t-1}$	-1.360^{***}	-1.136^{***}
	(0.073)	(0.078)
$\text{SIG}_{f,c,t-1} \times Flow_{f,t}$	-0.257	1.962^{**}
	(0.195)	(0.823)
$\mathbb{I}\{\text{SIG}_{f,c,t-1} > \text{OWN}_{f,c,t-1}\}$	-0.066^{***}	-0.030
	(0.025)	(0.027)
$\mathbb{I}\{\text{SIG}_{f,c,t-1} > \text{OWN}_{f,c,t-1}\} \times Flow_{f,t}$	0.278^{***}	0.980^{***}
	(0.081)	(0.283)
Observations	155,274	165,712
R^2	0.035	0.033
Quarter FE	Y	Y
Fund FE	Y	Y

Table 1.A.1: This table presents what the past and current investments predict about the response of a fund to inflows and outflows. The dependent variable ($\Delta \log Inv_{f,c,t}$) is the change in the log of par-value investment of fund f in the bonds issued by county government c at the end of quarter t. The investments are aggregated at government level. 1$ is added to all investments to avoid infinite values for zero positions. $Flow_{f,t}$ is the quarterly flow to fund f between $t-1$ and t, as a fraction of the merger-adjusted total net asset value at the end of quarter $t-1$. $OWN_{f,c,t}$ is the fraction of total par-value investment of mutual funds in government c at t that is owned by fund f. $SIG_{c,f,t}$ is the maximum value of $OWN_{c,f,t'}$ in the twelve quarters ending at t.

	Inflow Sample			Outflow Sample		
	$\Delta\log(Inv^{Par}_{f,c,t}+10)$	$\Delta\log(Inv^{Par}_{f,c,t}+10^2)$	$\Delta\log(Inv^{Par}_{f,c,t}+10^3)$	$\Delta\log(Inv^{Par}_{f,c,t}+10)$	$\Delta\log(Inv^{Par}_{f,c,t}+10^2)$	$\Delta\log(Inv^{Par}_{f,c,t}+10^3)$
$Flow_{f,t}$	−0.001	−0.001	−0.001*	0.317**	0.258**	0.198**
	(0.001)	(0.001)	(0.001)	(0.134)	(0.109)	(0.085)
$SIG_{f,c,t-1}$	−0.141**	−0.115**	−0.089**	−0.251***	−0.202***	−0.152***
	(0.059)	(0.048)	(0.037)	(0.063)	(0.052)	(0.040)
$SIG_{f,c,t-1} \times Flow_{f,t}$	0.412***	0.337***	0.261***	1.838***	1.550***	1.259***
	(0.132)	(0.107)	(0.082)	(0.668)	(0.545)	(0.423)
Observations	155,274	155,274	155,274	165,712	165,712	165,712
R^2	0.032	0.032	0.032	0.030	0.030	0.030
Quarter FE	Y	Y	Y	Y	Y	Y
Fund FE	Y	Y	Y	Y	Y	Y

Table 1.A.2: This table presents the relationship between the fund flows and investment adjustments for different definitions of the investment change. $Flow_{f,t}$ is the quarterly flow to fund f between $t-1$ and t, as a fraction of the merger-adjusted total net asset value at the end of quarter $t-1$. $SIG_{c,f,t}$ is the maximum ownership in the twelve quarters ending at t.

	$Flow_{f,t}$		
	(1)	(2)	(3)
$Ret_{f,t-1}$	−0.951	−1.309	−0.177
	(2.518)	(2.508)	(4.149)
$Ret_{f,t-2}$	0.624	0.294	−1.261
	(2.401)	(2.349)	(4.005)
$Ret_{f,t-3}$	−1.752	−1.885	0.046
	(2.472)	(2.401)	(4.247)
$Ret_{f,t-4}$	−1.005	−1.044	−2.615
	(2.472)	(2.419)	(4.142)
Observations	84,887	84,887	84,887
Projected R^2	0.00001	0.00001	0.00001
Fund FE	N	Y	Y
Quarter FE	N	N	Y

Table 1.A.3: This table presents how the past performance of a fund impacts its current flow. $Flow_{f,t}$ is the flow that fund f receives at quarter t as a fraction of its total assets. $Ret_{f,t}$ is the fund return on its assets in quarter t. $Flow_{f,t}$ is the quarterly flow to fund f between $t-1$ and t, as a fraction of the merger-adjusted total net asset value at the end of quarter $t-1$. The fund flows are constructed from the fund data available at CRSP.

	$Flow_{f,t}$			
	(1)	(2)	(3)	(4)
$Flow_{f,t-1}$	-0.0001	0.001	-0.0002	0.001
	(0.002)	(0.003)	(0.002)	(0.003)
$Flow_{f,t-2}$		0.001		0.001
		(0.001)		(0.001)
$Flow_{f,t-3}$		0.00001		0.00003
		(0.0002)		(0.0002)
$Flow_{f,t-4}$		0.00000		-0.00000
		(0.0001)		(0.0001)
Constant	0.414**	0.041***		
	(0.163)	(0.011)		
Observations	77,542	74,285	77,542	74,285
R^2	0.000	0.00001	0.001	0.001
Quarter FE	N	N	Y	Y

Table 1.A.4: This table examines the persistency of fund flows for municipal mutual funds, between 2009Q1-2019Q4. $Flow_{f,t}$ is the quarterly flow to fund f between $t-1$ and t, as a fraction of the merger-adjusted total net asset value at the end of quarter $t-1$. The fund flows are constructed from the fund data available at CRSP.

Appendix B: Proofs and Mathematical derivations

1.B.1 Proof of Lemma 1

I derive the optimal portfolio for the direct and indirect investors separately. Let $V_t^{D,R}(w^i)$ be the value function of a direct investor that rebalances at t. Likewise, $V_t^{D,NR}(w^i;\alpha^i)$ is the value function for a non-rebalancing investor with legacy portfolio share α in municipal bonds. Therefore, the following relationships between these two value functions hold:

$$\begin{aligned} V_t^{D,R}(w^i) = \quad \max_{\alpha} \quad & \delta \log\{R^F + \alpha(R^D - R^F)\} \\ & + (1-\delta)\Big\{(1-x)\log\{R^t + \alpha(R_t - R^F)\} \\ & + x(1-\lambda)V_{t+1}^{D,R}(w^i(R^t + \alpha(R_t - R^F))) \\ & + x\lambda V_{t+1}^{D,NR}(w^i(R^t + \alpha(R_t - R^F));\alpha)\Big\} \end{aligned} \tag{1.B.1}$$

$$\begin{aligned} V_t^{D,NR}(w^i;\alpha) = \quad & \delta \log\{R^F + \alpha(R^D - R^F)\} \\ & + (1-\delta)\Big\{(1-x)\log\{R^t + \alpha(R_t - R^F)\} \\ & + x(1-\lambda)V_{t+1}^{D,R}(w^i(R^t + \alpha(R_t - R^F))) \\ & + x\lambda V_{t+1}^{D,NR}(w^i(R^t + \alpha(R_t - R^F));\alpha)\Big\} \end{aligned} \tag{1.B.2}$$

By combining the equations above, one can guess and verify that the value functions are logarithmic in wealth, and the optimal portfolio for the rebalancing investors at t can be characterized as below:

$$\alpha_t^{D-Reb} = \arg\max_{\alpha} \frac{\delta}{1-\nu} \log\{R^F + \alpha(R^D - R^F)\} + (1-\delta)\sum_{s=0}^{\infty} \nu^s \log\{R^F + \alpha(R_{t+s} - R^F)\} \tag{1.B.3}$$

,where $\nu = x\lambda(1-\delta)$. The only difference for the indirect investors is that by investing fraction α_t^i in the funds at t, their return is $R^F + \alpha_t^i \alpha^F(R_t)(R^D - R^F)$ if the government defaults and $R^F + \alpha_t^i \alpha^F(R_t)(R_t - R^F)$ otherwise. Since the default risk is the only source of uncertainty before the government's default, the funds' return has a perfectly predictable exposure of $\alpha^F(R_{t+s})$ to the bond return in future. That said, the optimal allocation to the funds for the indirect investors is:

$$\begin{aligned} \alpha_t^{ID-Reb} = \arg\max_{\alpha} \delta \sum_{s=0}^{\infty} \nu^s \log\{R^F + \alpha\alpha^F(R_{t+s})(R^D - R^F)\} \\ + (1-\delta)\sum_{s=0}^{\infty} \nu^s \log\{R^F + \alpha\alpha^F(R_{t+s})(R_{t+s} - R^F)\} \end{aligned} \tag{1.B.4}$$

1.B.2 Derivation of Equation 1.22

Let y be the partial impact of the inflow on the wealth ratio between the indirect and direct investors, that is

$$\lim_{t\to\infty} \frac{W_t^{ID}}{W_t^D} = \frac{W_0^{ID}}{W_0^D}(1 + y f_0) \tag{1.B.5}$$

Moreover, let α_{SS} is the optimal allocation to municipal bonds in the new steady-state. Therefore, by dividing both sides of (1.21) by W_t^R and considering the limiting case of $t \to \infty$, we get:

$$
\begin{aligned}
\phi^{-\gamma^{-1}} {P_{SS}}^{1+\gamma^{-1}} &= \alpha_{SS} \frac{(1+f_0)(1+yf_0)\frac{W_0^{ID}}{W_0^D}+1}{(1+yf_0)\frac{W_0^{ID}}{W_0^D}+1} \\
&\simeq (1+\frac{\frac{W_0^{ID}}{W_0^D}}{(1+yf_0)\frac{W_0^{ID}}{W_0^D}} f_0)\alpha_{SS} \simeq (1+S^F f_0)\alpha_{SS}
\end{aligned} \tag{1.B.6}
$$

In the approximations above, I use the fact that f_0 is small.

1.B.3 Log-linearizations

Bond return (Equation 1.27)

$$
\begin{aligned}
R_t = \frac{D+P_{t+1}}{P_t} &= \frac{D+P_{SS}(1+\hat{p}_{t+1})}{P_{SS}(1+\hat{p}_t} \\
= \frac{D}{P_{SS}}(1-\hat{p}_t)+\hat{p}_{t+1}-\hat{p}_t &= R_t + \underbrace{\hat{p}_{t+1}-(1+dp)\hat{p}_t}_{r_t}
\end{aligned} \tag{1.B.7}
$$

Portfolio allocations (Equations 1.30 and 1.31)

$$
\begin{aligned}
\alpha_{t+1}^D = \alpha_{SS}^*(1+\hat{\alpha}_{t+1}^D) &= \lambda\alpha_{SS}^*(1+\hat{\alpha}_t^D)+(1-\lambda)\alpha_{SS}^*(1+\hat{\alpha}_t^{D-Reb}) \\
\Rightarrow \hat{\alpha}_{t+1}^D &= \lambda\hat{\alpha}_t^D+(1-\lambda)\hat{\alpha}_t^{D-Reb}
\end{aligned} \tag{1.B.8}
$$

Now, I show that $\hat{\alpha}_t^{D-Reb} = (1-\nu)\eta\sum_{s=0}^{\infty}\nu^s r_{t+s}$. Define $r^D = \frac{R^D}{R^F}-1$ and $\pi_t = \frac{R_t}{R^F}-1$. Therefore, the first-order condition implies:

$$
\frac{\delta}{1-\nu}\frac{r^D}{1+\alpha_t^{D-Reb}r^D} + (1-\delta)\sum_{s=0}^{\infty}\nu^s\frac{\pi_{t+s}}{1+\alpha_t^{D-Reb}\pi_{t+s}} = 0 \tag{1.B.9}
$$

The equation above can be approximated around the steady-state values as below:

$$-\frac{\delta}{1-\nu}\frac{\alpha_S S^* {r^D}^2}{(1+\alpha_{SS} r^D)^2}\hat{\alpha}_t^{D-Reb} - (1-\delta)\sum_{s=0}^{\infty}\frac{\alpha\pi_{t+s}^2}{(1+\alpha\pi_{t+s})^2}\hat{\alpha}_t^{D-Reb} + (1-\delta)\sum_{s=0}^{\infty}\nu^s\Delta\pi_{t+s} \tag{1.B.10}$$

,where $\Delta\pi_{t+s} \equiv \frac{R_{t+s}}{R^F} - \frac{R_{SS}}{R^F} = \frac{r_{t+s}}{R^F}$. By some rearrangements, we get:

$$\hat{\alpha}_t^{D-Reb} = (1-\nu)\underbrace{\frac{(1-\delta)\frac{1}{(1+\alpha\pi_{SS})^2}}{\delta\frac{\alpha_{SS} {r^D}^2}{(1+\alpha_{SS} r^D)^2} + (1-\delta)\frac{\alpha_{SS}\pi_{SS}^2}{(1+\alpha_{SS}^*\pi_{SS})^2}}}_{\eta R^F}\sum_{s=0}^{\infty}\nu^s\Delta\pi_{t+s} = (1-\nu)\eta\sum_{s=0}^{\infty}\nu^s r_{t+s} \tag{1.B.11}$$

The proof is similar for the case of the indirect investors.

Initial conditions (Equation 1.32)

By the definition of η in (1.23), we have:

$$\hat{\alpha}_{0-}^D = (\frac{\alpha_{0-}^D}{\alpha_{SS}^D} - 1) \simeq \eta(R_{0-} - R_{SS}) \simeq \eta dp(\frac{P_{SS}}{P_{0-}} - 1) = \frac{\eta S^F dp}{\eta dp + \gamma^{-1} + 1} f_0 \tag{1.B.12}$$

Similarly, for the indirect investors:

$$\alpha_{0-}^{ID} = \frac{\alpha_{0-}^*}{\alpha_{0-}^F} \Rightarrow \hat{\alpha}_{0-}^{ID} = \hat{\alpha}_{0-}^D - \eta^F(R_{0-} - R_{SS}) = \frac{(\eta - \eta^F)S^F dp}{\eta dp + \gamma^{-1} + 1} f_0 \tag{1.B.13}$$

Market clearing (Equation 1.33)

As stated in (1.21), the market clearing condition is

$$\alpha_t^F(R_t)(1+f_0)\alpha_t^{ID}W_t^{ID} + W_t^D\alpha_t^D = P_tQ_t = P_tq_tW_t^R \tag{1.B.14}$$

Define $w_t^{ID} \equiv \frac{W_t^{ID}}{W_t^R}$, and $w_t^D \equiv \frac{W_t^D}{W_t^R}$. By log-linearizng (1.B.14) around the steady state values, we get:

$$S^F\hat{\alpha}_t^F + S^F\hat{\alpha}_t^{ID} + S^F\hat{w}_t^{ID} + S^D\hat{\alpha}_t^D + S^D\hat{w}_t^D = \hat{p}_t + \hat{q}_t \tag{1.B.15}$$

Note that $S^F = w_0^{ID}$ and $S^D = w_0^D$ by definition. Therefore:

$$w_t^{ID} + w_t^D = 1 \Rightarrow S^F\hat{w}_t^{ID} + S^D\hat{w}_t^D = 0 \tag{1.B.16}$$

The intuition for the equation above is that the government absorbs the fluctuations in the residents' wealth that are caused by the inflow. We get the provided market clearing condition by plugging this result into (1.B.15).

Price dynamics (Equations 1.36 and 1.34)

By expanding Equation 1.33 by using Equations 1.30 and 1.31, we get:

$$\begin{aligned} S^F\hat{\alpha}_t^F + S^F(\lambda\hat{\alpha}_{t-1}^{ID} + (1-\lambda)(1-\nu)(\eta-\eta^F)\sum_{s=0}^{\infty}\nu^s r_{t+s}) \\ + S^D(\lambda\hat{\alpha}_{t-1}^D + (1-\lambda)(1-\nu)\eta\sum_{s=0}^{\infty}\nu^s r_{t+s}) = \hat{p}_t + \hat{q}_t \end{aligned} \tag{1.B.17}$$

We can employ the market-clearing condition at $t-1$ (provided $t \geq 1$), and bond supply equation (1.26), to substitute for $\hat{\alpha}^D_{t-1}$, $\hat{\alpha}^{ID}_{t-1}$, and $\hat{q}_t$. Furthermore, we know that $\hat{\alpha}^F_t = \eta^F r_t$. Therefore:

$$S^F \eta^F r_t + M \sum_{s=0}^{\infty} \nu^s r_{t+s} + \lambda(1+\gamma^{-1})\hat{p}_{t-1} - \lambda S^F \eta^F r_{t-1} = (1+\gamma^{-1})\hat{p}_t \tag{1.B.18}$$

,where $M = (1-\lambda)(1-\nu)(\eta - S^F \eta^F)$. We can solve for $M \sum_{s=1}^{\infty} \nu^s r_{t+s}$ by writing the same equation for time $t+1$. After some rearrangements, it yields us the equation below:

$$\begin{aligned} &\nu(1+\gamma^{-1})\hat{p}_{t+1} - (1+\lambda\nu)\hat{p}_t + \lambda(1+\gamma^{-1})\hat{p}_{t-1} - S^F \eta^F \nu r_{t+1} \\ &+ (\lambda\nu S^F \eta^F + S^F \eta^F + M) r_t - \lambda S^F \eta^F r_{t-1} = 0 \end{aligned} \tag{1.B.19}$$

Equation 1.34 can be obtained by using the fact that $r_t = \hat{p}_{t+1} - (1+dp)\hat{p}_t$.

1.B.4 Proof of Lemma 2

Let $C(z)$ be the characteristic polynomial of system 1.34. $C(z)$ can be written as below:

$$C(z) = S^F \eta^F (1-\nu z)(z-\lambda)(z-1-dp) + Mz(z-1-dp) - (1+\gamma^{-1})(1-\nu z)(z-\lambda) \tag{1.B.20}$$

It is straightforward to check $C(z) > 0$ for $z < 0$. Furthermore, $C(1) < 0$, $C(\max\{1+dp, \nu^{-1}\}) > 0$, and $C(z)$ converges to $-\infty$ as z goes to ∞. Therefore, $C(z)$ should have one root in each one of $(0,1)$, $(1, \max\{1+dp, \nu^{-1}\})$, and $(\max\{1+dp, \nu^{-1}\}, \infty)$. It completes the proof. In fact, one can show that the smallest root should be less than λ as well.

1.B.5 Derivation of short-run demand elasticity

In the derivation of the long-run demand elasticity, we consider the impact of a permanent price change on the steady-state bond demand. To make the short-term demand elasticity comparable, I consider the impact of the same permanent price change on the bond demand at $t = 0$. Specifically, let the bond price increase permanently by $P_{SS}\Delta p$ from its steady-state value P_{SS}. Therefore, the bond return falls by $dp\Delta p$ in every period. As such, the change in the bond demand at $t = 0$ is

$$\begin{aligned}\Delta q^{demand} =& \eta^F S^F r_0 + M \sum_{s=0}^{\infty} \nu^s r_s + \lambda(\hat{\alpha}_{0-}^{D} + \hat{\alpha}_{0-}^{ID}) \\ =& -(\eta^F S^F + \frac{M}{1-\nu}) dp\Delta p + \lambda(\hat{\alpha}_{0-}^{D} + \hat{\alpha}_{0-}^{ID}) \\ =& -((1-\lambda)\eta + \lambda\eta^F S^F) dp\Delta p + \lambda(\hat{\alpha}_{0-}^{D} + \hat{\alpha}_{0-}^{ID}).\end{aligned} \tag{1.B.21}$$

When the investors frequently rebalance (i.e., $\lambda = 0$), the short-run and long-run demand elasticities coincide. Otherwise, the short-run demand elasticity is strictly smaller, provided $\eta^F = \eta$.

Appendix C: Additional discussions

1.C.1 Price and quantity dynamics with stochastic fund flows and funding need shocks

Our baseline setup only examines the impact of a one-time inflow on the market outcomes, featuring no uncertainty about the future flows or shocks to the government's funding need.

This section relaxes these assumptions. Specifically, suppose the government's objective function is modified as below:

$$\max_{\{Q_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta_G^t \{(1+u_t)P_t(\frac{Q_t}{W_t^R}) - \frac{\phi}{1+\gamma}(\frac{Q_t}{W_t^R})^{1+\gamma}\} \tag{1.C.1}$$

u_t is a supply shifter that represents fluctuations in the government's need for funding. For instance, u_t is higher in times of natural disasters, which intensify the government's need for borrowing. In the baseline case, since the focus is on the impact of fund flows on the market outcomes, I assume that $\tilde{u}_t$'s are i.i.d across the periods and normally distributed with mean zero and variance σ_u^2, prior to the government's default. Once the government defaults, all investors reoptimize their portfolio, and no shock happens afterward. Furthermore, before the default, the allocation of the indirect investors to the funds follows the process below:

$$\alpha_t^{ID} = \lambda \alpha_{t-1}^{ID} + (1-\lambda)\frac{\alpha_t^*}{\alpha_t^F} + f_t \alpha_{t-1}^{ID} \tag{1.C.2}$$

In (1.C.2), f_t is the random flow to mutual funds, which is i.i.d and normally distributed: $\tilde{f}_t \sim \mathcal{N}(0, \sigma_f^2)$. α_t^* and α_t^F are the optimal portfolio shares to municipal bonds for the investors and funds, respectively. In particular, suppose both investors and funds are myopic and maximize $\mathbb{E}_t v^J(R_t^J)$, $J \in \{I, F\}$, respectively, where:

$$\begin{aligned} v^J(R_t^J) &= \frac{1}{1-\zeta_J} R_t^{J\,1-\zeta_J} \qquad J \in \{I, F\} \\ R_t^J &= R^F + \alpha(R_t - R^F) \end{aligned} \tag{1.C.3}$$

If the government does not default at $t{+}1$, the return on its bond is R_t, it is R^D otherwise. One could show that the bond return conditional on no default is normally distributed with a constant variance. Furthermore, similar to the baseline case, the deviations of the optimal portfolios from the steady-state ones can be characterized as below:

$$\begin{aligned} \hat{\alpha}_t^* &= \eta \mathbb{E}_t[\hat{p}_{t+1}] - (1+dp)\hat{p}_t \\ \hat{\alpha}_t^F &= \eta^F \mathbb{E}_t[\hat{p}_{t+1}] - (1+dp)\hat{p}_t \end{aligned} \tag{1.C.4}$$

Therefore, the equations that characterize the dynamics should be modified as below:

$$\hat{q}_t = \gamma^{-1}\hat{p}_t + \gamma^{-1}u_t \tag{1.C.5}$$

$$S^F\hat{\alpha}_t^F + S^F\hat{\alpha}_t^{ID} + S^D\hat{\alpha}_t^D = \hat{p}_t + \hat{q}_t \tag{1.C.6}$$

$$\hat{\alpha}_t^F = \eta^F E_t[r_t] \tag{1.C.7}$$

$$\hat{\alpha}_t^D = \lambda\hat{\alpha}_{t-1}^D + (1-\lambda)\eta E_t[r_t] \tag{1.C.8}$$

$$\hat{\alpha}_t^{ID} = \lambda\hat{\alpha}_{t-1}^{ID} + (1-\lambda)(\eta - \eta^F)E_t[r_t] + f_t \tag{1.C.9}$$

$$E_t[r_t] = \mathbb{E}_t[\hat{p}_{t+1}] - (1+dp)\hat{p}_t \tag{1.C.10}$$

Proposition 1. *The deviation of the bond price from its steady-state value is given by:*

$$\hat{p}_t = \kappa\hat{p}_{t-1} + C_1 u_t + C_t u_{t-1} + Df_t \tag{1.C.11}$$

$$\begin{aligned} D &= \frac{S^F/K}{1+dp+\frac{1+\gamma^{-1}}{K} - \kappa} \\ C_2 &= \frac{\lambda\gamma^{-1}/K}{1+dp+\frac{1+\gamma^{-1}}{K} + \frac{\lambda S^F\eta^F}{K} - \kappa} \\ C_1 &= \frac{C_2 - \gamma^{-1}/K}{1+dp+\frac{1+\gamma^{-1}}{K} - \kappa} \end{aligned} \tag{1.C.12}$$

Proof. By combining equations 1.C.5-1.C.10, we get:

$$\mathbb{E}_t[\hat{p}_{t+1}] = (1 + dp + \frac{1+\gamma^{-1}}{K})\hat{p}_t - (\frac{\lambda(1+\gamma^{-1})}{K} + \frac{\lambda S^F \eta^F (1+dp)}{K})\hat{p}_{t-1} + \frac{\lambda S^F \eta^F}{K}\mathbb{E}_{t-1}[\hat{p}_t]$$
$$-\frac{\lambda\gamma^{-1}}{K}u_{t-1} + \frac{\gamma^{-1}}{K}u_t - \frac{S^F}{K}f_t \tag{1.C.13}$$

One can verify that Equation 1.C.11 satisfies (1.C.13).

□

Note that there is no forward-looking element in Equation 1.C.11 since the shocks are i.i.d. This implies that current shocks do not impact the investors' belief about the future flows or the government's funding need shocks. If the shocks followed an AR(1) process, we would have an additional term capturing that element in the price formation. Moreover, note that the bond quantity can be inferred from (1.C.5).

In a model with fully attentive investors, the equilibrium bond price and quantity should only depend on the current and expected future perturbations. Equation 1.C.11 reveals that is not the case here when investors exhibit sluggish portfolio behavior. It causes the previous perturbations to impact the current bond price and quantity, captured by the lag terms in Equation 1.C.11. An implication of this observation is that a temporary shock to the government's funding need propagates over time since the demand is not resilient enough to absorb the shock immediately.

1.C.2 Examination of municipal bonds' credit spread when the investors are inattentive

In this section, I present a model of bond pricing with investor inattention. The purpose of the model is to understand how the investor inattention impacts the credit spreads on municipal bonds. The model features a government that decides about its size of bond issuance, and the prices are determined endogenously. The bond is defaultable. The government faces two groups of investors: Attentive investors, who are present to invest since the time of issuance, and some inattentive investors, who learn about the new bond issue with a delay, and trade the bond dynamically afterwards. The delayed learning, in fact, is the only point of departure of the model from a full-participation benchmark, in which all investors actively participate throughout the bond's lifetime.

We see that the credit spread is larger in the presence of inattentive investors, since it results in a limited risk-sharing between inattentive and attentive investors. In fact, shortly after the issuance, when the bond is mostly held by the attentive investors, they should bear the entire default risk. As a result, they need a higher compensation for being over-exposed to the default risk of the issuing government. In particular, the model predicts that younger bonds command a higher risk-premium, since the limited risk-sharing friction is the most severe for those bonds. In fact, since the younger bonds a longer time to maturity all else equal, this prediction is consistent with the empirical observation that the term structure of credit spread is upward-sloping for municipal bonds Cestau, Hollifield, Li, and Schürhoff (2019).

1.C.3 Setup

Time (t) is continuous. A government issues Q units of zero-coupon bonds at $t = 0$ that mature at $t = T$. Each unit of bond pays one unit of wealth at maturity. The bond is defaultable. The government defaults on the bond according to a Poisson process, with parameter $\delta > 0$. δ captures the riskiness of the government's bond; There higher δ, the government is more likely to default prior to the maturity. Alternative to the government bond, risk-free assets are available with short-term return $r > 0$. The price of the government bond is denoted by $P(t)$, and solved endogenously.

The economy comprises of two types of investors: attentive and inattentive investors. Attentive investors are aware of the new issuance at $t = 0$. In contrast, inattentive investors gradually become aware of the issuance. The learning process is modeled by a Poisson process with parameter λ. Thus, by $t \geq 0$, fraction $1 - e^{-\lambda t}$ of the inattentive investors have become aware of the new issue, and continuously trade in both the government and risk-free bond market. The rest of the inattentive investors only invest in the risk-free asset prior to learning about the government bond. At $t = 0$, the overall wealth of attentive and inattentive investors are W_0^A and W_0^{IA}, respectively. A key implication of the gradual learning by inattentive investors is that attentive investors buy the bonds first at $t = 0$, and gradually resell the bonds to inattentive investors.[23]

Note that the model lies between two extreme well-studied cases: The classical full-participation case, in which all investors are aware of the bond since $t = 0$, and the limited-participation case (e.g., Basak and Cuoco (1998)) in which inattentive investors never learn about the bond ($\lambda \to 0$).[24]

The objective of all investors is to maximize their utility over their terminal wealth at T.[25] The utility function is assumed to be logarithmic for tractability reasons.

23. Duffie (2010) postulates a similar prediction about asset allocation dynamics when investors are heterogeneously attentive.

24. Note that the extreme case of $\lambda \to \infty$ does not correspond to the full-participation case, as even at the limit, only attentive investors participate at $t = 0$.

25. It is innocuous to assume that the terminal point is the same as the bond maturity since there is no uncertainty afterwards.

1.C.4 Equilibrium price and risk-premium

An investor's optimization problem can be characterized as below for the period he is aware of the bond:

$$\delta V = \max_{\alpha} \quad V_t + (rw + \alpha w \underbrace{(\frac{dP}{P} - r)}_{\text{risk premium}})V_w + \delta \underbrace{\log((1-\alpha)e^{r(T-t)}w)}_{\text{Utility upon default}} \tag{1.C.14}$$

In (1.C.14), $V(w,t)$ denotes the investor's value function, and V_t and V_w represent the partial derivatives with respect to time and wealth respectively. α is the fraction of wealth invested in the government bonds, which is the investor's only control variable. For inattentive investors, prior to them learning about the bond, $\alpha = 0$.

Proposition 2. *The value function $V(w,t)$ is logarithmic in wealth. Specifically, there exists $v(t)$ such that*

$$V(w,t) = v(t) + \log w \tag{1.C.15}$$

Proof. We can prove the proposition by guessing and verifying that the value function is in form of (1.C.15). It implies that $V_w = w^{-1}$ and $V_t = v'(t)$. The first-order condition for the optimal α, denoted by $\alpha^*(t)$, is as below:

$$\frac{dP}{P} - r = \frac{\delta}{1 - \alpha^*(t)} \tag{1.C.16}$$

By substituting (1.C.16) in (1.C.14), a dividing both sides by δ, we get:

$$V(w,t) = v'(t) + r\delta^{-1} + \frac{\alpha^*(t)}{1-\alpha^*(t)} + \log(1-\alpha^*(t)) + r(T-t) + \log w \tag{1.C.17}$$

It proves the conjecture about the functional form, and provides a characterization for $v(t)$.

□

By applying (1.C.15) to (1.C.14), we see that all investors that are aware of the bond at t invest the same fraction (α^*) of their wealth in the government bonds, where α^* satisfies the following condition:

$$\mathbb{E}[\frac{dP}{P}] - r = \frac{\delta}{1-\alpha^*} - \delta \tag{1.C.18}$$

Note that when the bond size is small compared to the overall wealth of investors aware of the bond, α^* is small, implying the risk-premium is close to zero. Therefore, the model generates a risk-premium larger than the frictionless benchmark since all investors do not pay attention, leading to an imperfect risk-sharing among the investors.

Proposition 3 provides the equilibrium price and risk-premium:

Proposition 3. *Provided the government does not default on its bond by t, the price of the government bond is:*

$$P(t) = e^{(r+\delta)(t-T)}x(t) \tag{1.C.19}$$

,where

$$x(t) = (1 + \int_t^T \frac{\delta Q e^{-(r+\delta)T} e^{\delta s}}{W_0^A - P(0)Q + (1-e^{-\lambda s})W_0^{IA}} ds)^{-1} \tag{1.C.20}$$

The implied risk-premium is:

$$\mathbb{E}[\frac{dP}{P} - r] = \frac{\delta Q e^{-rt} P(t)}{W_0^A - P(0)Q + (1 - e^{-\lambda t})W_0^{IA}} \tag{1.C.21}$$

Proof. Equation 1.C.16 implies that attentive investors and the inattentive investors that are aware of the bond at t have the same α^*, that is they invest the same fraction of their wealth in the government bond. Define $W(t)$ as the total wealth of these two groups of investors at t. Therefore, $\alpha^* = \frac{QP(t)}{W(t)}$. By substituting this in (1.C.16), and some rearrangements, we get (Time indices are dropped for brevity, unless necessary):

$$P' - rP = \delta P(1 - \frac{QP}{W})^{-1} \tag{1.C.22}$$

,where $P'(t) = \frac{dP}{dt}$. One can solve the ODE and obtain the expressions provided in (1.C.19) and (1.C.20) by following the steps below:

$$\begin{aligned} P' - rP &= \delta P(1 - \frac{QP}{W})^{-1} \\ P' - (r+\delta)P &= \delta P^2 Q(W - PQ)^{-1} \\ P' e^{-(r+\delta)(t-T)} - (r+\delta)Pe^{-(r+\delta)(t-T)} &= e^{-(r+\delta)(t-T)}\delta P^2 Q(W - PQ)^{-1} \end{aligned} \tag{1.C.23}$$

Define $y = e^{(r+\delta)(t-T)}P^{-1}$, and substitute in (1.C.23):

$$y' = -\delta e^{(r+\delta)(t-T)} Q m^{-1} \tag{1.C.24}$$

,where $m = W - PQ$. Note that

$$\begin{aligned} dm &= dW - QdP = QdP + rmdt + \lambda e^{(r-\lambda)t} W_0^{IA} dt - QdP \\ \Rightarrow m &= e^{rt}(W_0^A - P(0)Q + (1 - e^{-\lambda t}) W_0^{IA}) \end{aligned} \tag{1.C.25}$$

In (1.C.25), I use the fact that at $t = 0$, only attentive investors participate, that is $W = W_0^A$. Now, we can solve for $y(t)$, and consequently $P(t)$, by substituting (1.C.25) in (1.C.24), integrating over $[t, T]$, and noting the fact that $y(T) = P(T)^{-1} = 1$:

$$\begin{aligned} y(t) &= 1 + \int_t^T \frac{\delta Q e^{-(r+\delta)T} e^{\delta s}}{W_0^A - P(0)Q + (1 - e^{-\lambda s}) W_0^{IA}} ds \\ P(t) &= e^{(r+\delta)(t-T)} y^{-1}(t) = e^{(r+\delta)(t-T)} x(t) \end{aligned} \tag{1.C.26}$$

To derive the risk-premium formula (1.C.21), we should note that the pricing formula is conditional on the government not defaulting on the bond. As a result the risk-premium is

$$\begin{aligned} \mathbb{E}[\frac{dP}{P} - r] &= \frac{dP}{P} - r - \delta \\ &= y^{-2} e^{(r+\delta)(t-T)} P^{-1} y' \\ &= \frac{\delta Q e^{-rt} P}{W_0^A - P(0)Q + (1 - e^{-\lambda t}) W_0^{IA}} \end{aligned} \tag{1.C.27}$$

□

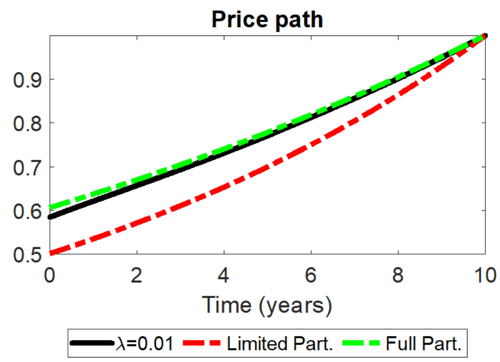
In (1.C.19), the first term is the bond price in a benchmark case in which the credit spread is equal to the default probability (Duffie and Singleton, 1999). The second term, $x(t)$, captures how the investor inattention impacts the bond price. $x(t)$ is close to one when the attentive capital (W_0^A) is large compared to the issue size ($W_0^A \gg Q$), or when the inattentive capital (W_0^{IA}) is large and the inattentive investors rapidly become aware of the bond after its issuance ($W_0^{IA} \gg Q, \lambda \to \infty$). In the specification of $x(t)$, $P(0)$ also appears, which can be found by combining (1.C.19) and (1.C.20).

Figure 1.C.1a depicts the price path for the case with investor inattention, along with two extreme cases: Full participation, corresponding to a case that all investors are aware of the bond at the issuance time, and limited participation, corresponding to the case with $\lambda = 0$.[26]

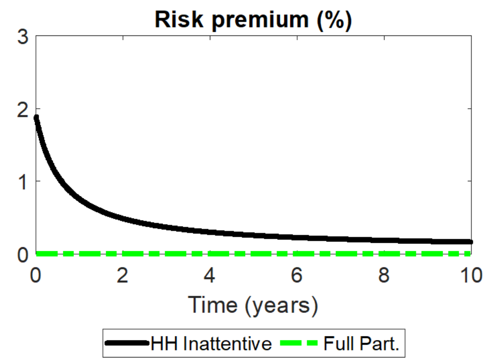
The risk-premium is provided by Equation (1.C.21), and its evolution is depicted in Figure 1.C.1b, along with the two extreme cases of full and limited participation. As shown in Figure 1.C.1b, the risk-premium goes down as more investors become aware of the bond, due to the improved risk-sharing between the attentive and inattentive investors. Interestingly, we see that the risk-premium is not simply between the values for the two extreme cases. To see the intuition, note that, in the beginning, the price reflects the fact that more investors will join the market in the future, thus the initial price is higher than the limited participation case. However, at the same time, the ownership is still concentrated, implying that the attentive investors need to allocate a larger fraction of their wealth to the bond compared to the limited participation case, thus they require a larger compensation for the default risk.

A long-standing puzzle in the literature of municipal bond markets is the large credit spreads implied by the prices, despite the low historical default rates (Cestau, Hollifield, Li, and Schürhoff, 2019; Schwert, 2017). Schultz (2013) and Li and Schürhoff (2019) suggest that the high credit spreads are attributable to the fact that municipal bond markets are segmented due to the tax advantage of in-state investors, which limits arbitrageurs' ability

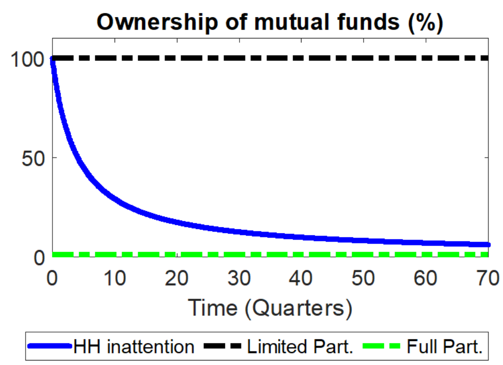
26. Note that the mode does not converge to the full participation case when $\lambda \to \infty$ because for all values of λ, only attentive investors are aware of the bond at $t = 0$. Full participation requires all investors to be participate in the government bond market from $t = 0$ onward.



(a)



(b)



(c)

Figure 1.C.1: The evolution of the price, risk-premium, and ownership throughout the bond's lifetime. In all figures, the solid black line represents the case with investor inattention. Red and green dashed liens represent the cases with limited and full participation. Figures (a), (b), and (c) respectively display the government bond price, risk premium on the bond ($\mathbb{E}[\frac{dP}{P}] - r$), and the fraction held by the attentive investors. The parameter values used are: $\lambda = 0.01$, risk-free rate (annual) $= 2\%, \delta = 3\%$, Maturity $= 10 years, Q = \$60M, W_0^A = \$90M, W_0^{IA} = \$9B$

to inject liquidity. Relatedly, Babina, Jotikasthira, Lundblad, and Ramadorai (2021) find evidence that the demand for the bonds is more inelastic in states with small out-of-state ownership, which are mostly the states with large relative tax-advantage of in-state investors. Note that this type of market segmentation hinders risk-sharing across states. However, investor inattention limits the risk-sharing between more attentive investors, such as mutual funds, and less attentive investors within the same state.

CHAPTER 2

PERSUASION IN RELATIONSHIP FINANCE

2.1 Introduction

What benefits do entrepreneurs get by raising capital from intermediaries such as banks or venture funds instead of issuing securities in public markets? A large literature on relationship finance reveals that intermediaries mitigate informational asymmetry and moral hazard by providing initial funding and forming relationships with entrepreneurs to become "insiders" (e.g., Ramakrishnan and Thakor, 1984; Fama, 1985; Diamond, 1991). Researchers recognize how insider financiers' information monopoly may hold up entrepreneurs (e.g., Sharpe, 1990; Rajan, 1992), but largely ignore the endogenous nature of information production and design. Yet, in reality, entrepreneurs' actions and experimentation not only shape project cash flows, but also alter the informational environment.[1]

Several questions naturally arise. Does the endogenous production of information matter for relationship formation and sequential fund-raising? Can it help rationalize puzzling empirical observations such as the non-monotone relationship between bank orientation and competition (e.g., Degryse and Ongena, 2007) that earlier models cannot explain? What are the implications for designing securities for venture investors at various stages?

To address these questions, we model interim experimentation in relationship financing, contracting, and security design as a Bayesian persuasion game with contingent transfers and

1. Pharmaceutical firms can affect the FDA's and investors' decisions by providing additional information and tests (e.g., "Guidance for Industry and FDA Staff. Post market Surveillance Under Section 522 of the Federal Food, Drug and Cosmetic Act" May 16, 2016). Software startups decide on different markets for beta launches because the media attention generated is different, and thus information communicated to potential users ("Why Is Canada Such A Good Testing Ground For Game Releases?" Forbes Tech. Nov 27, 2012); entrepreneurs choose the specific prototype or trial market to work on which produces disparate forms of information; career concerns and managerial choices of projects exhibit similar features.

differentially informed receivers. Specifically, in our baseline model we consider a capital-constrained (male) entrepreneur with a project requiring two rounds of financing. The first round requires a fixed investment that enables the entrepreneur to "experiment"—broadly interpreted as conducting early-stage activities such as hiring key personnel, acquiring initial users, and developing product prototypes—to produce interim information to persuade investors for continued financing. The key friction rests in that the entrepreneur lacks ex-ante commitment to specific experiments and thus to the interim information production. However, by monitoring and having access to the entrepreneur's team through the first-round investment, a relationship financier (henceforth referred to as a female "insider" or "insider investor") observes and can verify interim signals from the experiment. Because the signals are informative of the eventual profitability, the insider has an advantage relative to arm's-length investors (henceforth referred to as "outsiders"), consistent with assumptions in the literature and observations in practice concerning relationship finance (e.g., Rajan, 1992).

After forming the financing relationship and conducting interim experiments, the entrepreneur raises capital in a second round by issuing securities to the insider and potentially outsiders. The insider enjoys an informational advantage relative to outsiders who learn from the insider's decision on continuation or termination of the project. We capture the entrepreneur's and investors' (sender's and receivers') divergent interim objectives by modeling the entrepreneur's private benefit of continuation (which is difficult to verify or contract upon), limited liability, and endogenous security choice.

Our key finding is that the entrepreneur's endogenous information production reduces a relationship financier's rent from her interim informational advantage, thus inefficiently holding up her initial investment in the relationship — a phenomenon we call *Information Production Hold-up* (IPH). We show that the entrepreneur follows a threshold strategy for experimentation to produce information such that the insider is indifferent between termination and continuation. This reduces the insider's information monopoly, rendering the insider incapable of recovering the initial investment in forming the relationship. Good projects thus may fail to get initial financing. We are the first to make these observations, which are in stark contrast to theories on bank monitoring and hold-ups of entrepreneurial effort, taking informational environments as exogenous.

Endogenous information productions and their associated hold-ups have two immediate implications. First, IPH alters the conventional understanding of the roles of investors' information production/acquisition and interim competition in relationship finance. In industries requiring less entrepreneur-specific knowledge, "sophisticated" insiders who use their own information technology to evaluate projects' prospects can extract a positive interim rent and partially restore the feasibility of the initial relationship formation. Relationship financing also becomes viable with moderate interim competition, because selling to competitive outsiders encourages more efficient information production by the entrepreneur. Investors' interim competition (reflected through the insider's interim bargaining power) and sophistication (captured by the informativeness of her independent signal) jointly impact the dependence of relationship financing on competition, which is non-monotone in general. In particular, for intermediate levels of investor sophistication, the ease of relationship formation can therefore depend on interim competition with a U-shaped pattern, consistent with the empirical regularities (Elsas, 2005; Degryse and Ongena, 2007) that other models cannot explain.

The second implication of endogenous information production and IPH is how it affects contracting and security design for sequential investors. Even with general investor sophistication and interim competition, entrepreneurs' optimal contracts fully restore efficient information production: the entrepreneur optimally promises the early insider investor convertible securities at a pre-specified price and quantity, and upon the insider's continued financing, issues residual securities to the outsiders.[2]

Intuitively, the entrepreneur gets all the ex-ante surplus when facing competitive investors. Upon forming a financing relationship, he wants to but cannot commit to efficient interim information production, resulting in an inter-temporal wedge between the ex-ante entrepreneur (who shares the objective of a social planner) and the interim entrepreneur. The mix of inside and outside finance at the interim date then affects not only how surplus is shared (as in Rajan, 1992) but also how much information is produced. The ex-post expropriation (hold-up) generates suboptimal investment decisions by distorting entrepreneurial incentives to produce information. The first-best security design for the entrepreneur therefore should align both his interim information production incentive and the insider's continuation incentive with the social planner's.

This general contractual problem involves an infinitely-dimensional nested optimization, prompting us to take a novel constructive-proof approach: we first propose a set of contracts restoring social efficiency and then show that they are the only optimal contracts. Outsiders are competitive and pay the entrepreneur fair prices, effectively rendering the entrepreneur the residual claimant of the project's social surplus regardless of the information revealed during the interim. Therefore, giving the insider debt-like securities in all bad states of the world (when the entrepreneur tends to overcontinue) fully exposes the entrepreneur to the cost of inefficient continuation. Meanwhile, relationship financing is feasible as long as the

2. Our findings do not rely on the entrepreneur's ability to commit to a disclosure policy or designing information to be arbitrarily informative. Instead, we require that the entrepreneur can design simple experiments such as one producing binary, threshold signals, and that within a financing relationship an insider verifies interim experiment outcomes better than outsiders.

contract yields the insider enough interim rent to recover her initial investment, leaving the security design only partially determinate in good states of the world.

The optimal designs derived are broadly consistent with real-life observations in entrepreneurial finance: a variety of convertible securities for insiders and residual securities for arm's-length outsiders. For example, venture capitalists indeed routinely take convertible debts while arm's-length outsiders receive equities at subsequent public offerings.

Finally, we discuss the robustness of our findings by allowing a restricted experimentation space, partial observability and commitment to disclosure, general investor sophistication, scalable investments, and investors' security design, among others. IPH manifests itself under various security forms and the economic mechanism applies even beyond relationship lending and staged venture financing. We help underscore and formalize this practical issue, and then develop potential contractual solutions. From a theory perspective, our study also provides insights on Bayesian persuasion games with contingent transfers and sequential receivers facing discriminatory disclosures, while deepening our understanding of contracting under endogenous information production.

Literature — Our theory foremost contributes to the large literature on relationship finance. Theoretical studies on relationship banking focus on interim information and control (e.g., Diamond, 1984, 1991; Fama, 1985). While relationship finance can improve financing efficiency (e.g., Petersen and Rajan, 1994), it naturally induces an information monopoly (Berger and Udell, 1995; Petersen and Rajan, 2002), potentially holding up the entrepreneur's effort in relationship lending (e.g., Rajan, 1992; Santos and Winton, 2008; Schenone, 2010) and venture capital (e.g., Fluck, Garrison, and Myers, 2006; Ewens, Rhodes-Kropf, and Strebulaev, 2016). We inform the debate by endogenizing the informational environment and analyzing the IPH problem.[3]

Empirically, the effect of competition on bank orientation has received much attention. Elsas (2005) and Degryse and Ongena (2007) document a puzzling U-shaped effect of market concentration on relationship lending. Extant theories predict either opposing monotone patterns (e.g., Petersen and Rajan, 1995; Boot and Thakor, 2000; Dinc, 2000; Dell'Ariccia and Marquez, 2004) or suggest a hump-shaped pattern (e.g., Yafeh and Yosha, 2001; Anand and Galetovic, 2006). Our theory offers an information-based explanation for the empirical regularities.

We also elucidate the role of intermediaries and security design in entrepreneurial finance.[4] We add to earlier studies on optimal securities in relationship finance or venture capital (e.g., Gompers, 1997; Kaplan and Strömberg, 2004; Hellmann, 2006) by endogenizing general information production and incorporating flexible designs in sequential security issuance and experimentation. The optimality of convertible securities in our paper does not rely on ex-ante informational asymmetry between the issuer and investors (e.g., Stein, 1992; Brennan and Schwartz, 1988), hidden manipulations of signals under exogenously given information structures (Cornelli and Yosha, 2003), or investors' information acquisition (Yang and Zeng, 2018). We are the first to show that first issuing convertible securities to insiders and then equities to outsiders is optimal and robust to investor sophistication and interim competition. We also underscore that the presence of competitive outsiders crucially affects

3. In this regard, our paper broadly relates to incomplete contracting and hold-up problems (e.g., Hart and Moore, 1988; Aghion, Dewatripont, and Rey, 1994) and whether long-term contracts can mitigate investment inefficiencies (Von Thadden, 1995; Nöldeke and Schmidt, 1998). We differ primarily in endogenizing interim information production and deriving the optimal design without requiring contractibility of the entrepreneurs' bias of continuation.

4. Our paper broadly relates to seminal studies examining how agents' actions alter the distribution of cash flows (e.g., Holmstrom, 1979; Innes, 1990). The agent in our setting shapes the informational environment and securities are issued to heterogeneous agents. While first-best outcomes are typically unattainable in conventional settings, our contractual solution restores efficient information production and investment.

optimal contract designs.

From a theory perspective, our paper contributes to the field of information design, especially Bayesian persuasion (e.g., Kamenica and Gentzkow, 2011; Dworczak and Martini, 2019; Ely, 2017; Guo and Shmaya, 2019). We take a linear programming approach similar to Bergemann and Morris (2016), but allow infinite payoff-relevant states and different types of informed receivers. We do not require that the sender's utility from a message completely depend on the expected state (Kolotilin, 2018) or the payoff over the receiver's actions to depend on the state linearly (Gentzkow and Kamenica, 2016). More importantly, we incorporate into our model security design that endogenizes the dependence of the sender and receivers' payoffs on the state, allowing interactions of multiple asymmetrically informed receivers.

By so doing, our paper advances the emerging applications of information design in finance. Despite having been adopted to address issues in banking regulation, online advertising, entertainment, etc., most Bayesian-persuasion models do not allow contingent transfers, which are prevalent in security design and contracting. Relationship finance provides a natural setting for endogenous information design and contingent transfers.[5] Most closely related is Szydlowski (2019), a pioneering study applying Bayesian persuasion to corporate finance, which obtains an irrelevance result of security choice when the entrepreneur jointly designs disclosure and security. Other related studies concern topics on government intervention (Cong, Grenadier, and Hu, 2020), market for advice with heterogeneous agents (Chang and Szydlowski, 2020), and stress tests (Bouvard, Chaigneau, and Motta, 2015; Goldstein and

5. The commitment assumption has been a major challenge in the field, but here the dynamic financing relationship naturally allows the insider to observe interim experiment and signals. In fact, it is the *lack of commitment* to information design during entrepreneurs and investors' initial interactions that calls for security design to mitigate inefficient information production.

Leitner, 2018; Orlov, Zryumov, and Skrzypacz, 2019; Inostroza and Pavan, 2018).[6] Our paper instead constitutes a first examination of relationship finance and contracting under endogenous information production.

2.2 A model of relationship finance and information

2.2.1 Financing of projects

We consider a three-period economy with time index $t = 0, 1, 2$. There is no time discounting. A risk-neutral entrepreneur has a project that requires a fixed investment $I \in (0, 1)$ at $t = 1$ and produces an uncertain cash flow $X \in [0, 1]$ at $t = 2$ with a prior distribution denoted by a continuous and atomless pdf $f(X)$. The entrepreneur can raise I by issuing securities at $t = 1$ to competitive, risk-neutral investors to finance the project.

In addition, an investor may invest K in the project at $t = 0$ to become a "relationship financier" (the "insider"). The initial investment enables the entrepreneur to experiment and generate interim information about the distribution of the cash flow. One may also view $K + I$ as the total investment needed but raised in stages whereby early experimentation generates interim information (Kerr, Nanda, and Rhodes-Kropf, 2014). For simplicity, we normalize the cash flow from the seed investment to zero, which is innocuous (see, e.g., Online Appendix OA1).

6. In a related literature, researchers examine reduced-form parametrized disclosure rules and typically relies on informational asymmetry and signaling or cheap talk. Trigilia (2019), for example, studies company transparency and capital structure in the presence of heterogeneous investors.

2.2.2 Experimentation and information production

The interim experimentation essentially allows the entrepreneur (the sender) to choose messages from a compact metric space $\mathcal{Z}$ and mapping $\pi : [0,1] \to \Delta(\mathcal{Z})$, where $\Delta(\mathcal{Z})$ is the set of Borel probabilities on the message space. For notational simplicity and without loss of generality, we assume $\mathcal{Z}$ is finite, which implies that the experiment simply involves a mapping described by the conditional probabilities $\pi(z|X)$.[7]

Following the relationship finance literature, we assume that even though $(\mathcal{Z}, \pi)$ is common knowledge, with probability $\mu \in [0,1]$ only the insider observes the outcome of the experiment and with probability $1-\mu$ all investors publicly observe the outcome. μ succinctly captures the extent of the insider's informational advantage through close monitoring and repeated interactions between the entrepreneur and insiders (Megginson and Weiss, 1991); we can interpret $1-\mu$ as a reduced-form measure of the "interim competition" between the insider and the outsider commonly modeled in the relationship lending literature (Petersen and Rajan, 1995). $\mu = 0$ corresponds to perfect interim competition and $\mu = 1$ corresponds to information monopoly by the insider.

In our setting, it is not necessary for the entrepreneur to communicate or use a sophisticated disclosure policy. We only need the insider to observe or verify an experiment's outcomes (e.g., reception of a prototype, a trial run, a beta launch, etc.) more easily than the outsiders do. The extent of this advantage is captured by μ. For example, the experiment could be the building and testing of a prototype, over which the entrepreneur has full control. All we require is that an insider understands the prototype test and observes the outcome better.

7. As we show in Appendix A1 and Online Appendix OA1, all our results go through had we assumed $\mathcal{Z}$ to be countable or of a continuum, or a more restrictive space of information design.

2.2.3 Misalignment of incentives

As the project develops after the initial funding, the entrepreneur enjoys a private benefit from continuing the project. Specifically, he receives $\varepsilon \in (0, \bar{\varepsilon})$ if the project is financed at $t = 1$, where $\bar{\varepsilon}$ and K satisfy $\mathbb{E}\left[(X - I)\mathbb{I}_{\{X \geq I - \bar{\varepsilon}\}}\right] \geq K$ so that the project is always financed through relationship financing (positive NPV ex ante to the financier) under efficient interim information production. ε has an atomless distribution described by the pdf $g(\cdot)$ ex ante ($t = 0$) and its realization at the start of $t = 1$ is common knowledge.

As we show below, the entrepreneur's limited liability drives most of our results, and therefore we could have alternatively allowed negative values of X that the entrepreneur does not need to bear (e.g., Rajan, 1992). That said, using a private benefit ε to capture the misalignment of incentives is standard in the literature and it constitutes a realistic source of agency conflict and eases our subsequent exposition of optimal security designs.

To best illustrate our economic mechanism and match reality for early business start-ups, we assume that $\mathbb{E}[X - I + \bar{\varepsilon}] < 0$, which implies the absence of direct financing by arm's-length investors ex ante. This innocuous assumption allows us to focus on the case in which relationship financing and informational considerations are indispensable.

2.2.4 Contracting environments

In addition to the informational advantage, an insider investor can potentially contract with the entrepreneur upon forming a financing relationship at $t = 0$. We consider two main contracting environments.

First, to highlight the IPH, we assume in Section 2.3 an exogenously given security $s(X)$, limited liabilities of the entrepreneur and insider investor ($s(X) \in [0, X]$), and doublemonotonicity for the security (both $s(X)$ and $X - s(X)$ are weakly increasing in X).[8] After z is realized, the insider makes a take-it-or-leave-it offer (TIOLI) to purchase a λ fraction of the security $s(\cdot)$, i.e., $s_I(X) = \lambda s(X)$, at a total price p^I.[9] The entrepreneur decides whether to accept the offer and then if he still needs financing, he sells the remaining securities, $s_O(X) = (1-\lambda)s(X)$, to outsiders who offer a competitive total price p^O. The investment takes place if and only if I is successfully raised, otherwise the pledged capital is returned to investors.[10]

Second, in Section 2.4 we allow the entrepreneur to contract at $t = 0$ on both λ and the design of securities to be offered at $t = 1$ to the insider financier and arm's-length investors. This is equivalent to allowing any form of contracts over X, including one that promises the insider $s_0(X)$ at $t = 0$. [11]

We do not allow contracting on the experiment at $t = 0$ for two reasons. First, the security design literature typically makes security payoffs only contingent on cash flows and not on experimentation. Second, in practice a start-up constantly evolves and it is almost impossible to contract at a seed round or angel round what exact actions the entrepreneurs

8. See, for example, Nachman and Noe (1994), DeMarzo and Duffie (1999), DeMarzo, Kremer, and Skrzypacz (2005), and Cong (2017). If such monotonicity is violated, either the entrepreneur or the investor can be better off destroying some surplus for some state X, as Hart and Moore (1995) point out.

9. The TIOLI assumption, albeit natural, is not crucial and one can alternatively assume that the insider can renegotiate during the interim. The insider's proposing a different security can be viewed as a form of renegotiation. Our main results are robust to the possibility of renegotiation after the signal realization, because in that case a dominant strategy for the insider is to ask for the whole output X for price p^I. This offer would be equivalent to the specific case of the original security being $s(X) = X$. We discuss this further in Online Appendix OA2.

10. This scheme is often referred to as "all-or-nothing." Regarding its wide applications and impact on project implementation and information aggregation, see Cong and Xiao (2018).

11. The insider's getting a total of $s_0(X) + s_I(X)$ upon continuation and nothing otherwise is equivalent to getting some other $s_I(X)$ alone. To highlight the importance of security design, in Online Appendix OA2 we consider a third contracting environment wherein the entrepreneur can contract with the insider at time $t = 0$ only on the fraction λ of security issuance at $t = 1$ that the insider can purchase before it is offered to arm's-length outsiders.

should take when the founding team is not yet complete, not to mention that investors do not have the same technical expertise or control as the entrepreneurs at such an early stage (e.g., Gompers, Gornall, Kaplan, and Strebulaev, 2020) and interim signals are often not well-defined and too costly to verify to be useful for security contracts (Kaplan and Strömberg, 2003).

2.2.5 Interim payoffs and relationship formation

In general, the players' interim payoffs after the formation of a financing relationship are:

$$
\begin{aligned}
p^O &= \mathbb{E}[s_O(X)|\mathcal{F}^O], &(2.1)\\
u^E(X;p^I,p^O,\varepsilon) &= \Big(\varepsilon + X - s_I(X) - s_O(X) + p^I + p^O - I\Big)\,\mathbb{I}_{\{p^I+p^O\geq I\}}, &(2.2)\\
u^I(X;p^I,p^O) &= \Big(s_I(X) - p^I\Big)\,\mathbb{I}_{\{p^I+p^O\geq I\}}, &(2.3)
\end{aligned}
$$

where $\mathcal{F}^O$ denotes the outsiders' information set after having observed the project's continuation or termination (not the insider's offer per se as the entrepreneur may reject that).[12]

If the project is not financed, all players receive outside options, which are normalized to zero. Intuitively, relationship financing is feasible only if the insider can recover in expectation at least the initial investment K, i.e.,

$$\mathbb{E}[u^I(X;p^I,p^O)] \geq K. \tag{2.4}$$

Finally, we assume the project would always be funded through relationship financing

12. One can interpret the signal as being either public or private as the receiver's private type. Note that the entrepreneur's experimentation affects both the insider and outsiders' actions, different from Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), who examine the case of a single receiver with private types.

(positive NPV ex ante to the financier) if the interim information production is socially efficient. This holds under endogenous security design because $\bar{\varepsilon}$ satisfies $\mathbb{E}\left[(X-I)\mathbb{I}_{\{X\geq I-\bar{\varepsilon}\}}\right]=K$, but requires $\mathbb{E}[s(X)-I|X\geq I-\bar{\varepsilon}]\geq K$ in the baseline when $s(\cdot)$ is exogenous. This assumption allows us to focus on failures of financing relationship formation purely driven by the entrepreneur's endogenous information production.

2.2.6 Investor sophistication and information production

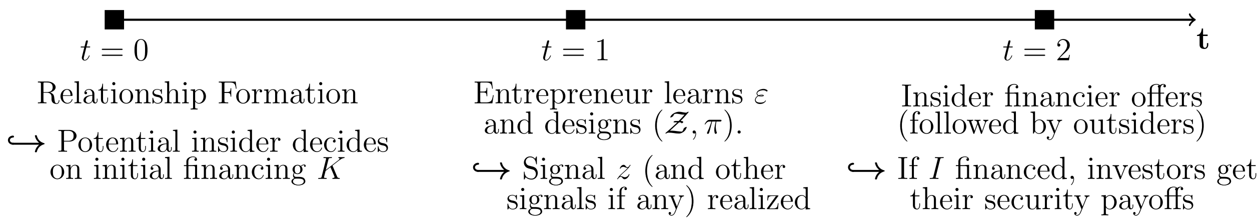
In reality, an insider may dictate the entrepreneur's information production activities, or receive additional information besides what the entrepreneur produces. For example, the insider may experiment herself, specify what the entrepreneur must do, or use her proprietary business experience and expertise to predict the market demand or project valuation in future financing rounds. Importantly, the insider can set milestones in the initial contract, in which case the insider conditions the next round of funding on pre-specified achievements and accomplishments. We collectively refer to the insider's ability to utilize such information production technology exogenous to the entrepreneur's design as "investor sophistication."

To best understand IPH, Sections 2.3 and 2.4 assume that only the entrepreneur has the relevant skill and expertise to design $(\mathcal{Z},\pi)$ after raising K. This happens when the lender either has no previous experience on the project or it is too costly for him to extract information (e.g. the firm is located in a hardly accessible location, or the investor has no relevant expertise to generate independent signals). In Sections 2.5 and 2.6.1, we relax the assumption and incorporate investor sophistication by allowing the insider to use a technology to produce an interim signal (not observable to the outsiders) about X.

2.3 Equilibrium and information production hold-up

To highlight the stark effect of IPH and the way it drastically alters our understanding of relationship finance, we abstract away from contracting in this section and take the security design as exogenous, before allowing contracting at $t = 0$ and highlighting the role of security design in Section 2.4. Figure 2.1 summarizes the timeline of the baseline game.

Figure 2.1: Timeline of the game.



To characterize the equilibrium, we work backward by first analyzing the interim persuasion game after the formation of the financing relationship. For any given $(\mathcal{Z}, \pi)$, if signal z is privately observed by the insider, then the insider offers $p^I = I$ to finance the project entirely ($\lambda = 1$ endogenously) when $\mathbb{E}[s(X)|z] \geq I$, assuming any indifference (when $\mathbb{E}[s(X)|z] = I$) is resolved by the insider's full financing; otherwise when $\mathbb{E}[s(X)|z] < I$, the insider terminates the project, leading the outsiders to negatively update their valuations and to not invest. The insider's information monopoly essentially gives her full bargaining power over the contractible interim surplus generated, $\max\{\mathbb{E}[s(X) - I|z], 0\}$, which corresponds to the well-known information hold-up in earlier models such as Rajan (1992).

In the case where z is publicly observable (which happens with probability $1 - \mu$), then both the insider and the outsiders would offer the competitive $p^I = p^O = \mathbb{E}[s(X)|z]$ when $\mathbb{E}[s(X)|z] - I \geq 0$. The entrepreneur extracts the whole interim surplus in this case.

The entrepreneur's expected payoff for a given realization of ε is thus

$$\begin{aligned} U^E(\mathcal{Z}, \pi; \varepsilon) &= \mathbb{E}[u^E] = \mu \int_0^1 \sum_{z \in \mathcal{Z}^+} (\varepsilon + X - s(X))\pi(z|X)f(X)dX \\ &\quad +(1-\mu)\int_0^1 \sum_{z \in \mathcal{Z}^+} (\varepsilon + X - s(X) + \mathbb{E}[s(X)|z] - I)\pi(z|X)f(X)dX \\ &= \int_0^1 \sum_{z \in \mathcal{Z}^+} (\varepsilon + X - I)\pi(z|X)f(X)dX - \mu \int_0^1 \sum_{z \in \mathcal{Z}^+} (s(X) - I)\pi(z|X)f(X)dX, \end{aligned} \tag{2.5}$$

where $\mathcal{Z}^+ = \{z \in \mathcal{Z} | \mathbb{E}[s(X)|z] \geq I\}$ corresponds to the set of signals inducing continued investment. Equation (2.5) follows from applying the law of iterated expectations to $\mathbb{E}[\mathbb{E}[s(X)|z]|z \in \mathcal{Z}^+]$. Correspondingly, the insider's payoff is:[13]

$$U^I(\mathcal{Z}, \pi) = \mathbb{E}[u^I] = \mu \int_0^1 \sum_{z \in \mathcal{Z}^+} (s(X) - I)\pi(z|X)f(X)dX. \tag{2.6}$$

From Equation (2.5), the entrepreneur solves the following maximization problem:

$$\max_{(\mathcal{Z}, \pi)} \mathbb{E}[(\varepsilon + X - \mu s(X) - (1-\mu)I)\mathbb{I}_{\{\mathbb{E}[s(X)|z] \geq I\}}]. \tag{2.7}$$

Proposition 4 (Entrepreneur's Optimal Experimentation). *The entrepreneur conducts an optimal experiment that entails two signals, i.e.,* $|\mathcal{Z}| = 2$. *A signal* h *induces investment if* $X \geq \max\{\bar{X}, \hat{X}(\mu)\}$, *and a signal* l *induces termination otherwise, where* $\bar{X}$

13. Equations (2.5) and (2.6) reveal that for a given experiment, the entrepreneur's expected payoff is decreasing and the insider's expected payoff is increasing in μ — a measure of the insider's information monopoly (opposite to competition). By juxtaposing equations (2.4) and (2.6), we see that financing relationship is feasible only when the interim rent is sufficiently high, i.e., $\mu\mathbb{E}[(s(X) - I)\mathbb{I}_{\{\mathbb{E}[s(X)|z] \geq I\}}] \geq K$. These observations confirm the results in Petersen and Rajan (1995) that less interim competition leads to the greater possibility of financing relationship.

and $\hat{X}(\mu)$ *solve:*

$$\begin{aligned}
\mathbb{E}[s(X)|X \geq \bar{X}] - I &= 0 && (2.8)\\
\varepsilon + \hat{X}(\mu) - \mu s(\hat{X}(\mu)) - (1-\mu)I &= 0 \quad \textit{if} \;\; \varepsilon - (1-\mu)I < 0 && (2.9)\\
\hat{X}(\mu) &= 0 \quad \textit{if} \;\; \varepsilon - (1-\mu)I \geq 0. && (2.10)
\end{aligned}$$

Moreover, all optimal experiments lead to the same investment and payoffs, rendering the equilibrium essentially unique.

Proposition 4 characterizes the optimal experimentation after relationship formation. Equation (2.8) indicates that the continuation based on $X \geq \max\{\bar{X}, \hat{X}(\mu)\}$ makes the insider financier at least break even; Equation (2.9) indicates that if the private benefit is small relative to the competition, the entrepreneur rationally induces the continuation at X if and only if he can break even; Equation (2.10) just states that if the private benefit is large relative to interim competition, the entrepreneur always benefits from the continuation, and the threshold is again pinned down by $\bar{X}$ in Equation (2.8).

Note that the threshold structure of the optimal experimentation is consistent with Szydlowski (2019), who first derives it for a lumpy investment under any given security design. As will become clear below, our main incremental contributions are to show how competition (captured by $1-\mu$) and the insider's information production affect the experiment's informativeness (characterized by the threshold), as well as how endogenous information production leads to a novel form of hold-up in relationship finance.

In equilibrium, while all profitable projects receive continued financing, some negative NPV ones do as well due to the entrepreneur's persuasion. When the insider privately observes the experiment's outcome, she bears the cost of inefficient continuation and the entrepreneur would like to lower the threshold for investment for his private benefit; but as μ decreases, the entrepreneur's chance of getting a competitive price becomes higher, helping him better internalize the cost of inefficient continuation. These trade-offs determine the optimal threshold. In the extreme case, $\mu = 0$, the entrepreneur sends a high signal for $X \geq I - \varepsilon$, which induces the socially efficient outcome; at the other extreme, $\mu = 1$, the entrepreneur decreases the threshold to make the insider indifferent between investment and termination ($\mathbb{E}[s(X)|z = h] = I$). The proposition reflects the general phenomenon in persuasion games that a sender can "squeeze" a receiver's rent.

Two questions are particularly relevant for relationship finance. First, what does this endogenous information design imply for relationship formation? Corollary 1 reveals that IPH may severely preclude (relationship) financing, both when the insider enjoys a strong information monopoly over the interim signal and when she faces intense interim competition.

Corollary 1 (**Information Production Hold-up** (IPH)). *For any K, information production and investment are socially inefficient for all μ. In particular, there exists $0 < \mu^l < \mu^h < 1$, such that when $\mu \in [0, \mu^l) \cup (\mu^h, 1]$, relationship financing is infeasible.*

Recall from Section 2.2.5 that with efficient information production, the financing relationship is feasible and socially efficient. But endogenous information production is inefficient. Obviously too much interim competition prevents initial investment from potential insiders, because they do not accrue enough interim rent to cover the initial investment K. Surprisingly, with full information monopoly, the insider financier could also be held-up.[14] Because the entrepreneur produces imperfect information to inefficiently continue projects,

14. Our results are robust to allowing agents to renegotiate the security after $t = 0$, as long as the renegotiation protocols do not depend on agents' posterior beliefs on X (otherwise it is equivalent to contracting on experiment outcome). For any informational environment, the entrepreneur may extract the full interim surplus no matter what the security is. Therefore, the entrepreneur always pools lower types of X with higher ones to make the insider break even.

the insider's payoff is not monotone in μ. In particular, when the insider enjoys information monopoly (μ approaches 1), her initial investment to form the relationship cannot be recouped. This reverse hold-up is in sharp contrast with the traditional hold-up in an exogenous information setting, in which the possibility of the relationship financing is the highest when $\mu = 1$ and the entrepreneur's effort is held up instead.

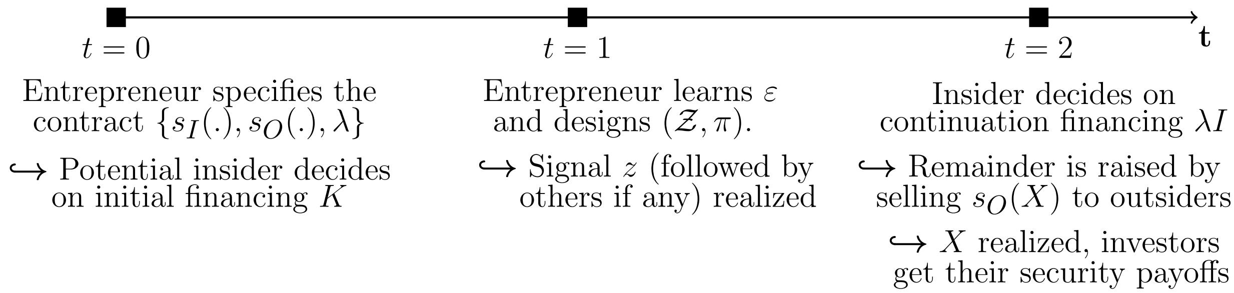
Given this IPH, the second question pertinent to finance is whether contingent transfers in the form of security payments would solve the problem. Contrasting Corollary 1 with Equations (2.5) and (2.6), it should also be apparent that taking information production as exogenous in relational financing is not innocuous—whether information production is endogenous determines whether the entrepreneur or the investor has interim bargaining power.

2.4 Contractual solution for IPH

So far, we have assumed that the entrepreneur only issues securities at $t = 1$. In this section, we allow long-term contracts at $t = 0$ that specify the security payoffs and the amount of the securities the insider can purchase at $t = 1$. In other words, the entrepreneur can contract on λ, the fraction of investment I to be financed from the insider financier in the second round, and the corresponding payment to the insider $s_I(\cdot)$.[15] Outsiders observe the insider's decision and finance the remaining $(1 - \lambda)I$ by purchasing security $s_O(\cdot)$ at a competitive price p^O. Recall that the security payoffs can depend on X but not $(\mathcal{Z}, \pi)$ or the interim signals, as described in Section 2.2.4. Furthermore, similar to the baseline setup, the outcome of the entrepreneur's experimentation is observed privately by the insider with

15. The insider faces no competition from the outsiders regarding $s_I(\cdot)$ and thus pays exactly λI upon continuation and getting $s_I(\cdot)$.

Figure 2.2: Timeline of the game with optimal security design.



probability μ and publicly by both the insider and outsiders with probability $1 - \mu$.

Every contract can then be summarized by a triplet $\{s_I(\cdot), s_O(\cdot), \lambda\}$. Because the contract specifies both how the cost I is shared and how the contingent payments depend on future cash flows, it constitutes a general contracting space. Figure 2.2 displays the timing of the interactions.

Security design and long-term contracting. Here we endogenize the security design $s_I(X)$ for the insider and $s_O(X)$ for the outsiders, and show that the first-best outcomes are restored. Online Appendix OA2 derives optimal long-term contracts with exogenous security types and shows that IPH is robust to contracting on the verifiable cash flow X.

First, note that the entrepreneur's expected payoff from the contract (s_I, s_O, λ) is then given by:

$$U^E = \int_0^{\bar{\varepsilon}} \int_0^1 (\varepsilon + X - s_I(X) - s_O(X) + p^O(\varepsilon) - (1 - \lambda)I)\mathcal{I}(X; \varepsilon) f(X) g(\varepsilon) dX d\varepsilon, \quad (2.11)$$

where p^O is the amount raised from the outsiders by selling the security $s_O(X)$, and $\mathcal{I}$ is the probability of investment at state $X \in [0, 1]$, when ε is realized. The outsiders' information set contains the public signal z (if any) they receive with probability $1 - \mu$, along with the inference from the insider's action of continuation or termination.

Outsiders, being competitive, pay a "fair price" p^O given by:

$$p^O(\varepsilon) = \frac{1}{\mathbb{E}[\mathcal{I}(X;\varepsilon)]} \int_0^1 s_O(X)\mathcal{I}(X;\varepsilon)f(X)dX. \tag{2.12}$$

Combining Equations (2.11) and (2.12) we have:

$$U^E = \int_0^{\bar{\varepsilon}} \int_0^1 M(X; s_I, s_O, \lambda, \varepsilon)\mathcal{I}(X;\varepsilon)f(X)g(\varepsilon)dXd\varepsilon, \tag{2.13}$$

where:

$$M(X; s_I, s_O, \lambda, \varepsilon) = \varepsilon + X - s_I(X) - (1-\lambda)I. \tag{2.14}$$

The entrepreneur's optimal contract design then corresponds to the following maximization problem:

$$\begin{aligned} &\max_{s_I(.), s_O(.), \lambda} \quad \mathbb{E}[M(X; s_I, s_O, \lambda, \varepsilon)\mathcal{I}^*(X;\varepsilon)] \\ &s.t. \quad \mathbb{E}[(s_I(X) - \lambda I)\mathcal{I}^*(X;\varepsilon)] \geq K \quad and \quad s_I(X) + s_O(X) \leq X \quad \forall X \in [0,1], \end{aligned} \tag{2.15}$$

where the optimization is over the set of designs and the option to walk away from the financing relationship, and the constraints are the incentive-compatibility condition of the insider to form a relationship and the entrepreneur's limited liability. $\mathcal{I}^*(\cdot;\varepsilon)$ is the equilibrium investment function under the optimal experiment $(\mathcal{Z}^*(\varepsilon), \pi^*(\varepsilon))$, and is given by:

$$\mathcal{I}^*(X;\varepsilon) = \sum_{z \in \mathcal{Z}^*(\varepsilon)} \pi^*(z|X,\varepsilon)\mathbb{I}_{\{\mathbb{E}[s_I(X) - \lambda I|z] \geq 0\} \cap \{\mathbb{E}[s_O(X) - (1-\lambda)I|\mathcal{F}^O] \geq 0\}} \tag{2.16}$$

and the optimal experiment given the contract $\{s_I(\cdot), s_O(\cdot), \lambda\}$ and ε solves the following:

$$\max_{(\mathcal{Z},\pi)} \int_0^1 M(X; s_I, s_O, \lambda, \varepsilon)\mathcal{I}(X;\varepsilon)f(X)dX$$
$$\text{where} \quad \mathcal{I}(X;\varepsilon) = \sum_{z\in\mathcal{Z}^*(\varepsilon)} \pi(z|X,\varepsilon)\mathbb{I}_{\{\mathbb{E}[s_I(X)-\lambda I|z]\geq 0\}\cap\{\mathbb{E}[s_O(X)-(1-\lambda)I|\mathcal{F}^O]\geq 0\}}. \tag{2.17}$$

Optimal solutions. The contracting problem involves infinitely-dimensional nested optimization, and the solution methodologies in conventional contracting and security design problems do not easily apply. We instead take a constructive-proof approach by conjecturing the optimal designs and show that this set of designs uniquely achieve the first-best outcome and are indeed optimal in the sense that they maximize the entrepreneur's ex-ante payoff for all ε.

Proposition 5. *(**Optimal Design**) An optimal contract exists and implements the first-best social outcome. All optimal contracts induce experiments that generate a continuation if and only if $X \geq I - \varepsilon$. Moreover, they essentially all involve the use of convertible securities and are described by $\{s_I(\cdot), s_O(\cdot), \lambda\}$ satisfying the following conditions:*

$$\lambda \in \left[0, \frac{I-\bar{\varepsilon}}{I}\right] \tag{2.18}$$

$$s_I(X) = \min\{\lambda I, X\}, \quad \forall\, X < I \tag{2.19}$$

$$\mathbb{E}[(s_I(X) - \lambda I)\mathbb{I}_{\{X\geq I\}}] = K \tag{2.20}$$

$$s_O(X) = X - s_I(X), \quad \forall\, X \in [0,1]. \tag{2.21}$$

Equation (2.18) reflects the partial indeterminacy of the optimal design because λ can take on a range of values. Equation (2.19) requires that in the bad states of the world, the security is debt-like. In fact, the shape of the securities in the region $X < I - \bar{\varepsilon}$ is indeterminate, but in terms of payoffs, they are equivalent, thus the use of word "essentially" in the proposition. Equation (2.20) ensures that the insider breaks even ex ante, but leaves the security shape indeterminate. Notice that when we achieve the first-best social outcome, the insider gets paid only when the project is continued $X \geq I - \varepsilon$. But the insider's payoff is zero for $X < I$ anyway because ε is a $t = 0$ random variable whose realizations are assumed to be non-contractible. Therefore, the indicator function in Equation (2.20) is independent of ε. Finally, Equation (2.21) is driven by the entrepreneur's limited liability: As we argue shortly, she can use outsiders to commit herself to internalizing the cost inefficient continuation, but still cannot give outsiders more than $X - s_I(X)$.

Proposition 5 essentially states that entrepreneurs optimally use convertible securities for insiders and residuals for outsiders.[16] Figure 2.3 provides a concrete illustration of the optimal contract using convertible notes for the insider and equities for the outsiders. The dotted line indicates the threshold for continuation versus termination signals.

We now provide the intuition for Proposition 5. Because the inefficiency lies in the continuation of bad projects ($X + \varepsilon < I$), what matters is how the security makes the information producer (the entrepreneur) internalize the cost of such inefficiency (partially through outsiders that are present). From the insider's perspective, efficient continuation then entails her paying the fair value for the security λI when the investment is socially marginal (zero-NPV), lest there is either a form of debt overhang resulting in underinvestment (when $s_I(I - \varepsilon) > \lambda I$ and the insider gets more than her fair share of the surplus) or a subsidy from the insider that leads to overinvestment (when $s_I(I - \varepsilon) < \lambda I$).

16. In Online Appendix OA3, we show equity or debt could be optimal when the entrepreneur is constrained to issue the same security for the insider and the outsiders ($s_I(X) = \lambda s(X)$, $s_O(X) = (1 - \lambda)s(X)$, for some $s(X)$) due to regulatory reasons. In Online Appendix OA4, we discuss how our contractual setting relates to the classical literature on contracting and moral hazard.

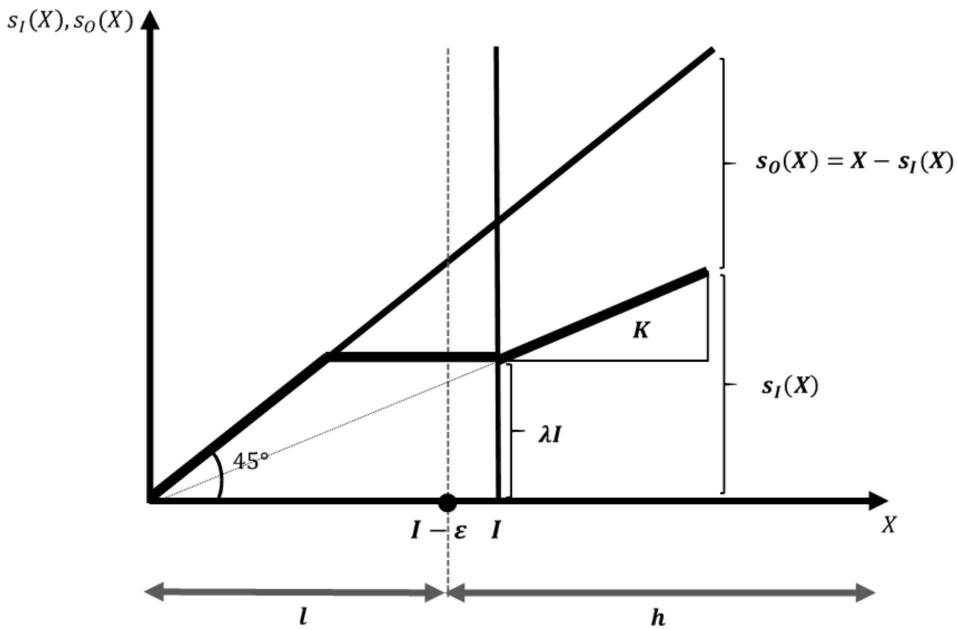


Figure 2.3: Illustration of optimal securities under flexible design.

At first, it seems that an optimal security only requires $s_I(I-\varepsilon) = \lambda I$, which makes the entrepreneur (i) indifferent between continuing and not continuing at $X = I-\varepsilon$ (social NPV is zero), (ii) strictly preferring continuation when $X > I - \varepsilon$, and (iii) strictly preferring termination when $X < I - \varepsilon$. Note that (ii) and (iii) follow from the monotonicity of $M(X; s_I, s_O, \lambda, \varepsilon)$ in X. However, ε represents an uncertainty resolved at $t = 1$, which then affects the entrepreneur's information production, and the entrepreneur cannot design the security at $t = 0$ to be dependent on one particular value of ε.[17]

As a result, committing to a large enough $1 - \lambda$ to expose the entrepreneur's payoff to inefficient continuation for the entire region of $X = I - \varepsilon$, $\varepsilon \in (0, \bar{\varepsilon})$ yields the debt-like flat

17. As we discuss in Online Appendix OA1, this is a general phenomenon when the information design space (from experimentation) differs from the contracting space (which is based on the cash flow only), not an artifact of our baseline assumption on ε.

region in Figure 2.3.[18] Meanwhile, the security design also needs to ensure that the insider earns enough from the second round to cover the initial investment K. This restricts the security somewhat, but as long as the area under s_I but above the horizontal line λI reaches K, its shape to the right of the debt-like region is indeterminate. Nevertheless, the endogenous experimentation leads to a determinate informational environment and investment decision that is also socially optimal.

We have demonstrated that optimal designs have to entail some form of debt-likeness and convertibility, consistent with real practice. Note that the debt-likeness is not driven by risk-bearing capacity or informational asymmetry, but by the entrepreneur's bias for continuation and contract incompleteness on information production. It aligns the entrepreneur's incentives in information production with that of the social planner and his ex-ante self.

Overall, Proposition 5 indicates that (i) conclusions from earlier studies are robust to introducing endogenous and flexible information production, and (ii) the optimal security for the insider not only has a debt-like region, but can also rationalize the use of a large class of securities in real life — an empirical observation other models do not fully account for.[19]

18. We prove in Online Appendix OA1 that even with a deterministic ε at $t = 0$, an optimal security design entails this debt-like flat region as long as there is some uncertainty at $t = 0$ (about the misalignment or final project cash flow) that resolves through the course of the financing relationship (starting from $t = 1$), because then pinning down a single point of the security based on ε does not trivially correct the divergence of incentives.

19. Our goal is not to introduce an alternative mechanism or competing theory for the predominant use of convertible securities in venture capital (Kaplan and Strömberg, 2003) or to proclaim the information design channel to be the most dominant. Besides proving the optimality of convertible securities, we emphasize the role of arm's-length outsiders and the need for the joint optimal security design for both the insider and outsiders, which extant theories do not discuss (Inderst and Vladimirov, 2019, is a notable exception).

2.5 Relationship finance under investor sophistication

In this section, we show how investor sophistication—the insider's ability to utilize a given information production technology other than the entrepreneur's experimentation to generate interim information—mitigates IPH. Interestingly, its interaction with interim competition helps us rationalize puzzling empirical patterns in the formation of lending relationships. We also prove that the contractual solution in Proposition 5 is robust to the level of the insider's sophistication. By doing so, we essentially develop the solution for a Bayesian persuasion game involving multiple receivers facing discriminatory disclosures. Finally, we discuss contracting on interim events (milestones).

We incorporate the insider's information production by endowing her with an exogenous technology to produce a proprietary signal about X during the financing relationship. In particular, the insider uses an experiment $(\mathcal{Y}, \omega_q)$, where $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$ is a finite set with m signals. $\omega_q(y_i|X)$ represents conditional distributions and we assume the posterior distribution induced by an experiment to be atomless with a full support. $q \in [0, 1]$ is an index we employ to rank the informativeness of different experiments, with $q = 0$ for an uninformative experiment. For $q > q' \in [0, 1]$, the experiment $(\mathcal{Y}, \omega_q)$ is more informative than $(\mathcal{Y}, \omega_{q'})$ in the sense of Blackwell (1953).

Moreover, for every q, we assume the signals in $\mathcal{Y}$ can be ranked: For every $m \geq i > i' \geq 1$, distribution $f(X|y_i)$ dominates distribution $f(X|y_{i'})$ in the sense of first order stochastic dominance (i.e. for every $X \in (0, 1)$, we have $F(X|y_i) < F(X|y_{i'})$). This assumption directly implies that $\mathbb{E}_q[s(X)|y_i, z] > \mathbb{E}_q[s(X)|y_{i'}, z]$ for every security $s(\cdot)$ and signal z with a non-degenerate distribution. The following example illustrate the points above.

Example 1. *The insider's experiment generates a binary signal, i.e.,* $\mathcal{Y} = \{\tilde{h}, \tilde{l}\}$*, with the* *following information structure for* $(\{\tilde{h}, \tilde{l}\}, \omega_q)$*:*

$$\omega_q(\tilde{h}|X) = \begin{cases} \frac{1+q}{2} & I \leq X \leq 1 \\ \frac{1-q}{2} & 0 \leq X < I. \end{cases} \tag{2.22}$$

The investor receives a signal $y = \tilde{h}$ with probability $\frac{1+q}{2} \geq \frac{1}{2}$ if the project is profitable, and $y = \tilde{l}$ with probability $\frac{1-q}{2} \leq \frac{1}{2}$ otherwise. Clearly $\mathbb{E}_q[s(X)|\tilde{h}] \geq \mathbb{E}_q[s(X)|\tilde{l}]$ for every $q > 0$ and security $s(\cdot)$. Moreover, if $1 > q > q' > \frac{1}{2}$, then the following inequalities also hold:

$$\mathbb{E}_q[s(X)|\tilde{h}] > \mathbb{E}_{q'}[s(X)|\tilde{h}] > \mathbb{E}_{q'}[s(X)|\tilde{l}] > \mathbb{E}_q[s(X)|\tilde{l}]. \tag{2.23}$$

The inequality means that experiments with higher values of q generate relatively more extreme signals.

Given the insider's information technology, the entrepreneur now designs an experiment $(\mathcal{Z}\times\mathcal{Y}, \pi)$, where $\pi(.,.|X) : \mathcal{Z}\times\mathcal{Y} \to [0,1]$ is the joint conditional probability of observing the signals. Note that the marginal distributions for $y \in \mathcal{Y}$ must be consistent with the insider's experiment, i.e., $\sum_{z\in\mathcal{Z}} \pi(z, y|X) = \omega_q(y|X)$ for every $X \in [0,1]$ and $y \in \mathcal{Y}$. Moreover, we allow the signals in $\mathcal{Z}$ and $\mathcal{Y}$ to be correlated conditional on the true state of the world. Similar to the baseline case, the outsiders observe signal z with probability $1-\mu$, while signal y is always privately observed by the insider financier.

2.5.1 Information production with investor sophistication

We start by characterizing in Lemma 3 the optimal experiment under investor sophistication. We denote the signal indicating whether $X \geq \max\{\hat{X}(\mu), \bar{X}(y_i)\}$ by x_H^i, where $\bar{X}(y_i)$ is the solution to $\mathbb{E}_q[s(X) - I|y_i, X \geq \bar{X}(y_i)] = 0$ if a solution exists and is zero otherwise;

we denote the opposite signal indicating whether $X < \max\{\hat{X}(\mu), \bar{X}(y_i)\}$ by x^i_L.

Lemma 3 (**Endogenous Information under Investor Sophistication**). *An optimal experiment exists and requires at most* $2|\mathcal{Y}| = 2m$ *signals. For every signal* $y_i \in \mathcal{Y}$, *the entrepreneur sends either* $z^h_i = \{y_i, x^i_H\}$ *or* $z^l_i = \{y_i, x^i_L\}$.

Lemma 3 implies that the entrepreneur can split the information design problem into m separate problems each characterized by Proposition 4.[20] Intuitively, the entrepreneur prefers to reveal y to the outsiders in order to level the playing field by informing them of y and eliminating the insider's informational advantage. Since the optimal experiment in Proposition 4 has at most two signals, the optimal experiment in the presence of a sophisticated investor has at most $2m$ signals.

Proposition 6. ***(a)*** *For every* $\mu \in [0, 1]$, *the insider's expected payoff* $U^I(\mu; q)$ *is weakly increasing in* q.

(b) *For any given* $(\mathcal{Y}, \omega_q)$, *the equilibrium investment decision is not ex-post socially optimal with a positive probability.*

Part (a) in Proposition 6 states that the insider's interim payoff increases with the informativeness of his endowed signal, which follows from Lemma 3 and Blackwell's theorem (Blackwell, 1953).[21] Part (b) derives from combining Corollary 1 with Lemma 3. The

20. A priori, it is not obvious that the entrepreneur should design the experiment to correspond with y type by type. For instance, if the insider and outsiders bid for the security simultaneously, the entrepreneur might find it optimal to only partially disclose the insider's signal to the outsiders. The simple design follows only after taking into consideration that outsiders learn from the insider's interim action.

21. Both Kolotilin (2018) and Guo and Shmaya (2019) derive non-monotone results for the case that the sender's and the receiver's experiments must generate *independent* signals conditional on X. We differ in that we do not restrict the insider's and the entrepreneur's information production to be independent. In Online Appendix OA5, we illustrate how our results extend to the setting with such independence requirement.

entrepreneur still tends to induce overinvestment, as he does not internalize the cost of experimentation K. Moreover, even in the case that the insider's signal is highly informative, the insider tends to underinvest since she does not internalize the entrepreneur's private benefit from the investment.

Overall, the proof further reveals that, while improving the entrepreneur's information production, the first-best outcomes cannot be achieved through investor sophistication alone. As such, initial long-term contracts with the right security design are integral to achieving the socially optimal investment, as we discuss in Section 2.5.3.

2.5.2 Relationship formation, sophistication, and competition

The theoretical predictions on the effect of competition on bank orientation have been ambiguous. According to the investment theory (e.g., Petersen and Rajan, 1995; Dell'Ariccia and Marquez, 2004), as the credit market concentration decreases, firms' borrowing options expand, rendering banks less capable of recouping their initial investments during the lending relationship, which hinders relationship banking. According to the strategic theory (e.g., Boot and Thakor, 2000; Dinc, 2000), fiercer interbank competition drives local lenders to take advantage of their competitive edge and reorient lending activities towards relational-based lending to small, local firms, which strengthens relationship banking. Others (e.g., Yafeh and Yosha, 2001; Anand and Galetovic, 2006) suggest that competition can have ambiguous effects on lending relationships, but typically predict an inverted U-shaped pattern.

Yet empirically Elsas (2005) and Degryse and Ongena (2007) document a U-shaped relationship between the likelihood of a lending relationship and the level of competition in the credit market.[22] Proposition 7 offers an explanation.

22. These two studies stand out because they measure relationship banking directly in terms of duration

Proposition 7 (**Relationship and Competition**). *$\exists\ \underline{\mu}(q) \in (0,1)$ such that for $\mu \in [\underline{\mu}(q), 1]$, the insider's payoff from the relationship financing, $U^I(\mu; q)$, is increasing in the level of interim competition $(1-\mu)$ for unsophisticated investors $(q = 0)$, decreasing for sophisticated investors (sufficiently large q), and U-shaped for investors with intermediate sophistication.*

On the one hand, for a fixed level of private benefit of continuation, lower levels of competition increase the insider's share of the surplus and are preferred by more sophisticated investors who can produce their own information. On the other hand, higher levels of competition can encourage more efficient information production from the entrepreneur, which increases total surplus. Thus, it is preferred by the less sophisticated investors who have no other means of obtaining information rent. For intermediate values of sophistication, competition hurts the insider's profit until it replaces the investor's independent information as her main source of interim rent, leading to the local U-shape.

Figure 2.4(a) illustrates the relationship between $U^I(\mu; q)$ and μ, the *inverse* measure of competition in our model. In particular, when q takes intermediate values and μ is exogenous, our model thus helps rationalize the findings of Elsas (2005) and Degryse and Ongena (2007). As is expected and shown in Figure 2.4(b), relationship formation eventually decreases with competition when the market becomes extremely competitive (μ gets closer to 0).

(a) The U-shaped relationship between U^I and $1-\mu$.

(b) Non-monotone relationship between $U^I(\mu; q)$ and $1-\mu$.

Figure 2.4: Illustration of the equilibrium capacity of the financing in the initial round as a function of the level of ex-post competition

and scope of interactions, thus improve upon and complement indirect measures such as the loan rate (Petersen and Rajan, 1995) or credit availability over firms' lifetime (Black and Strahan, 2002), for which the impact of competition could be ambiguous in equilibrium (Boot and Thakor, 2000).

2.5.3 Optimal contracting and security design

Next, we examine whether security design and contracting at $t = 0$ fully mitigate IPH under investor sophistication. Proposition 8 shows that Proposition 5 is robust to insider sophistication and still achieves the first-best outcome.

Proposition 8 (**Optimal Contracts with Investor Sophistication**). *Regardless of what* $(\mathcal{Y}, \omega_q)$ *the insider is endowed with, all optimal securities are characterized by Equations (2.19)-(2.21).*

The intuition behind Proposition 8 is the following: Because of the flat part of the security, the insider's security entails no risk when the entrepreneur reveals that $X > I - \varepsilon$. Therefore, the insider always continues the project when $z = h$, regardless of her own signal. In other words, the insider's action only reveals whether $X > I - \varepsilon$, and consequently, the outsiders invest if and only if the insider invests.

In contrast to the case without long-term contracts (Lemma 3), the entrepreneur's experiment under the optimal long-term contract does not depend on the insider's information production. In other words, the entrepreneur might find it optimal to hide some of the insider's information from the outsiders. Consequently, information asymmetry exists between the insiders and the outsiders, but it does not prevent implementation of the socially optimal investment, because the contractual solution resolves the inefficiencies.

2.5.4 Setting milestones

Contracting on interim events is related to "milestones" used in venture financing. For example, the entrepreneur can commit to reaching a pre-specified customer base before seeking additional financing. Would that solve the IPH problem? Perhaps surprisingly, the insider cannot increase her expected payoff by contracting on interim events.

Note that contracting on milestones is different from contracting on $(\mathcal{Z}, \pi)$. We use $\mathcal{Y}^b \subset \mathcal{Y}$ to denote binding signals, following which the insider commits to either continue or terminate financing. We denote the set of non-binding signals by $\mathcal{Y}^{nb} = \mathcal{Y} \setminus \mathcal{Y}^b$.

Corollary 2 (**Milestone Futility**). *The insider cannot gain from setting milestones. In particular, the insider's expected payoff is maximized when* $\mathcal{Y}^b = \emptyset$.

Corollary 2 helps explain why milestones are seldom binding in practice and early-stage projections are rarely made or enforced. The entrepreneur can flexibly control how the insider updates her prior for every $y \in \mathcal{Y}^{nb}$, regardless of the choice of $\mathcal{Y}^b$. Therefore, the insider's information set following signals in $\mathcal{Y}^{nb}$ does not change, while she potentially makes suboptimal decisions following signals in $\mathcal{Y}^b$ due to the binding commitment.

2.6 Discussions and extensions

Our main findings are robust under alternative model specifications. In the online appendix, we discuss (i) IPH and security design under restricted experimentation space (OA1); (ii) contracting and security design under IPH when the entrepreneur can only use one type of security (OA3); (iii) how IPH can also reduce the effort distortion introduced by Rajan (1992) (OA6); (iv) what happens when we allow a continuum of actions, i.e., scalable projects (OA7); and (v) the case of allocating security design right to the insider (OA8).

We next discuss how costly and endogenous information production or acquisition by the insider affects the entrepreneur's information production (Section 2.6.1), as well as the roles of the commitment to and the observability of experimentation (Section 2.6.2), both of which are fundamental issues in the literature of information design.

2.6.1 Costly information production by the insider

In our baseline setup (Section 2.3), only the entrepreneur has the expertise to experiment. Although in Section 2.5 we endow the insider with a specific information production technology, one may question what happens if the insider can also endogenously produce or acquire information. We now show that the insider's endogenous and costly experiment improves the entrepreneur's information production and consequently facilitates the relationship finance. Nevertheless, it does not fully resolve IPH and the information production is still socially inefficient.

Specifically, suppose an insider can pay $c > 0$ to observe X.[23] For illustration and simplicity, we assume $\mu = 1$, i.e., the insider faces no interim competition. Then following the realization of z, the insider incurs the cost and learns X if and only if the down-side risk implied by the signal is large enough.

$$\mathbb{E}[\{s(X) - I\}^{+}|z] - c \geq \mathbb{E}[s(X) - I|z] \iff -c \geq \mathbb{E}[\{s(X) - I\}^{-}|z], \tag{2.24}$$

where $\{x\}^{+} \equiv \max\{x, 0\}$ and $\{x\}^{-} \equiv \min\{x, 0\}$.

Therefore, the entrepreneur, aiming to maximize the probability of the investment, should choose a threshold larger than the baseline threshold $\bar{X}$ when c is sufficiently small:

$$-c > \mathbb{E}[\{s(X) - I\}^{-}|X \geq \bar{X}]. \tag{2.25}$$

23. Obviously, the insider would fully learn X if $c = 0$ and the problem becomes trivial. $c > 0$ reflects the fact that entrepreneur still produces information more efficiently than the investors.

In this case, the entrepreneur's optimal threshold is $\bar{X}^c > \bar{X}$, which satisfies:

$$-c = \mathbb{E}[\{s(X) - I\}^- | X \geq \bar{X}^c]. \tag{2.26}$$

We see that the insider's access to the costly information production technology can discipline the entrepreneur's experimentation by imposing an implicit constraint on the set of entrepreneur's experiments (Equation (2.24) in our setting). The constraint typically alleviates IPH and generally depends on the specific information production technology available to the investors.

Equation (2.26) also provides insights about which types of costly signals and security forms discipline the entrepreneur's experimentation more. As for the former, suppose the insider has access to all potential threshold experiments by paying c, i.e., she learns whether $X < T$ for a threshold level $T \in [0, 1]$ she pre-specifies. By repeating the arguments above, we see that a less informative signal induces the same entrepreneur's experiment, with the same threshold $\bar{X}^c$. Therefore, what really matters for the entrepreneur's information production is how cheaply the insider can assess the project's downside risks, which lies at the core of the misalignment of interests.

Regarding the role of the form of the security $s(\cdot)$, according to Equation (2.24), securities that allocate more downside risk to the insider are more effective in disciplining the entrepreneur. Corollary 3 formalizes this point.

Corollary 3. *For securities $s^1(\cdot)$ and $s^2(\cdot)$, suppose $s^1(X) - I$ and $s^2(X) - I$ are negative over the same range of values $[0, X^s]$, for some $X^s \in (0, 1)$, and the following condition holds:*

$$s^1(X) - I \leq s^2(X) - I \leq 0 \qquad \forall\ X \leq X^s \tag{2.27}$$

Then the entrepreneur chooses experiments with a threshold structure under both securities; however, the threshold is larger for $s^1(\cdot)$ than $s^2(\cdot)$. In other words, the entrepreneur's experiment induces a more socially efficient investment with security $s^2(\cdot)$ than with security $s^1(\cdot)$

The intuition for Corollary 3 is that the insider is more likely to acquire the costly signal when she faces a larger potential loss. Consequently, a security that allocates more downside risk to the insider forces the entrepreneur to choose a more informative experiment, in order to crowd out the insider's information production, in line with the finding of Yang and Zeng (2018) about the prevalent use of equity.

To summarize, we see the insider's access to costly information production technology improves the entrepreneur's information production and increases the possibility of relationship financing. Furthermore, the technology is more effective when the insider is more exposed to potential losses, as well as when the technology delivers a more precise assessment of the downside risks. The results here complement those in Section 2.5 by explaining the role of insider's flexibility in information production instead of the precision of her endowed signals. Both lead to more socially efficient investment outcomes, while none of them fully resolves the IPH issue.

2.6.2 Commitment to information design and partial observation

Ex-ante commitment to experimentation. Commitment issues are common in applications of information design. Our setup does not require that the entrepreneur commit to an experiment when raising K, but allows the investor to observe the experiment after becoming an insider. In fact, the lack of commitment and contractibility of the experimentation at $t = 0$ in our setting (partially) breaks the indeterminacy in Szydlowski (2019) and prompts the debt-like region of the optimal security design as a response. As shown above, the choice of security ex ante affects the choice of information design ex post.

Partial observation of experiments. In the baseline model, we assume that not only does the investor's monitoring technology rule out misreporting, but the investor also commits to perfectly monitoring the entrepreneur's experiment, even though she might be better off randomizing between monitoring and not monitoring. We relax these assumptions and find that while the first-best outcome becomes no longer achievable, our results on IPH and optimal contracting and security design still hold. Specifically, we assume:

1. With probability $\alpha \in [0, 1]$, the entrepreneur can misreport the signal z without getting caught even while being monitored by the investor.

2. At time $t = 0$, the investor commits to verifying the signal realization with probability $\beta \in [0, 1]$.[24]

Note that our baseline model corresponds to $\alpha = 0$ and $\beta = 1$. In general, the following variant of Corollary 1 and Proposition 5 holds.

Proposition 9 (**Partial Observation and Commitment**).

(a) *Absent long-term contracts, the insider receives no interim rent for extreme values of μ regardless of the values of $\alpha, \beta \in [0, 1]$.*

(b) *For $\alpha > 0$ and $\beta < 1$, there exists no contractual solution that implements the first-best outcome. For small enough values of α and large enough values of β, the convert-*

24. Note that if the investor monitors the experiment, she makes the continuation decision based on both the reported and the monitored signal. In this case, we allow the investor to commit to punishing the entrepreneur by not investing in the case of misreporting. We can easily allow the verification to incur a small cost, and the extension would share the spirit of, for example, Trigilia (2019). The key difference from the literature on costly-state-verification (CSV) is again the endogenous design of the information.

ible securities are still optimal for all ε, provided an optimal contract exists. Relationship financing is infeasible for large enough values of α and low enough values of β.

Proposition 9 validates our prior knowledge that the insider's monitoring is essential to relationship financing. But Part (a) also shows that the IPH is robust to partial monitoring when the competition is too low or too high. Furthermore, note that in contrast to costly state verification models, where the insider optimally randomizes between monitoring and not monitoring, Part (b) shows that the full observation by the investor is required to implement the socially optimal outcome. The main difference is that the investor has no way to castigate the entrepreneur for misreporting, as the latter is protected by his limited liability.

2.7 Conclusion

We model the dynamic financing of projects by relationship and arm's-length investors as a mechanism design problem with an embedded Bayesian persuasion game whereby the entrepreneur endogenously produces interim information to seek continued financing. We show that the entrepreneur's (sender's) endogenous experimentation typically reduces the insider investor's (receiver's) information-monopoly rent, holding up that relationship financier's incentives to form the relationship in the first place. Investor sophistication and interim competition can mitigate the problem, and they interact to produce non-monotone patterns of relationship formation and interim competition. We then derive optimal sequential securities to resolve the IPH problem: the entrepreneur contracts with investors in the initial round to allow them to purchase convertible securities in a later round, and issues residual claims to competitive outsiders later.

Our theory broadly applies to hold-up problems in persuasion games with contingent transfers and multiple receivers of different types. It is immediately relevant for at least two major areas of finance. First, it reveals the impact of endogenous information production in relationship lending and clarifies its interactions with investor sophistication and competition, rationalizing the puzzling U-shaped link between bank orientation and interim competition documented. Second, it highlights how IPH can rationalize the use of a large variety of convertible securities in venture capital. Given that the solutions to many of the world's biggest problems such as Alzheimer's disease, global warming, and fossil fuel depletion require initial funding for experimentation and reliable financing relationships (Nanda and Rhodes-Kropf, 2013; Hull, Lo, and Stein, 2019), the cost of inefficient information production could be tremendous. Our study constitutes a first attempt to underscore, formalize, and then examine potential contractual solutions for this practical issue.

Appendix: Proofs and Mathematical derivations

2.A.1 A technical lemma

The entrepreneur endogenously designs the experiment to maximize his payoff, subject to the insiders' second-round participation constraint. With a finite state space, the signal space as the range of a deterministic mapping from the state space is necessarily finite (Bergemann and Morris, 2016). Consequently, we can apply the method of Lagrange multipliers directly. But alas, we are dealing with infinite dimensional state space and unrestricted signal generation space. Technically speaking there is no guarantee that one can apply the method of Lagrange multipliers without additional regularity conditions.

We tackle this issue in two parts in Lemma A1. First, we show for any experiment, there exists a binary experiment that yields the entrepreneur the same expected payoff. Therefore, without loss of generality, we can restrict our attention to experiments generating binary signals of continuation versus termination. Had we assumed $\mathscr{Z}$ to be a continuum, our discussions remain the same except that notation-wise we have to mix summations with integrals and allow delta functions in probability densities. Because the experiments we look at are conditional probabilities mappings from the state space to $[0, 1]$, they live in a Banach space. With this insight and Part (a) of the lemma, we prove a mathematical result in Part (b) which allows us to use the method of Lagrange multipliers in the proofs of our lemmas and propositions (see also Ito (2016) for an abstract generalization).

[**A1**] **(a)** Consider a Bayesian Persuasion game à la Kamenica and Gentzkow (2011), with metric state space Ω, ex-ante probability measure μ_Ω, compact metric signal space S, with induced probability measure μ_S, and receiver's action space $a \in \{0, 1\}$. Suppose the sender's and receiver's payoff from action a in space $w \in \Omega$ are $a \times u(w)$ and $a \times v(w)$ respectively, for some real-valued measurable functions $u, v : \Omega \to \mathfrak{R}$. Then, for any experiment, denoted by measurable conditional probability functions $\pi(s|w)$ over S, there exists binary experiment $(\{h, l\}, \pi')$ that implements the same mapping from the states to actions, and thus the same expected payoff for both sender and receiver.

(b) Suppose $w_i(x), m_i(x) : [0, 1] \to \mathfrak{R}$ $(1 \leq i \leq N)$ are continuous and bounded functions.

Suppose the following maximization problem has a solution:

$$\begin{aligned}&\max_{\alpha_i(.)\in\mathcal{A}}\int_0^1\sum_{i=1}^{N}w_i(x)\alpha_i(x)dx\\ s.t.\quad &\int_0^1 m_i(x)\alpha_i(x)dx\geq 0\quad\forall\ 1\leq i\leq N,\quad and\quad \sum_{i=1}^{N}\alpha_i(x)\leq 1\quad\forall x\in[0,1],\end{aligned}\tag{2.A.1}$$

where $\mathcal{A}$ is the set of all measurable functions over $[0,1]$ that take value from $[0,1]$. Then, there exist non-negative real numbers $\{\mu_i\}_{i=1}^{N}$, such that the solution to (2.A.1) is a solution to the following maximization problem:

$$\max_{\alpha_i(.)\in\mathcal{A}}\int_0^1\sum_{i=1}^{N}(w_i(x)+\mu_i m_i(x))\alpha_i(x)dx\quad s.t.\quad\sum_{i=1}^{N}\alpha_i(x)\leq 1\quad\forall x\in[0,1]\tag{2.A.2}$$

Proof. **Proof of Part (a)**

Let S^+ be the set of signals that implement action $a=1$. Therefore, signal s belongs to S^+ iff the following condition holds (assuming the ties are broken in favor of action $a=1$):

$$\int_\Omega v(w)\pi(s|w)d\mu_\Omega\geq 0\tag{2.A.3}$$

Note that S^+ is a measurable subset of S, because it is a super-level set of function $\int_\Omega v(w)\pi(s|w)d\mu_\Omega$. That said, define experiment $(\{h,l\},\pi')$ as follows:

$$\pi'(h|w)=\int_{S^+}\pi(s|w)d\mu_S\quad and\quad \pi'(l|w)=1-\pi'(h|w)\qquad\forall w\in\Omega\tag{2.A.4}$$

In experiment $(\{h,l\},\pi')$, only signal h induces action $a=1$, because

$$\begin{aligned}\int_\Omega v(w)\pi'(h|w)d\mu_\Omega&=\int_\Omega\int_{S^+}v(w)\pi(s|w)d\mu_S d\mu_\Omega\geq 0,\\ \int_\Omega v(w)\pi'(l|w)d\mu_\Omega&=\int_\Omega\int_{S\setminus S^+}v(w)\pi(s|w)d\mu_S d\mu_\Omega<0.\end{aligned}\tag{2.A.5}$$

Equation 2.A.6 shows for any $w\in\Omega$, the unconditional probabilities of action $a=1$ taken by the sender in experiments (S, π) and $(\{h, l\}, \pi')$ are equal. That directly implies that the expected payoff of the sender and receiver are also the same under these two experiments. It completes the proof.

$$Prob(a = 1|w; (S, \pi)) = \int_{S^+} \pi(s|w) d\mu_S = \pi(h|w) = Prob(a = 1|w; (\{h, l\}, \pi')) \quad (2.A.6)$$

Proof of Part (b)

Let $\tilde{\mathcal{A}}^N$ be the set of all $N-$tuples of functions $(\alpha_1(.), \ldots, \alpha_N(.))$ in $\mathcal{A}$ that satisfy $\sum_{i=1}^N \alpha_i(x) \leq 1$, for every $x \in [0, 1]$.

Since all functions are bounded and measurable, it is easy to check that $\tilde{\mathcal{A}}^N$ constitutes a closed set in ${\mathcal{L}^1}^N$. The following maximization problem is then well-defined.

$$\max_{(\alpha_1(.), \ldots, \alpha_N(.)) \in \tilde{\mathcal{A}}^N} \int_0^1 \sum_{i=1}^N w_i(x) \alpha_i(x) dx \quad s.t. \quad \int_0^1 m_i(x) \alpha_i(x) dx \geq 0 \quad \forall \ 1 \leq i \leq N \quad (2.A.7)$$

Suppose $a^* \in \tilde{A}^N$ is the solution to the problems (2.A.1) and (2.A.7). It is easy to see that Slater condition and strong duality hold. Therefore, there exists a vector of non-negative real numbers $\{\mu_i\}_{i=1}^N$ such that a^* also solves the following maximization problem:

$$\max_{(\alpha_1(.), \ldots, \alpha_N(.)) \in \tilde{\mathcal{A}}^N} \int_0^1 \sum_{i=1}^N w_i(x) \alpha_i(x) dx + \sum_{i=1}^N \mu_i \int_0^1 m_i(x) \alpha_i(x) dx \quad (2.A.8)$$

Note that (2.A.2) is equivalent to (2.A.8), which completes the proof. □

2.A.2 Proof of Proposition 4

Optimality of a binary experiment with threshold structure. According to (2.5) and (2.6), we can appeal to Lemma A1(a), by considering state space $[0, 1]$ and the following utility functions:

$$u(X) \equiv \varepsilon + X - \mu s(X) - (1-\mu)I$$
$$v(X) \equiv \mu(s(X) - I) \tag{2.A.9}$$

Therefore, the entrepreneur optimally chooses a binary experiment. Then, employing Lemma A1(b) for $N = 1$, the entrepreneur solves the following optimization problem.

$$\max_{\pi(h|X)} \int_0^1 [\varepsilon + X - \mu s(X) - (1-\mu)I]\pi(h|X)f(X)dX \tag{2.A.10}$$

$$s.t. \int_0^1 (s(X) - I)\pi(h|X)f(X)dX \geq 0, \quad and \quad \pi(h|X) \in [0,1]$$

In (2.A.10), the entrepreneur maximizes his expected payoff given the participation constraint for the investors. Note that $\mathbb{E}[s(X)|X \geq I - \varepsilon] \geq I$. Therefore, $s(X)$ exceeds I with a positive probability. Furthermore, the set of measurable functions satisfying the constraint in (2.A.10) is a closed and bounded subset of $\mathcal{L}^1$. As a result, (2.A.10) has a solution.

The participation constraint in (2.A.10) could be either binding or non-binding. When it is non-binding, the optimal experiment sets $\pi^*(h|X) = 1$ for all values of X for which the value in the bracket is non-negative. It corresponds to the set $\{X \geq \hat{X}(\mu)\}$. When the constraint is binding, we apply Lemma A1 with $N = 1$. Let $\hat{\lambda}$ be the corresponding multiplier. Then the optimal experiment $\pi^*(h|X)$ solves:

$$\max_{\pi(h|X)} \int_0^1 [\varepsilon + X - I + (\hat{\lambda} - \mu)(s(X) - I)]\pi(h|X)f(X)dX \tag{2.A.11}$$

The term in the bracket is strictly increasing in X, because

$$\frac{d_+}{d_+X}[\varepsilon + X - I + (\hat{\lambda} - \mu)(s(X) - I)] = 1 + (\hat{\lambda} - \mu)\frac{d_+}{d_+X}s(X) > 1 - \mu\frac{d_+}{d_+X}s(X) \geq 1 - \mu \geq 0, \tag{2.A.12}$$

where $\frac{d_+}{d_+X}$ denotes the right derivative. Therefore, the optimal experiment has a threshold scheme, where the threshold $\bar{X}$ satisfies $\int_{\bar{X}}^1 (s(X) - I)f(X)dX = 0$. Since the constraint in this case is binding, $\hat{X}(\mu) \leq \bar{X}$ whereas the opposite holds in the first case. The entrepreneur thus always follows a threshold strategy where the threshold is given by $\max\{\bar{X}, \hat{X}(\mu)\}$. The uniqueness of the payoffs follows directly from Lemma A1(a).

Uniqueness and mixed strategies. So far, we have assumed that the insider follows a pure strategy. As follows, we prove that even if the insider can randomize between continuation and termination following some interim signals, the same essentially unique equilibrium ensues.

To see this, we show that there exists no mixed-strategy Nash Equilibrium in which the insider financier terminates the project with a positive probability when she is indifferent between continuation and termination. Suppose the contrary that the entrepreneur uses experiment $(\mathcal{Z}', \pi')$ and the insider uses the investment function $i'(.) : \mathcal{Z}' \to [0,1]$. If the insider randomizes following some signal realization $z' \in \mathcal{Z}'$, i.e. $i'(z') \in (0,1)$, it implies that the insider should be indifferent between continuation and termination after observing z', i.e. $\mathbb{E}[s(X) - I|z'] = 0$. If z' realizes with a positive probability, then there exists $X_{z'} \in (0,1)$ such that $P(X \geq X_{z'}|z') > i'(z')$. Then, consider an alternative experiment $(\mathcal{Z}'', \pi'')$ that splits signal z' to signals z'_h and z'_l, where $\mathcal{Z}'' = (\mathcal{Z}' \setminus \{z'\}) \cup \{z'_h, z'_l\}$ and $\pi''(z'_h|X) = \pi'(z'|X)\mathbb{I}_{\{X \geq X_{z'}\}}$, $\pi''(z'_l|X) = \pi'(z'|X)\mathbb{I}_{\{X < X_{z'}\}}$ and $\pi''(z''|X) = \pi'(z''|X)$ for all $z'' \in \mathcal{Z}' \setminus \{z'\}$. We have:

$$\begin{aligned} \mathbb{E}[s(X) - I|z'_h] > \mathbb{E}[s(X) - I|z'] = 0 > \mathbb{E}[s(X) - I|z'_l], \quad \textit{and} \\ U^E(\mathcal{Z}'', \pi'') - U^E(\mathcal{Z}', \pi') = P(z')[P(X \geq X_{z'}|z')\mathbb{E}[\varepsilon + X - s(X)|z', X \geq X_{z'}] \\ -i'(z')\mathbb{E}[\varepsilon + X - s(X)|z']] > 0, \end{aligned} \tag{2.A.13}$$

because $P(X \geq X_{z'}|z') > i'(z')$ by construction and $\mathbb{E}[\varepsilon + X - s(X)|z', X \geq X_{z'}] \geq \mathbb{E}[\varepsilon + X - s(X)|z']$ due to the monotonicity of $X - s(X)$. It contradicts the optimality of $(\mathcal{Z}', \pi')$. As such, any outcome that involves randomization by the insider cannot emerge as an equilibrium outcome.

2.A.3 Proof of Corollary 1

We remind the readers that the corollary does not assert that relationship financing is always feasible in $[\mu^l, \mu^h]$. The key message is that there are regions in which relationship financing breaks down.

The insider's equilibrium interim payoff is as follows:

$$U^I\left(\{h,l\},\pi^*(\mu);\mu\right) = \mu\mathbb{E}\left[(s(X)-I)\mathbb{I}_{\{X\geq\max\{\bar{X},\hat{X}(\mu)\}\}}\right], \qquad (2.A.14)$$

where $\pi^*(\mu)$ denotes the optimal experiment for μ. It is easy to see that the insider's payoff is zero for $\mu = 0$ and $\mu = 1$. Given that U^I is continuous in μ, the insider's expected payoff is less than K for a neighborhood around $\mu = 0$ and $\mu = 1$. The corollary follows.

2.A.4 Proof of Proposition 5

A design implements the socially optimal outcome when the investment takes place iff $X \geq I - \varepsilon$. We introduce a security that maximizes the entrepreneur's expected payoff and implements the socially optimal outcome. Then, we characterize the set of optimal designs that achieves the first-best regardless of the realization of ε.

Social optimality of optimal designs. Note that the social surplus from the relationship financing for a given ε is bounded by

$$U_{FB}(\varepsilon) = \mathbb{E}\left[(\varepsilon + X - I)\mathbb{I}_{\{\varepsilon+X-I\geq 0\}}\right] - K. \qquad (2.A.15)$$

We show that this bound is achievable for a contract that satisfies the constraints in (2.18)-(2.21). Note that (2.19) implies that $M(X; s_I, s_O, \lambda, \varepsilon) = \varepsilon + X - I$ for all $X \in [\lambda I, I]$, where $M(\cdot)$ is defined in (2.14). Since $M(\cdot)$ is increasing in X and $M(I-\varepsilon; s_I, s_O, \lambda, \varepsilon) = 0$, the entrepreneur sends a high signal for $X \geq I - \varepsilon$, provided the security can cover the investment cost for the insiders and outsiders. Condition (2.20) ensures that is the case for the insider. Furthermore, the condition ensures the entrepreneur can raise $(1-\lambda)I$ from the outsiders by issuing $s_O(\cdot)$, since:

$$
\begin{aligned}
\int_{I-\varepsilon}^{1}(s_O(X)-(1-\lambda)I)f(X)dX &= \int_{I-\varepsilon}^{1}(X-s_I(X)-(1-\lambda)I)f(X)dX \\
&= \int_{I-\varepsilon}^{1}(X-I)f(X)dX-\int_{I-\varepsilon}^{1}(s_I(X)-\lambda I)f(X)dX \quad (2.A.16) \\
&= \int_{I-\varepsilon}^{1}(X-I)f(X)dX-K=\mathbb{E}[(X-I)\mathbb{I}_{\{X\geq I-\varepsilon\}}]-K\geq 0 \\
\Rightarrow p^O(\varepsilon)-(1-\lambda)I &= \mathbb{E}\left[s_O(X)-(1-\lambda)I|X\geq I-\varepsilon\right]\geq 0 \quad \forall\varepsilon\in(0,\bar{\varepsilon}),
\end{aligned}
$$

where we used the conditions (2.19) and (2.20) in the third equation. As a result, the entrepreneur receives expected payoff $U_{FB}(\varepsilon)$ for all $\varepsilon \in (0, \bar{\varepsilon})$ and the socially optimal outcome is implemented. We have proven that all optimal designs implement the socially optimal outcome.

The set of optimal designs. We now argue that the set of contracts specified in (2.18)-(2.21) are the only optimal designs. In order for a contract to implement the socially efficient outcome for all values of ε, we need to have $M(I-\varepsilon; s_I, s_O, \lambda, \varepsilon) = 0$ for all $\varepsilon \in (0, \bar{\varepsilon})$. Therefore, we should have $s_I(I-\varepsilon) = \lambda I$ for all $\varepsilon \in (0, \bar{\varepsilon})$. It proves the necessity of condition (2.19) for $X > I-\bar{\varepsilon}$. Furthermore, there would be no investment for $X < I-\bar{\varepsilon}$, as it would be inefficient for any value of ε. It implies the contingent transfers for these states are irrelevant.

According to (2.19), we need to have $\lambda I = s_I(I-\varepsilon) \leq I-\varepsilon$ for all $\varepsilon \in (0, \bar{\varepsilon})$, which implies that (2.18) needs to hold as well. (2.20) ensures that the insider breaks even over the course of the relationship. Finally, the inequality in (2.A.16) becomes equality as ε goes to $\bar{\varepsilon}$. Therefore, the limited liability condition has to bind in order to implement the socially optimal investment decision when the entrepreneur's private benefit is large.

2.A.5 Proof of Lemma 3

We first prove a useful lemma:

Lemma 4. *The optimal (Z, π) is at least as informative as (q, ω_q), in the Blackwell sense.*

In other words, the outsiders perfectly infer $y \in \mathcal{Y}$ by observing $z \in \mathcal{Z}$ from the endogenous experimentation.

Proof. We show that the entrepreneur designs the experiment in a way that the realized signal z fully reveals the insider's signal y. In other words, for a given signal $z \in \mathcal{Z}$ in the optimal experiment, there exists signal $y \in \mathcal{Y}$ such that $P(y|z) = \frac{\int_0^1 \pi(z,y|X)f(X)dX}{\sum_{y' \in \mathcal{Y}} \int_0^1 \pi(z,y'|X)f(X)dX} = 1$.

Consider experiment $(\mathcal{Z}, \pi)$ and signal $z \in \mathcal{Z}$. Suppose there are $l \geq 2$ distinct signals $\tilde{\mathcal{Y}}(z) = \{y_1, y_2, \ldots, y_l\} \subset \mathcal{Y}$ such that $P(y_i|z) > 0$ for all $1 \leq i \leq l$. We show that the entrepreneur can increase his expected payoff by splitting signal z into signals $z_1, z_2, \ldots, z_l$, where $\pi(z_i, y_j|X) = \pi(z, y_j|X)\mathbb{I}_{i=j}$.

It should be apparent that the insider either chooses $\lambda = 1$ or $\lambda = 0$ because her expected payoff is linear in her amount of investment. Suppose the insider makes the investment for a subset $\tilde{\mathcal{Y}}^+(z) \subsetneq \tilde{\mathcal{Y}}(z)$, following signal z. When signal z is public and $\lambda = 0$, the outsiders offer $p^O = \mathbb{E}[s(X)|z, y \in \tilde{\mathcal{Y}}(z) \setminus \tilde{\mathcal{Y}}^+(z)]$ if it exceeds the cost of investment I, otherwise they do not make any offer. We argue that $p^O < I$, that is the outsiders never make any offer when the insider has a strictly more informative signal.

Consider the contrary and suppose $p^O = \mathbb{E}[s(X)|z, y \in \tilde{\mathcal{Y}}(z) \setminus \tilde{\mathcal{Y}}^+(z)] \geq I$. Then there should exist $\tilde{y} \in \tilde{\mathcal{Y}}(z) \setminus \tilde{\mathcal{Y}}^+(z)$ such that $\mathbb{E}[s(X)|z, \tilde{y}] \geq I$. It implies that the insider should also continue after $(z, \tilde{y})$, or equivalently, $\tilde{y}$ needs to be in $\tilde{\mathcal{Y}}^+(z)$ as well, which is a contradiction. Therefore, the insider never pays more than I following such signal z, even if z is publicly observed.

In contrast, if the entrepreneur splits the signals into $z_1, z_2, \ldots, z_l$ as mentioned above, he gets strictly more than I for the realization of z_i and y_i that $\mathbb{E}[s(X)|z_i, y_i] > I$ when z_i is publicly observed. Therefore, the entrepreneur would indeed be better off by the split. □

Lemma 4 essentially says that the entrepreneur optimally conveys the insider's private signal to the outsiders. This does not matter when z is only observed by the insider (with probability μ). But when z is public (with probability $1-\mu$), the entrepreneur prefers to level the playing field by informing the outsiders of y and eliminating the insider's informational advantage. A priori, one might think this could hurt the entrepreneur's payoff because a negative y signal may decrease the probability of continued financing from outsiders. But obfuscating the signal is not helpful here because the insider's termination action upon seeing a negative y already conveys the information to the outsiders.[25]

As such, the entrepreneur essentially faces m different experiment design problems each specified by (2.7) with the priors $f(X|y_i)$. Proposition 4 then leads us to the optimal experiments under investor sophistication. The only exceptions are the cases in which either $P(s(X) \geq I|y_i) = 0$ or $\mathbb{E}[X|y_i] > 0$, where $\bar{X}(y_i)$ does not exist. In the first case, there would be no investment by the insider, and consequently the outsiders (according to Lemma 4), regardless of the entrepreneur's choice of signals. In the second case, the entrepreneur optimally induces investment only when $X \geq \hat{X}(\mu)$.

2.A.6 Proof of Proposition 6

Proof of Part (a)

As follows, we show that the insider financier earns a higher expected payoff from a more informative experiment. Consider two experiments $(\mathcal{Y}, \omega_q)$ and $(\mathcal{Y}, \omega_{q'})$ with $q' > q$. According to Blackwell (1953), there exists an $m \times m$ Markovian matrix T such that $f_q(X|y_i) = \sum_{j=1}^{m} T_{ij} f_{q'}(X|y_j)$. Moreover, we can write the insider's expected payoff from experiment $(\mathcal{Y}, \omega_q)$ as:

$$U^I(\mu; q) = \mu \sum_{y \in \mathcal{Y}} P_q(y_i) \mathbb{E}\left[(s(X) - I)\mathbb{I}_{\{X \geq \max\{\hat{X}(\mu), \bar{X}(y)\}\}}\right]. \tag{2.A.17}$$

According to the definition of $\bar{X}(y)$ introduced in Lemma 3, $\bar{X}(y) > 0$ implies that $\mathbb{E}[(s(X) - I)\mathbb{I}_{\{X \geq \bar{X}\}}] = 0$. We can thus rewrite (2.A.14) as

$$U^I(\mu; q) = \mu \sum_{i=1}^{m} P_q(y_i) \max\left\{\int_{\hat{X}(\mu)}^{1} (s(X) - I) f_q(X|y_i) dX, 0\right\} \tag{2.A.18}$$

25. Note the lemma relies on the entrepreneur's ability in designing an experiment that nests the insider's experiment. This is not crucial for the results to follow, as we demonstrate in Online Appendix OA5.

Substituting $f_q(X|y_i)$ by $\sum_{j=1}^m T_{ij} f_{q'}(X|y_j)$, we have

$$\begin{aligned} U^I(\mu;q) &= \mu \sum_{i=1}^m P_q(y_i) \max\left\{ \int_{\hat{X}(\mu)}^1 (s(X)-I) \sum_{j=1}^m T_{ij} f_{q'}(X|y_j) dX, 0 \right\} \\ &\leq \mu \sum_{i=1}^m P_q(y_i) \sum_{j=1}^m T_{ij} \max\left\{ \int_{\hat{X}(\mu)}^1 (s(X)-I) f_{q'}(X|y_j) dX, 0 \right\} \\ &= \mu \sum_{i=1}^m P_{q'}(y_i) \max\left\{ \int_{\hat{X}(\mu)}^1 (s(X)-I) f_{q'}(X|y_j) dX, 0 \right\} = U^I(\mu;q'), \end{aligned}$$

where the last inequality follows from the identity $P_{q'}(y_j) = \sum_{i=1}^m T_{ij} P_q(y_i)$.

Proof of Part (b)

Suppose the contrary that there exists an insider's experiment $(\mathcal{Y}, \omega_q)$ that leads to the socially optimal investment decisions. To implement the socially optimal outcome, the threshold for all m signals should be $I - \varepsilon$. We thus need to have $\max\{\hat{X}(\mu), \bar{X}(y)\} = I - \varepsilon$ for all $y \in \mathcal{Y}$. Since $\hat{X}(\mu) < I - \varepsilon$ for all $\mu > 0$, then we need to have $\bar{X}(y) = I - \varepsilon$ for all signals in $\mathcal{Y}$. Therefore, by definition of $\bar{X}(y)$, we have $\mathbb{E}[s(X) - I | y_i, X \geq I - \varepsilon] = 0$ for all $1 \leq i \leq m$. Therefore, even though the optimal experiment is socially efficient, the insider receives zero interim expected payoff, failing to recover the initial cost K. Then the insider would not start the relationship financing in the first place, contradicting the outcome being socially optimal.

2.A.7 Proof of Proposition 7

The derivative of (2.A.17) with respect to μ (when it exists) is

$$\begin{aligned} \frac{d}{d\mu} U^I(\mu;q) = &\sum_{y \in \mathcal{Y}} P_q(y) \mathbb{E}[(s(X)-I) \mathbb{I}_{\{X \geq \max\{\hat{X}(\mu), \bar{X}_q(y)\}\}}] + \\ &\mu \sum_{y \in \mathcal{Y}} P_q(y) (s(\hat{X}(\mu)) - I) f(\hat{X}(\mu)|y) \mathbb{I}_{\{\hat{X}(\mu) > \bar{X}_q(y)\}} \end{aligned} \tag{2.A.19}$$

To derive the relation between the insider's expected payoff and μ, fix q and consider two cases:

1. Suppose $\mu \geq 1 - \frac{\varepsilon}{I}$, which implies $\hat{X}(\mu) = 0$. Then (2.A.14) implies that $U^I(\cdot; q)$ is weakly increasing in μ for $\mu \in [1 - \frac{\varepsilon}{I}, 1]$.

2. Suppose $\mu < 1 - \frac{\varepsilon}{I}$, then $\hat{X}(\mu) > 0$. In this range of values of μ, if $\bar{X}_q(y) < \hat{X}(\mu)$ for some $y \in \mathcal{Y}$, then the first term in the right-hand side of (2.A.19) is positive and the second term is negative. For small enough values of μ the derivative is strictly positive, since the first term dominates the second term. Moreover, the derivative is weakly decreasing, since both of the terms are decreasing in μ. It implies the insider's expected payoff is concave in μ for $\mu \in [0, 1 - \frac{\varepsilon}{I}]$.

Denote $\bar{\mu} \in [0, 1 - \frac{\varepsilon}{I}]$ the maximizer of $U^I(\cdot; q)$. If $\bar{\mu} < 1 - \frac{\varepsilon}{I}$, then the insider's expected payoff is U-shaped in μ for $\mu \in [\bar{\mu}, 1]$, which completes the proof.

2.A.8 Proof of Proposition 8

For every insider's experiment $(\mathcal{Y}, \omega_q)$, (2.19)-(2.21) characterize the set of optimal long-term contracts. In particular, we show that under these conditions the entrepreneur optimally designs a binary experiment that sends a high signal if $X \geq I - \varepsilon$, which induces investment. First, suppose the entrepreneur chooses this experiment. (2.19) implies the insider always invests if she learns that $X \geq I - \varepsilon$. (2.20) and (2.21) together imply the outsiders also invest if and only if the entrepreneur's experiment sends a high signal. The reason is that the insider's action is binary and it only reveals the signal of the entrepreneur's experiment. Therefore, the project is invested if and only if $X \geq I - \varepsilon$. Moreover, by an argument similar to the proof of Proposition 5, it is the optimal experiment for the entrepreneur and these contracts yield the entrepreneur the whole social surplus. Now to show that the optimal design has to satisfy (2.19)-(2.21), we can use the argument almost verbatim in the proof of Proposition 5.

2.A.9 Proof of Corollary 2

It is easy to show that Lemma 4 still holds: the entrepreneur chooses an experiment strictly more informative than $(\mathcal{Y}, \omega_q)$. It means the entrepreneur still solves m independent information design problem for every signal in $\mathcal{Y}$ to determine the additional information to reveal.

This independence implies that the entrepreneur does not choose a more informative experiment for signals in $\mathcal{Y}^{nb}$, compared to the benchmark case without setting milestone. Moreover, the insider's action following signals in $\mathcal{Y}^{b}$ is weakly dominated by that without the commitment, because in the latter she can optimally respond to the additional information the entrepreneur provides. Therefore, the insider does not gain from setting milestones.

2.A.10 Proof of Corollary 3

Similar to the earlier discussions, one can show the entrepreneur chooses a binary experiment with a threshold structure for both cases. Denote the threshold values for securities $s_1(X)$ and $s_2(X)$ by $\bar{X}_1^c$ and $\bar{X}_2^c$ respectively. Equation 2.26 directly implies that:

$$\mathbb{E}[\{s_1(X)-I\}^-|X \geq \bar{X}_1^c] = -c = \mathbb{E}[\{s_2(X)-I\}^-|X \geq \bar{X}_2^c] \geq \mathbb{E}[\{s_1(X)-I\}^-|X \geq \bar{X}_2^c] \Rightarrow \bar{X}_1^c \geq \bar{X}_2^c \tag{2.A.20}$$

2.A.11 Proof of Proposition 9

Proof for Part (a)

Since the insider's payoff is continuous in μ, we only need to prove that the insider receives zero expected payoff for $\mu = 0$ and $\mu = 1$. For $\mu = 0$, the insider has no information rent and clearly gets zero expected payoff. We now discuss the case of $\mu = 1$.

Secret manipulation. First suppose the entrepreneur can secretly change the signal realization with probability $\alpha > 0$. Similar to Proposition 4, the entrepreneur follows a threshold strategy, i.e., there exists $\bar{X}_\alpha \in [0, 1]$ such that the experiment generates a high signal for $X \geq \bar{X}_\alpha$. The high signal induces investment if the investor receives a non-negative expected payoff from the investment following the high signal, which is equivalent to

$$\alpha \int_0^{\bar{X}_\alpha} (s(X) - I) f(X) dX + \int_{\bar{X}_\alpha}^1 (s(X) - I) f(X) dX \geq 0. \tag{2.A.21}$$

The first term in (2.A.21) shows the probability that the experiment generates a low signal, but the entrepreneur finds the chance to send a high signal. Note that for $\alpha = 1$, the inequality does not hold because:

$$\mathbb{E}[s(X) - I] < \mathbb{E}[\varepsilon + X - I] < 0 \tag{2.A.22}$$

Therefore, there exists $\bar{\alpha} \in [0, 1]$ above which the investment is not feasible, because the entrepreneur's commitment problem to truthful reporting of the signal is sufficiently serious. However, for $\alpha \leq \bar{\alpha}$, the entrepreneur chooses $\bar{X}_\alpha$ such that the inequality (2.A.21) binds, which implies the investor becomes indifferent between investment and not investment after receiving the high signal. As a result, the investor receives zero interim rent for all values of $\alpha \in [0, 1]$.

Random monitoring. Now consider the case that the insider verifies the signal realization with probability $\beta < 1$. This case involves two subcases:

First, the investor cannot commit to punishing the entrepreneur for misreporting. In this subcase, there is no signal such as h that always induces investment, because otherwise the entrepreneur would optimally always report h, which leads to negative expected payoff for the investor when she does not monitor. As such, the investor only invests when she monitors, which makes the entrepreneur's reporting strategy irrelevant. Hence the entrepreneur would follow the threshold strategy at $\bar{X}$, in which the investor does not any get interim rent.

Second, consider the subcase that the investor commits to punish misreporting. In this case, the entrepreneur might use four different kinds of signals in his experiment: 1. low signals, such as l_0, that never induce investment. 2. Low signals, such as l_1, that only induce investment when the investor does not monitor. 3. high signals, such as h_0, that only induce investment if they are verified. 4. High signals, such as h_1, that always induce investment. The probability of investment is $(1-\beta)P(l_1) + \beta P(h_0) + P(h_1)$. We next check the incentive constraints for truthful reporting for the entrepreneur.

In particular, we show that there is no equilibrium that the insider invests without monitoring. Consider the contrary. If $\beta < \frac{1}{2}$, then types l_0 and h_0 prefer to report h_1 instead of truthfully reporting because the probability of investment strictly for l_0 and h_0 increases from 0 and β respectively to $1-\beta$. Note that in all cases, the insider pays exactly I upon continuation. Therefore, only the probability of investment affects the entrepreneur's payoff at every state X. Because of these, we should have $P(l_0) = P(h_0) = 0$, which implies that the investor always invests when she is not monitoring. Thus, she would be better off by not investing at all when she does not monitor, as her investment has a negative NPV conditional on not monitoring. It implies $P(l_1) = P(h_1) = 0$ as well. It contradicts that the equilibrium involves investment with a positive probability in absence of monitoring.

For $\beta \geq \frac{1}{2}$, $P(l_0) = 0$, because he would be strictly better off by reporting h_1. If the investor receives a positive payoff from investment following $h_1 \cup l_1$, she should make a loss by investing following h_0, since $P(h_0)+P(l_1)+P(h_1) = 1$, and the project has an unconditional negative NPV. Therefore, the investor would be better off by not investing following h_0, even after verifying it, which is a contradiction with its definition. Consequently, the investor never invests without monitoring.

Proof for Part (b)

Secret manipulation. We first show that the socially optimal outcome cannot be implemented for $\alpha > 0$, then we show that the optimal contracts feature securities with a flat region for the insider, i.e., condition (2.19) holds.

Note that once the entrepreneur raises K, he always wants to continue the project, since he receives a strictly positive payoff from continuation. Consequently, the entrepreneur misreports the bad signals whenever it is possible, which leads to inefficient continuation. Therefore, the socially optimal outcome is not implementable for $\alpha > 0$.

Now we solve for the optimal contract. If the project is invested with a positive probability, then the expected payoff of the entrepreneur from the contract $\{s_I(.), s_O(.), \lambda\}$ is

$$U_\alpha^E(s_I(\cdot), s_O(\cdot), \lambda, \varepsilon) = \mathbb{E}[M(X; s_I, s_O, \lambda, \varepsilon)(\alpha + (1-\alpha)\mathcal{I}(X; \varepsilon))], \tag{2.A.23}$$

where $\mathcal{I}(\cdot)$ is the investment function for the case that the entrepreneur cannot secretly manipulate the signal. If $\mathbb{E}[\max\{X - I + \varepsilon, 0\}] > K$, then with an argument similar to the proof of Proposition 5, the following set of convertible securities are optimal and independent of ε, for small enough values of α:

$$\begin{aligned}
&\lambda I \leq I - \bar{\varepsilon}, \\
&s_I(X) = \min\{\lambda I, X\} \ \ \forall\, X < I, \\
&\mathbb{E}[(s_I(X) - \lambda I)(\alpha + (1-\alpha)\mathbb{I}_{\{X \geq I - \varepsilon\}})] = K, \\
&\mathbb{E}[(s_O(X) - (1-\lambda) I)(\alpha + (1-\alpha)\mathbb{I}_{\{X \geq I - \varepsilon\}})] = 0, \\
&0 \leq s_O(X) \leq X - s_I(X) \ \ \forall\, X \in [0,1]
\end{aligned} \tag{2.A.24}$$

Note that the design might not be robust to the insider's experiment $(\mathcal{Y}, \omega_q)$ because the insider's payoff becomes sensitive to the downside realization of the final cash-flow. Moreover, for big enough values of α, no security can satisfy condition (2.A.24). Therefore, relationship financing is infeasible for such big values of α.

Random monitoring. Consider the case that the investor cannot credibly threaten the entrepreneur to terminate the project when he misreports. The argument for the other case is similar. As it is discussed earlier, in equilibrium, the entrepreneur always reports a high signal and the investor invests if and only if she verifies the signal is truthfully reported. Therefore, the socially optimal outcome cannot be implemented when $\beta < 1$. Moreover, the following convertible securities are optimal.

$$\begin{aligned}
\lambda I \leq I - \bar{\varepsilon}, \quad s_I(X) = \min\{\lambda I, X\} \ \forall\, X < I, \quad \beta\mathbb{E}[(s_I(X) - \lambda I)\mathbb{I}_{\{X \geq I - \varepsilon\}}] = K, \\
\mathbb{E}[(s_O(X) - (1-\lambda) I)(\alpha + (1-\alpha)\mathbb{I}_{\{X \geq I - \varepsilon\}})] = 0, \quad 0 \leq s_O(X) \leq X - s_I(X) \ \forall\, X \in [0,1]
\end{aligned} \tag{2.A.25}$$

Clearly, these securities are implementable for large enough values of β. For the case of credible punishments, an optimal design might not exist because the optimal experimentation involves three signals for large values of ε, while it involves two signals for the smaller values. However, for smaller values of ε, the convertible securities specified above are optimal and the equilibrium outcomes are similar.

CHAPTER 3
INVESTMENT SOPHISTICATION AND WEALTH INEQUALITY

3.1 Introduction

Wealth inequality has been rising for decades (Piketty, 2014; Saez and Zucman, 2016). A key contributing factor is that households earn substantially different returns on their savings (Fagereng, Guiso, Malacrino, and Pistaferri, 2020). This return heterogeneity partly stems from differences in households' information about the risk and return associated with investment opportunities, which shapes households' saving decisions and, consequently, their wealth accumulation.[1] In this paper, building on this intuitive interplay between information asymmetries about asset returns and wealth inequality, I develop a model to understand and make empirical predictions about how wealth distribution interacts with information environment, market microstructure, and asset prices. The model offers specific policy implications on what types of information subsidies or regulations on the wealth management industry could help reduce wealth inequality.

Particularly, I construct a dynamic model of financial markets with investors heterogeneously informed about future returns. Endowed with different information, the investors trade assets, hold different portfolios, and receive different returns on their savings. Thus, asset price movements, coupled with portfolio heterogeneities induced by the asymmetric information, is the key source of wealth share dynamics.[2] To obtain stationarity, tax and death processes are also introduced.

Even though the investors are rational and asymmetrically informed, the wealth dynamics

1. The information asymmetries among households can be due to differences in educational attainment and background, attentiveness to market trends, or access to sound financial advice. For instance, Fagereng, Guiso, Malacrino, and Pistaferri (2020) find a college degree increases the average annual rate of return on net worth by 60 bps, after controlling for portfolio compositions into broad asset classes. Furthermore, they find that economics and business degrees are associated with an additional 15 bps average annual return on net worth. Overall, they argue that there are persistent differences in expected returns that reflect "heterogeneities in sophistication, and financial information, as well as entrepreneurial talent."

2. Relatedly, Hubmer, Krusell, and Smith Jr (2020) find that dynamics of wealth inequality can be well explained by asset price movements, due to vast portfolio heterogeneities across households.

and asset prices are tractable and derived in closed-form. I obtain tractability by modifying a dynamic version of classical models of rational expectation with asymmetrically informed investors (e.g., Grossman and Stiglitz (1980); Diamond and Verrecchia (1981)) in two ways: First, I employ logarithmic preferences, instead of CARA preferences. Second, I assume the "noise traders," introduced to make the prices partially revealing, have misspecified beliefs about the underlying data generating process, instead of exogenously specifying their asset demand as in Grossman and Stiglitz (1980). The tractability is obtained due to two properties that hold under logarithmic preferences: First, the wealth ratios move exactly with the probabilities the investors assign to future events (Blume and Easley, 1992). Second, the wealth-weighted average belief is sufficient to solve for the asset prices. Putting differently, the prices are equivalent to those of an economy in which the heterogeneous investors are replaced by a hypothetical representative investor that holds the wealth-weighted average belief of the heterogeneous investors.[3]

The unconditional distribution of wealth shares has a thick right tail populated by the most informed investors. Two opposite forces govern the thickness of the tail, which is a standard measure of top wealth inequality (Benhabib and Bisin, 2018), and best captures top wealth shares (Gabaix, 2009). Taxes and redistributions push down the thickness of the tail. Conversely, the thickness increases with the expected return of the most informed investors net of what the market portfolio delivers. In fact, any change in the information environment or market structure impacts the thickness of the tail only through the expected return of the most informed investors.

The model implies that changes in the information environment can impact wealth inequality in a non-trivial manner. For example, disseminating a public and free signal about asset returns can increase wealth inequality. The intuition is that investors, facing less uncertainty, react more aggressively to their own private signals, provided that their signals are orthogonal to the public signal. The investors trading more aggressively can magnify the portfolio heterogeneity, and thus the return heterogeneity, which increases wealth inequality. Therefore, depending on whether this channel, the stronger reaction to private signals, dom-

3. It resembles the irrelevance result of Grossman and Shiller (1982). They find conditions under which only the preferences and beliefs of a representative investor matter for asset prices.

inates the opposite channel, the reduction in belief heterogeneity, the availability of more accurate public information about asset returns might increase or decrease wealth inequality.

Along with the information environment, financial markets have evolved drastically in their size, liquidity, and complexity over the past decades. Regarding their distributional consequences, I find that wealth inequality increases with the liquidity of financial markets and the magnitude of non-fundamental noise in the prices. More generally, the model implies that price informativeness plays a crucial role in mitigating wealth inequality since asset prices act as endogenous signals (Fama, 1970; Grossman, 1976). Hence, any change in the market structure that hampers information dissemination through the prices also increases wealth inequality, and more specifically, top wealth shares. For instance, the recent reduction in the transaction costs associated with trading individual stocks could increase the volume of noise trading, which amplifies both return heterogeneity and wealth inequality.

Asymmetric information impacts the dynamics in the wealth distribution by inducing return heterogeneity across investors. The return heterogeneity partly reflects the belief heterogeneity among households, which is bounded by the magnitude of noise in the prices. However, it also depends on how aggressively investors trade given their beliefs. By operating through the latter, availability of more precise public signals can interestingly result in a greater return heterogeneity, and thus faster dynamics in the wealth distribution. It puts forward a potential explanation for the recent rapid dynamics observed in the US top wealth shares, which is hard to square with existing models of wealth dynamics (Gabaix, Lasry, Lions, and Moll, 2016).

Letting investors acquire information at a cost or delegate their investment decisions to the most informed investors do not make the tail disappear. In fact, the mere presence of a costly source of information about asset returns is enough to generate a fat-tailed unconditional distribution of wealth shares among ex-ante homogeneous investors. Moreover, the thickness of the tail is generally increasing with the cost of information acquisition.[4] Delegating investment decisions to more informed investors, such as fund managers, can increase wealth inequality if the market for wealth management is not perfectly competitive.

4. Relatedely, Mihet (2019) finds that capital income inequality between market participants and non-participants might increase when the cost of information acquisition is reduced, provided market participation is costly.

This result indicates that the recent surge in the size of delegated investments is likely to have contributed to the rise of super-rich fund managers (See Figure 3.B.1). Nevertheless, my model implies that a further increase in these delegations could reduce wealth inequality since it would improve price informativeness, and hence mitigate asymmetric information, lowering the return heterogeneity.

Lastly, in an empirical assessment of the model, I find wealthier households indeed have more precise macroeconomic beliefs, consistent with earlier empirical findings (See, e.g., Das, Kuhnen, and Nagel (2017)). Moreover, I obtain the model-implied dynamics for top wealth shares by feeding data on belief dynamics to the model. Interestingly, the model-implied dynamics match well with the empirically documented dynamics. It lends support to the empirical relevance of the framework presented here.

The rest of the paper is organized as follows: Section 3.2 reviews the related literature. Section 3.3 presents a dynamic model of financial markets with heterogeneously informed investors. Its equilibrium characterization is provided in Section 3.4. Section 3.5 characterizes the unconditional distribution of wealth and its dynamics. I discuss the cases with endogenous information acquisition, delegated investments, and multiple risky assets in Sections 3.6.1, 3.6.2, and 3.6.3 respectively. Then, I empirically assess the dynamic implications of the model in Section 3.6.4. Internet Appendix contains other extensions and further empirical verifications of the model.

3.2 Related Literature

With the recent surge in wealth inequality, its causes and consequences are currently at the center of political and academic debates.[5] Earlier studies focused more on the mechanisms relating wealth inequality to income inequality (Aiyagari, 1994; Castaneda, Diaz-Gimenez, and Rios-Rull, 2003). Nonetheless, they achieved limited success, especially in explaining top wealth inequality (Fagereng, Guiso, Malacrino, and Pistaferri, 2020). Alternatively, a new strand of theoretical and empirical studies have examined the role of return

5. Wilkinson, Pickett, and Cato (2009) and Wilkinson and Pickett (2019) argue that inequality creates social problems such as illiteracy, crime, and poor health. Pastor and Veronesi (2018) discuss its role in the recent upsurge of populism.

heterogeneities in wealth inequality.[6] To the best of my knowledge, this study is the first to examine the distributional implications of information heterogeneities and public information about asset returns within a dynamic framework.

Providing evidence on the presence of persistent information heterogeneities, Fagereng, Guiso, Malacrino, and Pistaferri (2020) find a substantial heterogeneity in the Sharpe ratios of portfolios held by Norwegian households, and discover the Sharpe ratios are highly correlated with the wealth percentiles. This outcome coincides with the prediction of my model. Additionally, Grinblatt, Keloharju, and Linnainmaa (2011, 2012), Korniotis and Kumar (2013), Barber, Lee, Liu, and Odean (2014), Gargano and Rossi (2018), and Clark, Lusardi, and Mitchell (2017) provide evidence that high-IQ, skilled, and attentive investors hold portfolios with a higher Sharpe ratio. Bartscher, Kuhn, and Schularick (2018) and Girshina (2019) emphasize the role of education in portfolio choice and generating higher risk-adjusted returns from risky assets.

Apart from investment sophistication, the returns are heterogeneous due to differences in investment opportunity sets (Guvenen, 2009), risk preferences (Veronesi, 2018; Gomez, 2019), and entrepreneurial skills (Cagetti and De Nardi, 2006).[7] However, none of those studies justify the over-representation of highly sophisticated and arguably talented investors among the wealthiest, as documented by Kaplan and Rauh (2013).[8]

This paper methodologically contributes to the theoretical literature that examines asset prices and wealth distribution within a heterogeneous-agent framework, by examining the role of heterogeneities in information.[9] The framework studied here is the closest to that of Blume and Easley (1992) and Mailath and Sandroni (2003). However, they examine

6. Supporting this channel, Bach, Calvet, and Sodini (2016) and Fagereng, Guiso, Malacrino, and Pistaferri (2020) document a vast heterogeneity in returns in Sweden and Norway, respectively. Benhabib, Bisin, and Zhu (2011) and Benhabib, Bisin, and Luo (2015, 2019) demonstrate, theoretically and quantitatively, the importance of return heterogeneity to match the empirical wealth distribution. However, they do not discuss the root causes of return heterogeneity.

7. Pástor and Veronesi (2016) study a model of occupational choice with agents heterogeneous in their entrepreneurial skills and risk-aversion.

8. They find hedge fund managers, private equity managers, money managers, and venture capital managers constitute 20% of billionaires on Forbes 400 list of wealthiest in US at 2011 (See Figure 3.B.1).

9. For a review, see Panageas (2020).

asymptotic survival instead of the unconditional distribution of wealth shares or its dynamics. Additionally, I analyze the role of public signals, market liquidity, cost of information acquisition, delegated investments, among others, in wealth inequality and its dynamics that are absent the previous studies.

This paper is foremost related to the nascent literature on the role of sophistication in wealth inequality.[10] Lei (2019) examines wealth inequality among agents who dynamically adjust their investment in a privately-owned risky asset, as they learn about its return. Peress (2003), Kacperczyk, Nosal, and Stevens (2018) and Mihet (2019) study the interaction between information acquisition and wealth inequality in a finite-period setting. Specifically, Peress (2003) shows that wealth inequality increases with the emergence of costly signals about asset returns, which I confirm in the infinite-horizon framework presented here. Kacperczyk, Nosal, and Stevens (2018) find that capital income inequality might increase with the aggregate information capacity when investors have heterogeneous information processing ability. Mihet (2019) finds that in the presence of market participation costs, cheaper access to information could increase wealth inequality. My paper contributes by developing a dynamic framework that illuminates how equilibrium level of wealth inequality is impacted by changes in the precision of public information provided by authorities, market liquidity, cost of access to information, and the structure of the wealth management industry.

3.3 A Dynamic Model of Financial Markets

Consider an endowment economy with discrete periods, i.e., $t = 1, 2, \ldots$. A tree pays out d_t units of consumption good at the beginning of each period. The dividend in the first period, d_1, is normalized to one. The dividend growth $g_t \equiv \log d_t - \log d_{t-1}$ $(t \geq 2)$ follows

10. An, Bian, Lou, and Shi (2019), Jorring (2018), and Campbell, Ramadorai, and Ranish (2018) provide evidence on the role of investment and financial sophistication in wealth inequality.

a two-state Markov process, as follows:

$$
\begin{gathered}
g_t \in \{g^h, g^l\}, \qquad g^h > g^l \\
\tilde{z}_t^I \sim U[-\bar{z}^I, \bar{z}^I] \quad \tilde{z}_t^N \sim U[-\bar{z}^N, \bar{z}^N] \quad \tilde{z}_t^P \sim U[-\bar{z}^P, \bar{z}^P] \\
P(\tilde{g}_{t+1} = g^h | \tilde{g}_t = g^j, z_t^I, z_t^N, z_t^P) = q(g^j) + z_t^P + z_t^I \quad j \in \{h, l\} \quad \forall t \\
\bar{z}^N, \bar{z}^I < \min\{1 - q(g^h), q(g^l)\} - \bar{z}^P
\end{gathered}
\tag{3.1}
$$

In (3.1), $\tilde{z}_t^N$, $\tilde{z}_t^P$ and $\tilde{z}_t^I$ are exogenous signals based on which the agents make their prediction about the next period's growth rate g_{t+1}. Following the literature, the variables with $\sim$ are random variables. Note that, in this specification, $\tilde{z}_t^I$ and $\tilde{z}_t^P$ are informative about the future growth rate, while $\tilde{z}_t^N$ is not. $\tilde{z}_t^P$ is a public signal, that is its realization is observed and incorporated in investment decisions by all investors. $\tilde{z}_t^I$ is not a public signal. In fact, its realization is heterogenously observed by the agents in the economy, who are described later. Lastly, $\tilde{z}_t^N$ is not informative about future growth rates, but some "naive investors" mistakenly believe it is informative and incorporate its realization in their investment decisions. In fact, the signal introduces non-fundamental noise into the prices and prevents them from being perfectly revealing.[11] $q(g^j)$, $j \in \{g, h\}$ is the conditional probability of a high growth rate realizing in period $t+1$ ($\tilde{g}_{t+1} = g^h$) when $g_t = g^j$, without knowing z_t^P, z_t^I and z_t^N. For tractability reasons, I also assume the unconditional probability of remaining in the same growth state is constant, that is:

$$
P(\tilde{g}_{t+1} = \tilde{g}_t | \tilde{g}_t = g_t^h) = P(\tilde{g}_{t+1} = \tilde{g}_t | \tilde{g}_t = g_t^l) \Rightarrow q(g^h) = 1 - q(g^l). \tag{3.2}
$$

Agents. The economy is populated by a unit measure of investors. The investors are Bayesian, die with probability $1 - \delta \in (0, 1)$, and have time-separable and logarithmic preferences over their consumption stream. Therefore, an investor alive at s aims to maximize the following objective function:

$$
U = \mathbb{E}[\sum_{t=s}^{\infty} (\delta\beta)^{t-s} \log c_t] \tag{3.3}
$$

11. The details are explained once the economic agents and their preferences are described.

The economy has $S+1$ types of atomistic investors, where $S > 1$. S types of investors have consistent beliefs with the underlying data-generating process. They are referred as "sophisticated investors" throughout the paper. The population measure of type i-investors is denoted by $b_i \in (0,1)$, for $1 \leq i \leq S$.

The sophisticated investors are different in their information set. At the beginning of period t, an i.i.d. random variable $\tilde{\lambda}_t \in \{1,,\ldots,S\}$ is drawn with cumulative distribution $F(\cdot) : \{1,2,\ldots,S\} \to [0,1]$. Once λ_t is realized, any sophisticated investor with type $i \geq \lambda_t$ perfectly observes z_t^I (informed investors), and sophisticated investors with type $i < \lambda_t$ do not receive any signal (uninformed investors). Therefore, all sophisticated investors observe z_t^I when $\lambda_t = 1$, and only type-S investors observes the realization when $\lambda_t = S$. If a sophisticated investor remains uninformed about z_t^I at t, she forms a posterior about the distribution of $\tilde{z}_t^I$ based on the equilibrium prices by appealing to Bayes' rule. I denote the posterior belief of the informed and uninformed investors (after potentially observing the signal realization and equilibrium prices) by $q_t^I = P_t^I(\tilde{g}_{t+1} = g^h)$ and $q_t^U = P_t^U(\tilde{g}_{t+1} = g^h)$ respectively, where P_t^J, $J \in \{U, I\}$, represents the probability measure that an J-type investor assigns at time t to the future events.

Note that sophisticated investors are only heterogeneous in the probability they become informed, i.e., learn z_t^I. Specifically, a sophisticated investor with type i ($1 \leq i \leq S$) becomes informed with probability $F(i)$.[12] [13] That said, triplet $(S, \{b_i\}_{i=1}^N, F(\cdot))$ governs the information environment among sophisticated investors, which takes a highly flexible form.

The primary goal here is to study the dynamics of wealth distribution among a set of asymmetrically informed investors. However, to motivate the trade and have a non-trivial wealth dynamics, we need to introduce an exogenous and stochastic supply of liquidity (Milgrom and Stokey, 1982), as done in different forms by previous theoretical studies of rational expectation model (e.g., Grossman and Stiglitz (1980); Diamond and Verrecchia

12. It is not crucial that the investors are fully ranked based on their information set. In Section 3.6.3, I consider an extension in which the investors are asymmetrically informed about the return of multiple risky assets, and the rankings could be different from one asset to another.

13. Note that once λ_t is realized at t, all informed and uninformed investors have the same information about the future growth rates for the rest of the period. The assumption is made to simplify the exposition and avoid unnecessary complications in dealing with different information sets.

(1981); Kyle (1985)). In this model, the liquidity is provided by a set of "naive" investors, who constitute a measure $b_N \equiv 1 - \sum_{i=1}^{S} b_i$ set.[14] The key assumption about naive investors is that they falsely believe that the growth rate follows the following Markov process, instead of the actual one provided in (3.1):

$$P^N(\tilde{g}_{t+1} = g^h | \tilde{g}_t = g^j, z_t^I, z_t^N, z_t^P) = q(g^j) + z_t^P + z_t^N \quad j \in \{h, l\} \tag{3.4}$$

The only difference between naive and sophisticated investors is that naive investors mistakenly believe signal $\tilde{z}_t^N$, instead of $\tilde{z}_t^I$, is informative about the next period's growth rate. However, both groups agree on public signal $\tilde{z}_t^P$ being informative about the future growth rates. All naive investors perfectly observe z_t^N at the beginning of each period, hence no asymmetric information exists among them. As a result, they do not update their beliefs based on the equilibrium prices, and they just act based on the realization of $\tilde{z}_t^N$ and $\tilde{z}_t^P$. In fact, $\tilde{z}_t^N$ works as a non-fundamental uncertainty in the economy that distorts the prices, and hence, prevents z_t^I from being perfectly revealed to the uninformed investors. $q_t^N \equiv P_t^N(\tilde{g}_{t+1} = g^h)$ denotes the posterior belief of naive investors.

In my model, what is exogenously specified is the belief of naive investors, instead of their demand schedule (e.g., as in Grossman and Stiglitz (1980); Kyle (1985)). This modeling choice yields substantial analytical tractability because, as shown in Section 3.4, under logarithmic preferences, wealth ratios across the investors move with the ratios in their probability assignments to realized events, as in Blume and Easley (1992). Lemma 9 describes the wealth dynamics and this property in detail.

I assume each type of investors comprises a continuum of dynasties of investors. Each dynasty in the economy is indexed by a pair (f, k), where $k \in \{1, \ldots, S\} \cup \{N\}$ indicates their type and $f \in [0, 1]$ is their index within type-k investors. In each period, only one investor from each dynasty is alive. More specifically, each deceased investor is replaced with her "child," who inherits the parent's type.[15] The presence of infinite dynasties of

14. The presence of noise trading could be microfounded by introducing uninsurable endowment or labor income shocks, as done by Diamond and Verrecchia (1981), Vayanos (1999) and Bond and Garcia (2018). However, it would add to the analyses' complexity without generating new economic insight into the interaction between wealth distribution and information asymmetries.

15. This assumption implies that the sophistication level is perfectly persistent across all generations of

investors within each type, instead of a single dynasty, helps decouple the dynamics of each types' wealth share from the death shock realizations.[16]

Assets. Two assets are available for trading. One is the shares of the tree, which is in unit supply and traded at equilibrium price p_t, in the unit of consumption goods at period t. In addition, the economy has a risk-free asset with a net-zero supply. In particular, at period t, the investors trade consumption claims for period $t+1$, at an equilibrium substitution rate R^F_t. That is, at time t, they can trade one unit of consumption goods at time t for R^F_t units of consumption goods at time $t+1$. I denote the share of the tree held by the investor of family (f,k) at the beginning of period t by $x^k_{f,t}$. Likewise, $s^k_{f,t}$ is defined as the consumption claims of period t (traded at $t-1$) owned by the investor.Thus, her wealth at time t, in the unit of consumption goods, is

$$W^k_{f,t} = (p_t + d_t)x^k_{f,t} + s^k_{f,t}. \tag{3.5}$$

Note that the financial markets here are dynamically incomplete. For instance, no long-term risk-free bond is available. In Appendix 3.B.2, I prove even if the set of available securities fully spans the space of dividend realizations, the asset allocations and wealth dynamics would be exactly the same as those in the restricted case I analyze here.

Inheritance. To prevent the most informed investors, i.e., type-S investors, from dominating the economy, I assume an Estate tax is imposed on inherited wealth. Specifically, upon an investor's death at period t, her wealth is taxed at rate $\tau \in (0,1)$. As such, the child receives fraction $1-\tau$ of the parent's assets, and the taxed assets are distributed evenly among the newborns at time t. Equation 3.6 specifies the inheritance process.

$$\begin{aligned} x^k_{f,t} &= (1-\tau)\hat{x}^k_{f,t} + \tau(x^N_t + \int_0^1 x^i_t di) = (1-\tau)\hat{x}^k_{f,t} + \tau \\ s^k_{f,t} &= (1-\tau)\hat{s}^k_{f,t} + \tau(s^N_t + \int_0^1 s^i_t di) = (1-\tau)\hat{s}^k_{f,t}. \end{aligned} \tag{3.6}$$

a family. I relax this assumption in Appendix 3.B.4 and show that it reduces the right-skewness of the unconditional distribution of wealth.

16. It is crucial since we will see that the wealth share of each type emerges as a state variable, and due to this assumption, we do not need to keep track of the death shock realizations in their transitions.

Therefore, the initial wealth of the replacing child is $W^k_{f,t} = (1-\tau)\hat{W}^k_{f,t} + \tau(p_t + d_t)$, where $\hat{W}^k_{f,t}$ is the child's before-tax wealth.

Optimization problem. The optimization problem of the investor of dynasty (f,k), $(f \in [0,1], k \in \{1,\ldots,S\} \cup \{N\})$, at time t is as follows:

$$\begin{aligned} &\max_{\{c^k_{f,t+j}, x^k_{f,t+j+1}, s^k_{f,t+j+1}\}_{j=0}^{\infty}} \mathbb{E}^k_t[\sum_{j=0}^{\infty} (\beta\delta)^j \log c^k_{f,t+j}] \\ s.t. \quad & c^k_{f,t+j} + p_t x^k_{f,t+j+1} + \frac{1}{R^F_{t+j}} s^k_{f,t+j+1} \le (p_t + d_t) x^k_{f,t+j} + s^k_{f,t+j} \quad \forall j \ge 0 \\ & P^k_t(W^k_{f,t+j} > 0) = 1 \qquad \forall j \ge 0 \end{aligned} \tag{3.7}$$

,where the operator $\mathbb{E}^k_t[\cdot]$ takes expectation with respect to the information set and belief of type-k investors at time t. Likewise, $P^k_t(\cdot)$ represents the probability assignments by a type-k investor. The first constraint describes the budget constraint. It indicates that at any period, the sum of consumption and savings, captured by the left-hand side, cannot exceed the wealth, which is the total value of assets held at the beginning of period t. The second constraint in (3.7) reflects a borrowing constraint that ensures the existence of a solution, although it is not binding on the equilibrium path.[17]

3.4 Equilibrium

In this section, I solve for the equilibrium consumptions, asset allocations, prices, and posterior beliefs in closed-form, given the growth and signal realizations. The goal of this section is to characterize the investors' wealth dynamics.

3.4.1 Definition of Equilibrium

To simplify the expositions, it is useful to introduce aggregated variables $c^k_t, x^k_t, s^k_t, W^k_t$ for each type $k \in \{1,\ldots,S\} \cup \{N\}$:

17. In fact, to avoid Ponzi-like consumption schemes, I assume an investor with a negative wealth cannot borrow. Thus, all investors optimally maintain a positive wealth in all states of the world.

$$
\begin{aligned}
c_t^k &= \int_0^1 c_{f,t}^k df \\
x_t^k &= \int_0^1 x_{f,t}^k df \\
s_t^k &= \int_0^1 s_{f,t}^k df \\
W_t^k &= \int_0^1 W_{f,t}^k df.
\end{aligned}
\tag{3.8}
$$

[Dynamic Competitive Equilibrium] Define

$$
\mathcal{F}_t^P \equiv (\{x_t^k\}_{k\in\{1,\ldots,S\}\cup\{N\}}, \{s_t^k\}_{k\in\{1,\ldots,S\}\cup\{N\}}, \lambda_t, g_t),
$$

the sequence of demand functions $\{c_t^k(p_t, R_t^F, q_t^k; \mathcal{F}_t^P), x_{t+1}^k(p_t, R_t^F, q_t^k; \mathcal{F}_t^P), s_{t+1}^k(p_t, R_t^F, q_t^k; \mathcal{F}_t^P)\}_{t=1}^{\infty}$ for $k \in \{1,\ldots,S\}\cup\{N\}$ and the price functions $\mathrm{p}(z^P, z^I, z^N, \mathcal{F}^P), \mathrm{R}(z^P, z^I, z^N, \mathcal{F}^P)$, where $p_t = \mathrm{p}(z_t^P, z_t^I, z_t^N, \mathcal{F}_t^P)$ and $R_t^F = \mathrm{R}(z_t^P, z_t^I, z_t^N, \mathcal{F}_t^P)$, constitute a Markov perfect equilibrium if

1. [**Optimality**] the demand functions constitute a solution to the optimization problem (3.7) given beliefs $\{q_t^k\}_{k\in\{1,\ldots,S\}\cup\{N\}}$ and $\mathcal{F}_t^P$ for every t.

2. [**Bayesian Inference**] all investors make a Bayesian inference based on their signals and equilibrium prices p_t and R_t^F at time t. In particular, if $g_t = g^j$, $j \in \{h, l\}$:

$$
q_t^k = \begin{cases} q^j + z_t^P + z_t^I & \lambda_t \le k \le S \\ q^j + z_t^P + z^U(z_t^I, z_t^N; \mathcal{F}_t^P) & k < \lambda_t \\ q^j + z_t^P + z_t^N & k = N, \end{cases}
\tag{3.9}
$$

where $z^U(\cdot,\cdot,\cdot;\cdot)$ is obtained from Bayes' rule.[18]

3. **[Market Clearing]** For every t, the prices p_t and R_t^F are such that the market for consumption goods, risky assets, and riskless assets clear:

$$\sum_{i=1}^{S} c_t^i + c_t^N = d_t$$
$$\sum_{i=1}^{S} x_{t+1}^i + x_{t+1}^N = 1 \tag{3.10}$$
$$\sum_{i=1}^{S} s_{t+1}^i + s_{t+1}^N = 0.$$

3.4.2 Optimal Intertemporal Decisions

In this section, I derive the optimal consumption and portfolio of a type-k investor with wealth $W_t^k = (d_t + p_t)x_t^k + s_t^k$, given posterior belief q_t^k, $k \in \{1,\ldots,S\} \cup \{N\}$. To solve for the investors' optimal consumption and portfolio decisions, I first characterize their Bellman equation. Note that the state variable here is $\mathcal{F}_t \equiv (\mathcal{F}_t^P, z_t^P, z_t^I, z_t^N)$, which includes the state of asset allocations at t across different types, the growth rate g_t, λ_t, and signal realizations z_t^P, z_t^I and z_t^N. Therefore, the value function can be written as follows:

$$V^k(W_t;\mathcal{F}_t) = \max_{c_t, x_{t+1}, s_{t+1}} \log c_t + \beta\delta\mathbb{E}^k[V^k(W_{t+1};\mathcal{F}_{t+1})] \tag{3.11}$$
$$s.t. \quad c_t + p_t x_{t+1} + \frac{1}{R_t^F} s_{t+1} \leq W_t$$
$$W_{t+1} = (d_{t+1} + p_{t+1})x_{t+1} + s_{t+1}$$
$$Prob(W_{t+1} > 0) = 1$$
$$p_t = \mathrm{p}(z_t^P, z_t^I, z_t^N, \mathcal{F}_t^P) \qquad R_t^F = \mathrm{R}(z_t^P, z_t^I, z_t^N, \mathcal{F}_t^P).$$

18. More specifically,

$$z^U(z_t^I, z_t^N; \mathcal{F}_t^P) = \frac{\int_{-\bar{z}^N}^{\bar{z}^N}\int_{-\bar{z}^I}^{\bar{z}^I} z_t^I \mathbf{1}\{p_t = \mathrm{p}(z_t^I, z_t^N, \mathcal{F}_t^P), R_t^F = \mathrm{R}(z_t^I, z_t^N, \mathcal{F}_t^P)\} dz_t^I dz_t^N}{\int_{-\bar{z}^N}^{\bar{z}^N}\int_{-\bar{z}^I}^{\bar{z}^I} \mathbf{1}\{p_t = \mathrm{p}(z_t^I, z_t^N, \mathcal{F}_t^P), R_t^F = \mathrm{R}(z_t^I, z_t^N, \mathcal{F}_t^P)\} dz_t^I dz_t^N}.$$

Lemma 5. ***a)*** *For every type* $k \in \{1, \ldots, S\} \cup \{N\}$*, function* $a^k(\cdot)$ *exists such that*

$$V^k(W_t; \mathcal{F}_t) = \frac{1}{1-\beta\delta} \log W_t + a^k(\mathcal{F}_t).$$

b) *All investors consume a fixed fraction of their wealth. In particular,* $c_t = (1-\beta\delta)W_t$*.*
c) *For every* $t \geq 1$*,* $p_t = \frac{\beta\delta}{1-\beta\delta} d_t$*.*

Lemma 5 describes the optimal consumption and saving behavior of the investors. Part (a) shows the value functions are logarithmic in wealth. Part (b) states that the investors always save a fixed fraction of their wealth, regardless of their beliefs about the future growth rates.[19] Thus, since the optimal consumption is independent of future expected returns, the investors optimally maximize their expected log-wealth in the next period, or equivalently, maximize their expected log-return.

Part (c) of Lemma 5 states that the price-dividend ratio in this economy is constant. It implies the price of the tree contains no information about future growth rates. We later see the only price that partially reflects the informed investors' information is the risk-free rate.

A constant price-dividend ratio implies that the return on the risky asset is always proportional to the growth rate. Therefore, the investors decide to what extent they want to expose their savings to the random growth rate $\tilde{g}_{t+1}$. The optimal exposure depends on risk-free rate R_t^F and one's belief about the distribution of $\tilde{g}_{t+1}$. Lemma 6 provides a closed-form solution to the optimal portfolio problem.

Lemma 6. *Suppose type-k investors,* $k \in \{1, \ldots, S\} \cup \{N\}$*, have posterior belief* $q_t^k \in (0,1)$*, i.e.,* $q_t^k = P_t^k(\tilde{g}_{t+1} = g^h)$*. Let* $\mu_t^k \equiv \frac{p_t x_{t+1}^k}{p_t x_{t+1}^k + s_t^k}$ *be the fraction of their saving invested in the risky asset at time t. Then*

$$\mu_t^k = 1 + q_t^k \frac{1}{\beta\delta R_t^F e^{-g_l} - 1} + (1 - q_t^k) \frac{1}{\beta\delta R_t^F e^{-g_h} - 1}. \tag{3.12}$$

19. The intuition lies in the Cobb-Douglas nature of the preferences. It is well-known that agents with such preferences allocate their wealth to different components of their consumption bundle in fixed fractions, determined by substitution elasticities parameters. In my model, the discount rate (β) and survival rate (δ) govern the households' saving ratio.

Lemma 6 states that the share invested in the risky asset linearly changes with the probability assigned to having a high growth rate at $t+1$.[20] As we see next, an implication of this result, when combined with the market-clearing conditions, is that all investors can infer the wealth-weighted average belief of the other investors from the equilibrium prices.

3.4.3 Posterior Belief of Uninformed Investors

To solve for uninformed investors' posterior, I first extract the following identity from the market-clearing conditions.

$$1 = \int_0^{\lambda_t} \mu_t^i w_t^i di + \int_{\lambda_t}^1 \mu_t^i w_t^i di + \mu_t^N w_t^N. \tag{3.13}$$

Equation 3.13 reflects the fact that the investors' investment in the risky asset sums up to the value of the tree. By substituting μ_t^k's in (3.13) with the expression in (3.12), and after some rearrangements, we get:

$$\begin{aligned} &(q_t^U w_t^U + q_t^I w_t^I + q_t^N w_t^N)\frac{1}{\beta\delta R_t^F e^{-g_l} - 1} \\ &+((1-q_t^U)w_t^U + (1-q_t^I)w_t^I + (1-q_t^N)w_t^N)\frac{1}{\beta\delta R_t^F e^{-g_h} - 1} = 0. \end{aligned} \tag{3.14}$$

Based on Equation 3.12, a hypothetical investor with posterior belief $\bar{q}_t \equiv w_t^U q_t^U + w_t^I q_t^I + w_t^N q_t^N$ should have μ of one, meaning she invests all of her savings in the risky asset, or equivalently, does not borrow or lend. For the rest of the paper, I refer to this hypothetical investor as the "representative investor."

An important implication of (3.14) is that all investors can infer $\bar{q}_t$ from the risk-free rate R_t^F. Therefore, $w_t^I q_t^I + w_t^N q_t^N$, and consequently $w_t^I z_t^I + w_t^N z_t^N$ are observable to the uninformed investors. This point is formalized in Lemma 7.

Lemma 7. *For any* $(z_t^I, z_t^N, \mathcal{F}_t^P)$, $w_t^I z_t^I + w_t^N z_t^N$ *can be inferred from equilibrium prices.*

20. The optimal portfolio, specified in (3.12), is discontinuous at the extreme certainty points $q_t^k \in \{0,1\}$. The reason is that as long as each state has a positive probability, the investors do not take an extreme position to avoid falling into negative wealth values.

Building on this result, Lemma 8 provides a closed-form expression for the uninformed investors' posterior, as a function of z_t^I, z_t^N, and other publicly observable objects.

Lemma 8. $q_t^U = q(g_t) + z_t^P + z_t^U$, *where*

$$z_t^U = \frac{1}{2}[\max\{-\bar{z}^I, z_t^I + \frac{w_t^N}{w_t^I}(z_t^N - \bar{z}^N)\}, \min\{\bar{z}^I, z_t^I + \frac{w_t^N}{w_t^I}(z_t^N + \bar{z}^N)\}]. \tag{3.15}$$

Lemma 8 provides the uninformed investors' posterior as a function of the wealth distribution at t and exogenous signals z_t^I and z_t^N. To elicit insights from (3.15), consider a case in which $\frac{w_t^N}{w_t^I}$ is sufficiently small, so that

$$-\bar{z}^I \leq z_t^I + \frac{w_t^N}{w_t^I}(z_t^N - \bar{z}^N) \leq z_t^I + \frac{w_t^N}{w_t^I}(z_t^N + \bar{z}^N) \leq \bar{z}^I. \tag{3.16}$$

In this case, the uninformed investors' posterior boils down to $q_t^U = q_t^I + \frac{w_t^N}{w_t^I} z_t^N$. We see the difference between the informed and uninformed investors' posterior is proportionate to the wealth ratio between the naive and informed investors, $\frac{w_t^N}{w_t^I}$. Therefore, the wealth distribution plays an important role in the informativeness of the prices: The larger the informed investors' wealth share, the more informative the prices are about future growth rates. This relationship is intuitive because the informed investors push the prices more strongly when they hold a larger wealth share, making the prices more informative.

3.4.4 The Risk Free Rate and Risk Premium

The last step for the equilibrium characterization is to solve for the risk-free rate R_t^F. Rearranging (3.14) yields

$$R_t^F = (\beta\delta)^{-1} \frac{1}{\bar{q}_t e^{-g_h} + (1 - \bar{q}_t) e^{-g_l}}. \tag{3.17}$$

Consequently, we can derive the risk premium:

$$\bar{E}_t[\frac{p_{t+1}+d_{t+1}}{p_t} - R_t^F] = \bar{q}_t(1-\bar{q}_t)\frac{(e^{\frac{g_h-g_l}{2}} - e^{-\frac{g_h-g_l}{2}})^2}{\bar{q}_t e^{-g_h} + (1-\bar{q}_t)e^{-g_l}}. \tag{3.18}$$

,where expectation $\bar{E}_t$ is taken with respect to the representative investor's belief. One can see that the risk premium, computed in (3.18), is concave in $\bar{q}_t$. Therefore, as wealth becomes more concentrated in the hand of more informed investors, the representative belief becomes more precise, which reduces the the risk-premium. This result lends an informational justification to the negative relationship between top wealth shares and risk-premium documented by Toda and Walsh (2014).

3.5 Unconditional Distribution and Dynamics of Wealth Shares

This section contains the main results relating the information environment to wealth distribution. Lemma 9 and Corollary 5 describe the evolution of wealth shares, given the posteriors and growth rates. Then, I provide some characterizations for the unconditional distribution of wealth shares in Proposition 10. The characterization is used to study how wealth inequality responds to changes in the information environment or market liquidity. Section 3.5.3 discusses the speed of wealth share dynamics.

Lemma 9. *Suppose $w_{f,t}^k$ is the wealth share of the investor of dynasty (f,k) ($f \in [0,1]$ and $k \in \{1,\ldots,S\} \cup \{N\}$) at t. Then her wealth share at $t+1$, if alive, is:*

$$\log w_{f,t+1}^k - \log w_{f,t}^k = \begin{cases} \log \frac{q_t^k}{\bar{q}_t} & \textit{if } g_{t+1} = g^h \\ \log \frac{1-q_t^k}{1-\bar{q}_t} & \textit{if } g_{t+1} = g^l \end{cases} \quad \forall \;\; 1 \le l \le t. \tag{3.19}$$

Lemma 9 specifies the dynamics of an investor's wealth share, as a function of her posterior belief q_t^k, the representative belief $\bar{q}_t$, and the realized growth rate. The right-hand side in (3.19) reflects the log-ratio of the probabilities assigned at t by the investor and representative investors to the realized growth rate at $t+1$. In other words, it states that an investor's wealth share increases if and only if her posterior belief is more consistent with the realized growth rate compared to that of the representative investor. For instance, if

$g_{t+1} = g^h$, the wealth share of the investor increases if and only if $q_t^k > \bar{q}_t$; that is, at t, the investor was more optimistic than the representative investor about having a high growth rate at $t+1$, and hence, allocated a larger fraction of her savings to the risky asset.

A few remarks are in order regarding Equation 3.19. First, note that the equation uses no equilibrium price or asset allocation, which drastically simplifies our analysis, both analytically and numerically. Second, Equation 3.19 is applicable for any belief process, and we do not need the economic agents to be Bayesian to obtain the equation. Putting differently, the proof of Lemma 9 works for any distribution of beliefs and it only requires the investors to have logarithmic preferences. In Section 3.6.4, by employing Equation 9, I construct the model-implied wealth dynamics with beliefs constructed from data and compare it against the empirical dynamics in wealth shares.

3.5.1 Unconditional Distribution of Wealth Shares

In this economy, the distribution of wealth shares moves with the growth rate, acting as an aggregate state. Thus, the distribution of wealth shares, itself, fails to be stationary. However, in this section, I prove that the wealth share of every family has a type-dependent stationary distribution. It is enough to be able to study the unconditional distribution of wealth shares across the families. In particular, I show that the unconditional distribution of wealth shares has a thick right tail and provide some characterizations for its thickness, which can be used as a measure of inequality (Gabaix, 2009).

To verify the existence of a stationary distribution for the wealth shares, we first need to introduce the Markov chain specifying the dynamics of the model. In this regard, I show the distribution of wealth shares across the types, an $S+1$-dimensional positive valued object, constitute a Markov chain.

Lemma 10. *Let Ω_t be the vector of wealth shares held by different types of investors:*

$$\Omega_t \equiv \{\{w_t^i\}_{i \in \{1,\ldots,S\}}, w_t^N\} \in [0,1]^{S+1}. \tag{3.20}$$

Then, there exists transition probability function $Q(\cdot,\cdot)$ such that $Q(\Omega_t, \Omega_{t+1})$ denotes the probability distribution function of Ω_{t+1} given the state at t is Ω_t.

Note that the vector of wealth shares at $t+1$ (Ω_{t+1}) is uniquely determined if values of Ω_t, growth states g_t and g_{t+1}, and signal realizations z_t^I, z_t^P, z_t^N are known. That said, $Q(\Omega_t, \Omega_{t+1})$ is the probability density of Ω_{t+1} realizing at $t+1$, without knowing the realizations of $\tilde{g}_t, \tilde{g}_{t+1}, \tilde{z}_t^I, \tilde{z}_t^P, \tilde{z}_t^N$. Therefore, a stationary distribution for $Q(\cdot,\cdot)$ is a distribution over the set of Ω's, such as $\mu : [0,1]^{S+1} \to \mathbb{R}_+$, that satisfies the following condition:

$$\mu(\Omega) = \int_{\Omega'} Q(\Omega', \Omega) d\mu(\Omega') \tag{3.21}$$

Lemma 11. ***a)*** *A stationary distribution for operator $Q(\cdot,\cdot)$ exists.*

b) *For every $k \in \{1,\dots,S\} \cup \{N\}$, a stationary distribution for sequence $\{q_t^k\}_{t=1}^{\infty}$ exists.*

c) *For every family (f,k), $f \in [0,1], k \in \{1,\dots,S\} \cup \{N\}$, the sequence of the family's wealth share $\{w_{f,t}^k\}_{t=1}^{\infty}$ has a stationary distribution.*

Lemma 11(c) states that the wealth share of every family is a stationary variable. Fix a stationary distribution of $\{\Omega_t\}_{t=1}^{\infty}$ and denote the corresponding stationary distribution for a family of type $k \in \{1,\dots,S\} \cup \{N\}$ by $G^k(\mathbf{w}) = Prob(w_{f,t}^k < \mathbf{w})$. That said, the unconditional distribution of wealth shares among sophisticated investors can be written as follows

$$G(\mathbf{w}) = \frac{1}{1-b_N} \sum_{i=1}^{S} b_i G^i(\mathbf{w}).$$

Proposition 10 provides a characterization of the unconditional distribution and type composition among the top groups.

Proposition 10. ***a)*** *If $\tau > 0$ and $\delta < 1$, $G(\cdot)$ has a thick right-tail and its tail parameter is the unique positive solution to the following equation:*

$$\lim_{T\to\infty} \mathbb{E}[(\Pi_{t=1}^{T}(\frac{(q_t^I)^{\gamma+1}}{\bar{q}_t^{\gamma}} + \frac{(1-q_t^I)^{\gamma+1}}{(1-\bar{q}_t)^{\gamma}}))]^{\frac{1}{T}} = (\delta + (1-\delta)(1-\tau)^{\gamma})^{-1} \tag{3.22}$$

Putting differently, $\lim_{w\to\infty} \frac{1-G(w)}{w^{-\gamma}}$ exits and is stricly positive for some $\gamma > 1$.[21]

b) *[Dominance of type-S investors in top wealth groups] Suppose $F(S-1) < 1$, that is type-S investors are the only group that always become informed. Then for an econometrician*

21. In fact, γ is the infimum value for which $A > 0$ exists, such that $G(\mathbf{w}) > 1 - A\mathbf{w}^{-\gamma}$ for all $\mathbf{w} > 0$.

who can only observe the wealth shares, not the types, we have:

$$\lim_{\mathbf{w}\to\infty} P(\tilde{i} = S | w_{f,t}^{\tilde{i}} > \mathbf{w}) = 1. \tag{3.23}$$

Proposition 10(a) shows that under very general conditions, the unconditional distribution has a thick right-tail. The thickness of the tail is governed by the extent of asymmetric information between the most informed investors, i.e., type-S investors, and the hypothetical representative investor, according to (3.22). Therefore, the information environment impacts the thickness parameter γ only through the extent to which the most informed investors' belief is more precise than that of the representative investor. This simple criterion can be used to gauge the impact of a change in the information environment on wealth inequality. Building on this result, Corollary 4 relates the tail parameter to the excess return of the most informed investors.

Corollary 4. *Let r_t^I be the log-return of the most informed investors, i.e., type-S investors. Then, γ is the unique positive solution to the following equation:*

$$\lim_{T\to\infty} \mathbb{E}[e^{\gamma \sum_{t=1}^{T}(r_t^I - g_t + \log\beta)}]^{\frac{1}{T}} \times (\delta + (1-\delta)(1-\tau)^{\gamma}) = 1. \tag{3.24}$$

Corollary 4 shows that the tail parameter is related to $r_t^I - g_t$: the return of the most informed investors minus the growth rate, which is also the return of the market portfolio. The intuition is that $r_t^I - g_t$ captures how faster the wealth of the most informed investors grows compared to the whole economy. This result resembles the $r - g$ rule put forward by Piketty (2014), with only one exception. In Piketty's rule, *the return* is the average return on capital, whereas here, *the return* is the return of the most informed investors on their savings. That said, Corollary 4 underscores the role of asymmetric information in top inequality and enriches Piketty's rule to account for such heterogeneities.

The proposition also indicates the tax rate τ and the death rate $1-\delta$ also play a counterbalancing role in top wealth inequality. Consistent with intuition, an increase in the tax rate or death probability reduces wealth inequality, as the former increases the size of redistributions, and the latter makes them more frequent.

Part (b) in Proposition 10 specifies the composition of the wealthiest. It states that the fraction of investors with a type less than S and wealth share exceeding $\mathbf{w}$ goes to zero as $\mathbf{w}$ is taken to infinity. In other words, as we focus on a smaller set of top wealthiest investors, less informed ones asymptotically vanish from the set. Therefore, the tail is populated by highly sophisticated investors. It provides an explanation for the substantial representation of highly sophisticated individuals in the finance sector among the ultra-rich (Kaplan and Rauh, 2013).

3.5.2 Comparative Statics

This section presents how the tail parameter changes with the precision of the public signal and market liquidity.

Proposition 11. *The tail parameter is not monotone in the precision of the public signal. Putting differently, there exist $\bar{z}_1^P > \bar{z}_2^P$ and parameter values $\bar{z}^I$ and $\bar{z}^N$ such that they induce tail parameters γ_1 and γ_2 respectively, and $\gamma_1 < \gamma_2$.*

Proposition 11 states that wealth inequality is not monotonically decreasing in the precision of the public signal, governed by $\bar{z}^P$.[22] The intuition is that the investors, endowed with more accurate signals and facing less uncertainty, react more aggressively to their beliefs. To see this in the model, we can obtain the portfolio sensitivities, by combining 3.12 with 3.17, and examine how it changes with the public signal precision.

$$\begin{aligned} \mu_t^k &= 1 + \frac{q_t^k - \bar{q}_t}{\bar{q}_t(1-\bar{q}_t)}\left(\frac{e^{g_h}}{e^{g_h} - e^{g_l}} - \bar{q}_t\right) \\ \Rightarrow \mathbb{E}\left[\frac{\partial \mu_t^k}{\partial q_t^k}\right] &= \left(\frac{e^{g_h}}{e^{g_h} - e^{g_l}} - \frac{1}{2}\right)\mathbb{E}\left[\frac{1}{\bar{q}_t(1-\bar{q}_t)}\right] \end{aligned} \tag{3.25}$$

Note that $\bar{q}_t(1-\bar{q}_t)$ is an inverse measure of price informativeness since $\bar{q}_t(1-\bar{q}_t) = Var[\mathbf{1}_{\{\tilde{g}_{t+1}=g^h\}}|p_t, R_t^F]$, which is directly related to the conditional variance of future outcomes implied by current equilibrium prices. That said, Equation 3.25 states that the investors' portfolios become more sensitive to their beliefs as a result of any change that makes

22. One can show that with an increase in $\bar{z}^P$, the public signal becomes more informative in Blackwell sense.

the prices more informative on average. An example of this change is an increase in the precision of public signals. The increase can widen the portfolio heterogeneity, and hence the return heterogeneity, which results in larger wealth inequality. Figure 3.1 illustrates this point in a numerical example. It is an example in which the top wealth shares do not decrease with the precision of the public signal.

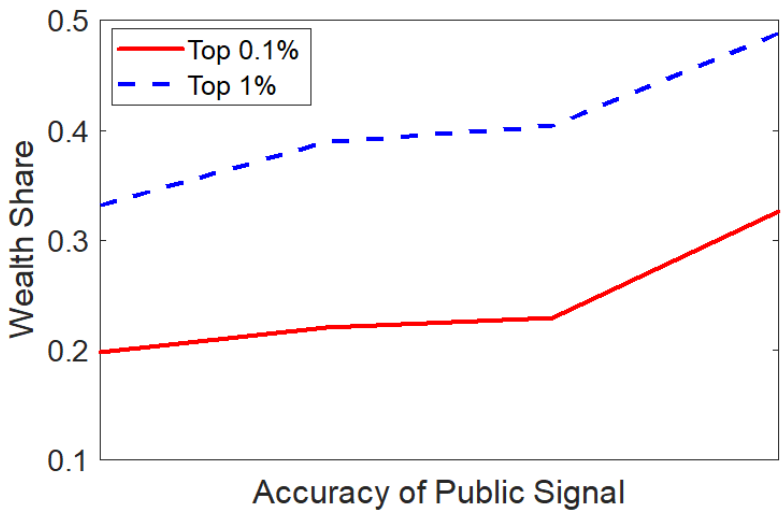


Figure 3.1: Y-axis displays the top wealth shares in the unconditional wealth distributions for different values of $\bar{z}^P$, which governs the precision of the public signal.

Next proposition discusses the role of market liquidity in wealth inequality.

Proposition 12. *Top wealth inequality monotonically increases with market liquidity. In other words, for any set of parameters $q(\cdot), \bar{z}^P, \bar{z}^I, \bar{z}^N$, the tail parameter is strictly decreasing with b.*

Proposition 12 states that top wealth inequality increases with market liquidity, which is captured by the measure of naive investors in this model.[23] The intuition for this result is that with an increase in the measure of naive investors, the vector of equilibrium prices becomes less informative about the future growth rates, which widens the gap in the expected returns

23. Note that for a higher value of b_N, any size of transactions by sophisticated investors has a smaller impact on the prices. As such, b_N can be reasonably used as an exogenous variable governing market liquidity in this model.

between the informed and uninformed investors. Putting differently, when markets are more liquid, more informed investors can exploit their information advantage better because they can trade based on their information with making less impact on the equilibrium prices. As such, they reveal less of their information to the rest of the investors. Therefore, they can generate a higher expected return in more liquid markets, making the unconditional wealth distribution more right-skewed. Figure 3.2 illustrates this point.

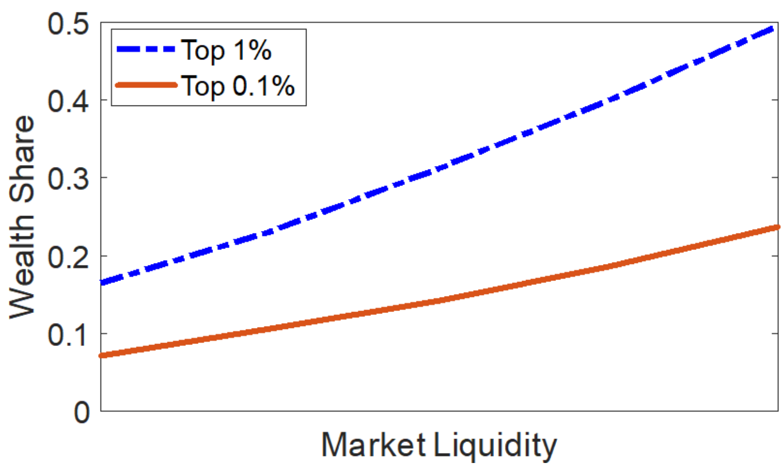


Figure 3.2: Impact of market liquidity on top wealth shares. The graph displays top wealth shares, in the unconditional distribution, for different values of b_N, keeping the ratios between $b_1, \ldots, b_S$ constant.

One important insight generated by the model is that price informativeness plays a key role in the wealth distribution. Equilibrium prices, acting as endogenous signals, help reduce the return heterogeneity. For instance, when few discover a profitable investment opportunity, it is the resulting price appreciation that informs other investors about the opportunity. A direct implication of this insight is that market friction that impairs price informativeness could increase wealth inequality.

A policy implication of the model is that wealth inequality can be reduced by improving price efficiency. As a result, the recent reduction in the transaction costs of trading individual stocks could increase the volume of noise trading, which amplifies both return heterogeneity and wealth inequality. That said, informing investors about the risks associated with holding individual stocks reduces wealth inequality not only by improving their wealth management but also making stock prices more informative about their fundamentals.

3.5.3 Asymmetric Information and Wealth Distribution Dynamics

This section examines dynamic implications of asymmetric information for the distribution of wealth shares.

Proposition 13 (Bounds on the speed of wealth distribution dynamics)**.** *Suppose sophisticated investors in dynasties (f_1, i_1) and (f_2, i_2) $(1 \leq i_1 < i_2 \leq S)$ have wealth shares $w^{i_1}_{f_1,t}$ and $w^{i_2}_{f_2,t}$ at t respectively. Furthermore, assume the parameters are such that $q^I_t \leq 1 - \frac{\kappa b_N}{1-\kappa b_N}\bar{z}^N$ with probability one. Then, provided neither of them receive the death shock at $t+1$, there exist constant values $\underline{A} < \bar{A} < 0$ such that:*[24]

$$\underline{A} < \frac{\mathbb{E}_t[\log \frac{w^{i_1}_{f_1,t+1}}{w^{i_2}_{f_2,t+1}}] - \log \frac{w^{i_1}_{f_1,t}}{w^{i_2}_{f_2,t}}}{(\mathbb{E}_t[\frac{1}{\tilde{q}^I_t} + \frac{1}{1-\tilde{q}^I_t}|g_t, z^P_t])(F(i_2) - F(i_1))(\bar{z}^N)^2} < \bar{A} \tag{3.26}$$

Inequality 3.26 provides a lower and upper bound for the speed of wealth divergence among the sophisticated investors.[25] The speed of the dynamics increases with the magnitude of noise in the equilibrium prices, captured by $\bar{z}^N$, and the extent of asymmetric information among the investors, captured by CDF $F(\cdot)$, since they increase the return heterogeneity. The precision of the public signal also plays a role in the speed of the dynamics, as it impacts the belief precision of the informed investors. When public signals are more precise, they take more aggressive positions based on their beliefs since they face less uncertainty. It increases the return heterogeneity, and thus the speed of wealth share dynamics.

24. In the denominator, $\tilde{q}^I_t$ is a random variable denoting the belief of informed investors, conditional on the growth state g_t and public signal realization z^P_t.

25. Note that the divergence is eventually counteracted by the taxation and redistribution processes embedded in the model, which act as the stabilizing forces.

3.6 Discussion

In the rest of the paper and Internet Appendix, I demonstrate the key results are robust to several extensions of the model. Specifically, Section 3.6.1 endogenizes information acquisition, Section 3.6.2 examines the role of delegated investment. Section 3.6.3 presents an extension of the baseline model with multiple risky assets and imperfectly ordered information sets. Section IA2 expands the security space, and Section IA3 studies the impact of bequest motives for saving. In these two sections, I show the extensions virtually have no impact on the dynamics and characterizations derived thus far. Section IA4 examines a case in which children imperfectly inherit their parent's type.

Two sections provide quantitative analysis on the role of investment sophistication in wealth inequality. Section 3.6.4 examines belief heterogeneity in data. The beliefs are used to construct model-implied dynamics for top wealth shares, which display substantial consistency with the actual dynamics. In a calibration exercise, Section IA5 tests the model's implications regarding the portfolio composition and its dynamics for the top 0.01% US households.

3.6.1 Endogenous Information Acquisition

Economic agents are heterogeneously informed about future returns due to the costly nature of information acquisition. The cost is partly pecuniary, like that of obtaining access to datasets, or the cost of education, which is paid to earn skills needed to process available signals (Lusardi, Michaud, and Mitchell (2017)). In this section, I endogenize the types by allowing the newborns to choose their type at a cost. Therefore, investors select their probability of becoming informed before they start trading. It can be thought of as the decision to acquire permanent human capital to increase the expected returns on savings.[26] In short, I show that top wealth inequality, captured by the tail parameter, increases with the cost of information acquisition.

In this section, I modify the baseline setup in two aspects. First, I assume there are

26. Fagereng, Guiso, Malacrino, and Pistaferri (2020) and Girshina (2019) find empirical evidence supporting the impact of higher education in earning risk-adjusted return on risky investments.

only two types of sophisticated investors: One type is always informed investors and learn the realization $\tilde{z}_t^I$, and the other type is always uninformed. Second, I assume all newborns choose their type before they start trading. Thus information acquisition here is a one-time and irreversible decision. In particular, they become an informed investor for their entire lifetime by paying cost $(\beta\delta)^{-1}\chi p_t$ out of their endowed wealth, where $\chi > 0$, and they remain uninformed otherwise. All of information acquisition payments at t are evenly redistributed among all the newborns at t.[27] Lemma 12 characterizes the newborns' type acquisition decision.

Lemma 12. *A newborn with inherited wealth* $w_{f,t} = (1-\tau)\tilde{w}_{f,t} + \tau$ *becomes informed iff*

$$\frac{1}{1-\beta\delta}\log(w_{f,t} - \chi + h_t\chi) + \sum_{m=1}^{\infty}\frac{(\beta\delta)^m}{1-\beta\delta}\mathbb{E}_t[q_{t+m-1}^I \log\frac{q_{t+m-1}^I}{q_{t+m-1}^U} + (1-q_{t+m-1}^I)\log\frac{1-q_{t+m-1}^I}{1-q_{t+m-1}^U}] \geq \frac{1}{1-\beta\delta}\log(w_{f,t} + h_t\chi), \tag{3.27}$$

,where h_t *is the fraction of newborns who decide to become informed at* t.

Lemma 12 exhibits the trade-off a newborn faces in her information acquisition decision. On the one hand, becoming informed increases the expected return on savings and increases her lifetime expected utility. On the other hand, a fixed cost is associated with becoming informed, justified only for wealthy enough newborns. In fact, a threshold value $\bar{w}_t$ exists such that a newborn becomes informed iff $w_{f,t} \geq \bar{w}_t$. By inspecting (3.27), we see an increase in the cost of becoming informed, χ, increases the threshold $\bar{w}_t$, and consequently, tends to reduce the fraction of informed investors. As such, a higher χ leads to a smaller fraction of informed investors, resulting in less incorporation of informative signal $\tilde{z}_t^I$ into the equilibrium prices. It widens the belief gap between the informed and uninformed investors. Proposition 14 shows that top wealth inequality increases boundlessly with the cost of information acquisition.

27. This assumption directly implies that the total wealth is not affected by the newborns' information acquisition decision. The assumption is made for tractability reasons. Particularly, if costly information acquisition affects the total wealth level, it would also impact the current prices, which complicates the analysis.

Proposition 14. *Denote $\gamma(\chi)$ is the tail parameter of the unconditional distribution of wealth shares corresponding to cost parameter χ.*[28] *Furthermore, suppose signal $\tilde{z}_t^I$ is sufficiently informative so that the following condition holds:*

$$\mathbb{E}[q_t^I \log \frac{q_t^I}{q(g_t)} + (1 - q_t^I) \log \frac{1 - q_t^I}{1 - q(g_t)}] + (1 - \delta) \log(1 - \tau) > 0, \tag{3.28}$$

where $q(g_t) = P(g_{t+1} = g^h | g_t = q^j)$, $j \in \{h, l\}$. Then, $\gamma(\chi) \to 1$, as χ goes to infinity.

Proposition 14 states that as the cost of information acquisition increases, so does the inequality, provided the informative signal is sufficiently informative (condition 3.28). Note that $\gamma = 1$ designates the highest level of inequality. The main force leading to this result is the relationship between χ and the extent of asymmetric information between informed and representative investors. As the cost of information acquisition increases, a smaller and smaller fraction of the investors become informed. Therefore, the equilibrium prices continue to reflect less of the informative signal, which pushes up the expected return of informed investors and the speed at which their wealth diverges from the uninformed investors. At the same time, the redistributive power of the taxation is fixed by parameters τ and δ. Therefore, the result indicates a positive relationship between top wealth inequality and the cost of information acquisition.[29]

3.6.2 Delegated Investment

Thus far, I have assumed that investors cannot delegate their investment decisions to other more informed investors. In this section, I relax this assumption to examine the role of delegated investments in wealth inequality.

Particularly, I consider the following modification of the baseline model. Suppose three types of investors are present, in addition to the naive investors: Informed investors, who observe z_t^I in every period (they can represent hedge fund or active mutual fund managers); "delegating" investors, who represent the funds' clients; and uninformed investors, who

28. If there are multiple tail parameters, I take the smallest one.

29. The fraction of informed investors is always positive since condition 3.28 ensures the existence of arbitrarily wealthy investors.

manage their wealth by their own. Delegating and uninformed investors never observe the signal realization z_t^I.

The excess returns generated from the delegated investments are divided between the client and her fund manager based on a Nash bargaining protocol, where the fund manager receives fraction $\psi \in [0,1]$ of the surplus. More specifically, if the fund manager generates return R_t^I and uninformed investors obtain R_t^U, the client's effective return on her wealth is $R_t^U + (1-\psi)(R_t^I - R_t^U)$. The rest of the return accrues to the fund manager. ψ captures the competitiveness of the market for wealth-management. I denote the measure of informed, delegating, and uninformed investors by b^I, b^D, b^U. Note that $b^I + b^D + b^U = 1 - b_N$.

Proposition 15. *The wealth share of the informed investors is non-monotone in the fraction of delegating investors (b^D).*

As illustrated in Figure 3.3, Proposition 15 states that an interior value of the clientele size maximizes the expected wealth share of the fund managers. As for the increasing part, it is straightforward that the managers benefit from managing a larger wealth share, because they receive a fraction of the return they generate for their clients. However, because the fund managers collectively manage a larger wealth share of the economy, they impact the prices more, making them more informative. It shrinks the return gap between informed and informed investors. Particularly, as the size of delegated investment increases, two forces hurt the fund managers' wealth accumulation: first, the fund managers obtain a lower rent from their management service, due to the reduction in $\mathbb{E}[R_t^I - R_t^U]$. Second, their own expected return, $\mathbb{E}[R_t^I]$, also decreases as more information is disseminated through the equilibrium prices. Due to these forces, the fund managers' expected wealth share does not always increase with their fund size, leading to the non-monotonicity result pointed by Proposition 15.

The finding suggests that the growth of the wealth-management industry could have contributed to the rise of super-rich fund managers, and consequently, wealth inequality. Nevertheless, a further increase in delegated wealth management might mitigate, not aggravate, the inequality. In order to reduce inequality, the result suggests removing the frictions in accessing profitable funds, such as minimum investment requirements, which is prevalent in the industry.

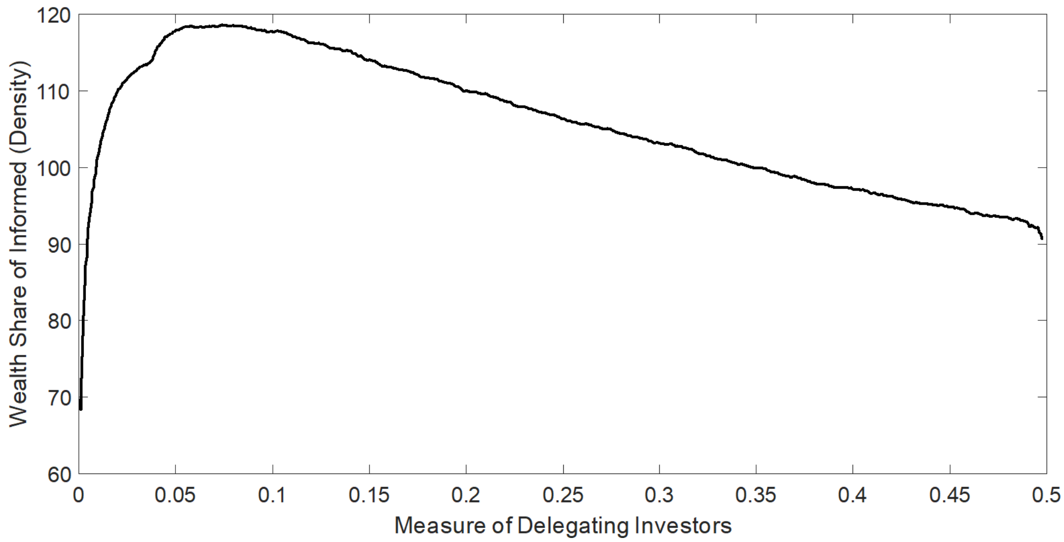


Figure 3.3: The expected wealth share of the most informed investors for different values of their clientele size

3.6.3 The Case of Multiple Risky Assets

In the baseline setup, the investors are asymmetrically informed only about future growth rates. However, in practice, some investors and fund managers engage in market-neutral strategies to minimize their exposure to aggregate risks. In this section, I show the results carry over for various forms of asymmetric information, and the model implications are not restricted to the specific form of asymmetric information studied in the baseline. Moreover, we see that the results do not hinge on the single risky asset assumption in the baseline setup.

I expand the model by introducing growth-independent state $\tilde{\varepsilon}_t \in \{0, 1\}$. Suppose the investors at period t can trade securities contingent on the realization of ε_{t+1}. Specifically, two new securities, with zero net-supply, are introduced: a security that pays one unit of consumption good at time $t+1$ if $\varepsilon_{t+1} = 1$, and a security that pays one share of the tree if $\varepsilon_{t+1} = 1$.[30]

The information environment around ε_{t+1} is as follows:

30. Note that these two assets are different. The first asset is only exposed to the growth-independent shock, while the second asset is exposed to both shocks in the model.

$$
\begin{gathered}
\varepsilon_t \in \{0,1\} \\
q^{\varepsilon} \in (0,1) \\
\tilde{y}_t^I \sim U[-\bar{y}^I, \bar{y}^I] \quad \tilde{y}_t^N \sim U[-\bar{y}^N, \bar{y}^N] \quad \bar{y}^N, \bar{y}^I < \min\{1-q^{\varepsilon}, q^{\varepsilon}\} \\
P(\tilde{\varepsilon}_{t+1} = 1 | g_t, z_t^I, z_t^N, z_t^P, y_t^I, y_t^N) = q^{\varepsilon} + y_t^I \quad \forall t.
\end{gathered}
\tag{3.29}
$$

Equation 3.29 indicates that the only signal informative about the value of ε_{t+1} is y_t^I. However, similar to the baseline case, the naive investors believe y_t^N is informative about ε_{t+1}, and y_t^I is not.

Furthermore, I assume that sophisticated investors are further divided into $S' > 1$ types regarding their information set about $\tilde{\varepsilon}_{t+1}$. The type composition is independent of the previous S types. Therefore, sophisticated investors are divided into $S \times S'$ types, meaning that only a subset of the most sophisticated investors for one risky asset are also among the most sophisticated group for the other asset.

More specifically, each sophisticated investor is indexed by $(f, (i, i^{\varepsilon}))$, where $(i, i^{\varepsilon}) \in \{1, \ldots, S\} \times \{1, \ldots, S'\}$ and $f \in [0,1]$. In every period $\lambda_t^{\varepsilon} \in \{1, \ldots, S'\}$ from CDF $F^{\varepsilon} : \{1, \ldots, S'\} \to [0,1]$ is drawn and all sophisticated investors with type $i^{\varepsilon} \geq \lambda_t^{\varepsilon}$ learn y_t^I. Similar to the baseline case, $\lambda_t \in \{1, \ldots, S\}$ is drawn and all sophisticated investors with $i \geq \lambda_t$ learn the realization of z_t^I.[31] Overall, the information environment regarding to g_{t+1} is the same as the baseline case.

Proposition 16 extends the result of Proposition 10 to the case with asymmetric information in multiple dimensions.

Proposition 16. *Suppose $\tau > 0$ and $\delta < 1$ and let $P^I(\cdot)$ be the probability assignments by the most informed investors, who are always informed about the realizations of $\tilde{z}_t^I$ and $\tilde{y}_t^I$. Moreover, $\bar{P}(\cdot)$ represents the belief of the representative investor. Then, the unconditional distribution has a thick right tail, with tail parameter γ that satisfies the following:*

31. Note that according to (3.29), $\tilde{\varepsilon}_{t+1}$ and g_{t+1} are independent.

$$\lim_{T\to\infty}\mathbb{E}[(\Pi_{t=1}^{T}\sum_{g^j\in\{g^h,g^l\},\epsilon\in\{0,1\}}\frac{P^I(\tilde{g}_{t+1}=g^j,\tilde{\varepsilon}_{t+1}=\epsilon)^{\gamma+1}}{\bar{P}(\tilde{g}_{t+1}=g^j,\tilde{\varepsilon}_{t+1}=\epsilon)^{\gamma}})]^{\frac{1}{T}}=(\delta+(1-\delta)(1-\tau)^{\gamma})^{-1}. \tag{3.30}$$

The key message of Proposition 16 is that the form of asymmetric information does not matter for the form of unconditional wealth distribution. Whether about the growth state or other growth-independent states, asymmetric information about future payoff-relevant states impacts the wealth distribution in the same way and has the same implications. Moreover, Equation 3.30 confirms the baseline result that asymmetric information contributes to top wealth inequality only through the extent of asymmetric information between the most informed investors and representative investor. The robustness of other baseline results, such as Propositions 11 and 12, follows from this observation.

3.6.4 Model-Implied Evolution of Wealth Shares

Data reveals households hold vastly different beliefs about future macroeconomic conditions (Das, Kuhnen, and Nagel, 2017). My model can inform how the belief heterogeneity has contributed to the recent dynamics of wealth inequality. In particular, my model delivers Equation 3.A.6 that can be used to connect belief dynamics to wealth share dynamics.

I construct the belief by exploiting a sample of 218,559 individuals, who were asked about their macroeconomic expectation in Michigan Survey Data, between 1978q1-2018q1. The survey includes several questions, but for belief construction, I only use the respondents' expectation about 1-year-ahead economic growth, captured by variable "BUS12."[32] That said, we can find the model-implied evolution of wealth shares by feeding the beliefs obtained from the survey into Equation 3.A.6. However, because the data does not have information about the respondents' wealth, I use their income as a proxy. Figure 3.4 depicts average forecast error in beliefs about next year's growth rates for different income groups. The

32. More specifically, it asks "Now turning to business conditions in the country as a whole, do you think that during the next 12 months we'll have good times financially, or bad times, or what?" The data also includes information about households' income and their investment in the stock market.

negative relationship observed in the figure confirms high-income households holding more precise beliefs about future growth rates, consistent with the findings of Das, Kuhnen, and Nagel (2017).[33]

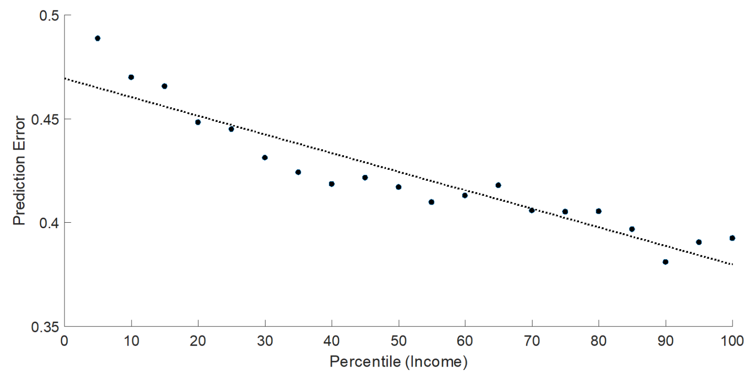


Figure 3.4: Prediction error of the growth rate for different levels of (a) income (b) total investment in the publicly listed stocks. The predictions are binary and the realized growth rates are divided into "high" or "low" growth rate, with a cutoff equal to the annual rate of 2.4%. The figure shows $\mathbb{E}[(respond_{i,t} - \mathbb{I}_{g_t+g_{t+1}+g_{t+2}+g_{t+3}\geq 2.4})^2]$, where i is the respondent indicator and t is the quarter indicator and g_t is the quarter-to-quarter growth at t.

Figure 3.5 plots the model-implied and empirically-documented time-series of the wealth-ratio between the top 1% and bottom 90%. For the model-implied series, I use the average belief among the top 1% and the bottom 90% in income among the respondents. The time series of the wealth-ratio is taken from Saez and Zucman (2016). I use their estimated wealth-ratio in 1978 as my initial value. I set $\delta = 0.975, \tau = 0.2$.

Figure 3.5 exhibits a remarkable co-movement between the model-implied and empirical dynamics in the wealth-ratios, especially in terms of the trend. The model-implied series is more volatile than the empirical counterpart. The logarithmic preferences used in the model, which imply a relatively low risk-aversion parameter, can be responsible for the excess volatility. Conceptually, the figure speaks to the ongoing discussions on the role of

33. To construct the forecast errors, I compare the respondents' answers with the realized growth in the following year (Including the survey's quarter). I only use responses with a clear, good, or bad expectation, which constitute more than 75% of the total responses. Moreover, I consider an annual growth rate above its historical mean as "high", and "low" otherwise. Then, I calculate and report the average value of $(respond_{i,t} - \mathbb{I}_{g_t+g_{t+1}+g_{t+2}+g_{t+3}\geq \bar{g}})^2$ for each of twenty half-deciles. $respond_{i,t}$ is a binary variable indicating whether the respondent has a positive growth prospect.

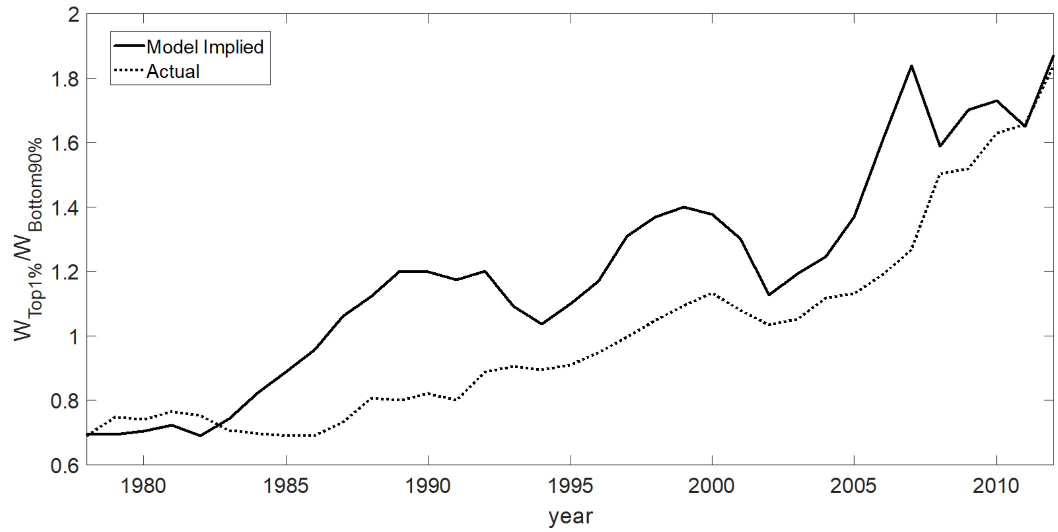


Figure 3.5: Model implied versus empirically documented wealth-ratio between the top 1% and bottom 90% in income, based on the Michigan Survey Data for 1978-2012. The empirical time series is from Saez and Zucman (2016). To estimate the beliefs, I use income as a proxy for wealth. The implied evolution is based on equation (3.19) for $\delta = 0.975$ and $\tau = 0.2$. The empirically estimated wealth-ratio at 1978 used as the initial condition in computing the model-implied dynamics.

portfolio heterogeneity in wealth inequality dynamics (Hubmer, Krusell, and Smith Jr, 2020), by suggesting that the belief heterogeneity observed in data can account for the portfolio heterogeneity needed to explain recent dynamics in the US wealth distribution.

3.7 Conclusion

I study a general equilibrium model with a continuum of long-lived and heterogeneously informed investors, who trade risky and risk-less claims of a Lucas tree. I show that the unconditional distribution of wealth shares features a thick right-tail, populated by the best-informed investors.

My work has important policy implications in the areas of consumer finance and wealth management industry. First, all forms of public information subsidies do not reduce wealth inequality. Thus, government agencies need to ensure that their information subsidies indeed reduce asymmetric information among households, if they aim to mitigate wealth inequality through improving households' saving decisions. That said, policymakers should help households make more informed investment decisions by providing them understandable information about the risk and return prospects of different investment opportunities. Doing so increases the households' confidence in entering riskier asset classes and could alleviate wealth inequality.

Second, frictions in the wealth management industry can exacerbate wealth inequality by enabling fund managers to extract rents from their wealth management services. In the presence of such frictions, wealth inequality might increase with the size of delegated investments in the economy. My results indicate that resolving the frictions and removing the barriers in entering profitable funds result in more price informativeness and less return heterogeneity, reducing wealth inequality. Furthermore, pension fund managers, who manage a substantial fraction of households' wealth (36% according to Smith, Zidar, Zwick, et al. (2020)), play a crucial role in reducing inequality. Therefore, authorities should ensure pension fund managers exert the necessary due diligence in their investment decisions, and their incentives are best aligned with those of their clients in wealth accumulation.

Appendix 3.A: Proofs and Mathematical derivations

3.A.1 Proof of Lemma 5

a) The optimal strategy is scale-invariant, that is investors with the same type and different wealth levels consume and save in the same proportion. The scalability implies:

$$V^k(W_t;\mathcal{F}_t) = V^k(1;\mathcal{F}_t) + \sum_{l=0}^{\infty}(\beta\delta)^l \log W_t = \frac{1}{1-\beta\delta}\log W_t + V^k(1;\mathcal{F}_t)$$

,which proves part (a).

b) Let R_t^k be the random return on the optimal portfolio. Therefore, (3.11) can be rewritten as follows:

$$\begin{aligned} V^k(W_t;\mathcal{F}_t) &= \max_{c_t} \log c_t + \beta\delta\mathbb{E}_t^k[\frac{1}{1-\beta\delta}\log((W_t - c_t)R_t^k) + a^k(\mathcal{F}_{t+1})] \\ &= \max_{c_t} \log c_t + \frac{\beta\delta}{1-\beta\delta}\log(W_t - c_t) + \beta\delta\mathbb{E}_t^k[\frac{1}{1-\beta\delta}\log R_t^k + a^k(\mathcal{F}_{t+1})] \end{aligned} \tag{3.A.1}$$

Note that the last term in (3.A.1) is independent of W_t. Therefore, the first order condition for c_t implies:

$$\frac{1}{c_t} = \frac{\beta\delta}{1-\beta\delta}\frac{1}{W_t - c_t} \Rightarrow c_t = (1-\beta\delta)W_t$$

c) According to the market clearing condition for the good market:

$$\begin{aligned} d_t &= c_t^N + \int_0^1 c_t^i di = (1-\beta\delta)(W_t^N + \int_0^1 W_t^i di) \\ &= (1-\beta\delta)((p_t + d_t)(x_t^N + \int_0^1 x_t^i di) + s_t^N + \int_0^1 s_t^i di) = (1-\beta\delta)(p_t + d_t) \end{aligned}$$

After a few rearrangements, we get:

$$p_t = \frac{\beta\delta}{1-\beta\delta}d_t$$

3.A.2 Proof of Lemma 6

According to Lemma 5, the investors choose μ_t^k to maximize their expected log-return. Thus, μ_t^k is the solution to the following optimization problem:

$$\max_{\mu} \quad \mathbb{E}_t^k[\log(\mu \frac{d_{t+1} + p_{t+1}}{p_t} + (1-\mu)R_t^F)] \tag{3.A.2}$$

According to Lemma 5(c), we can rewrite the optimization problem (3.A.2) as follows:

$$\max_{\mu} \quad q_t^k\{\log e^{g^h}(\mu + (1-\mu)\beta\delta e^{-g^h} R_t^F))\} + (1-q_t^k)\{\log e^{g^l}(\mu + (1-\mu)\beta\delta e^{-g^l} R_t^F))\} \tag{3.A.3}$$

Due to the strict concavity of the objective function, we only need to solve for μ that satisfies the first condition. One can verify that μ_t^k specified in (3.12) indeed satisfies the condition.

3.A.3 Proof of Lemma 7

Equation 3.14 directly implies that $q_t^I w_t^I + q_t^N w_t^N$ lies in the information set of the uninformed investors. Since $q(g_t) = q_t^I - z_t^I = q_t^N - z_t^N$ and $w_t^I + w_t^N = 1 - w_t^U$, $w_t^I z_t^I + w_t^N z_t^N$ lies in the same information set, i.e., it is inferable from the equilibrium prices.

3.A.4 Proof of Lemma 8

$$q_t^U = \mathbb{E}[\mathbb{I}_{\{g_{t+1}=g^h\}}|\mathcal{F}_t^P, \bar{z}_t^U = w_t^I z_t^I + w_t^N z_t^N] = \mathbb{E}\big[\mathbb{E}[\mathbb{I}_{\{g_{t+1}=g^h\}}|z_t^I]|\mathcal{F}_t^P, \bar{z}_t^U\big] = \mathbb{E}[z_t^I|\mathcal{F}_t^P, \bar{z}_t^U]$$
$$= \frac{1}{2}[\max\{-\bar{z}^I, z_t^I + \frac{w_t^N}{w_t^I}(z_t^N - \bar{z}^N)\}, \min\{\bar{z}^I, z_t^I + \frac{w_t^N}{w_t^I}(z_t^N + \bar{z}^N)\}]$$

I apply the law of iterative expectations and then the Bayes' rule in the equations above.

3.A.5 Proof of Lemma 9

The evolution of wealth for type-k investors, who invest fraction μ_t^k of their saving in the risky asset, is given by:

$$W_{t+1}^k = \beta\delta W_t^k(\mu_t^k \frac{d_{t+1} + p_{t+1}}{p_t} + (1-\mu_t^k)R_t^F)$$

Therefore, by applying Lemma 5(c), we get:

$$w_{t+1}^k = w_t^k(\mu_t^k + (1-\mu_t^k)\beta\delta R_t^F e^{-g_{t+1}})$$

Now, by some rearranging and substituting (3.12) for μ_t^k, we have:

$$\begin{aligned} w_{t+1}^k &= w_t^k + (\mu_t^k - 1)(1 - \beta\delta R_t^F e^{-g_{t+1}})w_t^k \\ &= w_t\big[1 + (\frac{q_t^k}{\beta\delta R_t^F e^{-g^l} - 1 + \kappa} + \frac{1-q_t^k}{\beta\delta R_t^F e^{-g^h} - 1})(1 - \beta\delta R_t^F e^{-g_{t+1}})\big] \end{aligned} \tag{3.A.4}$$

When $g_{t+1} = g^h$, we have:

$$\frac{w_{t+1}^k}{w_t^k} = 1 - q_t^k \frac{\beta\delta R_t^F e^{-g^h} - 1}{\beta\delta R_t^F e^{-g^l} - 1} - 1 + q_t^k = q_t^k \frac{\beta\delta R_t^F (e^{-g^h} - e^{-g^l})}{\beta\delta R_t^F e^{-g^l} - 1} \tag{3.A.5}$$

Note that Equation 3.A.5 holds for any investor, including the hypothetical representative investor with belief $\bar{q}_t \equiv w_t^N q_t^N + w_t^I q_t^I + w_t^U q_t^U$. Furthermore, the wealth share of this representative agent is fixed at one. Equations 3.14 and 3.12 imply that such representative agent always chooses $\mu_t = 1$. Therefore, if we plug the belief and the wealth of the representative agent into (3.A.5), we get:

$$\bar{q}_t \frac{\beta\delta R_t^F (e^{-g^h} - e^{-g^l})}{\beta\delta R_t^F e^{-g^l} - 1} = 1$$

Therefore, we can rewrite (3.A.5) as follows:

$$\frac{w_{t+1}^k}{w_t^k} = \frac{q_t^k}{\bar{q}_t}$$

For $g_{t+1} = g^l$, we can similarly show:

$$\frac{w^k_{t+1}}{w^k_t} = \frac{1-q^k_t}{1-\bar{q}_t}$$

The specification in the Lemma directly follows from the equations above.

3.A.6 Proof of Lemma 10

The following Corollary provides the dynamics of state variable Ω_t.

Corollary 5. *Suppose $\kappa = (1-\delta)\tau$. Then, the wealth share in the hand of type-k investors evolves according to the following equation of motion:*

$$w^k_{t+1} = \begin{cases} w^k_t \frac{q^k_t}{\bar{q}_t}(1-\kappa)+\kappa & \textit{if } g_{t+1} = g^h \\ w^k_t \frac{1-q^k_t}{1-\bar{q}_t}(1-\kappa)+\kappa & \textit{if } g_{t+1} = g^l \end{cases} \quad \forall k \in \{1,\ldots,S\} \cup \{N\}. \tag{3.A.6}$$

Note that in Equation 3.A.6, q^k_t and $\bar{q}_t$ are random variables, which depend on state variable Ω_t and i.i.d. random variables $\tilde{\lambda}_t$, $\tilde{z}^I_t$, and $\tilde{z}^N_t$. Therefore, the transition equation induces a random mapping between Ω_t and the distribution of wealth shares at $t+1$.

3.A.7 Proof of Lemma 11

a) Note that it is essentially a fixed point problem. The state variable takes value from a finite dimensional compact space, and the random operator maps elements of this compact space to itself. Hence, it is intuitive that an invariant probability measure, i.e., a stationary distribution exists. To formalize the point, note that due to continuinity of the mapping, the process satisfies the Feller property, which ensures the existence of a stationary distribution.[34] However, the stationary distribution might not be unique.[35]

34. For the formal statement of the theorem used here and its proof, see Theorem 5.4.1 at `https://stanford.edu/class/msande321/Handouts/05%20Feller%20Chains.pdf`

35. Note that Λ is not totally ordered, therefore we cannot directly apply the existence and uniqueness theorem of Hopenhayn and Prescott (1992).

b) Note that sophisticated investors with identical and independent probabilities become either informed or uninformed. Thus, based on Equation 3.9, we only need to show $z^U(z_t^I, z_t^U; \mathcal{F}_t^P)$ accepts a stationary distribution. According to (3.15), this is equivalent to $\frac{w_t^N}{w_t^I}$ being stationary, which is straightforward from part (a).

c) Note that the wealth share of family (f, k) follows the following Markov process:

$$w_{f,t+1}^k = \begin{cases} b_t^k w_{f,t}^k & w.p. \quad \delta \\ b_t^k(1-\tau)w_{f,t}^k + \tau & w.p. \quad 1-\delta \end{cases} \tag{3.A.7}$$

,where

$$b_t^k = \begin{cases} \frac{q_t^k}{\bar{q}_t} & g_{t+1} = g^h \\ \frac{1-q_t^k}{1-\bar{q}_t} & g_{t+1} = g^l \end{cases} \tag{3.A.8}$$

According to part b, the coefficients in (3.A.7) are stationary. Note that the process cannot be divergent with a positive probability, as it would imply unboundedness of the total wealth share for a type, which contradicts with total wealth share being capped by one. That said, Prokhorov's theorem ensures that the wealth share of every family has a stationary distribution.

3.A.8 Proof of Proposition 10

a)

Lemma 13. *For any* $i \in \{1, \ldots, S\}$, *we have* $\mathbb{E}[q_t^I \log(\frac{q_t^i}{\bar{q}_t}) + (1-q_t^I)\log(\frac{1-q_t^i}{1-\bar{q}_t})] \leq -\log(1-\kappa)$, *where* $\kappa \equiv (1-\delta)\tau$.

Proof. Note that Corollary 5 implies the following for the dynamics of wealth share in the hand of type-k investors, $1 \leq i \leq S$:

$$\log w_{t+1}^i - \log w_t^i > (\log(\frac{q_t^i}{\bar{q}_t}))\mathbb{I}_{\{g_{t+1}=g^h\}} + (\log(\frac{1-q_t^i}{1-\bar{q}_t}))\mathbb{I}_{\{g_{t+1}=g^l\}} + \log(1-\kappa)$$

If the inequality in Lemma 13 does not hold for some $i \in \{1, \ldots, S\}$, then the wealth share of the investors with type above i would unboundedly increase, which is in contradiction with the wealth shares being bounded. □

According to (3.A.7) and (3.A.8), we can write wealth share dynamics of family (f, i) in the form of $w^i_{f,t+1} = B^i_t w^i_{f,t} + A^i_t$, where (B^i_t, A^i_t) follows a Markov process. In Lemma 13, we see that $\mathbb{E}[\log B^i_t] \leq 0$. Therefore it is a Kesten process (Kesten (1973)) and its stationary distribution has a thick right tail if $P(B_t > 1^i) > 0$ (Roitershtein et al. (2007)), with the tail parameter $\gamma(i)$ that satisfies:

$$\begin{aligned}
&\lim_{T\to\infty}(\mathbb{E}[\Pi_{t=1}^T B_t^{\gamma(i)}])^{\frac{1}{T}} = \lim_{T\to\infty}\mathbb{E}[\Pi_{t=1}^T (b^i_t)^{\gamma(i)}]^{\frac{1}{T}}(\delta + (1-\delta)(1-\tau)^{\gamma(i)} = 1 \\
&\Rightarrow \lim_{T\to\infty}\mathbb{E}[q^I_t(\frac{q^i_t}{\bar{q}_t})^{\gamma(i)} + (1-q^I_t)(\frac{1-q^i_t}{1-\bar{q}_t})^{\gamma(i)}]^{\frac{1}{T}}(\delta + (1-\delta)(1-\tau)^{\gamma(i)} = 1
\end{aligned} \tag{3.A.9}$$

, where in (3.A.9), I used the fact that $(\tilde{\lambda}_t, z^I_t, z^N_t)$ are drawn independently across the periods and $q^h_t = 1 - q^l_t$. For $\gamma(i)$ that satisfies (3.A.9) we have $P(w^i_{f,t} > \mathbf{w}) > Q(i)\tilde{w}^{-\gamma(i)}$, for some $Q(i) > 0$ and sufficiently large $\mathbf{w}$.

Note that $\gamma(i)$ is decreasing in i, because for any $\gamma > 1$ and $i_2 > i_1$, we have:

$$\begin{aligned}
\mathbb{E}[q^I_t(\frac{q^I_t}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-q^I_t}{1-\bar{q}_t})^{\gamma}] &> \mathbb{E}[q^I_t(\frac{q^U_t}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-q^U_t}{1-\bar{q}_t})^{\gamma}] \\
\Rightarrow \mathbb{E}[(b^{i_2}_t)^{\gamma}] &= F(i_2)\mathbb{E}[q^I_t(\frac{q^I_t}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-q^I_t}{1-\bar{q}_t})^{\gamma}] \\
&+(1-F(i_2))\mathbb{E}[q^I_t(\frac{q^u_t}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-q^U_t}{1-\bar{q}_t})^{\gamma}] \\
> F(i_1)\mathbb{E}[q^I_t(\frac{q^I_t}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-q^I_t}{1-\bar{q}_t})^{\gamma}] &+ (1-F(i_1))\mathbb{E}[q^I_t(\frac{q^u_t}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-q^U_t}{1-\bar{q}_t})^{\gamma}]
\end{aligned}$$

,where the first inequality is achieved from the fact that $q^I_t(\frac{x}{\bar{q}_t})^{\gamma} + (1-q^I_t)(\frac{1-x}{1-\bar{q}_t})^{\gamma}$ is a convex function of γ and the distribution of q^I_t is a mean-preserving spread of that of q^U_t.

Therefore:

$$1 - G^i(\mathbf{w}) = \int_0^1 P(w^i_{f,t} \geq \mathbf{w})di > \int_0^1 Q(i)\mathbf{w}^{-\gamma(i)}di \geq \tilde{Q}\mathbf{w}^{-\gamma(S)} \quad \exists\tilde{Q} > 0$$

where the last inequality is obtained from the fact that $\gamma(S) \leq \gamma(i)$, for $1 \leq i < S$, and Equation 5 in Gabaix (2009). Therefore, $1 - G(\mathbf{w}) \sim \mathbf{w}^{-\gamma(S)}$, which completes the proof.

Part b. By using the derivations in the previous part and applying Bayes' rule, we get:

$$P(\tilde{i} < i | w^{\tilde{i}}_{f,t} > \mathbf{w}) = \frac{\sum_{j=1}^{i-1} b_j P(w^j_{f,t} > \mathbf{w})}{\sum_{j=1}^{S} b_j P(w^j_{f,t} > \mathbf{w})} < \frac{L_1 \mathbf{w}^{-\gamma(i-1)}}{L_2 \mathbf{w}^{-\gamma(S)}} \tag{3.A.10}$$

,where L_1 and L_2 are some positive real numbers. Since $\gamma(S) < \gamma(i-1)$, the statement can be proven by taking $\mathbf{w}$ in (3.A.10) to infinity.

3.A.9 Proof of Proposition 11

Suppose under parameter values $\bar{z}^I, \bar{z}^P, \bar{z}^N$, the induced tail parameter is γ. One can easily verify that the expression inside the expectation in (3.22) is strictly convex in $\bar{q}_t$, where the global minimum is achieved at $\bar{q}_t = q^I_t$.

Note that $\bar{q}_t = q(g^j_t) + \bar{z}^P \nu_t + w^I_t z^I_t + w^U_t z^U_t + w^N_t z^N_t$ and $q^I_t = q(g^j_t) + \bar{z}^P \nu_t + z^I_t$, where $\tilde{\nu_t} \sim U[-1, 1]$. Also note that $w^N_t > \varepsilon \equiv b\tau(1-\delta)$, since ε is the amount transferred to the naive newborns. Now define $\hat{q}_t \equiv q(g^j_t) + \bar{z}^P \nu_t + (1-\varepsilon) z^I_t + \varepsilon z^N_t$. Therefore, due to the abovementioned convexity, one can show:

$$\mathbb{E}[\frac{(q^I_t)^{\gamma+1}}{(\hat{q}_t)^{\gamma}} + \frac{(1-q^I_t)^{\gamma+1}}{(1-\hat{q}_t)^{\gamma}}]M < 1$$

,where $M \equiv \delta + (1-\delta)(1-\tau)^{\gamma}$. Similarly, a sufficient condition that a more informative public signal, with $\bar{z}^P_1 > \bar{z}^P$, induces a smaller tail parameter is that

$$H(\bar{z}^P_1) \equiv \mathbb{E}[\frac{(q(g^j_t) + \bar{z}^P_1 \nu_t + z^I_t)^{\gamma+1}}{(q(g^j_t) + \bar{z}^P_1 \nu_t + (1-\varepsilon) z^I_t + \varepsilon z^N_t)^{\gamma}} + \frac{(1 - (q(g^j_t) + \bar{z}^P_1 \nu_t + z^I_t))^{\gamma+1}}{(1 - (q(g^j_t) + \bar{z}^P_1 \nu_t + (1-\varepsilon) z^I_t + \varepsilon z^N_t))^{\gamma}}]M > 1 \tag{3.A.11}$$

Since by increasing $\bar{z}^P_1$, the denominators in (3.A.11) take values arbitrarily close to zero with a positive probability, then $H(z^P_1)$ can become arbitrarily large. It proves the existence of such $\bar{z}^P_1$.

3.A.10 Proof of Proposition 12

One can show that the expected value in (3.22), which captures the gap in beliefs between the most informed and representative investors, increases with the wealth share of naive investors (w_t^N). In fact, a higher w_t^N makes the representative belief more tilted toward the belief of naive investors, which increases the expected return the informed investors generate. Therefore, we only need to show w_t^N increases with b_N conditional on a given history of shock realizations ($\{g_s, z_s^P, z_s^I, z_s^N, \lambda_s\}_{s=1}^{t-1}$). It can be shown by an induction on t. To see this, when $g_t = g^h$, we have:

$$w_t^N = \frac{w_{t-1}^N q_{t-1}^N (1-\kappa)}{w_{t-1}^N q_{t-1}^N + w_{t-1}^I q_{t-1}^I + w_{t-1}^U q_{t-1}^U} + b_N \kappa$$

,where $\kappa = (1-\delta)\tau$. There are four cases for q_{t-1}^U, depending on the realizations of z_{t-1}^I and z_{t-1}^U. I prove the case $q_{t-1}^U = q_{t-1}^I + \frac{w_{t-1}^N}{w_{t-1}^I} z_{t-1}^U$. The proof for the other cases is similar.

$$q_{t-1}^U = q_{t-1}^I + \frac{w_{t-1}^N}{w_{t-1}^I} z_{t-1}^U \Rightarrow w_t^N = \frac{q_{t-1}^N (1-\kappa)}{q_{t-1}^N + \frac{w_{t-1}^I}{w_{t-1}^N} q_{t-1}^I + w_{t-1}^U \left(\frac{1}{w_{t-1}^N} q_{t-1}^I + \frac{1}{w_{t-1}^I} z_{t-1}^N\right)} + b_N \kappa \tag{3.A.12}$$

Clearly w_t^N is increasing in b_N and w_{t-1}^N for any given realization of z_{t-1}^I, z_{t-1}^N, z_{t-1}^P and λ_{t-1}. Therefore, w_t^N is increasing in b_N given a sequence of shock realizations. It completes the proof.

3.A.11 Proof of Lemma 12

The expected utility of a newborn from choosing to be informed at time t is:

$$\begin{aligned}
V_t^I(w_{f,t}) &= \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \log((1-\beta\delta)\frac{d_{t+j}}{1-\beta\delta}(w_{f,t}-\chi+h_t\chi)\Pi_{j'=0}^{j-1}\frac{q^I_{t+j'}(g_{t+j'+1})}{\bar{q}_{t+j'}(g_{t+j'+1})}) \\
&= \sum_{j=0}^{\infty}(\beta\delta)^j \mathbb{E}_t[\log d_{t+j}] + \frac{1}{1-\beta\delta}\log(w_{f,t}-\chi+h_t\chi) \\
&+ \sum_{m=1}^{\infty}\frac{(\beta\delta)^m}{1-\beta\delta}\mathbb{E}_t[q^I_{t+m-1}\log\frac{q^I_{t+m-1}}{\bar{q}_{t+m-1}} + (1-q^I_{t+m-1})\log\frac{1-q^I_{t+m-1}}{1-\bar{q}_{t+m-1}}|\Omega_t]
\end{aligned} \tag{3.A.13}$$

,where $q^I_{t+j'}(g_{t+j'+1})$ denotes the probability that informed investors assigned at $t+j'$ to the growth rate being high at $t+j'+1$. $\bar{q}_{t+j'}(g_{t+j'+1})$ is similarly defined. Likewise, the expected payoff from remaining uninformed is:

$$\begin{aligned}
V_t^U(w_{f,t}) &= \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \log((1-\beta\delta)\frac{d_{t+j}}{1-\beta\delta}(w_{f,t}+h_t\chi)\Pi_{j'=0}^{j-1}\frac{q^U_{t+j'}(g_{t+j'+1})}{\bar{q}_{t+j'}(g_{t+j'+1})}) \\
&= \sum_{j=0}^{\infty}(\beta\delta)^j \mathbb{E}_t[\log d_{t+j}] + \frac{1}{1-\beta\delta}\log(w_{f,t}+h_t\chi) \\
&+ \sum_{m=1}^{\infty}\frac{(\beta\delta)^m}{1-\beta\delta}\mathbb{E}_t[q^I_{t+m-1}\log\frac{q^U_{t+m-1}}{\bar{q}_{t+m-1}} + (1-q^I_{t+m-1})\log\frac{1-q^U_{t+m-1}}{1-\bar{q}_{t+m-1}}|\Omega_t]
\end{aligned} \tag{3.A.14}$$

Inequality (3.27) can be verified by comparing the equations above.

3.A.12 Proof of Proposition 13

a) From Lemma 9, we have:

$$\begin{aligned}
\mathbb{E}[\log\frac{w^{i_1}_{f_1,t+1}}{w^{i_2}_{f_2,t+1}}] &= \mathbb{E}[\log\frac{w^{i_1}_{f_1,t}}{w^{i_2}_{f_2,t}} + q^I_t\log\frac{q^{i_1}_t}{q^{i_2}_t} + (1-q^I_t)\log\frac{1-q^{i_1}_t}{1-q^{i_2}_t}] \\
= \log\frac{w^{i_1}_{f_1,t}}{w^{i_2}_{f_2,t}} &+ \mathbb{E}[q^I_t\log\frac{q^U_t}{q^I_t} + (1-q^I_t)\log\frac{1-q^U_t}{1-q^I_t}|i_1<\lambda_t\le i_2](F(i_2)-F(i_1))
\end{aligned} \tag{3.A.15}$$

From Lemma 8, we can see $q_t^U|q_t^I, \frac{w_t^N}{w_t^I}$ is uniformly distributed in a subset of $[q(g_t) + z_t^P - \bar{z}^I, q(g_t) + z_t^P + \bar{z}^I]$ $(g_t = g^j)$ and, depending on the realizations, it could have a mass point at the extreme values or $q(g_t) + z_t^P$. To simplify the expositions, suppose the distribution does not have a mass point.[36] Therefore, let $\tilde{q} = q_t^U$ has a uniform distribution, i.e., $\tilde{q}|q_t^I, \frac{w_t^N}{w_t^I} \sim U[\underline{q}, \bar{q}]$, for some $\underline{q}, \bar{q} \in [0, 1]$. In the next step, I find an upper bound for the expression below:

$$\mathcal{H}(q^*; \underline{q}, \bar{q} \equiv \mathbb{E}[q^* \log \frac{\tilde{q}}{q^*} + (1 - q^*) \log \frac{1 - \tilde{q}}{1 - q^*}] \quad \tilde{q} \sim U[\underline{q}, \bar{q}] \tag{3.A.16}$$

By taking the expectation, we get:

$$\begin{gathered}\mathcal{H}(q^*; \underline{q}, \bar{q}) = \frac{1}{\bar{q} - \underline{q}}[q^*(\bar{q} \log \bar{q} - \underline{q} \log \underline{q} - (\bar{q} - \underline{q})) \\ +(1 - q^*)(1 - \underline{q} \log(1 - \underline{q}) - (1 - \bar{q}) \log(1 - \bar{q}) - (\bar{q} - \underline{q}))] \\ -q^* \log q^* - (1 - q^*) \log(1 - q^*)\end{gathered} \tag{3.A.17}$$

Now, by using the Taylor series to approximate $\mathcal{H}(q^*; \underline{q}, \bar{q})$ around q^* up to the third order. After some simplifications, we get:

$$\mathcal{H}(q^*; \underline{q}, \bar{q}) = -\frac{1}{6}(\frac{1}{q^*} + \frac{1}{1 - q^*})[(\bar{q} - q^*)^2 + (q^* - \underline{q})^2 + (\bar{q} - q^*)(q^* - \underline{q})] + o((\max\{\bar{q} - q^*, q^* - \underline{q}\})^3) \tag{3.A.18}$$

According to Lemma 8, the minimum possible value that the bracket above takes is $(\frac{w_t^N}{w_t^I}\bar{z}^N)^2$. Furthermore, note that according to Corollary 5, $w_t^N \geq \kappa b$ for all $t \geq 1$. Therefore, $\frac{w_t^N}{w_t^I} \geq \frac{w_t^N}{1-w_t^N} > \frac{\kappa b_N}{1-\kappa b_N}$. As a result:

$$\mathcal{H}(q^*; \underline{q}, \bar{q}) < -\frac{1}{6}(\frac{\kappa b_N}{1 - \kappa b_N})^2(\frac{1}{q^*} + \frac{1}{1 - q^*})\bar{z}^{N^2} \tag{3.A.19}$$

Similarly, one can show

36. The proof steps are exactly the same for the case with a mass point.

$$\mathcal{H}(q^*;\underline{q},\bar{q}) > -\frac{1}{2}(\frac{1-\kappa b_S}{\kappa b_S})^2(\frac{1}{q^*}+\frac{1}{1-q^*})\bar{z}^{N^2} \tag{3.A.20}$$

By substituting q^* by q_t^I and using expressions (3.A.15), 3.A.19, and (3.A.20), we can complete the proof.

3.A.13 Proof of Proposition 14

Denote the stationary distribution for a given χ by $G_\chi(\cdot)$. Consider the contrary and suppose the tail parameter converges to some $\gamma^* > 1$. Therefore, for sufficiently small values of $\varepsilon > 0$, there exists positive number $A(\varepsilon)$ such that:

$$1 - G_\chi(\mathbf{w}) < A(\varepsilon)\mathbf{w}^{-(\gamma^*-\varepsilon)} \tag{3.A.21}$$

Inequality 3.A.21 implies that the fraction of investors affording to pay the information cost χ goes to zero as χ goes to infinity. Thus, the wealth share of informed investors share should converge to zero. Therefore, the representative belief converges to $q(g_t)$, almost surely. Due to the convergence, according to condition (3.28), there exists a sufficiently large χ and positive number $\mathcal{A}$ such that

$$\mathbb{E}_t[q_t^I \log\frac{q_t^I}{\bar{q}_t} + (1-q_t^I)\log\frac{1-q_t^I}{1-\bar{q}_t}] + (1-\delta)\log(1-\tau) > \mathcal{A} > 0$$

Note that we have the following inequality for the wealth share of an informed investor, like f:

$$\mathbb{E}[\log w_{f,t+1}] - \log w_{f,t} > \delta\mathbb{E}_t[\log\frac{q_t^I(g_{t+1})}{\bar{q}_t(g_{t+1})}] + (1-\delta)\mathbb{E}_t[\log\frac{q_t^I(g_{t+1})}{\bar{q}_t(g_{t+1})}] + (1-\delta)\log(1-\tau)$$
$$= \mathbb{E}_t[q_t^I \log\frac{q_t^I}{\bar{q}_t} + (1-q_t^I)\log\frac{1-q_t^I}{1-\bar{q}_t}] + (1-\delta)\log(1-\tau) > \mathcal{A}$$

It means for such χ, the wealth share of the informed investors boundlessly increases, which is a contradiction.

3.A.14 Proof of Proposition 15

In an economy with no delegation, fraction b^I of informed, and fraction $1 - b_N - b^I$ of uninformed, let $Ew(b^I)^I$ and $Ex(b^I)^U$ be the expected wealth shares of an informed and uninformed investor respectively. The expected wealth share of the fund managers in an economy with fraction b^D of delegators is:

$$b^I Ew(b^I + b^D)^I + b^D \psi(Ew(b^I + b^D)^I - Ew(b^I + b^D)^U) \tag{3.A.22}$$

With delegation, wealth of both informed and delegating investors is managed by informed investors. Thus, their combined equilibrium wealth share , by definition, is $(b^I + b^D)w(b^I + b^D)^I$. The first term in (3.A.22) captures the share the informed investors get by managing their own wealth. The second term is their share from managing the wealth of delegating investors.

Suppose b^I and b_N are sufficiently small, so that the first term in (3.A.22) becomes sufficiently small. Consider two extreme cases of b^D. First, when b^D takes a relatively large value, which implies b^U is close to zero. In this case, an uninformed investor can infer the information of informed investors close to perfectly. Therefore, $Ew(b^I+b^D)^I - Ew(b^I+b^D)^U$, thus (3.A.22) can be arbitrarily small.

When b^D is very small, the informed investors have a substantial information advantage as they manage a small wealth share. In that case the second term in (3.A.22) increases with b^D. So, it can take a positive value larger than the very small value in the previous case. By juxtaposing these two examples, we see the wealth share of informed investors can be non-monotone for some values of b^I and b_N.

3.A.15 Proof of Proposition 16

First, I revise the wealth dynamics provided by Lemma 9 for the modified setup in Lemma 14. The remaining steps are similar to the proof of Proposition 10.

Lemma 14. *The wealth share of investor from family $f \in [0,1]$ and with type $k \in \{1, \ldots, S'\} \cup \{N\}$, if alive at both t and $t+1$, in the modified setup (in Section 3.6.3), is given by:*

$$\log w_{f,t+1}^k - \log w_{f,t}^k = \frac{P^k(g_{t+1}, \varepsilon_{t+1})}{\bar{P}(g_{t+1}, \varepsilon_{t+1})} \tag{3.A.23}$$

The proof strategy is similar to that of Lemma 9.[37]

Appendix 3.B: Additional Discussions and Results

3.B.1 Prominent investors among the super rich

Top wealth shares have increased substantially over the past decades (Saez and Zucman, 2016; Smith, Zidar, Zwick, et al., 2020). However, along with the distribution, the composition of the wealthiest individuals has drastically changed. For instance, Kaplan and Rauh (2013) find the number of fund managers in the Forbes 400 list has increased close to five times between 1982 and 2011. More strikingly, the number of hedge fund managers have increased from two in 1982 to thirty in 2011 (Figure 3.B.1). Among salient distinguishing features of hedge fund managers are their provision of wealth management services and direct access to the state-of-the-art technologies for information production and processing. By linking asymmetric information to wealth distribution, my model explains the presence of highly informed investors among the wealthiest. It further implies that the fund managers' wealth share increases with the emergence of new and expensive information-production technologies and rise of investment delegations, justifying the pattern in Figure 3.B.1.

3.B.2 Expanding the Security Space

So far, we have assumed that the markets are dynamically incomplete, meaning that the set of available securities do not span the whole space of potential contingent transfers. For instance, a log-term risk-free bond cannot be manufectured with the one-period risk-free bonds and the shares of the tree. A natural question is how expanding the set of securities

37. Similar to Blume and Easley (1992), one can show Equation 3.A.23 holds if the number of payoff relevant states does not exceed the number of independent assets when economic agents have logarithmic preferences.

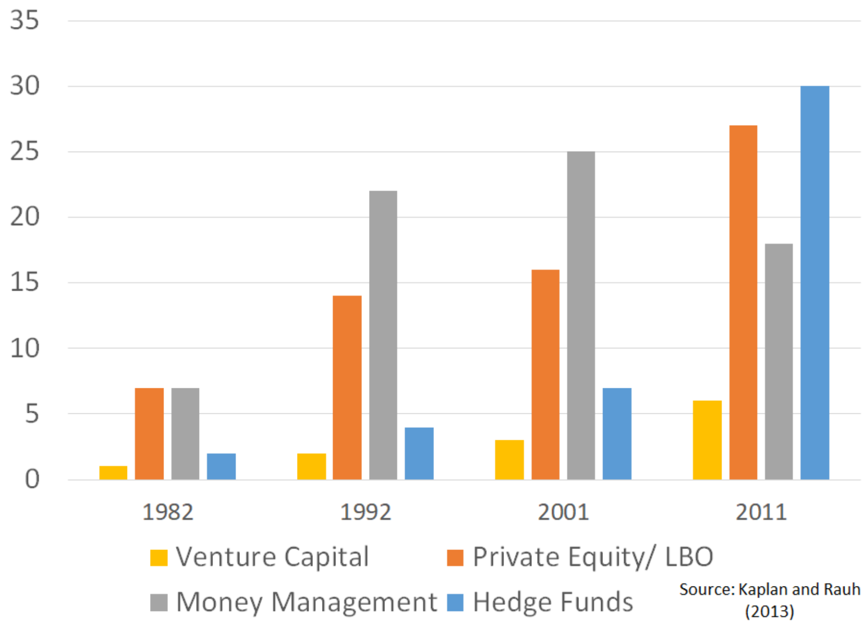


Figure 3.B.1: Number of fund managers in the Forbes 400 list of the US wealthiest individuals.

impact the allocations, and more specfically, the wealth distribution. Proposition 17 shows that the expansion has no impact on the wealth distribution.

Proposition 17. *In the baseline setup described in Section 3.3, suppose the markets are dynamically complete, keeping the information structure unchanged. Then, Equations 3.19 and 3.A.6 still specify the wealth dynamics.*

Proof. To prove the proposition, I show for any sequence of growth rate realizations $\mathcal{G} = (g^*_{t+1}, g^*_{t+2}, \ldots, g^*_{t+N})$, all investors assign the same price to the security that pays ε at beggining of $t+N$ following $\mathcal{G}$, under the allocation provided for the baseline case. In particular, we need to show:

$$(\beta\delta)^N \mathbb{E}^k[\frac{u'(c(\mathcal{G}))}{c_t}\mathbb{I}_{\{(g_{t+1},g_{t+2},\ldots,g_{t+N})=\mathcal{G}\}}] = (\beta\delta)^N \mathbb{E}^{k'}[\frac{u'(c(\mathcal{G}))}{c_t}\mathbb{I}_{\{(g_{t+1},g_{t+2},\ldots,g_{t+N})=\mathcal{G}\}}], \quad \forall k, k' \in [0,1] \cup \{N\} \tag{3.B.1}$$

,where $c(g_{t+1}, g_{t+2}, \ldots, g_{t+K})$ denotes the consumption under the baseline allocation after the realization $(g_{t+1}, g_{t+2}, \ldots, g_{t+K})$ and $u(\cdot) \equiv \log(\cdot)$. In the baseline, the allocations between any two consecutive periods are optimal, given the information available at the time of consumption and portfolio decision. Therefore, for any $1 \leq j \leq N-1$ and any type of investor $k \in [0,1] \cup \{N\}$, we have:

$$\beta\delta \frac{u'(c(\ldots, g^*_{t+j}, g^*_{t+j+1}))}{u'(c(\ldots, g^*_{t+j}))} P^k_{t+j}(\tilde{g}_{t+j+1} = g^*_{t+j+1}) = 1 \tag{3.B.2}$$

Now, by multiplying all of these Euler equations, we get:

$$(\beta\delta)^N \frac{u'(c(\mathcal{G}))}{u'(c(g^*_{t+1}))} \Pi_{j=1}^{N-1} P^k_{t+j}(\tilde{g}_{t+j+1} = g^*_{t+j+1}) = 1 \tag{3.B.3}$$

After taking time t expectation, we have:

$$(\beta\delta)^N \frac{u'(c(\mathcal{G}))}{u'(c(g^*_{t+1}))} \Pi_{j=1}^{N-1} P(\tilde{g}_{t+j+1} = g^*_{t+j+1} | g^*_{t+j}) = 1 \tag{3.B.4}$$

By multiplying the last equation with the Euler equation at time t, we have:

$$(\beta\delta)^N \frac{u'(c(\mathcal{G}))}{u'(c_t)} P^k_t(\tilde{g}_{t+1} = g^*_{t+1}) \Pi_{j=1}^{N-1} P(\tilde{g}_{t+j+1} = g^*_{t+j+1} | g^*_{t+j}) = (\beta\delta)^N \frac{u'(c(\mathcal{G}))}{u'(c_t)} P^k_t(\mathcal{G}) = 1 \tag{3.B.5}$$

Equation 3.B.5 shows that all investors assign the same price to the security that pays contingent on $\mathcal{G}$. Since this security is in zero net supply, it would not be traded. Therefore, the dynamic incompleteness imposed in the baseline does not change the allocations.

□

Proposition 17 states expanding the set of securities does not change the wealth distribution and its dynamics. The intuition is as follows: Standing at time t, all investors assign the same probability to the events after time $t+1$, once it is conditioned on the next period's growth, g_{t+1}. In other words, the only state that they have disagreement about is g_{t+1}. In fact, in the proof, I show that all investors price all long-term assets (the ones that their value depends on a subset of $g_{t+2}, g_{t+3}, \ldots$) in the same way since all investors agree on the relative prices between any two consecutive periods. By combining this property in the baseline allocations and the abovementioned point, I conclude that all investors should assign the same price to all long-term assets, given the baseline allocation. Since the long-term assets are in zero net supply, except the tree, they are not traded in equilibrium. Therefore, the allocations remain intact.

3.B.3 Bequest Motives

An extensive literature highlights the role of bequest motive in wealth inequality (De Nardi (2004)). In this section, I allow the investors to have bequest motive for their intertemporal decisions. In short, I find the wealth dynamics remain unchanged if the investors are homogeneous in their bequest motive, that is all investors trade off their own and their child expected utility similarly. The intuition being that it impacts the consumption and saving decision of all investors in the same way and it does not affect the portfolio compositions.

I modify the utility function in (3.3) by adding a term $\Phi(\cdot)$ capturing the bequest motive. Equation (3.B.6) presents the modified utility function. In fact, $\Phi(W)$ is the utility an investor gets from leaving W units of consumption goods, before tax, to his child.[38]

$$U(\{c^k_{f,t}\}_{t=s_1}^{s_2-1}, W_{s_2}) = \sum_{t=s_1}^{s_2-1} \beta^{t-s_1} \log c^k_{f,t} + \beta^{s_2-s_1}\Phi(W_{s_2}) \qquad k \in [0,1] \cup \{N\} \qquad (3.B.6)$$

To maintain tractability, I assume Φ is logarithmic in the bequest wealth, i.e. $\Phi(W) = \phi \log W$. The following proposition shows how the bequest motives changes the consumption and saving behaviors and wealth dynamics.

Proposition 18. ***a)*** *For every $t \geq 1$, all investors consume fraction $\frac{1-\beta\delta}{1+\beta\phi(1-\delta)}$ of their wealth, i.e. $c^k_{f,t} = \frac{1-\beta\delta}{1+\beta\phi(1-\delta)} W^k_{f,t}$, where $f \in [0,1]$ and $k \in [0,1] \cup \{N\}$.*

b) *The distribution of the beliefs and evolution of wealth shares, specified in (3.19), are the same for all real values of ϕ.*

Proof. For the proof, we only need to incorporate the bequest motive into the baseline value function, specified in (3.A.1). Similar to the proof of Proposition 5, we can guess and verify

38. Note that there is a one-to-one mapping between before-tax and after-tax inherited wealth.

the value function is logarithmic in wealth and is in the following form:

$$V_{\phi}^{k}(W_t;\mathcal{F}_t) = \frac{1+\beta(1-\delta)\phi}{1-\beta\delta}\log W_t + a_{\phi}^{k}(\mathcal{F}_t) \tag{3.B.7}$$

The optimal consumption can be similarly solved and verified that $c_t = \frac{1-\beta\delta}{1+\beta\phi(1-\delta)}$. Therefore, similar to the baseline case, the investors optimally consume a fixed fraction of their wealth in every period. Moreover, due to the logarithmic form of the value function, the investors optimally maximize the expected log-return of their investment, similar to the baseline case. Therefore, a change in ϕ does not affect the composition of portfolios, and as a result, does not impact the information dissemination through the prices. Therefore, by following the procedure carried out in the baseline case, we can find the wealth and belief dynamics and prove part (b).

□

Proposition 18(a) shows that the bequest motive changes the saving behavior of all investors uniformly. It means as ϕ increases, all investors scale their savings at the same proportion. This increase in demand for saving pushes up the asset prices, and subsequently, lowers the expected returns. Overall, this general equilibrium effect fully offsets the higher demand for saving, which renders ϕ irrelevant for the wealth distribution. This point is formalized in Part (b) of the proposition.

Note that Proposition 18 states the bequest motive has no role in wealth distribution when there is no heterogeneity in this motive for saving. However, a heterogeneity in the bequest motive can exacerbate or ameliorate wealth inequality, which is relegated to future studies.

Imperfect Type Inheritance

Data suggests billionaires' descendants are not as good as their fathers in wealth creation and maintenance. In fact, there is a tendency to mean in the average returns that different family generations get and this mean-reversion has amplified over the past decades. For example Kaplan and Rauh (2013) find that only one-third of the richest individuals in Forbes 400 list at 2011 have grown wealthy, compared to two-third at 1988. Relatedly, Fagereng, Guiso, Malacrino, and Pistaferri (2020) find an economically small intergenerational correlation in returns to financial wealth and net worth. These findings suggest that the investment sophistication is imperfectly transmitted to the next generations.

Motivated by these findings, in this section, I analyze a modification of the model in which the newborns imperfectly inherit their parents type. Particularly, suppose there are only two types of sophisticated investors: informed and uninformed, similar to Section 3.6.1. However, a newborn inherits the type of his parent with probability $\rho \in (0, 1)$. Therefore, the child of an informed investor is also informed with probability ρ and becomes uninformed with probability $1 - \rho$. The following proposition shows the wealth inequality is larger for higher values of ρ.

Proposition 19. *The tail parameter $\gamma(\rho)$ is at least weakly decreasing in ρ.*

Proof. Note that the tail parameter is the solution to the following equation:

$$\lim_{T\to\infty} \mathbb{E}[(\frac{q_{f,t}(g_{t+1})}{\bar{q}(g_{t+1})})^{\gamma}]^{\frac{1}{T}}(\delta + (1-\delta)(1-\tau)^{\gamma}) = 1 \tag{3.B.8}$$

,where $q_{f,t}$ is the probability that the investor from family f assigned on the realized growth g_{t+1} at time t.[39] A higher persistence in types means that consecutive terms in sequence $\{q_{f,t}\}_{t=1}^{\infty}$ have a higher correlation. Therefore, one can show that the expectation of the consecutive terms in the sequence, $\mathbb{E}(\Pi_{j=0}^{T} q_{f,t+j}(g_{t+j+1}))$, is increasing in ρ by applying law of iterated expectation and noting the wealth share of the informed investors increases with ρ.

A higher ρ implies that the informed investors can save more and, eventually, claim a larger fraction of the wealth, which implies the representative belief becomes more accurate. It leads to an increase in $\mathbb{E}(\Pi_{j=0}^{T} \bar{q}_{t+j}(g_{t+j+1}))$ for every T. However, if the increase is so large that $\mathbb{E}[\Pi_{j=0}^{T} \frac{q_{f,t+j}(g_{t+j+1})}{\bar{q}_{t+j}(g_{t+j+1})}]$ decreases, it means the expected wealth share of the sophisticated investors should decrease in ρ, or equivalently, the wealth share of the naive investors should increase with ρ, which is not possible.

□

39. Note that, in this case, the investors are homogeneous, ex ante.

Higher persistence in the sophistication types increases the inequality since it amplifies the compounding effect, that is the ones with a higher return today are more likely to get a higher return tomorrow, which further increases the inequality. This result is another indication of the importance of education in reducing inequality. Chetty, Friedman, Saez, Turner, and Yagan (2017) find that after students attending the same college land in relatively similar income percentiles, regardless of their parents' income status. It underscores the role of higher education in social mobility.

3.B.4 A Calibration Exercise

In this section, I empirically assess the role of asymmetric information in wealth inequality. Clearly, asymmetric information impacts wealth inequality through investors' heterogeneous choice of portfolio. Therefore, it has implications for both portfolio dynamics and the degree of wealth inequality, which I examine their consistency with data. In particular, the Survey of Customer Finance contains information on risky asset holdings by the top 0.1% wealthiest in the US. I find under some reasonable parameters, the numbers generated by the model with multiple risky assets (presented in Section 3.6.3) match well with the average and variance of the share of risky assets in the financial wealth for the top households.

Table 3.B.1 provides the results for the selected parameter values. Assuming only the top 0.1% are informed, we see a set of parameters can generate the same expected top wealth share and generate patterns in the portfolio dynamics arguably similar to data, in terms of average and volatility. The average fraction in the model is smaller than in data, which indicates the model overestimates the risk-aversion of the top 0.1%. Interestingly, we see the model requires a low degree of asymmetric information about the growth state, and a high degree of asymmetric information regarding the growth-independent state to match the data.

The key message of this section is that the empirical patterns on the extent of wealth inequality and portfolio dynamics are consistent with the model predictions. The model suggests that the patterns observed in data correspond to a low degree of asymmetric information about future growth rates and a high degree of asymmetric information about some growth-independent states.

Parameters	Symbol	Value	Data
Discount rate	β	0.95	
Tax rate	τ	0.4	
Survival rate	δ	0.97	
Growth rates	g^h, g^l	3%,1%	
Transition probabilities	$q(g^h), q(g^l)$	0.8,0.4	
Measure of naives	b	0.2	
Precision of informative signal for growth rate	$\bar{z}^I$	0.0105	
Noise magnitude for growth rate	$\bar{z}^N$	0.0008	
Precision of informative signal for growth-unrelated states	$\bar{y}^I$	0.49	
Noise magnitude for growth-unrelated states	$\bar{y}^N$	0.49	
Outcomes			
Top 0.1% wealth share	$\mathbb{E}[w_t^I]$	0.15	0.15
Average fraction invested in risky	$\mathbb{E}[1-\mu_t^{IF}]$	1	1.15
Std of fraction invested in risky	$Std[1-\mu_t^{IF}]$	0.17	0.15

Table 3.B.1: Information-related parameters and outcome values. It is assumed that the top 0.1% are always informed about z_t^I and y_t^I, and the rest is always uninformed. μ_t^{IF} is the fraction of their financial assets invested in riskless assets.

References

Abel, Andrew B, Janice C Eberly, and Stavros Panageas, 2013, Optimal inattention to the stock market with information costs and transactions costs, Econometrica 81, 1455–1481.

Adelino, Manuel, Igor Cunha, and Miguel A Ferreira, 2017, The economic effects of public financing: Evidence from municipal bond ratings recalibration, The Review of Financial Studies 30, 3223–3268.

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey, 1994, Renegotiation design with unverifiable information, Econometrica: Journal of the Econometric Society 62, 257–282.

Agrawal, Ashwini K, and Daniel Kim, 2021, Municipal bond insurance and the us drinking water crisis, Available at SSRN 3813348.

Aiyagari, S Rao, 1994, Uninsured idiosyncratic risk and aggregate saving, The Quarterly Journal of Economics 109, 659–684.

An, Li, Jiangze Bian, Dong Lou, and Donghui Shi, 2019, Wealth redistribution in bubbles and crashes, .

Anand, Bharat N, and Alexander Galetovic, 2006, Relationships, competition and the structure of investment banking markets, Journal of Industrial Economics 54, 151–199.

Azarmsa, Ehsan, 2021, Asset pricing with heterogeneously attentive investors, Available at https://drive.google.com/file/d/1lWAkMHfR9HR6A9TkKa2kA8XgcGdSc6kF/view.

Babina, Tania, Chotibhak Jotikasthira, Christian Lundblad, and Tarun Ramadorai, 2021, Heterogeneous taxes and limited risk sharing: Evidence from municipal bonds, The review of financial studies 34, 509–568.

Bach, Laurent, Laurent E Calvet, and Paolo Sodini, 2016, Rich pickings? risk, return, and skill in the portfolios of the wealthy, .

Barber, Brad M, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2014, The cross-section of speculator skill: Evidence from day trading, Journal of Financial Markets 18, 1–24.

Bartscher, Alina K, Moritz Kuhn, and Moritz Schularick, 2018, The college wealth divide: Education and inequality in america, 1956-2016, .

Basak, Suleyman, and Domenico Cuoco, 1998, An equilibrium model with restricted stock market participation, The Review of Financial Studies 11, 309–341.

Benhabib, Jess, and Alberto Bisin, 2018, Skewed wealth distributions: Theory and empirics, Journal of Economic Literature 56, 1261–91.

———, and Mi Luo, 2015, Wealth distribution and social mobility in the us: A quantitative approach, Discussion paper, National Bureau of Economic Research.

———, 2019, Wealth distribution and social mobility in the us: A quantitative approach, American Economic Review 109, 1623–47.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu, 2011, The distribution of wealth and fiscal policy in economies with finitely lived agents, Econometrica 79, 123–157.

Bergemann, Dirk, and Stephen Morris, 2016, Information design, Bayesian persuasion, and Bayes correlated equilibrium, American Economic Review 106, 586–591.

Berger, Allen N, and Gregory F Udell, 1995, Relationship lending and lines of credit in small firm finance, Journal of Business 68, 351–381.

Bergstresser, Daniel, and Randolph Cohen, 2016, Changing patterns in household ownership of municipal debt, Hutchins Center Working Papers.

Black, Sandra E, and Philip E Strahan, 2002, Entrepreneurship and bank credit availability, Journal of Finance 57, 2807–2833.

Blackwell, David, 1953, Equivalent comparisons of experiments, The Annals of Mathematical Statistics 24, 265–272.

Blume, Lawrence, and David Easley, 1992, Evolution and market behavior, Journal of Economic theory 58, 9–9.

Bond, Philip, and Diego Garcia, 2018, The equilibrium consequences of indexing, Discussion paper, Working paper.

Boot, Arnoud WA, and Anjan V Thakor, 2000, Can relationship banking survive competition?, Journal of Finance 55, 679–713.

Bouvard, Matthieu, Pierre Chaigneau, and Adolfo De Motta, 2015, Transparency in the financial system: rollover risk and crises, Journal of Finance 70, 1805–1837.

Boyer, Charles, 2020, Public pensions and state government borrowing costs, Ph.D. thesis The University of Chicago.

Brennan, Michael J, and Eduardo S Schwartz, 1988, The case for convertibles, Journal of Applied Corporate Finance 1, 55–64.

Cagetti, Marco, and Mariacristina De Nardi, 2006, Entrepreneurship, frictions, and wealth, Journal of political Economy 114, 835–870.

Campbell, John Y, Tarun Ramadorai, and Benjamin Ranish, 2018, Do the rich get richer in the stock market? evidence from india, Discussion paper, National Bureau of Economic Research.

Castaneda, Ana, Javier Diaz-Gimenez, and Jose-Victor Rios-Rull, 2003, Accounting for the us earnings and wealth inequality, Journal of political economy 111, 818–857.

Cestau, Dario, Burton Hollifield, Dan Li, and Norman Schürhoff, 2019, Municipal bond markets, Annual Review of Financial Economics 11, 65–84.

Chang, Briana, and Martin Szydlowski, 2020, The market for conflicted advice, Journal of Finance 75(2).

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, 2017, Mobility report cards: The role of colleges in intergenerational mobility, Discussion paper, national bureau of economic research.

Chien, YiLi, Harold Cole, and Hanno Lustig, 2012, Is the volatility of the market price of risk due to intermittent portfolio rebalancing?, American Economic Review 102, 2859–96.

———, 2016, Implications of heterogeneity in preferences, beliefs and asset trading technologies in an endowment economy, Review of Economic Dynamics 20, 215–239.

Chodorow-Reich, Gabriel, Plamen T Nenov, and Alp Simsek, 2021, Stock market wealth and the real economy: A local labor market approach, American Economic Review 111, 1613–57.

Clark, Robert, Annamaria Lusardi, and Olivia S Mitchell, 2017, Financial knowledge and 401 (k) investment performance: a case study, Journal of Pension Economics & Finance 16, 324–347.

Cong, Lin William, 2017, Auctions of real options, Unpublished working paper. Chicago Booth School of Business.

———, Steven R Grenadier, and Yunzhi Hu, 2020, Dynamic interventions and informational linkages, Journal of Financial Economics 135, 1–15.

Cong, Lin William, and Yizhou Xiao, 2018, Information cascades and threshold implementation, Unpublished Working Paper. The Chinese University of Hong Kong and University of Chicago.

Cornaggia, Jess, Kimberly J Cornaggia, and Ryan D Israelsen, 2018, Credit ratings and the cost of municipal financing, The Review of Financial Studies 31, 2038–2079.

Cornelli, Francesca, and Oved Yosha, 2003, Stage financing and the role of convertible securities, Review of Economic Studies 70, 1–32.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, Journal of Financial Economics 86, 479–512.

Dagostino, Ramona, 2018, The impact of bank financing on municipalities' bond issuance and the real economy, Work. Pap., Univ. Rochester, Rochester, NY Google Scholar Article Location.

Das, Sreyoshi, Camelia M Kuhnen, and Stefan Nagel, 2017, Socioeconomic status and macroeconomic expectations, Discussion paper, National Bureau of Economic Research.

De Nardi, Mariacristina, 2004, Wealth inequality and intergenerational links, The Review of Economic Studies 71, 743–768.

Degryse, Hans, and Steven Ongena, 2007, The impact of competition on bank orientation, Journal of Financial Intermediation 16, 399–424.

Dell'Ariccia, Giovanni, and Robert Marquez, 2004, Information and bank credit allocation, Journal of Financial Economics 72, 185–214.

DeMarzo, Peter, and Darrell Duffie, 1999, A liquidity-based model of security design, Econometrica 67, 65–99.

DeMarzo, Peter M, Ilan Kremer, and Andrzej Skrzypacz, 2005, Bidding with securities: auctions and security design, American Economic Review 95, 936–959.

Diamond, Douglas W, 1984, Financial intermediation and delegated monitoring, The review of economic studies 51, 393–414.

——— , 1991, Monitoring and reputation: the choice between bank loans and directly placed debt, Journal of Political Economy 99, 689–721.

——— , and Robert E Verrecchia, 1981, Information aggregation in a noisy rational expectations economy, Journal of financial economics 9, 221–235.

Dinc, I Serdar, 2000, Bank reputation, bank commitment, and the effects of competition in credit markets, Review of Financial Studies 13, 781–812.

Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, The Journal of finance 65, 1237–1267.

——— , and Kenneth J Singleton, 1999, Modeling term structures of defaultable bonds, The review of financial studies 12, 687–720.

Dworczak, Piotr, and Giorgio Martini, 2019, The simple economics of optimal persuasion, Journal of Political Economy 127, 1993–2048.

Elsas, Ralf, 2005, Empirical determinants of relationship lending, Journal of Financial Intermediation 14, 32–57.

Ely, Jeffrey C, 2017, Beeps, American Economic Review 107, 31–53.

Ewens, Michael, Matthew Rhodes-Kropf, and Ilya A Strebulaev, 2016, Insider financing and venture capital returns, Unpublished working paper. California Institute of Technology, Harvard Business School, and Stanford University.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri, 2020, Heterogeneity and persistence in returns to wealth, Econometrica 88, 115–170.

Fama, Eugene F, 1970, Efficient capital markets: A review of theory and empirical work, The journal of Finance 25, 383–417.

——— , 1985, What's different about banks?, Journal of Monetary Economics 15, 29–39.

Fluck, Zsuzsanna, Kedran Garrison, and Stewart C Myers, 2006, Venture capital contracting: staged financing and syndication of later-stage investments, Unpublished working paper. National Bureau of Economic Research.

Gabaix, Xavier, 2009, Power laws in economics and finance, Annu. Rev. Econ. 1, 255–294.

——— , and Ralph SJ Koijen, 2020, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Available at SSRN 3686935.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, 2016, The dynamics of inequality, Econometrica 84, 2071–2111.

Gargano, Antonio, and Alberto G Rossi, 2018, Does it pay to pay attention?, The Review of Financial Studies 31, 4595–4649.

Gentzkow, Matthew, and Emir Kamenica, 2016, A Rothschild-Stiglitz approach to Bayesian persuasion, American Economic Review 106, 597–601.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five facts about beliefs and portfolios, American Economic Review 111, 1481–1522.

Girshina, Anastasia, 2019, Education and wealth accumulation: Evidence from sweden, Available at SSRN 3440282.

Goldstein, Itay, and Yaron Leitner, 2018, Stress tests and information disclosure, Journal of Economic Theory 177, 34–69.

Gomez, Matthieu, 2019, Asset prices and wealth inequality, Unpublished paper: Princeton.

Gompers, Paul A, 1997, Ownership and control in entrepreneurial firms: an examination of convertible securities in venture capital investments, Unpublished working paper. Harvard Business Schools.

——— , Will Gornall, Steven N Kaplan, and Ilya A Strebulaev, 2020, How do venture capitalists make decisions?, Journal of Financial Economics 135, 169–190.

Gormsen, Niels Joachim, and Eben Lazarus, 2019, Duration-driven returns, Available at SSRN 3359027.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa, 2011, Iq and stock market participation, The Journal of Finance 66, 2121–2164.

Grinblatt, Mark, Matti Keloharju, and Juhani T Linnainmaa, 2012, Iq, trading behavior, and performance, Journal of Financial Economics 104, 339–362.

Grossman, Sanford, 1976, On the efficiency of competitive stock markets where trades have diverse information, The Journal of finance 31, 573–585.

Grossman, Sanford J, and Robert J Shiller, 1982, Consumption correlatedness and risk measurement in economies with non-traded assets and heterogeneous information, Journal of financial Economics 10, 195–210.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, The American economic review 70, 393–408.

Guo, Yingni, and Eran Shmaya, 2019, The interval structure of optimal disclosure, Econometrica 87, 653–675.

Guvenen, Fatih, 2009, A parsimonious macroeconomic model for asset pricing, Econometrica 77, 1711–1750.

Hart, Oliver, and John Moore, 1988, Incomplete contracts and renegotiation, Econometrica: Journal of the Econometric Society 56, 755–785.

——— , 1995, Debt and seniority: an analysis of the role of hard claims in constraining management, American Economic Review 85, 567–585.

Haughwout, Andrew, Benjamin Hyman, and Or Shachar, 2021, The option value of municipal liquidity: Evidence from federal lending cutoffs during covid-19, Available at SSRN 3785577.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, American Economic Review 103, 732–70.

Hellmann, Thomas, 2006, IPOs, acquisitions, and the use of convertible securities in venture capital, Journal of Financial Economics 81, 649–679.

Holmstrom, Bengt, 1979, Moral hazard and observability, Bell Journal of Economics 10, 74–91.

Hopenhayn, Hugo A, and Edward C Prescott, 1992, Stochastic monotonicity and stationary distributions for dynamic economies, Econometrica (1986-1998) 60, 1387.

Hubmer, Joachim, Per Krusell, and Anthony A Smith Jr, 2020, Sources of us wealth inequality: Past, present, and future, in NBER Macroeconomics Annual 2020, volume 35 (University of Chicago Press).

Hull, John C, Andrew W Lo, and Roger M Stein, 2019, Funding long shots, Journal of Investment Management 17.

Inderst, Roman, and Vladimir Vladimirov, 2019, Growth firms and relationship finance: A capital structure perspective, Management Science 65, 4951–5448.

Innes, Robert D, 1990, Limited liability and incentive contracting with ex-ante action choices, Journal of Economic Theory 52, 45–67.

Inostroza, Nicolas, and Alessandro Pavan, 2018, Persuasion in global games with application to stress testing, Unpublished working paper. Northwestern University.

Ito, Kazufumi, 2016, Functional analysis and optimization, Unpublished working paper. North Carolina State University.

Jorring, A, 2018, The costs of financial mistakes: Evidence from us consumers, Discussion paper, Working Paper.

Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens, 2018, Investor sophistication and capital income inequality, Journal of Monetary Economics.

Kamenica, Emir, and Matthew Gentzkow, 2011, Bayesian persuasion, American Economic Review 101, 2590–2615.

Kaplan, Steven N, and Joshua D Rauh, 2013, Family, education, and sources of wealth among the richest americans, 1982-2012, American Economic Review 103, 158–62.

Kaplan, Steven N., and Per Strömberg, 2003, Financial contracting theory meets the real world: an empirical analysis of venture capital contracts, Review of Economic Studies 70, 281–315.

Kaplan, Steven N, and Per ER Strömberg, 2004, Characteristics, contracts, and actions: evidence from venture capitalist analyses, Journal of Finance 59, 2177–2210.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf, 2014, Entrepreneurship as experimentation, Journal of Economic Perspectives 28, 25–48.

Kesten, Harry, 1973, Random difference equations and renewal theory for products of random matrices, Acta Mathematica 131, 207–248.

Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, Journal of Political Economy 127, 1475–1515.

Kolotilin, Anton, 2018, Optimal information disclosure: a linear programming approach, Theoretical Economics 13, 607–635.

——— , Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li, 2017, Persuasion of a privately informed receiver, Econometrica 85, 1949–1964.

Korniotis, George M, and Alok Kumar, 2013, Do portfolio distortions reflect superior information or psychological biases?, Journal of Financial and Quantitative Analysis 48, 1–45.

Kyle, Albert S, 1985, Continuous auctions and insider trading, Econometrica: Journal of the Econometric Society pp. 1315–1335.

Lei, Xiaowen, 2019, Information and inequality, Journal of Economic Theory.

Li, Dan, and Norman Schürhoff, 2019, Dealer networks, The Journal of Finance 74, 91–144.

Li, Jiacui, 2021, What drives the size and value factors?, Available at SSRN 2909960.

Li, Jessica, and Rayhan Momin, 2020, The causal effect of the fed's corporate credit facilities on eligible issuer bonds, Banking & Insurance eJournal.

Lou, Dong, 2012, A flow-based explanation for return predictability, The Review of Financial Studies 25, 3457–3489.

Lou, Xiaoxia, and Albert Y Wang, 2018, Flow-induced trading pressure and corporate investment, Journal of Financial and Quantitative Analysis 53, 171–201.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S Mitchell, 2017, Optimal financial knowledge and wealth inequality, Journal of Political Economy 125, 431–477.

Mailath, George J, and Alvaro Sandroni, 2003, Market selection and asymmetric information, The Review of Economic Studies 70, 343–368.

Megginson, William L, and Kathleen A Weiss, 1991, Venture capitalist certification in initial public offerings, Journal of Finance 46, 879–903.

Mian, Atif, and Amir Sufi, 2014, What explains the 2007–2009 drop in employment?, Econometrica 82, 2197–2223.

Mihet, Roxana, 2019, Who benefits from innovations in financial technology?, Available at SSRN 3474720.

Milgrom, Paul, and Nancy Stokey, 1982, Information, trade and common knowledge, Journal of economic theory 26, 17–27.

Myers, Sean, 2019, Public employee pensions and municipal insolvency, Discussion paper, Working Paper.

Myers, Thomas E, and Douglas E Goodfriend, 2009, Bond Basics for Counties in New York State (Orrick, Herrington Sutcliffe LLP).

Nachman, David C, and Thomas H Noe, 1994, Optimal design of securities under asymmetric information, Review of Financial Studies 7, 1–44.

Nanda, Ramana, and Matthew Rhodes-Kropf, 2013, Innovation and the financial guillotine, Unpublished working paper. National Bureau of Economic Research.

Nöldeke, Georg, and Klaus M Schmidt, 1998, Sequential investments and options to own, RAND journal of Economics 29, 633–653.

Orlov, Dmitry, Pavel Zryumov, and Andrzej Skrzypacz, 2019, Design of macro-prudential stress tests, Unpublished working paper. Stanford University.

Panageas, Stavros, 2020, The implications of heterogeneity and inequality for asset pricing, Discussion paper, National Bureau of Economic Research.

Pástor, Luboš, and Pietro Veronesi, 2016, Income inequality and asset prices under redistributive taxation, Journal of Monetary Economics 81, 1–20.

Pastor, Lubos, and Pietro Veronesi, 2018, Inequality aversion, populism, and the backlash against globalization, Discussion paper, National Bureau of Economic Research.

Peng, Cameron, and Chen Wang, 2019, Factor demand and factor returns, Available at SSRN 3327849.

Peress, Joel, 2003, Wealth, information acquisition, and portfolio choice, The Review of Financial Studies 17, 879–914.

Petersen, Mitchell A, and Raghuram G Rajan, 1994, The benefits of lending relationships: evidence from small business data, Journal of Finance 49, 3–37.

——— , 1995, The effect of credit market competition on lending relationships, Quarterly Journal of Economics, 110, 407–443.

——— , 2002, Does distance still matter? The information revolution in small business lending, Journal of Finance 57, 2533–2570.

Piketty, Thomas, 2014, Capital in the twenty-first century (Harvard University Press).

Rajan, Raghuram G, 1992, Insiders and outsiders: The choice between informed and arm's-length debt, The Journal of Finance 47, 1367–1400.

Ramakrishnan, Ram TS, and Anjan V Thakor, 1984, Information reliability and a theory of financial intermediation, Review of Economic Studies 51, 415–432.

Roitershtein, Alexander, et al., 2007, One-dimensional linear recursions with markov-dependent coefficients, The Annals of Applied Probability 17, 572–608.

Saez, Emmanuel, and Gabriel Zucman, 2016, Wealth inequality in the united states since 1913: Evidence from capitalized income tax data, The Quarterly Journal of Economics 131, 519–578.

Santos, Joao AC, and Andrew Winton, 2008, Bank loans, bonds, and information monopolies across the business cycle, Journal of Finance 63, 1315–1359.

Schenone, Carola, 2010, Lending relationships and information rents: do banks exploit their information advantages?, Review of Financial Studies 23, 1149–1199.

Schultz, Paul, 2013, State taxes, limits to arbitrage and differences in municipal bond yields across states, Unpublished working paper. University of Notre Dame.

Schwert, Michael, 2017, Municipal bond liquidity and default risk, The Journal of Finance 72, 1683–1722.

Sharpe, Steven A, 1990, Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships, The journal of finance 45, 1069–1087.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, The Journal of Finance 41, 579–590.

Smith, Matthew, Owen Zidar, Eric Zwick, et al., 2020, Top wealth in america: New estimates and implications for taxing the rich, Discussion paper, .

Stein, Jeremy C, 1992, Convertible bonds as backdoor equity financing, Journal of Financial Economics 32, 3–21.

Szydlowski, Martin, 2019, Optimal financing and disclosure, Management Science, forthcoming.

Toda, Alexis Akira, and Kieran James Walsh, 2014, The equity premium and the one percent, The Review of Financial Studies.

Trigilia, Giulio, 2019, Optimal leverage, transparency, and the decision to go public, Unpublished working paper. University of Warwick.

Van Binsbergen, Jules, Wouter Hueskes, Ralph Koijen, and Evert Vrugt, 2013, Equity yields, Journal of Financial Economics 110, 503–519.

Van Binsbergen, Jules H, and Ralph SJ Koijen, 2017, The term structure of returns: Facts and theory, Journal of Financial Economics 124, 1–21.

Vayanos, Dimitri, 1999, Strategic trading and welfare in a dynamic market, The Review of Economic Studies 66, 219–254.

Veronesi, Pietro, 2018, Heterogeneous households under uncertainty, Discussion paper, Working paper.

Von Thadden, Ernst-Ludwig, 1995, Long-term contracts, short-term investment and monitoring, Review of Economic Studies 62, 557–575.

Wilkinson, Richard, and Kate Pickett, 2019, The inner level: How more equal societies reduce stress, restore sanity and improve everyone's well-being (Penguin Press).

———, and Molly Scott Cato, 2009, The spirit level. why more equal societies almost always do better, .

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, The Journal of Business 75, 583–608.

Yafeh, Yishay, and Oved Yosha, 2001, Industrial organization of financial systems and strategic use of relationship banking, Review of Finance 5, 63–78.

Yang, Ming, and Yao Zeng, 2018, Financing entrepreneurial production: security design with flexible information acquisition, Review of Financial Studies 32, 819–863.

Yi, Hanyi Livia, 2020, Finance, public goods, and migration, Discussion paper, Working Paper.