Essays on the Macroeconomic Implications of Financial Frictions

by

Yan Ji

B.Eng. Automation, Tsinghua University (2009)
S.M. Civil Engineering, Massachusetts Institute of Technology (2011)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author .... Signature redacted ....
Department of Economics
June 2017

Certified by Signature redacted ....
Robert M. Townsend
Elizabeth & James Killian Professor of Economics
Thesis Supervisor

Certified by ... Signature redacted ....
Alp Simsek
Rudi Dornbusch Career Development Associate Professor of Economics
Thesis Supervisor

Accepted by .... Signature redacted ....
Ricardo Caballero
Ford International Professor of Economics
Chairman, Departmental Committee on Graduate Theses

Essays on the Macroeconomic Implications of Financial Frictions

by

Yan Ji

Submitted to the Department of Economics
on June 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three chapters on the macroeconomic implications of financial frictions.

The first chapter investigates the implications of student loan debt on labor market outcomes. I begin by analytically demonstrating that individuals under debt tend to search less and end up with lower-paid jobs. I then develop and estimate a quantitative model with college entry, borrowing, and job search using NLSY97 data to evaluate the proposed mechanism under the fixed repayment plan and the income-based repayment plan (IBR). My simulation suggests that the distortion of debt on job search decisions is large under the fixed repayment plan. IBR alleviates this distortion and improves welfare. In general equilibrium, debt alleviation achieved through IBR effectively offers a tuition subsidy that increases college entry and encourages firms to post more jobs, further improving welfare.

The second chapter, joint with Winston Dou, proposes a dynamic corporate model in which firms face imperfect capital markets and frictional product markets. We highlight the importance of the endogeneity of the marginal value of liquidity in determining the interactions between investment, financing and product price setting decisions. The model implies several testable predictions: (1) financially constrained firms are more inclined to increase their desired markups of products; (2) firms facing larger price stickiness tend to issue less external equity and conduct less big payouts; and (3) a large part of the cost from price stickiness is induced by financial frictions. Lastly, we provide stylized facts consistent with our model's predictions.

The third chapter (joint with Era Dabla-Norris, Robert Townsend, and Filiz Unsal) develops a general equilibrium model with three dimensions of financial inclusion, depth, and intermediation efficiency. We find that the economic implications of financial inclusion policies vary with the source of frictions. In partial equilibrium, we show analytically that relaxing each of these constraints separately increases GDP. However, when constraints are relaxed jointly, the impacts on the intensive margin (increasing output per entrepreneur with access to credit) are amplified, while the impacts on the extensive margin (promoting credit access) are dampened. In general equilibrium, we discipline the model with firm-level data from six countries and quantitatively evaluate the policy impacts.

Thesis Supervisor: Robert M. Townsend
Title: Elizabeth & James Killian Professor of Economics

Thesis Supervisor: Alp Simsek
Title: Rudi Dornbusch Career Development Associate Professor of Economics

Acknowledgments

I am incredibly grateful to my committee chair Robert M. Townsend. I cannot find enough words in expressing how instrumental Rob has been in my development as an economist as well as a person. I met Rob for the first time while I was doing my master in Civil Engineering. I can never begin my journey in economics research without Rob's recommendation letter or his continuous support during my Ph.D. study. He has been always optimistic, encouraging, and supportive, making me feel enthusiastic after every meeting we had. He has spent uncountable hours helping me take my paper and presentation to a higher level. He read every sentence of my job market paper and gave me a detailed feedback. Rob was also generous with me in many other ways, supporting my research by putting me in connection with his previous students and researchers from IMF, by inviting me to present my work at conferences in Chicago and Montreal, and by funding me during every summer. It is my hope that this thesis marks the beginning of a new journey of interaction, collaboration, and learning.

I am very fortunate to have Alp Simsek as my thesis adviser. Alp has been very helpful throughout this thesis. His remarkable intuitions anticipated many findings of my paper. He is a gifted communicator who can always express abstract economic theory in a few simple examples. Alp sets very high standard for research and this allows me to think deeper and aim higher. Every time after our meeting, I feel like a new door is opened and there are many things to be done. This pushes me to work harder to get sharper results. Alp helped me tremendously not only in terms of writing this thesis but also in terms of developing my research taste. Without him, I will never learn what question is interesting or what theory is useful. Alp is a perfect person in many aspects—he is my idol.

I also want to express my sincere gratitude to Abhijit Banerjee for his help and support throughout this thesis. I cannot imagine a person who is as knowledgeable as Abhijit. Abhijit always presents new perspectives to look at problems. He is also open to unorthodox ideas and approaches—as long as the idea is ambitious and the approach is suitable. I am also indebted to Daron Acemoglu for his numerous suggestions and comments at each stage of this thesis. My knowledge and understanding of economics has been greatly influenced by him. I am also thankful to other faculty members at MIT including George-Marios Angeletos, David Autor, Ricardo Caballero, Anna Mikusheva, and James Poterba, who have helped me in many ways during my Ph.D. adventure.

Finally, I would like to extend my gratitude to my classmates and friends, notably Jie Bai, Vivek Bhattacharya, Winston Dou, Eric Huang, Ernest Liu, Hong Ru, Hai Wang, Hongkai Zhang. The years at MIT would have been less colorful without you.

Contents

Chapter 1

Job Search under Debt: Aggregate Implications of Student Loans

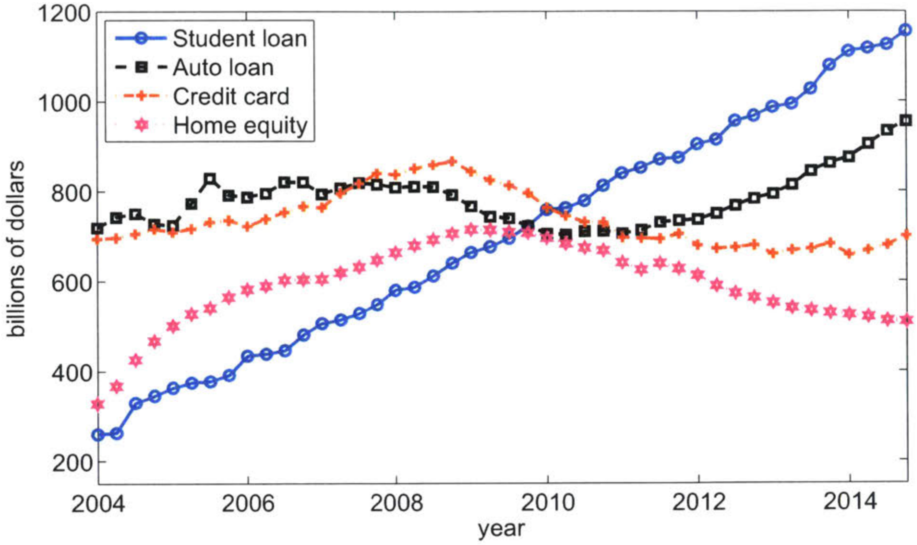
1.1 Introduction

Americans are more burdened by student loan debt than ever. Over the past decade, student loans have more than quadrupled to surpass $1.2 trillion, becoming the second largest type of consumer debt in the U.S. (see Figure 1-1). The increasing number of people facing difficulties paying off these debts has led many to wonder whether student loans might generate broader effects throughout the entire economy. Specifically, concerns about debt repayment presumably affect students' job search decisions after college. Despite the potential importance of student loans, little is known about their implications on labor market outcomes.

In this paper, I fill this gap by theoretically and quantitatively investigating the underlying channels and tradeoffs. I contribute to the existing literature in three ways: (1) by illustrating a mechanism through which student debt burden induces borrowers to search less and end up with lower-paid jobs; (2) by developing a structural model that incorporates college entry, borrowing, and job search to quantify the mechanism; (3) by applying the model to evaluate a realistic income-based repayment plan (IBR) introduced in 2009 and examine the potential general equilibrium implications.

My main quantitative exercise suggests that the distortion of student debt on job search decisions could be large under the fixed repayment plan, and that it is very much relieved by IBR. Debt alleviation achieved by IBR benefits poorer and more indebted borrowers more, and on average, it is equivalent to cutting student debt by half. One-third of debt alleviation is attributed to better job matches. Moreover, adopting IBR also brings two general equilibrium effects that encourage college entry and job postings, effectively offering a tuition subsidy that is much less costly to the government.

The mechanism is closest to that of Herkenhoff, Phillips and Cohen-Cole [129], who show that credit-constrained job seekers tend to be less picky in job search. In this paper, I emphasize



Source: Federal Reserve Bank of New York Consumer Credit Panel (a representative sample drawn from anonymized Equifax credit data).

Figure 1-1: Non-mortgage balances, 2004Q1-2014Q4.

that student debt repayment generates a similar risk and liquidity effect that induces borrowers to search less. My quantitative model incorporates college entry and borrowing decisions into an equilibrium quantitative search model [127, 155, 165, 166]. The model is developed with rich features to match a set of labor market characteristics, and it departs from most of the existing equilibrium search models along three dimensions. First, I introduce college entry and borrowing decisions to study the potential general equilibrium effects. Second, I model student loan debt as a distinct variable, instead of focusing on net worth, to study the implications of different repayment policies. Third, I introduce elastic labor supply for employed workers to capture the potential adverse incentive effects of IBR. As a result, workers and firms bargain over wage rates instead of wage income.

My first key result illustrates and quantifies the mechanism through which debt repayment influences job search decisions. I show that with fixed repayment, more indebted agents set lower reservation wages in job search. This result comes from the fact that search risks are not perfectly insured in an incomplete market. Intuitively, there is a *risk channel* due to the tradeoff between risks and returns because marginally raising the reservation wage increases both expected income and search risks. When debt is higher, the agent becomes more risk averse due to lower consumption, which pushes her to avoid search risks by setting a lower reservation wage. Moreover, because the credit market is imperfect, there also exists a *liquidity channel* from repayment. The liquidity channel reinforces the risk channel and further reduces the reservation wage, substantially increasing the effect of the debt burden.

To evaluate the quantitative implications of this mechanism, I estimate the quantitative model based on 1997-2013 panel data from NLSY97 using the Method of Simulated Moments (MSM). The model is able to capture the positive correlation between talent and debt, endogenous student debt distribution, and various labor market characteristics observed in the data. I validate the model by conducting two sets of out-of-sample tests related to the proposed mechanism. First, I check whether the model can reproduce the differential wage income between borrowers and non-borrowers observed in the data. Second, I check whether the model-implied structural estimates of several related elasticities are in line with the micro estimates in related literature.

I then use the estimated model to evaluate the life-cycle implications of student loans under the fixed repayment plan. On average, borrowers' unemployment duration is 7.6% (1.8 weeks) shorter and they earn 4.2% less ($2,008) annually in the first 10 years after graduation compared to non-borrowers due to inadequate job search. Even after debt has been paid off, borrowers still spend less time on job search and earn relatively less. This lasting effect is attributed to lower savings and the low job-to-job transition rate in the labor market. During the first 10 years after graduation, borrowers accumulate significantly less wealth compared to non-borrowers due to lower wage income and debt repayment. The lower savings would continue affecting borrowers' job search decisions through a similar mechanism in the following years. Moreover, borrowers are stuck at less productive jobs for a while because the estimated search intensity for employed workers is relatively low.

These results suggest that existing studies that do not consider endogenous job search might overestimate the welfare benefit of student debt on improving college entry. While providing student debt enables students from low-income families to attend college and realize the college wage premium, the debt burden after college would distort borrowers' job search decisions, reducing their wage income. According to my simulation, the average annual wage premium of college attendance is about $17,192 for a non-borrower, but the premium is reduced by $2,008 for an average borrower. The 11.7% difference reflects the after-college impact of the debt burden on job search decisions. The debt burden also potentially has aggregate implications on output and productivity by affecting job search decisions. As the simulation suggests, average output and match quality (measured by job productivity) among young borrowers are 3.8% and 2.9% lower compared with non-borrowers. My model suggests a relatively large aggregate effect precisely because of the mismatch in the timing of the benefits and the costs of college attendance, i.e., under the standard fixed repayment plan, student loans are due when borrowers have the lowest capacity to pay. One remedy is to insure risks using income contingent loans.

My second key result illustrates and quantifies the effect of IBR on the reservation wage and welfare. Theoretically I show that income contingency raises the reservation wage through both the risk channel and the liquidity channel. The potential costs of income contingency come from the distortion on labor supply, due to the canonical tradeoff between insurance and the incentive to work. I use my quantitative model to assess the implication of IBR introduced in 2009. Under this realistic plan, borrowers are eligible to repay 15% of their monthly discretionary income, and all the remaining outstanding debt will be forgiven after 25 years of repayment.

I first analyze what would happen in partial equilibrium. If borrowers are unexpectedly allowed to enroll in IBR, my simulation suggests that they would conduct more adequate job search and get matched with jobs that are 1.7% more productive. As a result, their output and wage income increase by 1.3% and 1.9% ($897) in the first 10 years after graduation. The higher wage income already nets out the adverse incentive effect on labor supply, which is small because IBR does not generate much debt forgiveness for borrowers due to the long repayment period. Therefore, instead of taxing borrowers' income, IBR merely restructures payments intertemporally. The average debt alleviation achieved by IBR is equivalent to cutting the amount of debt by half. One-third of the debt alleviation is attributed to better job matches, while the remainder is attributed to better consumption smoothing. There is a large distributional effect: borrowers who are poorer and more indebted benefit more after switching to IBR.

In general equilibrium, debt alleviation after college would also affect agents' college entry and borrowing decisions. After the economy adopts IBR, more students would choose to enter college by borrowing student debt. The college entry rate increases by 6.1% with an increase in the fraction of borrowers from 62.2% to 67.5%. I find that enrolling borrowers in IBR brings an effect on college entry similar to a tuition subsidy of $2,252. However, this tuition subsidy is much less costly from the government's perspective because most borrowers can repay their debt under IBR.

The increase in college entry rate in turn affects firms' job posting decisions. Since college graduates are also more productive, the increase in firms' profits incentivizes more job postings, increasing the equilibrium job contact rate by 6% for unemployed workers. My simulation suggests that switching from the fixed repayment plan to IBR would increase young agents' welfare by about 2.4%. Through a decomposition, I show that the increase in welfare attributed to better job search and insurance, more college entry and borrowing, and more job postings are 0.8%, 1.1%, and 0.5%, respectively.

Related Literature Existing studies have considered how individuals' job search decisions are affected by liquidity and risks. For example, an extensive body of literature investigates how unemployment benefits and private savings affect employment incentives [e.g., 4, 13, 39, 44, 59, 69, 122, 168, 211]. Recently, researchers have started considering the labor market implication of other consumption smoothing mechanisms such as intra-household insurance [117, 144], credit access [127, 129], housing market [45], mortgage modifications [32, 128, 186], and default arrangements [75, 76, 128]. My paper contributes to this research agenda by explicitly modeling and quantitatively evaluating the implication of student debt on job search behavior and the consumption smoothing mechanism offered by different repayment plans.

This paper contributes to the large literature on student loans [see 170, for a recent survey]. An extensive body of this literature focuses on the impact of financial aid during college [e.g., 1, 148]. However, much less is known about the impact of student loans on labor market outcomes after college. The empirical evidence is inconclusive.[1] In this paper, I take a structural approach to highlight one plausible mechanism that could influence indebted students' job search decisions. Abbott et al. [1] develop a rich general equilibrium model with heterogeneous agents to evaluate education policies. My model focuses less on college participation but more on job search decisions. Instead of analyzing further expansions of government-sponsored loan limits, I use the model to evaluate income-based repayment plans, which have been argued to offer risk-sharing benefits with minimal incentive costs [see 213, for a review]. My analyses elucidate the channels through which income contingency influences the outcome of job search. There are studies using structural models to assess income-driven repayment plans [72, 135, 136, 140], but none of them account for search risks in the labor market, which is the focus of my paper.[2]

This paper also relates to the burgeoning literature on the connection between household debt and labor market outcomes. To my knowledge, previous research has discussed three plausible mechanisms. First, household credit could affect the labor market via the aggregate demand channel [85, 116, 178, 180]. Second, households with mortgage debt engage in risk shifting by searching for higher-paid but riskier jobs because they are protected by limited liability [79]. Third, borrowers tend to work in high-paid industries [171, 206]. My paper proposes that borrowers are less picky and more likely to have lower earnings, consistent with recent evidence from Gervais and Ziebarth [100] and Weidner [221].

1.2 Program Description

In the U.S., student loans play a very significant role in higher education. About 60% of college students borrow to help cover costs. In 2014, the number of borrowers surpassed 43 million, with an average balance of about $27,000. Student loans are split into federal loans and private loans, with the former constituting 80% of the total volume. This paper focuses on federal loans because of their importance.

Student loans are arguably more burdensome compared to other loans because repayment usually starts immediately after students leave college, aside from a 6-month grace period offered by Federal Stafford Loans. Moreover, student loans can only be discharged through bankruptcy if borrowers prove "undue hardship" through a court determination. As a result, the insurance provided by consumer bankruptcy [167] is generally absent in student debt.

One prominent feature of student debt is that the federal student loan program allows

[1]For example, Minicozzi [181] suggests that higher educational debt is associated with higher initial wages and lower wage growth rates. Field [93] finds that debt induces NYU law school admits to work in private sector law jobs. Based on a natural experiment in an elite university, Rothstein and Rouse [206] find that indebted students receive higher initial wages as they are more likely to work in high-paid industries. Luo and Mongey [171] try to generalize these findings to a nationally representative sample. Recently, Gervais and Ziebarth [100] explore a regression kink design in need-based federal student loans and find a negative effect of student loans on earnings. Using data from NLSY97 and Baccalaureate and Beyond, Weidner [221] finds that indebted students tend to accept jobs quicker and select jobs in unrelated fields, leading to lower wage income.

[2]Luo and Mongey [171] develop a partial equilibrium model to account for search risks, but they focus on the tradeoff between wage and non-wage benefits.

borrowers to choose among different repayment plans.[3] The standard repayment plan is the default option for student loan borrowers. Under this plan, monthly payments are fixed and made for up to 10 years. As of 2013, 88% of federal direct loan borrowers repay their debt under the standard repayment plan [83]. In addition to the standard plan, the federal student loan program has been offering income-driven repayment plans. The income-based application now includes four different income-driven repayment plans. The main feature of these plans is that borrowers make payments contingent on their income instead of the balance of outstanding debt, and the remaining debt is forgiven after a certain number of payments.[4] Although the first income-driven repayment plan has been made available since 1994, the take-up rate was below 1% until 2008 due to various behavioral issues [83]. As suggested by The Executive Office of the President of the United States [214], continuing to expand enrollment in income-driven repayment plans remains a key priority for the administration.[5] A more detailed description of the federal loan program can be found in Online Appendix A.6.

1.3 Mechanism and Channels

In this section, I build a partial equilibrium model based on McCall [176] with several simplifying assumptions to shed light on the mechanism linking the debt burden to labor market outcomes under two stylized repayment contracts. These assumptions will be made more realistic when conducting quantitative analyses in the next section.

1.3.1 Environment

Consider an agent who is born at $t = 0$ and sequentially searches for a job. Time is discrete and there is no aggregate uncertainty. The agent maximizes lifetime utility from consumption, $E\sum_{t=1}^{\infty}\beta^t u(c(t))$ with subjective rate of time preference β. The per-period utility function, $u(x)$, is bounded from above, strictly increasing, concave, and twice continuously differentiable, i.e., $\lim_{x\to\infty} u(x) = M, u'(x) > 0, u''(x) < 0$.

[3]This paper seeks to understand the implications of the standard fixed repayment plan and the income-driven repayment plan on students' job search behavior. In reality, federal student loan borrowers also have the option to enroll in graduated repayment plan under which monthly payments start out low and increase every two years, or extended repayment plans which allow an extension of repayment period up to 30 years. This paper does not emphasize these plans because of the low take up rate.

[4]All of these plans are different from the first attempt at income contingent loans in the U.S. in 1971—the Yale Tuition Postponement Option (TPO). The main difference is that under these plans borrowers do not need to repay more than the amount borrowed. However, there is cross-subsidization under TPO as participants are required to make payments until the debt of an entire "cohort" is repaid.

[5]In fact, the Obama administration has used several tools to increase enrollment, such as behavioral "nudges", improved loan servicer contract requirements, efforts associated with the President's Student Aid Bill of Rights, a student debt challenge to gather commitments from external stakeholders, and increased and improved targeted outreach to key borrower segments who would benefit from income contingency. The participation rate in income-driven repayment plans has quadrupled over the last four years, from 5% in 2012 to 20% in 2016. In April 2016, the administration announced a series of new actions to further expand enrollment in income-driven repayment plans.

The agent can either be unemployed or employed. For now, suppose that the agent supplies one unit of labor inelasticly when being employed. Starting from $t = 1$, if the agent is unemployed, the agent receives UI benefits $\theta > 0$, and wage offers w from an exogenous cumulative distribution function $F(w)$ in each period, which is differentiable on the support $[\theta, \overline{w}]$.

The agent needs to decide immediately whether to accept the wage offer upon receiving it. There is no recall of past wage offers. Consumption is chosen after the realization of wage offers. If the agent rejects the offer, she continues to search. Otherwise, she gets employed at wage w forever.

The credit market is imperfect in the sense that savings are constrained to be non-negative, $s_t \geq 0$, for all t. The interest rate on savings is r. For simplicity, I assume $\beta(1+r) = 1$ so that the agent has no incentive to transfer wealth across periods.[6]

The agent is born with outstanding debt S whose repayment schedule is specified in the contract. The interest rate on debt is equal to the interest rate on savings. In the following, I analyze the implication of the debt burden on job search decisions for two stylized repayment contracts.

1.3.2 Fixed Repayment Contract

In this subsection, I analyze job search decisions under the fixed repayment contract. To obtain a stationary result, I consider indefinite fixed payment flows such that the present value of this perpetuity covers the initial outstanding debt S.

Definition 1. *The fixed repayment contract requires the agent to repay $s = rS$ in each period.*

For tractability, I assume that the agent cannot be delinquent on making payments. Therefore, to avoid the pathological case, I consider $S < \frac{\theta}{r}$ so that the agent can repay the loan, while at the same time maintaining positive consumption, even if she is permanently unemployed.[7]

Denote U as the value function of an unemployed agent, and $W(w)$ as the value function of an employed agent with wage w. Thus,

$$W(w) = \frac{u(w-s)}{1-\beta}. \tag{1.3.1}$$

When the agent rejects the wage offer, the income in the current period is θ and the value function U can be written as

$$U = u(\theta - s) + \beta \int_{\theta}^{\overline{w}} \max\{W(w), U\} dF(w). \tag{1.3.2}$$

[6] When the agent is unemployed, the agent does not save because she expects future income to be higher. When the agent is employed, the agent is indifferent about savings because wage income is flat and $\beta(1+r) = 1$.

[7] If $S > \frac{\theta}{r}$, the agent is involuntarily forced into delinquency either when she is unemployed or when she is employed at wage $w < rS$. Suppose the remaining income is garnished upon delinquency. Then we can show how the reservation wage varies with debt depends on whether there is an Inada condition on $u(\cdot)$. If utility is bounded from below when consumption approaches zero, we can show that the reservation wage increases with debt. This is because limited liability in debt repayment generates a risk shifting effect as in Donaldson, Piacentino and Thakor [79].

Equation (1.3.1) states that the agent accepts the wage offer if it provides a higher value than unemployment. Because $W(w)$ is increasing in w, the optimal job search decision follows a cutoff strategy, and the wage offer is accepted if $w > w^*_{FIX}$, where w^*_{FIX} is the reservation wage under the fixed repayment contract. The agent sets w^*_{FIX} to maximize her welfare, which happens when the value of staying unemployed is equal to the value of being employed at the reservation wage, i.e., $U = W(w^*_{FIX})$:

$$u(w^*_{FIX} - s) = u(\theta - s) + \frac{\beta}{1-\beta} \int_{w^*_{FIX}}^{\overline{w}} [u(w-s) - u(w^*_{FIX} - s)] dF(w). \tag{1.3.3}$$

The RHS of equation (1.3.3) captures the per-period utility of rejecting the wage offer. It states that rejecting the wage offer results in a lower current utility $u(\theta - s)$ but preserves the possibility of receiving a higher wage offer in the future. Setting a higher reservation wage implies a smaller chance of being employed but also generates a higher expected employment value. The optimal reservation wage is set to balance these two effects.

The Risk and Liquidity Channel of the Debt Burden

Job search is a risky investment that pays off in the future. The agent controls the reservation wage to manage risks, as setting a lower reservation wage allows the agent to accept a constant wage offer sooner and take fewer search risks. Therefore, we can think of the reservation wage characterized by equation (1.3.3) as the certainty equivalent payoff of continued job search. More risk-averse agents have a lower certainty equivalent valuation of any risky lotteries, thus they set a lower reservation wage in job search, which is formalized in Proposition 1.

Proposition 1. *Under the fixed repayment contract, the effect of debt depends on how risk aversion varies with consumption. With decreasing absolute risk aversion, w^*_{FIX} is decreasing in debt; with increasing absolute risk aversion, w^*_{FIX} is increasing in debt; with constant absolute risk aversion, w^*_{FIX} is unaffected by debt.*

Because decreasing absolute risk aversion is empirically plausible [95], Proposition 1 suggests that an indebted agent would set a lower reservation wage to avoid search risks. I discuss in the proof that this proposition holds even if the credit market is perfect. However, the quantitative implication would be much smaller because what would matter is the relative value of outstanding debt to total income instead of income in the current period. This implies that Proposition 1 incorporates both a risk channel and a liquidity channel.

It is worth noting that the risk channel and the liquidity channel result from two different tradeoffs in job search. First, job search is risky. Therefore, an agent who becomes more risk averse due to a higher level of debt would trade off risks and returns by adjusting the reservation wage. This is the risk channel. Second, job search encodes an option value that only pays off in the future, at the time of accepting the wage offer. Therefore, the reservation wage implicitly determines the wealth transfer across periods. When the credit market is imperfect, the agent faces an intertemporal tradeoff in job search because a lower reservation wage increases the chance of accepting a wage offer, and thus more wealth is transferred from future periods to the current period. This is the liquidity channel.

A lower reservation wage implies that the agent is taking fewer search risks in the labor market. Because uninsured search risks are compensated with a risk premium, this implies that indebted agents would have less expected income compared to non-borrowers (see Online Appendix A.4.2 for the proof).

1.3.3 Income-Based Repayment Contract

The main feature of IBR is that borrowers make payments contingent on their income instead of the balance of outstanding debt. Although a realistic IBR also incorporates other auxiliary features like debt forgiveness and repayment caps, my theoretical analysis for now does not explicitly consider them.[8] Instead, I consider IBR that allows the lender to recover all the outstanding debt in expectation conditional on the agent's endogenous job search decisions. Similar to the fixed repayment contract, I assume that the repayment period is indefinite.

Definition 2. *IBR requires the agent to repay a fraction α of her income. The repayment ratio α is set by the lender such that the expected present value of payment flows is just enough to cover the outstanding debt S:*

$$\alpha I(w^*_{IBR}) = \frac{S}{\beta}, \tag{1.3.4}$$

where w^*_{IBR} is the agent's optimal reservation wage under the income-based repayment contract:

$$u((1-\alpha)w^*_{IBR}) = u((1-\alpha)\theta) + \frac{\beta}{1-\beta}\int_{w^*_{IBR}}^{\overline{w}} \left[u((1-\alpha)w) - u((1-\alpha)w^*_{IBR})\right] dF(w). \tag{1.3.5}$$

I call equation (1.3.4) the lender's recoverability constraint. Expected repayment not only depends on the repayment ratio α but also on the agent's reservation wage w^*_{IBR}. Because the reservation wage is unobservable, IBR only specifies the repayment ratio α. The agent optimally chooses her reservation wage according to the indifference equation (1.3.5), which can be thought of as the incentive compatibility constraint.[9]

IBR provides insurance and risk sharing for job search, because the agent repays less when income is low. In fact, we can view the fixed repayment contract as a pure debt contract and IBR as an equity contract. Intuitively, the agent should set a relatively higher reservation wage if debt is repaid under IBR, because equity contracts encourage activities with high returns and high risks. This result is summarized in the following proposition.

Proposition 2. *With CRRA utility, the reservation wage under IBR is strictly higher, i.e., $w^*_{IBR} > w^*_{FIX}$.*

[8] I incorporate these features in section 1.4 when quantitatively evaluating IBR.

[9] Since the income-based repayment contract does not specify the reservation wage, this naturally introduces an inefficiency because the agent does not internalize the effect of her reservation wage on expected repayment. The welfare implication of this inefficiency is discussed in Online Appendix A.4.4.

Since CRRA utility has decreasing absolute risk aversion, Propositions 1 and 2 jointly imply that with CRRA utility, the fixed repayment of debt reduces the reservation wage and IBR alleviates this distortion.

Channels Determining the Reservation Wage

I now elucidate the exact channels through which IBR influences the reservation wage. In the next section, I will enrich the model with additional ingredients informed by these channels to develop a quantitative model for my empirical analysis.

Let us focus on the disposable wage, which is wage income net of debt repayment. Denote $F_{IBR}(w)$ and $F_{FIX}(w)$ as the disposable wage offer distribution under IBR and the fixed repayment contract; thus

$$F_{IBR}(w - \alpha w) = F_{FIX}(w - s) = F(w), \forall w \in [\theta, \overline{w}] \tag{1.3.6}$$

Denote $\tilde{w}^*_{IBR}$ and $\tilde{w}^*_{FIX}$ as the associated disposable reservation wages. By definition,

$$\tilde{w}^*_{FIX} = w^*_{FIX} - s, \tag{1.3.7}$$

$$\tilde{w}^*_{IBR} = (1 - \alpha) w^*_{IBR}. \tag{1.3.8}$$

IBR has a less risky disposable wage offer distribution because of better risk sharing. Using the single-crossing property of $F_{IBR}(w)$ and $F_{FIX}(w)$, I show that IBR is second-order stochastic dominant over the fixed repayment contract.

Lemma 1. *The disposable wage offer distribution under IBR, $F_{IBR}(w)$, strictly second-order stochastically dominates that under the fixed repayment contract, $F_{FIX}(w)$:*

$$\int_0^x F_{IBR}(w)dw \leq \int_0^x F_{FIX}(w)dw, \forall x. \tag{1.3.9}$$

By applying integration by parts twice, I decompose the difference in the disposable reservation wage under the two contracts into three channels:

Proposition 3. *The difference in the disposable reservation wage between IBR and the fixed repayment* *contract is characterized by the following decomposition:*

$$
\begin{aligned}
u(\tilde{w}^*_{IBR}) - u(\tilde{w}^*_{FIX}) = {} & \underbrace{(1-\beta)\left[u((1-\alpha)\theta) - u(\theta - s)\right]}_{\text{liquidity channel, (+)}} \\
& + \underbrace{\beta\left[\int_{\tilde{w}^*_{IBR}}^{\infty}\left(\int_0^w F_{IBR}(x)dx\right)u''(w)dw - \int_{\tilde{w}^*_{FIX}}^{\infty}\left(\int_0^w F_{FIX}(x)dx\right)u''(w)dw\right]}_{\text{risk channel, (+)}} \\
& + \underbrace{\beta\left[u'(\tilde{w}^*_{IBR})\int_0^{\tilde{w}^*_{IBR}} F_{IBR}(w)dw - u'(\tilde{w}^*_{FIX})\int_0^{\tilde{w}^*_{FIX}} F_{FIX}(w)dw\right]}_{\text{optionality channel, (-)}}. \qquad (1.3.10)
\end{aligned}
$$

The RHS of equation (1.3.10) consists of three channels. The first term captures a liquidity channel. The agent repays less during unemployment under IBR, thus $u((1-\alpha)\theta) > u(\theta - s)$. This implies that the first term is positive, contributing to a higher reservation wage.

The second term captures the risk channel. IBR generates a less risky wage offer distribution according to Lemma 1. Because the agent is risk averse, she would raise the reservation wage to pursue a higher expected return when there are fewer risks in job search. Therefore, the second term is also positive, contributing to a higher reservation wage.

The third term captures the difference in the option value of staying unemployed under the two repayment contracts. Intuitively, the agent has a larger option value of staying unemployed when the wage offer distribution is more dispersed. This is because lower wages would be turned down, and higher wages are more likely to be drawn from a more dispersed wage offer distribution. Thus the optionality channel contributes to a lower reservation wage.

1.4 Quantitative Model

In section 1.3, I developed a stylized framework to understand the channels through which the debt burden influences the agent's reservation wage under different repayment contracts. However, the model is not rich enough to match data and conduct quantitative analyses. In this section, I develop a quantitative model to address these issues. In order to highlight the implication of the debt burden on job search decisions, the model intentionally places more emphasis on the labor market relative to the decisions made during college study. In the following, I first give an overview of the model, then I introduce college entry and borrowing decisions. Next, I characterize the post-college life related to job search in the labor market and student loan repayment. Finally, I formulate the recursive dynamic programming problems and define the stationary competitive equilibrium.

1.4.1 Overview

Compared to the partial equilibrium theoretical framework, the quantitative model has the following additional ingredients.

(1). I introduce college entry decisions to capture the potential equilibrium effects through which IBR influences college entry and borrowing decisions by alleviating the debt burden after college.

(2). I introduce age-specific labor productivity to capture the hump-shaped life-cycle earnings profile. Under the fixed repayment plan, borrowers are required to repay debt immediately after college graduation while earnings are low. Thus capturing the life-cycle earnings profile will increase the effect of the debt burden through the liquidity channel.

(3). I introduce loan default that allows borrowers to delay repayment. Default provides some sort of insurance, which mitigates the effect of the debt burden through the risk channel and the liquidity channel.

(4). I introduce on-the-job search and job separation. Both features reduce the value of staying unemployed through the optionality channel. Thus without these ingredients, the model would over-estimate the effect of the debt burden on job search, as the importance of searching for jobs by staying unemployed would be exaggerated.[10]

(5). I introduce nonlinear income taxes. Introducing income taxation is important for the quantitative implication of IBR, because progressive taxation provides partial insurance, and the distortion on labor supply from income-contingency increases with the income taxes facing indebted agents.[11]

(6). I introduce vacancy posting to endogenize the matching rate and the wage offer distribution. I do this to capture potential general equilibrium responses on the firm side after a large-scale policy change on student debt.

1.4.2 College Entry and Borrowing

There is a continuum of agents of measure one in each cohort who live for T periods. In each period, the oldest cohort of agents dies at age T and a new cohort of agents is born with initial wealth b_0 and talent a randomly drawn from the cumulative distribution function $\mho(a, b_0)$.[12] In the following, I describe the agent's problem using age index t.

[10] In the extreme case, if searching during unemployment is as efficient as searching during employment, then the reservation wage is always equal to UI benefits [164], and the proposed mechanism is absent. Therefore, it is important to introduce these realistic features and ask the data to determine the relative efficiency.

[11] As Stiglitz [212] points out, the adverse incentive effects from IBR are likely to be small, so long as income tax rates and repayment rates combined are not too large.

[12] The assumption that all cohorts of agents are born with the same initial distribution of wealth and talent enables a stationary equilibrium, in which the distribution of agents at a given age is the same across cohorts, although different cohorts reach the same age in different periods.

At the beginning of the life, the agent decides whether to enter college after drawing a monetary cost k and a non-monetary utility benefit/cost e randomly from cumulative distributions $\Pi(k)$ and $\Upsilon(e)$. The monetary cost k captures the tuition fees and living expenses net of scholarships and parental transfers received during college study. The utility benefit/cost captures unobserved preference heterogeneity, efforts, or any psychic costs related to college study.

Agents who are wealth constrained (i.e., $b_0 < k$) can borrow an amount of $k - b_0$ student loan debt to pay the monetary cost. As a result, the agent who graduates from college has initial debt burden $s_1 = \max\{k - b_0, 0\}$. At $t = 1$, the agent enters the labor market as an unemployed worker, and her labor productivity z depends on her talent a, education levels ($n = 0, n = 1$), and age t. Specifically, the agent's labor productivity is determined by

$$z(a, n, t) = A_n a g(t), \tag{1.4.1}$$

The value of parameter A_n varies with education level. The difference $A_1 - A_0$ captures the college premium. $g(t)$ is a deterministic trend, which is the same across all agents and only depends on age t.[13] Following Bagger et al. [22], I assume the deterministic trend $g(t)$ to be cubic,

$$g(t) = \mu_0 + \mu_1 t + \mu_2 t^2 + \mu_3 t^3. \tag{1.4.2}$$

Parameters μ_0, μ_1, μ_2, μ_3 are estimated to match the life-cycle earnings profile. The assumption that labor productivity depends on age instead of the number of periods in employment greatly simplifies the problem as z_t is homogeneous within the same cohort conditional on the same talent.

As I show below, in the labor market, agents' characteristics can be summarized by five state variables: wealth b, student debt s, talent a, education level n, and age t. Denote $U(b, s, a, n, t)$ as the value of unemployed workers. The college entry decision is made by comparing the value of entering college, $U(\max\{b_0 - k, 0\}, \max\{k - b_0, 0\}, a, 1, 1) - e$, and the value of not entering college $U(b_0, 0, a, 0, 1)$ at $t = 1$.

1.4.3 Labor Market

Agents have per-period utility $u(c, l)$ and discount factor β. I model $u(c, l)$ using GHH preferences [113],

$$u(c, l) = \frac{1}{1-\gamma}\left(c - \phi \frac{l^{1+\sigma}}{1+\sigma}\right)^{1-\gamma}, \tag{1.4.3}$$

where c and l are consumption and labor supply.

Agents are matched pairwise to jobs, which are created by firms. Following the standard

[13] My model does not address the issues of on-the-job investment in skills emphasized by Heckman, Lochner and Taber [125]. Investigating the implication of student debt on on-the-job human capital accumulation is an interesting topic that is left for future research.

in the literature on search-theoretic models, each firm only creates one job vacancy, thus I do not distinguish between firms and jobs. Jobs are heterogeneous in productivity ρ. There are no productivity shocks, therefore job productivity is constant for a worker-job match until the match breaks up.

Jobs are either vacant or matched with workers and workers are either unemployed or matched with jobs. To simplify notations, I denote $\Omega = (b, s, a, n, t)$ as the worker's characteristic. Denote $\phi^u(\Omega)$ as the PDF of unemployed workers, $\phi^e(\Omega, \rho)$ as the PDF of employed workers matched with jobs whose productivity is ρ, and $v(\rho)$ as the PDF of vacancies. Denote $\Phi^u(\Omega)$, $\Phi^e(\Omega, \rho)$, and $V(\rho)$ as their CDFs. Denote N_v as the number of vacancies and $\overline{u}$ as the unemployment rate.

The number of type-ρ vacancies is

$$N_v(\rho) = N_v v(\rho). \tag{1.4.4}$$

Because each generation has measure one, and there are T overlapping generations, the number of type-ρ jobs in the economy is

$$N(\rho) = (1 - \overline{u})T \int \phi^e(\Omega, \rho) d\Omega + N_v(\rho). \tag{1.4.5}$$

The total number of jobs is

$$N = \int N(\rho) d\rho. \tag{1.4.6}$$

When a worker Ω is matched with a job ρ, they jointly produce a flow of output using the following production technology:

$$F = z(a, n, t)\rho l. \tag{1.4.7}$$

Matching Job search is a random matching process. Agents contact jobs at endogenous rates that depend on their search intensity and the number of vacancies. I allow for on-the-job search and assume that unemployed agents have search intensity h^u and employed agents have search intensity h^e following Lise and Robin [165].[14] Denote H as the aggregate level of search intensity contributed by both unemployed and employed agents:

$$H = h^u \overline{u} T + h^e (1 - \overline{u}) T. \tag{1.4.8}$$

The total number of meetings is determined by a Cobb-Douglas matching function,

$$M = \chi H^{\omega} N_v^{1-\omega}, \tag{1.4.9}$$

where χ and ω are two parameters governing the matching efficiency. From a vacancy's perspec-

[14]The assumption that search intensities are different during unemployment and employment is standard in the search literature. For example, Postel-Vinay and Robin [198] estimate a model with on-the-job search and find that job contact rates are uniformly higher during unemployment across a wide range of occupations.

tive, the probability of contacting a worker is

$$q = M/N_v. \tag{1.4.10}$$

The job contact rates for unemployed workers and employed workers are

$$\lambda^u = h^u M/H; \quad \lambda^e = h^e M/H. \tag{1.4.11}$$

Denote $W(\Omega, \rho, w)$ as the value of an employed agent Ω in job ρ at wage rate w, $U(\Omega)$ as the value of an unemployed agent Ω, and $J(\Omega, \rho, w)$ as the value of a filled job ρ that pays wage rate w. The value of a vacancy is zero due to the free entry condition. When an agent and a job meet each other, a match is formed if there exists wage rate w, such that the worker is willing to accept the job and the firm is willing to hire the worker. Thus the participation constraints are

$$W(\Omega, \rho, w) \geq U(\Omega) \text{ and } J(\Omega, \rho, w) \geq 0. \tag{1.4.12}$$

Matches break up at an exogenous rate κ. After job separations, workers flow into unemployment and jobs disappear. An unemployed worker receives UI benefits θ in every period. The wage income is given by the wage rate w specified in the contract multiplied by the units of labor supply l. Upon forming a worker-firm match, the wage rate is determined through Nash bargaining:

$$w^u(\Omega, \rho) = \underset{w}{\operatorname{argmax}}[W(\Omega, \rho, w) - U(\Omega)]^{\xi} J(\Omega, \rho, w)^{1-\xi}, \tag{1.4.13}$$

where ξ represents the worker's bargaining power.[15]

On-the-Job Search and Poaching I adopt the sequential auction framework of Postel-Vinay and Robin [198] to model the wage determination during on-the-job search. The firm's participation constraint (1.4.12) implies that the highest wage rate that firm ρ can offer to worker Ω is its marginal product of labor, $z\rho$. Because $W(\Omega, \rho, w)$ is increasing in the wage rate, $W(\Omega, \rho, z\rho)$ is the highest value that firm ρ can offer to worker Ω. I define this as the the maximal employment value.

Definition 3. *The maximal employment value, denoted by $\overline{W}(\Omega, \rho)$, is the value of worker Ω being employed by firm ρ when the wage rate is set equal to the marginal product of labor $z\rho$,*

$$\overline{W}(\Omega, \rho) = W(\Omega, \rho, z\rho). \tag{1.4.14}$$

The marginal product of labor increases with job productivity ρ, thus more productive firms can offer higher wage rates to workers. This implies that the maximal employment value that

[15] Note that the usual linear sharing rule [196] is no longer a solution to the Nash Bargaining problem due to the introduction of several features, e.g., risk-averse agents, labor supply, and on-the-job search. Therefore, the wage rate is determined by solving the full maximization problem.

a worker can obtain, $\overline{W}(\Omega,\rho)$, increases with job productivity ρ. Because on-the-job search is modeled based on Bertrand competition, the job with higher productivity will keep the worker. Therefore, on-the-job search may trigger job-to-job transitions or wage renegotiations, depending on the relative productivity of the two jobs competing for the worker.

To elaborate, consider a worker Ω working in a job with productivity ρ' and wage w', poached by a new job with productivity ρ. If the maximal employment value of the new job ρ is smaller than the current job's value, i.e., $\overline{W}(\Omega,\rho) < W(\Omega,\rho',w')$, then the worker will discard the new job offer and stay with the current job with the old wage w'.

If the new job can offer a higher job value, then the two jobs will compete to bid up the wage rate. The job with higher productivity is able to overbid the other job and thus keep the worker. There are two cases:

First, if $\rho > \rho'$, the worker currently employed at job ρ' will transfer to job ρ and the old job ρ' will become the negotiation benchmark due to Bertrand competition. This grants the worker an outside option value that is equal to the maximal employment value of ρ'. The new wage rate will be set according to

$$w^e(\Omega,\rho,\rho') = \underset{w}{\operatorname{argmax}}[W(\Omega,\rho,w) - \overline{W}(\Omega,\rho')]^{\xi} J(\Omega,\rho,w)^{1-\xi}, \tag{1.4.15}$$

where the worker's outside option is captured by the old job's productivity ρ'.

Second, if $\rho \leq \rho'$, the worker will stay with the current employer ρ', but job ρ will be used as the new negotiation benchmark for a wage rise. This grants the worker an outside option value that is equal to the maximal employment value of ρ. The new wage rate will be set to

$$w^e(\Omega,\rho',\rho) = \underset{w}{\operatorname{argmax}}[W(\Omega,\rho',w) - \overline{W}(\Omega,\rho)]^{\xi} J(\Omega,\rho',w)^{1-\xi}. \tag{1.4.16}$$

Reservation Productivity Equation (1.4.15) nests equation (1.4.13), if we treat an unemployed agent Ω as being employed in a fictitious job $\rho_u(\Omega)$, such that $\overline{W}(\Omega,\rho_u(\Omega)) = U(\Omega)$. Hence, the negotiation benchmark for an unemployed agent is $\rho_u(\Omega)$ and the wage rate satisfies

$$w^u(\Omega,\rho) = w^e(\Omega,\rho,\rho_u(\Omega)). \tag{1.4.17}$$

In fact, $\rho_u(\Omega)$ can be considered as the reservation productivity for an unemployed agent Ω, because she is indifferent between being employed at job $\rho_u(\Omega)$ or staying unemployed. On the other hand, job $\rho_u(\Omega)$ is also indifferent about hiring because it is offering the worker the maximal employment value. I define this formally as follows:

Definition 4. *The reservation productivity for an unemployed agent Ω is a fictitious job with productivity $\rho_u(\Omega)$ such that the agent is indifferent between accepting the job or staying unemployed, i.e.,*

$$\overline{W}(\Omega,\rho_u(\Omega)) = U(\Omega). \tag{1.4.18}$$

For any reservation productivity $\rho_u(\Omega)$, the corresponding marginal product of labor, $Az\rho_u(\Omega)$, can be considered as the reservation wage of an unemployed agent Ω. It is difficult to obtain a formal proof on how the reservation productivity changes with the level of student loan debt, but the intuition is exactly the same as what is discussed in section 1.3. Therefore, indebted agents set lower reservation productivity and search for a shorter time.

1.4.4 Repayment, Default, and Taxes

Repayment As noted in section 1.2, most federal student loan borrowers repay under the fixed repayment plan during my sample period. Therefore, I only consider the fixed repayment plan when estimating the model.

I assume that student loan borrowers make fixed payments every period after college graduation until the 10th year. This is consistent with the terms specified in the standard 10-year fixed repayment plan. The interest rate for the fixed repayment plan is variable before July 1, 2006, and fixed thereafter. For simplicity, I consider a fixed interest rate r^s. Hence, the annual payment is given by:

$$y_t^{fix} = \frac{r^s}{(1+r^s)\left[1 - \dfrac{1}{(1+r^s)^{10-(t-1)}}\right]} s_t, \quad \text{for } t <= 10. \tag{1.4.19}$$

When conducting my quantitative analyses, I consider what would happen if borrowers are allowed to enroll in IBR. In reality, the interest rate does not depend on repayment plans. IBR passed by Congress in 2009 has three main features: (1) borrowers are required to repay 15% of their discretionary income, which is defined as the difference between pre-tax income and 150% of the poverty guideline;[16] (2) the monthly payment is capped by the amount under the 10-year fixed repayment plan, based on the outstanding loan balance when the borrower initially entered IBR. This implies that the repayment under IBR is never more than the 10-year fixed repayment plan amount. (3) the repayment period is 25 years and the remaining balance will be forgiven at the end. To reflect these features, I model the annual payment under IBR by:

$$y_t^{ibr} = \min\left(0.15\max(w_t l_t - pov, 0), \quad y_1^{fix}, \quad s_t\right), \quad \text{for } t <= 25, \tag{1.4.20}$$

where minimizing over the term y_1^{fix} captures the repayment cap, and the term s_t ensures that the borrower will never repay more than the amount owed. I set the poverty guideline based on the average individual poverty level for the 48 contiguous states (excluding Hawaii and Alaska) and the District of Columbia. The inflation-adjusted poverty level is quite stable over time, and the 150% poverty level is set to be $pov = \$15,650$ corresponding to its average value between 1997-2013 measured in 2009 dollars.

[16]Borrowers are required to repay 10% of their discretionary income if they are new borrowers on or after July 1, 2014. The repayment period is 20 years for new borrowers.

Default Unlike other loans, student loans are practically non-dischargeable after default (and bankruptcy). I assume that borrowers incur a cost η if they default on their loans. In the year following the default, borrowers negotiate a new repayment plan that has the same repayment period as the fixed repayment plan.[17] Modeling default in this way ensures that default time is not a state variable. As a result, in my model default delays the repayment by one year, but the payment in the following years will increase. Moreover, I do not allow repeated default given the complexity of the current setup.[18] If agents default at time t_{def}, the annual payment thereafter is

$$y_t^{def} = \begin{cases} 0, & \text{for } t = t_{def}. \\ \dfrac{r^s}{(1+r^s)\left[1 - \dfrac{1}{(1+r^s)^{10-(t-1)}}\right]} s_t, & \text{for } t_{def} < t <= 10. \end{cases} \tag{1.4.21}$$

It is also possible that deeply indebted agents may not be able to honor the payment if they have been unemployed for a long time. While this is theoretically possible, it rarely happens in simulations because very few agents take on large debt. If this involuntary delinquency happens, I assume that agents have to repay all earnings (up to a consumption floor specified below) in every following period until all the past payments required under the fixed repayment plan are repaid.

Income Taxes Agents face progressive income taxes. Following Benabou [31] and Heathcote, Storesletten and Violante [124], I model after-tax income $\tilde{E}$ as:

$$\tilde{E} = \varkappa(wl)^{1-\tau}, \tag{1.4.22}$$

where wl is the pre-tax wage income.

In the U.S., UI benefits are also taxable, thus the formula for unemployed workers is $y = \varkappa\theta^{1-\tau}$. The fiscal parameters $\varkappa$ and τ are set to approximate the U.S. income tax system. The parameter $\varkappa$ determines the overall level of taxation. The parameter τ determines the rate of progressivity because it reflects the elasticity of after-tax income with respect to pre-tax income. When $\tau = 0$, the tax system has a flat marginal tax rate $1 - \varkappa$, and when $\tau > 0$, the tax system is progressive.

In the baseline simulation, I assume that the tax revenue is collected to finance the UI benefits and a non-valued public consumption good G:

$$(1-\overline{u})T \iint wl[1 - \varkappa(wl)^{-\tau}]\phi^e(\Omega,\rho)d\Omega d\rho = \overline{u}T \int \varkappa\theta^{1-\tau}\phi^u(\Omega)d\Omega + G. \tag{1.4.23}$$

[17] In reality, borrowers can get rehabilitation on their defaulted loans after consequently making several eligible payments. Then borrowers must agree with the U.S. Department of Education on a reasonable and affordable repayment plan. The repayment plans after default are set case by case. Generally, a monthly payment is considered to be reasonable and affordable if it is at least 1.0% of the current loan balance. Volkwein et al. [219] find that two out of three defaulters reported making payments shortly after the official default first occurred.

[18] In practice, loan rehabilitation is a one-time opportunity, and more severe punishments are imposed on borrowers who default repeatedly.

When conducting the quantitative analyses in section 1.6, I take the value of G from the baseline as exogenously given. When evaluating IBR, I adjust the parameter $\varkappa$ to balance the budget:

$$(1-\bar{u})T\iint wl[1-(\varkappa-\Delta\varkappa)(wl)^{-\tau}]\phi^{e}(\Omega,\rho)d\Omega d\rho=\bar{u}T\int(\varkappa-\Delta\varkappa)\theta^{1-\tau}\phi^{u}(\Omega)d\Omega+G+\text{Forgiveness}. \tag{1.4.24}$$

The implied value of $\Delta\varkappa$ captures the increase in overall tax level in order to finance the debt forgiveness.

1.4.5 Value Functions

The timing of events in the labor market is the following. At the beginning of age t, firms post vacancies at cost v and existing matched jobs separate at rate κ. Vacancies and agents meet each other at Poisson rates, λ^u, λ^e, and q. Agents then make default decisions (if not yet in default) and repay student loan debt. At the end of age t, unemployed agents receive UI benefits θ, and employed agents supply labor l and negotiate wage rates w with firms based on their negotiation benchmarks' productivity. After receiving income, agents pay income taxes and choose consumption c_t. Following Hubbard, Skinner and Zeldes [133], I introduce a consumption floor $\underline{c}$ to model means-tested benefits.

Instead of using the wage rate w as a state variable for an employed worker, the discussions in subsection 1.4.3 suggest that the negotiation benchmark's productivity is a natural state variable. Therefore, the state variables are worker characteristic Ω, job productivity ρ, and the negotiation benchmark's productivity ρ'. The value of an employed worker and the value of a job immediately after search and matching can be written as $W(\Omega,\rho,\rho')$ and $J(\Omega,\rho,\rho')$ before default. I add superscript d to represent value functions and variables after default. Below I present the value functions of each participant after default. The value functions for the non-default case are presented in Online Appendix A.5.1.

Unemployed Workers An unemployed worker who has defaulted has value

$$\begin{aligned}
U^{d}(\Omega_t)=\max_{c_t,l_t}\quad & u(c_t,l_t)+\beta\left[\lambda^{u}\int_{x\geq\rho_u^d}W^{d}(\Omega_{t+1},x,\rho_u^d)dV(x)+[1-\lambda^{u}+\lambda^{u}V(\rho_u^d)]U^{d}(\Omega_{t+1})\right],\\
\text{subject to}\quad & b_{t+1} = (1+r)(b_t-y_t^{def})+\varkappa\theta^{1-\tau}-c_t,\\
& s_{t+1} = (1+r^{s})(s_t-y_t^{def}),\\
& c_t \geq \underline{c},\\
& b_{t+1} \geq 0,
\end{aligned} \tag{1.4.25}$$

where r is the interest rate on deposit and ρ_u^d is the reservation productivity for the unemployed worker Ω_{t+1} who has defaulted. In the objective function, the term $u(c_t,l_t)$ represents the realized utility at age t; the first term in the squared bracket represents the expected value of entering the labor market at age $t+1$; and the second term represents the value of staying unemployed, which could happen when the productivity draw is less than the reservation productivity, $\rho_u^d(\Omega_{t+1})$.

Following Acemoglu and Shimer [5] and Krusell, Mukoyama and Sahin [155], I impose the borrowing constraint, $b_{t+1} \geq 0$, so that agents do not have access to other credit apart from student loans. Relaxing this constraint enables the model to parsimoniously capture other types of loans, e.g., consumption loans. I provide a sensitivity analysis for the credit limit in Online Appendix A.7.

Employed Workers The value of defaulted employed workers at job ρ, with negotiation benchmark ρ', is given by

$$
\begin{aligned}
W^d(\Omega_t,\rho,\rho') = \max_{c_t,l_t} \quad & u(c_t,l_t) + \beta\Bigg\{\kappa U^d(\Omega_{t+1}) + (1-\kappa)\Bigg[[1-\lambda^e+\lambda^e V(\rho')]W^d(\Omega_{t+1},\rho,\rho') \\
& +\lambda^e\Bigg(\int_{x\geq\rho} W^d(\Omega_{t+1},x,\rho)dV(x) + \int_{\rho'<x<\rho} W^d(\Omega_{t+1},\rho,x)dV(x)\Bigg)\Bigg]\Bigg\}, \\
\text{subject to} \quad & b_{t+1} = (1+r)(b_t - y_t^{def}) + \varkappa[w^{e,d}(\Omega_t,\rho,\rho')l_t]^{1-\tau} - c_t, \\
& s_{t+1} = (1+r^s)(s_t - y_t^{def}), \\
& c_t \geq \underline{c}, \\
& b_{t+1} \geq 0,
\end{aligned}
\tag{1.4.26}
$$

In problem (1.4.26), the first term in the curly bracket captures exogenous job separations at rate κ, in which case the worker becomes unemployed in period $t+1$, and receives $U^d(\Omega_{t+1})$. The job is maintained with probability $1-\kappa$, and the three cases resulting from on-the-job search are captured by the second term. With probability λ^e, the worker gets contacted by a new job x. If the new job's productivity x is larger than the current job ρ, the worker moves to the new job and her current job becomes the new negotiation benchmark. If she samples a job with productivity larger than the current negotiation benchmark but smaller than her current job's productivity, she will stay at the current job with an updated negotiation benchmark. Finally, she may stay with the current job with an unchanged negotiation benchmark either when she is not poached by a new job or the new job's productivity is lower than her current negotiation benchmark.

Filled Jobs and Match Surplus The value of a job filled by a worker who has defaulted is,

$$
\begin{aligned}
J^d(\Omega_t,\rho,\rho') &= [Az_t\rho - w^{e,d}(\Omega_t,\rho,\rho')]l^d(\Omega_t,\rho,\rho') \\
&+ \beta(1-\kappa)\Bigg[\lambda^e \int_{\rho'<x<\rho} J^d(\Omega_{t+1},\rho,x)dV(x) + [1-\lambda^e+\lambda^e V(\rho')]J^d(\Omega_{t+1},\rho,\rho')\Bigg],
\end{aligned}
\tag{1.4.27}
$$

where the first term in the square bracket represents the case in which the poaching job results in a wage increase by raising the negotiation benchmark. The second term represents the case in which the worker does not receive a competitive outside offer.

The match surplus relative to unemployment is given by

$$Surplus^d(\Omega_t, \rho, \rho') = W^d(\Omega_t, \rho, \rho') - U^d(\Omega_t) + J^d(\Omega_t, \rho, \rho'). \tag{1.4.28}$$

1.4.6 Stationary Competitive Equilibrium

To close the model, I describe the free entry condition and define the stationary equilibrium.

Free Entry Condition The cost of vacancy creation is ν. Following Lise, Meghir and Robin [166], I assume that once the firm pays the cost, a job is created with productivity ρ being randomly drawn from a CDF $F(\rho)$. Vacancies last for one period; thus if the created vacancy is not filled by a worker in the current period, the vacancy will be destroyed. This immediately implies that the equilibrium vacancy distribution $V(\rho)$ is the same as $F(\rho)$. In equilibrium, the free entry condition requires that the cost of vacancy creation is equal to its expected value,

$$\begin{aligned} \frac{\nu}{q} = \frac{\overline{u}Th^u}{H}&\left[\iint_{\rho>\rho_u^d} J^d(\Omega,\rho,\rho_u^d)\phi^u(\Omega,1)d\Omega dF(\rho) + \iint_{\rho>\rho_u} J(\Omega,\rho,\rho_u)\phi^u(\Omega,0)d\Omega dF(\rho)\right] \\ + \frac{(1-\overline{u})Th^e}{H}&\left[\iiint_{\rho>\rho'} J^d(\Omega,\rho,\rho')\phi^e(\Omega,\rho',1)d\Omega d\rho' dF(\rho) + \iiint_{\rho>\rho'} J(\Omega,\rho,\rho')\phi^e(\Omega,\rho',0)d\Omega d\rho' dF(\rho)\right], \end{aligned} \tag{1.4.29}$$

where $\phi^u(\Omega, d)$ represents the PDF conditional on whether unemployed agents have defaulted ($d = 1$) or not ($d = 0$). Thus, $\phi^u(\Omega) = \phi^u(\Omega, 0) + \phi^u(\Omega, 1)$. Similarly, for employed agents, $\phi^e(\Omega, \rho) = \phi^e(\Omega, \rho, 0) + \phi^e(\Omega, \rho, 1)$.

Equation (1.4.29) states that a new vacancy meets an agent with probability q. Conditional on a meeting, the vacancy meets an unemployed worker with probability $\overline{u}Th^u/H$ and is filled if the vacancy's productivity is above the reservation productivity, $\rho > \rho_u(\Omega)$. The vacancy meets an employed worker with probability $(1-\overline{u})Th^e/H$ and is filled if the vacancy's productivity is above the worker's current job's productivity, $\rho > \rho'$.

In the stationary equilibrium, the flows in and out of unemployment balance each other out. The unemployment rate $\overline{u}$ is determined by the following equation:

$$(1-\overline{u})\kappa = \overline{u}\lambda^u\left[\int [1 - V(\rho_u^d)]\phi^u(\Omega,1)d\Omega + \int [1 - V(\rho_u)]\phi^u(\Omega,0)d\Omega\right], \tag{1.4.30}$$

where the LHS represents the flow into unemployment due to exogenous separations of employed agents at rate κ, and the RHS represents the flow into employment when unemployed agents contact jobs whose productivity is above their reservation productivity.

Equilibrium Definition Below I define the stationary competitive equilibrium.

Definition 5. *The stationary competitive equilibrium consists of stationary distributions of unemployed agents, $\phi^u(\Omega)$, employed agents $\phi^e(\Omega, \rho)$, vacancies $V(\rho)$, the number of vacancies N_v, and unemployment rate $\overline{u}$, such that:*

(1). *The job contact rates for agents and firms are determined by the Cobb-Douglas meeting technology according to (1.4.10-1.4.11).*

(2). *All unemployed agents Ω make consumption and default decisions by solving problem (1.4.25) depending on their default status.*

(3). *All employed agents Ω at job ρ with negotiation benchmark ρ' receive wage income and make consumption, labor supply, and default decisions by solving problem (1.4.26) depending on their default status.*

(4). *Wage rates, $w^e(\Omega, \rho, \rho')$ and $w^{e,d}(\Omega, \rho, \rho')$, are determined by Nash bargaining specified in (1.4.15) and (1.4.17).*

(5). *The equilibrium number of vacancies N_v and the vacancy distribution $V(\rho)$ are determined by the free entry condition (1.4.29).*

(6). *The equilibrium unemployment rate $\overline{u}$ is determined to balance flows in and out of unemployment, as specified in (1.4.30).*

1.5 Data, Estimation, and Validation Tests

In this section, I first introduce the data. Then I present the estimation procedures of my quantitative model. Finally, I conduct two validation tests to check the external validity of the model.

1.5.1 Data

My empirical analysis uses panel data from the National Longitudinal Survey of Youth 1997 (NLSY97). This is a nationally representative survey conducted by the Bureau of Labor Statistics. In round 1, 8,984 youths were initially interviewed in 1997. Follow-up surveys were conducted annually. Almost 83% (7,423) of the round 1 sample were interviewed in round 15 (2011-2012). Youths were born between 1980 and 1984. Their ages ranged from 12 to 18 in round 1 and were 26 to 32 in round 15. The survey contains extensive information on each youth's labor market behavior and documents the amount of education loans borrowed during college, which makes NLSY97 an ideal data set for studying the implications of student loan debt on job search decisions.

My analysis focuses on high school and college graduates. I do not include college dropouts because it is not clear when they enter the labor market. I drop youths who have ever served in the military or attended graduate schools because they are not in the same position as the other youths in my sample when it comes to making labor market decisions. I also drop youths who received the bachelor's degree before 1997 due to the lack of labor market information upon college graduation. This leaves me with a sample of 1,721 high school graduates and 1261 college graduates. I construct the variables used in structural estimation following the steps illustrated in Online Appendix A.1.

1.5.2 Estimation

My estimation consists of three steps. First, I specify the parametric functional forms for several distributions in order to identify the model and match the data. Second, I determine the values of a set of parameters without running simulations. These parameters' values are either separately estimated or taken from existing literature. Finally, I discuss the identification of the model's remaining parameters and estimate their values using MSM.

Parametrization

I assume that the marginal distribution of initial wealth follows a flexible generalized Pareto distribution with location parameter $\underline{b}$, scale parameter ζ, and shape parameter φ:

$$\mho_{b_0}(b_0) = \frac{1}{\zeta}\left(1 + \varphi\frac{b_0 - \underline{b}}{\zeta}\right)^{-\frac{1+\varphi}{\varphi}}. \tag{1.5.1}$$

The marginal distribution of talent follows a flexible beta distribution with parameters f_1^a and f_2^a. To capture the potential correlation between initial wealth and talent, I use the Frank copula, where the single parameter ϑ governs the dependence between the CDF of the marginal distribution of wealth, $\mho_{b_0}(b_0)$, and the CDF of talent, $\mho_a(a)$[19]:

$$C(u,v) = \mathbb{P}(\mho_{b_0}(b_0) \leq u, \mho_a(a) \leq v) = -\frac{1}{\vartheta}\log\left[1 + \frac{(e^{-\vartheta u} - 1)(e^{-\vartheta v} - 1)}{e^{-\vartheta} - 1}\right]. \tag{1.5.2}$$

I assume that the monetary cost k and non-monetary utility cost e of college entry are drawn from a (truncated) normal distribution with parameters (μ_k, σ_k^2) and (μ_e, σ_e^2). Because monetary costs of college entry are non-negative, I set $k = 0$ for negative draws.

Following Lise, Meghir and Robin [166] and Jarosch [137], I assume that job productivity follows a flexible Beta distribution on support $[0,1]$ with parameters f_1^p, f_2^p.

[19]The use of Frank copula allows me to estimate the parameters governing the marginal distribution of wealth separately using MLE. The parameters governing the marginal distribution of talent along with the parameter ϑ are estimated with other internally-estimated parameters using MSM.

Externally Determined Parameters

Table 1.1 presents the values for externally determined parameters. The three parameters governing the initial wealth distribution, $(\underline{b}, \zeta, \varphi)$, are estimated directly using MLE to match the empirical distribution of wealth (see panel A of Figure 1-3).

Table 1.1: Parameters determined outside the model.

Parameter	Description	Value	Source
$\underline{b}$	Location parameter	0	Estimated from NLSY97
ζ	Scale parameter	223.0	Estimated from NLSY97
φ	Shape parameter	1.52	Estimated from NLSY97
$\varkappa$	Overall tax level	2.17	Estimated from March CPS 1997-2008
τ	Rate of tax progressivity	0.11	Estimated from March CPS 1997-2008
γ	Risk aversion	3	Hubbard, Skinner and Zeldes [133]
σ	Elasticity of labor supply	2.59	Keane [147], Frisch elasticity=0.33
r	Annual risk-free rate	4.5%	Real interest rate between 1997-2008
r^s	Interest rate on student loans	6.6%	Ionescu [135], risk premium=2.1%
β	Discount factor	0.96	Standard practice
ω	Meeting technology	0.5	Pissarides and Petrongolo [197]
ξ	Bargaining parameter	0.5	Hosios condition
θ	UI benefits	\$8,000	40% of average 6-month wage income
$\underline{c}$	Consumption floor	\$900	AFDC, food stamps, and WIC
T	Number of years working	38	Real-life working age of 23 to 60

The fiscal parameters $\varkappa$ and τ are identified using the regression coefficients obtained from regressing log individual after-tax earnings $\tilde{E}_i$ on log individual pre-tax earnings E_i:

$$log(\tilde{E}_i) = log(\varkappa) + (1-\tau)log(E_i) + \varepsilon_i \tag{1.5.3}$$

The pre-tax earnings data are obtained from March CPS 1997-2008. I use the NBER's TAXSIM program to compute after-tax earnings as earnings minus all federal and state taxes. The estimated values are $\varkappa = 2.17$ and $\tau = 0.11$.

I take advantage of the existing findings to determine the values of γ and σ. I choose γ according to the literature that is most closely related to this paper. In particular, I set $\gamma = 3$ consistent with the precautionary savings literature [e.g., 133]. Since the value of γ is the most important parameter that determines the quantitative implication of the debt burden, I provide a sensitivity analysis using other values of γ in Online Appendix A.7. The tax-modified Frisch elasticity of labor supply with respect to pre-tax wage rates is $(1-\tau)/(\sigma+\tau)$. Thus I set $\sigma = 2.59$, which implies that the tax-modified Frisch elasticity is 0.33, broadly consistent with microeconomic evidence [147]. The elasticity of labor supply determines the distortionary effect of IBR, I provide a sensitivity analysis using other values of σ in Online Appendix A.7.

I set the annual risk-free rate to be $r = 4.5\%$, corresponding to the average real interest rate in the U.S. between 1997-2008 (source: World Development Indicators). I set the interest rate on student loans to be $r_s = 6.6\%$, which implies a risk premium consistent with the annualized mark-up over the Treasury bill rate, 2.1%, set by the government for subsidized loans issued before 2006 [135]. Following the standard practice, I set the annual discount rate to be $\beta = 0.96$.

I set the matching parameter to be $\omega = 0.5$, which lies in the middle of existing estimates using information on the flow of hires and the stock of unemployment and job vacancies [197]. One way to determine the bargaining parameter is to follow the Hosios efficiency condition [155, 210], thus I set the bargaining parameter to have the same value $\xi = 0.5$.

In the U.S., UI benefits generally pay eligible workers between 40%-50% of their previous pay. The standard time-length of unemployment compensation is 6 months, although during the recent recession, Congress passed the emergency benefit program to extend the duration to 73 weeks. In my model, unemployed agents receive UI benefits every year. Therefore, I choose a relatively lower value of UI benefits to account for this discrepancy. I set $\theta = \$8,000$, which amounts to roughly 40% of the average 6-month income.

Means-tested benefits include Aid to Families with Dependent Children (AFDC), food stamps, and Women, Infants, Children (WIC). In my sample, the percent of youths who had ever received AFDC, food stamps, and WIC by 2013 are 1.3%, 8.4%, and 6.3%. About 11.5% of youths had ever received any means-test benefits during my sample period, with a median monthly benefit level of \$150. Because the take-up rate is far from universal, following Kaplan [144], the annual consumption floor is set to be \$900, half of the median value of means-tested benefits.

Between 2002-2012, the average retirement age is around 60. I set $T = 38$, which corresponds to a real-life working age of 23 to 60.

Internally Estimated Parameters

I now turn to the identification discussion of internally estimated parameters.

Labor Market Moments Parameters A_0 and A_1 in equation (1.4.1) determine the labor productivity of high school and college graduates. Their values are identified to match the average wage income among high school and college graduates.

The exogenous job separation rate κ is identified from the average duration of employment spells. In the NLSY97 sample, employment spells last for about 2.2 years on average, consistent with the calculations of Shimer [210] using CPS data.

The search intensity during employment h^e is normalized to be 1. The search intensity during unemployment h^u and the parameter governing matching efficiency χ are identified from the average unemployment duration and the average duration of job tenure. In the data, the average unemployment duration is 27.2 weeks and jobs last for about 1.5 years on average. Because job separations could either result in a transition into unemployment or a transition into another job, the small difference between the average employment duration and the average job tenure implies

Table 1.2: Model fit for targeted moments.

Labor Market Moments	**Model**	**Data**
Mean of wage income among high school graduates in first 5 years	\$26,364	\$26,736
Mean of wage income among college graduates in first 5 years	\$40,354	\$40,619
Mean of employment duration (year)	2.2	2.2
Mean of unemployment duration (week)	27.2	27.2
Mean of job tenure (year)	1.5	1.5
Variance of log wage income	0.180	0.155
Skewness of log wage income	0.068	-0.174
Mean of log wage increase upon job-to-job transitions	0.132	0.150
Variance of log wage increase upon job-to-job transitions	0.023	0.042
Vacancy to unemployment ratio	0.409	0.409
Average hours worked per year	1,731	1,729
Life-cycle earnings profile	see Figure 1-2	
College and Debt Moments	**Model**	**Data**
Fraction of agents with a bachelor's degree	41.4%	42.2%
Unexplained variance in college entry decisions ($1 - R^2$)	0.64	0.64
Correlation between talent and student debt	0.05	0.04
Default rate	9.65%	9.26%
Student debt distribution upon college graduation	see Figure 1-3	

Note: This table presents model fit for targeted moments. The life-cycle earnings profile is constructed using March CPS 1997-2008 data. The default rate is constructed by Yannelis [224] using a random 1% sample of NSLDS. The vacancy to unemployment ratio is constructed using JOLTS data between 2001-2013. The remaining moments are constructed using the sample from NLSY97.

that on-the-job search is much less efficient compared to searching during unemployment.[20]

As argued by Jarosch [137], the second and third moments of the cross-sectional log wage income distribution are informative about the distribution of job productivity. However, in my model the productivity of matched worker job pair is given by $z\rho$. The symmetric roles played by worker productivity z and job productivity ρ suggest that it is impossible to separately identify the parameters f_1^a, f_2^a governing the marginal distribution of talent and the parameters f_1^ρ, f_2^ρ governing the marginal distribution of vacancy's productivity if we only use moments from the cross-sectional log wage income distribution. Note that upon job-to-job transitions, worker productivity remains the same but job productivity increases. Therefore, the mean and variance of log wage increase upon job-to-job transitions are informative about the value of parameters f_1^ρ, f_2^ρ. In the data, there are unmodeled sources of variation that affect the dispersion of the log wage income distribution, thus I adjust for these sources of variation when constructing the variance and skewness (see Online Appendix A.1.3). The cross-sectional log wage income

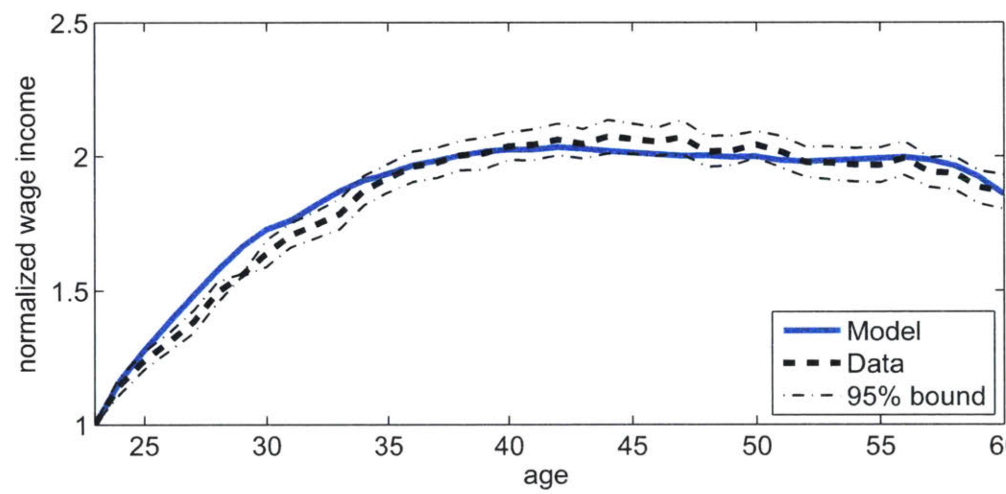
[20]In the extreme case where the average employment duration is equal to the average job tenure, there is no job-to-job transitions, which implies the absence of on-the-job search. On the other hand, if the average job tenure is much shorter than the average employment duration, it means most of the job separations are due to job-to-job transitions instead of employment-to-unemployment transitions.

residuals have variance 0.155 and skewness -0.174. The log hourly wage rate rises by about 15.0% upon job-to-job transitions on average with a variance of 0.042.

The flow cost of vacancy creation ν is identified from the vacancy to unemployment ratio. The Job Openings and Labor Turnover Survey (JOLTS) collected job openings information since December 2000 in the United States. I estimate the vacancy to unemployment ratio to be 0.409 using the data between 2001-2013. This estimate is smaller than the estimate of 0.539 provided by Hall [119], who uses data between 2001-2002.

Parameter ϕ is a scale factor of labor supply, which is identified from the average number of hours worked in each year. In the data, people with full-time jobs work for roughly 1,729 hours per year on average.

Parameters μ_0, μ_1, μ_2, and μ_3 are identified to match the average wage income in each year between ages 23-60. Because NLSY97 does not provide individual labor market histories at this length, I construct the life-cycle earnings profile using March CPS 1997-2008 data. I use 38 moments to capture the full life-cycle earnings profile (see Figure 1-2).

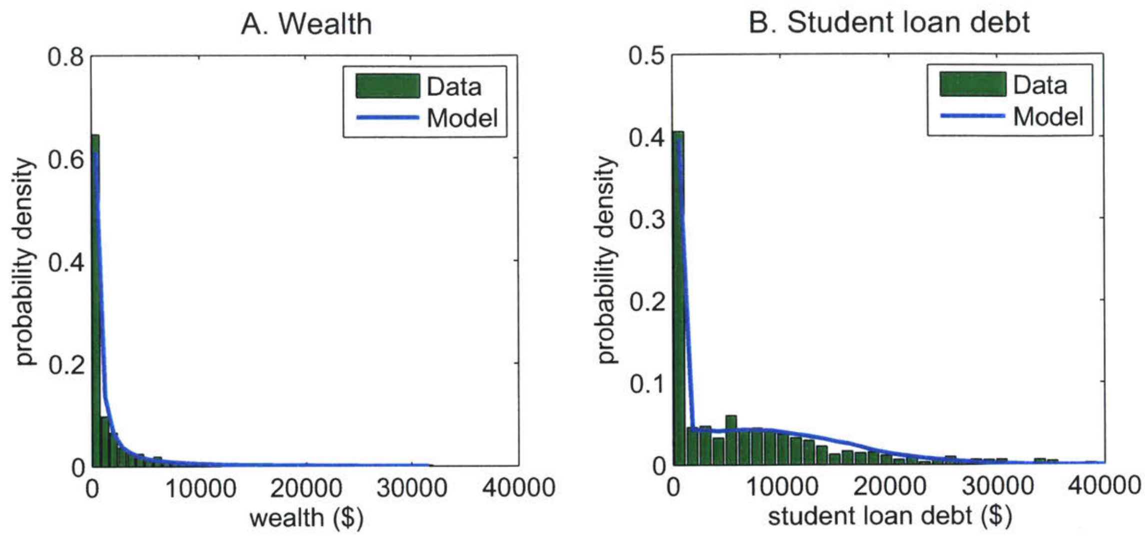


Note: This figure compares the targeted moments of life-cycle earnings profiles between model and data. The solid line represents the earnings profile generated by the model. The dashed line represents the earnings profile in the data, constructed using March CPS 1997-2008 data.

Figure 1-2: Comparing life-cycle earnings profiles between model and data.

College and Debt Moments The parameter μ_e is identified to match the average fraction of students with a bachelor's degree. The parameter σ_e is identified to match the variation in college entry decision not explained by individual talent and wealth. Specifically, I regress the college entry dummy on talent and initial wealth using the actual data and the simulated data. The value of parameter σ_e is identified to match the unexplained variance (i.e., $1 - R^2$).

The parameter ϑ captures the correlation between talent and initial wealth. A larger ϑ suggests that talented agents are wealthier and as a result, demand fewer student loans. Therefore, the



Note: This figure plots the model-generated marginal distribution of wealth and student loan debt. I assume that the exogenous marginal distribution of wealth follows a generalized Pareto distribution, whose parameters are directly estimated using MLE. The endogenous student debt distribution upon college graduation is generated by model simulations.

Figure 1-3: The distribution of initial wealth and student loan debt upon college graduation.

value of ϑ can be identified to match the correlation between individual AFQT score (a proxy of talent) and student debt upon college graduation. In the data, there is a slight positive correlation between AFQT and student debt, 0.04, after controlling for other characteristics.

The default cost η is identified from the equilibrium default rate on student loan debt. Using a random 1% sample of National Student Loan Data System (NSLDS), Yannelis [224] computes that the average two-year cohort default rate for undergraduate borrowers is 9.26% between 1997-2011.

The two parameters (μ_k, σ_k) capturing the net-monetary costs of college study are identified to match the distribution of student loan debt upon college graduation. In the data, about 61.6% of college graduates have outstanding student loans with a mean of $11,873. To fully account for the variation, I use 40 equally spaced moments to capture the empirical histogram of student debt distribution (see panel B of Figure 1-3).[21]

Estimation In total, I choose 93 moments that are sufficient to identify the 21 parameters. I estimate the set of parameters Ξ using MSM:

$$\hat{\Xi} = \underset{\Xi}{\operatorname{argmin}}\, L(\Xi) \tag{1.5.4}$$

[21] It is difficult to directly estimate these two parameters based on college tuition, because in principle students also receive parental transfers, scholarships, and incur living costs (consumption, housing, etc) during college study. My indirect inference suggests that the average total college cost is about $12,505. Data from IPEDS documents that during 2001-2004, the annual college tuition for a four-year college is between $989-$2,520 depending on state category and the national average cost of room and board is $6,532 [141]. This implies a total college cost of $10,488-$16,612.

Table 1.3: Parameters estimated jointly using MSM.

	Labor Market Parameters	Value	Standard Error
A_0	Productivity (high school)	25.1	3.1
A_1	Productivity (college)	40.0	2.6
κ	Exogenous job separation rate	0.31	0.04
h^u	Search intensity during unemployment	4.82	0.56
h^e	Search intensity during employment	1	N/A
χ	Matching efficiency	0.69	0.10
f_1^a	Talent distribution	1.52	0.35
f_2^a	Talent distribution	0.45	0.11
f_1^p	Vacancy productivity distribution	1.41	0.28
f_2^p	Vacancy productivity distribution	0.52	0.09
ν	Flow cost of vacancy creation	49,535	3,569
ϕ	Labor supply scaling factor	6.3×10^{-8}	0.3×10^{-8}
μ_0	Constant term in worker's ability	0.836	0.007
μ_1	Linear term in worker's ability	0.087	0.002
μ_2	Square term in worker's ability	-3.9×10^{-3}	-0.2×10^{-3}
μ_3	Cubic term in worker's ability	5.5×10^{-5}	0.4×10^{-5}
	College and Debt Parameters	**Value**	**Standard Error**
μ_e	Mean of non-monetary college cost	3.0×10^{-9}	0.6×10^{-9}
σ_e	Standard deviation of non-monetary college cost	5.1×10^{-8}	1.0×10^{-8}
ϑ	Correlation between talent and initial wealth	0.47	0.15
η	Default cost	3.0×10^{-8}	0.4×10^{-8}
μ_k	Mean of monetary college cost ($)	12,505	1,238
σ_k	Standard deviation of monetary college cost ($)	15,406	2,429

Note: This figure presents parameter values estimated jointly using MSM following the two-step estimation procedure detailed in Online Appendix A.3. Standard errors are computed by bootstrapping.

The objective function is given by

$$L(\Xi_2) = [\hat{m}_N - \hat{m}_S(\Xi)]^T \hat{\Theta}^{-1} [\hat{m}_N - \hat{m}_S(\Xi)]. \tag{1.5.5}$$

where $\hat{m}_N = \frac{1}{N}\sum_{i=1}^{N} m_i$ is the vector of moments computed in the data. $\hat{m}_S(\Xi)$ is the vector of moments generated by the model simulation in the stationary equilibrium. $\hat{\Theta}$ is a weighting matrix, constructed from the diagonal of the estimated variance-covariance matrix of $\hat{m}_N$ using bootstrapping. Estimates are not sensitive to alternative choices of weighting matrices because most moments are matched well (see Table 1.2).

The asymptotic variance-covariance matrix for MSM estimators $\hat{\Xi}$ is given by:

$$Q(\hat{\Theta}) = (\nabla^T \hat{\Theta} \nabla)^{-1} \nabla^T \hat{\Theta} \widehat{COV} \hat{\Theta}^T \nabla (\nabla^T \hat{\Theta}^T \nabla)^{-1}, \tag{1.5.6}$$

where $\widehat{COV}$ is the variance-covariance matrix of $\hat{m}_N$; and $\nabla = \frac{\partial \hat{m}_S(\Xi)}{\partial \Xi}|_{\Xi=\hat{\Xi}}$ is the Jacobian matrix of the simulated moments evaluated at the estimated parameters.[22] The first derivatives are calculated numerically by varying each parameter's value by 1%. The standard errors of $\hat{\Xi}_2$ are given by the square root of the diagonal elements of $Q(\hat{\Theta})$. Table 1.3 presents the internally estimated parameters.

1.5.3 Validation Tests

I conduct two validation tests to provide a type of out-of-sample evaluation of the structure imposed by the quantitative model. First, I check whether the model can replicate several non-targeted moments in the data. Second, I check whether the model can produce several elasticity measures that are consistent with micro estimates in related literature.

Non-Targeted Moments

As an initial exercise, I check whether the model can replicate the difference in average wage income between non-borrowers and borrowers observed in the data. Figure 1-4 compares the average wage income after college graduation for non-borrowers and all borrowers in the data and model.[23]

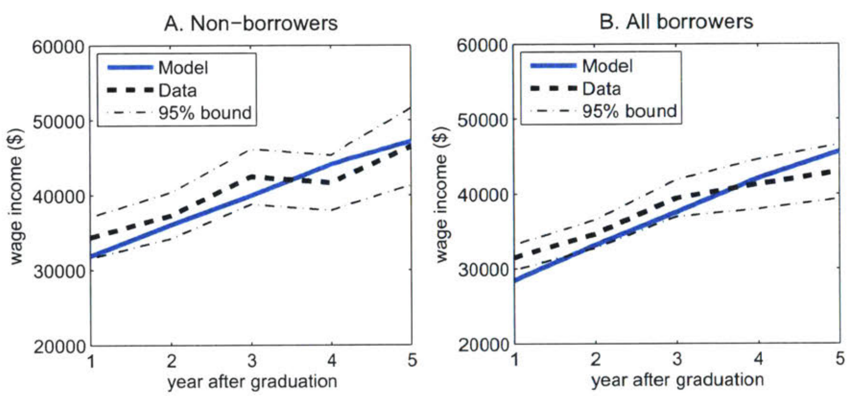


Figure 1-4: Comparing non-targeted moments: annual wage income in the first five years.

Next, I check whether the model can replicate the regression coefficients observed in the data. Based on the NLSY97 sample, I first explore the implication of debt on job search decisions by regressing the duration of the first unemployment spell (Dur_i) after college graduation on the amount of student loan debt (s_i) and control variables X_i including parental wealth, parental education, gender, race, AFQT score, marital status, the cubic age polynomials, and the county of

[22]In general, the formula should also incorporate simulation errors, thus the variance-covariance matrix for MSM estimators also depends on the number of simulated agents [107]. The formula I use does not consider this type of simulation errors because instead of simulating a number of agents, I adopt the histogram method by simulating the distribution of characteristics. Therefore, the simulated values of aggregate moments are not dependent on randomly drawn shocks.

[23]When constructing the moments in the data, I use adjusted data after partialling out the unmodelled covariates that are potentially correlated with both wage income and student loan debt (see Online Appendix A.1.3).

Table 1.4: Comparing reduced-form regression estimates: actual data vs simulated data.

	Uemp. duration First spell	First year	**Wage income** Second year	Third year
Actual data				
"Impact" coefficient	-2.08***	-2,067**	-2,152**	-2,619**
Standard error	(0.68)	(890)	(865)	(1,309)
Simulated data				
"Impact" coefficient	-1.83**	-2,411**	-2,122*	-1,810*
Standard error	(0.70)	(914)	(1,254)	(1,121)
Chow test p-value	0.81	0.83	0.85	0.83

Note: The "impact coefficient" is the coefficient on student loan debt, recorded in units of $10,000. The regressions using actual data also control for parental wealth, parental education, gender, race, AFQT score, marital status, the cubic age polynomials, and the county of residence in graduation year. The Chow test is used to test whether the coefficients from actual data and simulated data are equal to each other. ***, **, indicate significance at the 1 and 5 percent level. Full regression tables of actual data are in Online Appendix A.1.4.

residence in the graduation year:

$$Dur_i = \alpha + \beta_1 s_i + \beta_2 X_i + \varepsilon_i. \tag{1.5.7}$$

I then explore the implication of debt on wage income by regressing the annual wage income in the first three years after college graduation on the amount of student loans:

$$Wage_{i,t} = \alpha_t + \beta_{1,t} s_i + \beta_{2,t} X_{i,t} + \varepsilon_{i,t}, \quad \text{for } t = 1, 2, 3. \tag{1.5.8}$$

Turning to the model, I simulate the same number of college graduates over their life cycles based on the equilibrium policy functions, job contact rates, randomly drawn wage offers from $F(\rho)$, and separation shocks at rate κ. I do this 500 times to create 500 simulated datasets. I regress the duration of the first unemployment spell and annual wage income on the amount of student loan debt controlling for individual talent for each simulated dataset to construct the mean and standard errors of the estimates. Table 1.4 compares the regression results of the model to the data and shows that the model does generally quite well in replicating the results. In particular, a $10,000 increase in the amount of student loans reduces unemployment duration by about 1.8 weeks, and reduces the annual wage income by about $2,000 in the first three years after college graduation. The final row reports the p-value of the Chow test, where the null is no structural break between the actual and simulated data. The Chow test shows formally that the regression estimates from the model are statistically similar to those in the data at a 5% significance level.

Comparison to Micro Estimates

I now check whether the model can produce elasticity measures that are consistent with the micro estimates in related literature. When conducting the following experiments, I focus on partial-equilibrium counterfactual simulations in which the job contact rates and fiscal parameters

Table 1.5: Comparison to micro estimates.

	Model	Micro Estimates	Source
UI on unemp. dur.	0.49	0.35-0.9	Card et al. [53]
UI on res. wage	6.4%	4%	Feldstein and Poterba [92]
Credit on unemp. dur.	0.7 week	0.15-3 weeks	Herkenhoff, Phillips and Cohen-Cole [129]
Credit on reemploy. wage	1.4%	0.8%-1.7%	Herkenhoff, Phillips and Cohen-Cole [129]

Note: This figure compares the model-implied structural estimates with micro estimates. The elasticity of unemployment duration with respect to UI benefits is estimated by simulating the counterfactual with UI benefits θ being increased by 25%, from \$8000 to \$10000, corresponding to a 10% increase in UI replacement rate, from 40% of 6-month earnings to 50%. The effect of UI benefits on the reservation wage also estimated from this counterfactual. The duration and earnings replacement elasticities with respect to unused credit limit are estimated from newly laid off agents due to exogenous job separations in the model. The counterfactual simulation relaxes credit constraints for these agents by 10% of their wage income in previous jobs. The elasticity is estimated using the average difference in unemployment duration and wage income between the baseline economy and the counterfactual economy with relaxed credit constraints.

are fixed, so that the elasticities are estimated in a context consistent with the setting in which the micro estimates are obtained. Note that all the elasticities I structurally estimate are based on global elasticities, although some of the micro estimates are local elasticities (see Table 1.5).

I begin by examining whether the model-implied elasticity of unemployment duration with respect to UI benefits matches the micro estimates using U.S. data. The positive effect of unemployment insurance on unemployment duration is one of the most robust empirical findings. The effect of UI benefits is also delivered from a channel related to job seekers' liquidity constraints, as Chetty [59] argues that the liquidity effect accounts for 60% of the impact of UI. To estimate the elasticity, I simulate a counterfactual by increasing UI benefits θ by 25%, from \$8000 to \$10000, corresponding to a 10% increase in UI replacement rate, from 40% of 6-month earnings to 50%. I find that the average unemployment duration increases by about 3.3 weeks, implying that the elasticity of unemployment duration with respect to UI benefits is about 0.49. This elasticity is roughly in line with the estimate of Card et al. [53], who find that the elasticity is around 0.35 during the pre-recession period (2003-2007) and between 0.65 and 0.9 during the recession and its aftermath.

Next, I check whether the model-generated responses of the reservation wage and average wage income to UI benefits are in line with the micro estimates. The estimate of Feldstein and Poterba [92] indicates that a 10% increase in the UI replacement ratio raises the reservation wage by 4% for job losers who are not on layoff. My model generates a larger response in the reservation wage, 6.4%. The empirical evidence on the effect of UI benefits on reemployment wages is mixed. My model's simulation results indicate that reemployment wages increase by about 4.1% following a 10% increase in the UI replacement rate.

Finally, I use the model to estimate the elasticities of unemployment duration and earnings to relaxed borrowing constraints for displaced workers. I then compare these structural estimates with the micro estimates of Herkenhoff, Phillips and Cohen-Cole [129]. Using administrative data from TransUnion and Longitudinal Employment and Household Dynamics (LEHD), Herkenhoff, Phillips and Cohen-Cole [129] find that increasing credit limits by 10% of prior annual earnings would lead displaced workers to take 0.15 to 3 weeks longer to find a job. Among job finders, the replacement earnings increased by 0.8% to 1.7%.

To evaluate the impact of access to credit on job search and wage income, I isolate agents newly laid off due to exogenous job separations in the model. Denote their prior wage income as $Inc_{-1}(\Omega_{-1}, \rho_{-1}, \rho'_{-1})$ and the set of agents as I_{κ}. I then simulate these agents' over time until they find the next job, and obtain unemployment duration, $Dur(\Omega)$, and wage income, $Inc(\Omega, \rho, \rho')$. Finally, I run a counterfactual in partial equilibrium to obtain the unemployment duration, $Dur^{\Delta}(\Omega)$, and wage income, $Inc^{\Delta}(\Omega, \rho, \rho')$ if these agents were provided with 10% unused credit during unemployment, i.e., the borrowing constraint is relaxed from $b \geq 0$ to $b \geq -0.1 Inc_{-1}(\Omega_{-1}, \rho_{-1}, \rho'_{-1})$.

Following Herkenhoff, Phillips and Cohen-Cole [129], I estimate the duration and earnings elasticity using the following formulas:

$$\epsilon_{dur} = \sum_{I_{\kappa}} \frac{Dur^{\Delta}(\Omega) - Dur(\Omega)}{10\%}, \tag{1.5.9}$$

$$\epsilon_{inc} = \sum_{I_{\kappa}} \frac{[Inc^{\Delta}(\Omega, \rho, \rho') - Inc(\Omega, \rho, \rho')]/Inc_{-1}(\Omega_{-1}, \rho_{-1}, \rho'_{-1})}{10\%}. \tag{1.5.10}$$

The structural estimates of ϵ_{dur} and ϵ_{inc} are 0.13 year and 0.14. Therefore, the model predicts that in response to a 10% increase in unused credit, unemployed workers will take 0.7 week longer to find a job that on average pays 1.4% more wage income, roughly in line with the micro estimates of Herkenhoff, Phillips and Cohen-Cole [129].

1.6 Quantitative Analyses

In this section, I use the estimated model to conduct quantitative analyses. First, I illustrate the college entry and borrowing decisions under the fixed repayment plan. Second, I present the effect of student debt burden on labor market outcomes in partial equilibrium. Finally, I conduct the full general equilibrium analyses to shed light on the welfare implications of student loan debt.

1.6.1 College Entry and Borrowing

The model implies that more talented agents are more likely to attend college because of the higher college premium captured by equation (1.4.1). Among college graduates, the model is able to capture the small positive correlation between talent and student loan debt, consistent with the data. In terms of talent distribution, Figure 1-5 shows that the distribution of talent among college borrowers, college non-borrowers, and high school graduates can be ranked by first-order stochastic dominance, with the average group talent being 0.851, 0.843, and 0.823, respectively.

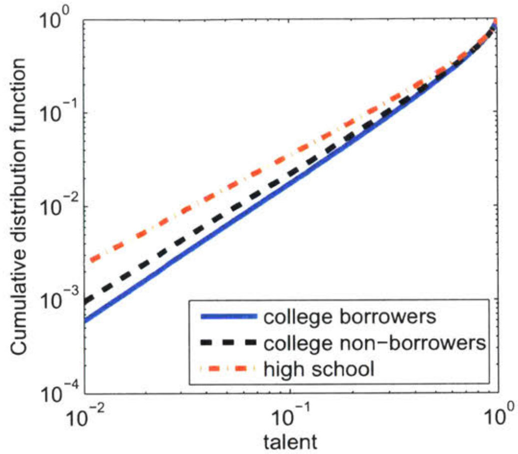


Figure 1-5: Model-implied talent distribution for high school and college graduates.

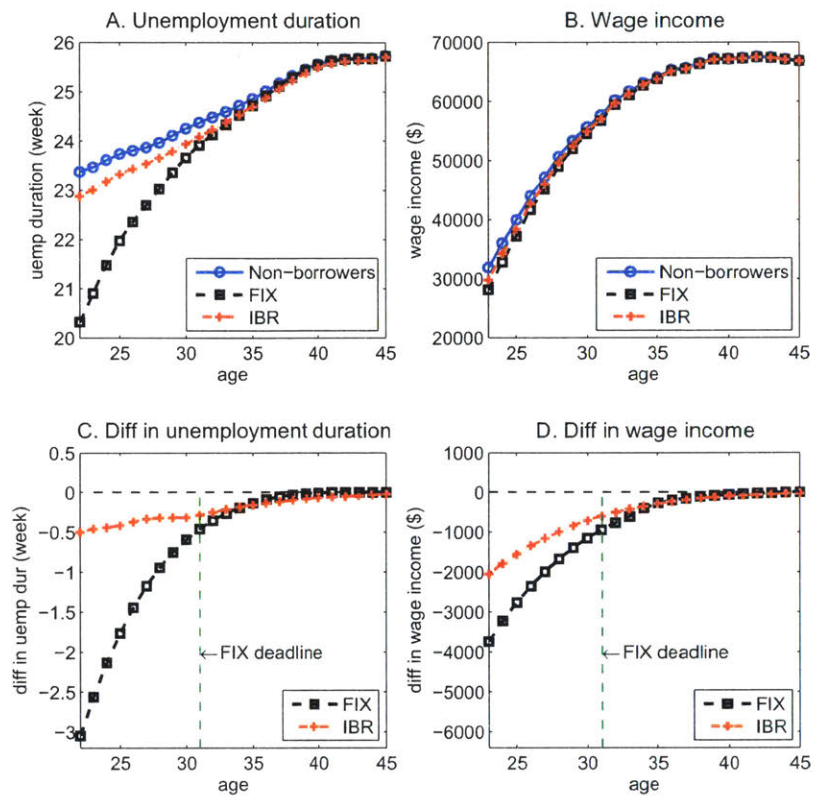
1.6.2 Student Debt on Labor Market Outcomes

In this subsection, I compare the labor market statistics between non-borrowers and normalized borrowers. I first use the model to look at the life-cycle implications of student loan debt under the two repayment plans. I then conduct two counterfactual analyses to evaluate and dissect the effects of IBR.

The difference in the composition of talent suggests that, all else equal, borrowers are also more likely to have higher earnings after leaving school because they are on average more talented. Therefore, the talent difference generates a selection effect that would confound the effect of the debt burden on earnings. To isolate the direct effect of student debt burden on labor market outcomes, in this subsection, I focus on borrowers with normalized talent (I call them normalized borrowers hereafter). Specifically, I normalize the talent of all student loan borrowers by a factor of 0.991 such that the mean talent of borrowers is equal to the mean talent of non-borrowers. Relatedly, when evaluating IBR, I consider what would happen to normalized borrowers if they are unexpectedly allowed to enroll in IBR to cleanly quantify the direct insurance effect of IBR. Therefore, the analysis in this section does not consider the three general equilibrium effects after the economy adopts IBR: (1) the change in college entry and borrowing decisions; (2) the change in firms' job posting decisions; (3) budget balancing adjustments in fiscal tax parameters. The endogenous difference in talent composition and the general equilibrium effects are considered in subsection 1.6.3.

Life-Cycle Outcomes

To evaluate the long-term effect of the debt burden, I simulate the model and track the movements of average unemployment duration and wage income over the life cycle for a single generation. In Figure 1-6, I plot these aggregate statistics for non-borrowers and normalized borrowers under the fixed repayment plan and IBR.



Note: This figure plots the life-cycle unemployment duration and wage income for non-borrowers and normalized borrowers under the two repayment plans. In panels A and B, the blue solid line plots the average unemployment duration and wage income for non-borrowers. The black dashed line plots the average unemployment duration and wage income for normalized borrowers under the fixed repayment plan. The red dash-dotted line plots the average unemployment duration and wage income for normalized borrowers under IBR. Panels C and D plot the difference in unemployment duration and wage income between non-borrowers and normalized borrowers under the two repayment plans.

Figure 1-6: Simulated life-cycle unemployment duration and wage income.

Panels A and C show that non-borrowers on average spend 3 weeks more when searching for their first jobs compared to normalized borrowers under the fixed repayment plan. The inadequate job search translates into a negative effect on wage income. Panels B and D indicate that non-borrowers on average earn about $3,700 more relative to normalized borrowers in the first year after college graduation.

Figure 1-6 also reveals that at age 32, even after debt has been paid off, normalized borrowers under the fixed repayment plan still spend less time on job search and earn relatively less. This long-term effect of the debt burden is attributed to lower savings. Between ages 22-31, borrowers accumulate significantly less wealth compared to non-borrowers due to lower wage income and debt repayment. The average wealth among normalized borrowers at age 31 is about $8,000 lower compared to that of non-borrowers. Although there no longer exists any pressure from debt repayment after age 32, the lower wealth would continue affecting borrowers' job search decisions through a mechanism similar to that of debt repayment. This asset accumulation channel is likely to be empirically relevant as Elliott, Grinstein-Weiss and Nam [88, 89] provide evidence that households with outstanding student loan debt have fewer assets. In addition, because of the low search intensity for employed workers, borrowers are stuck at their lower-paid jobs for a relatively longer time. In other words, "first jobs" matter precisely because job-to-job transitions are rare in both my model and data.

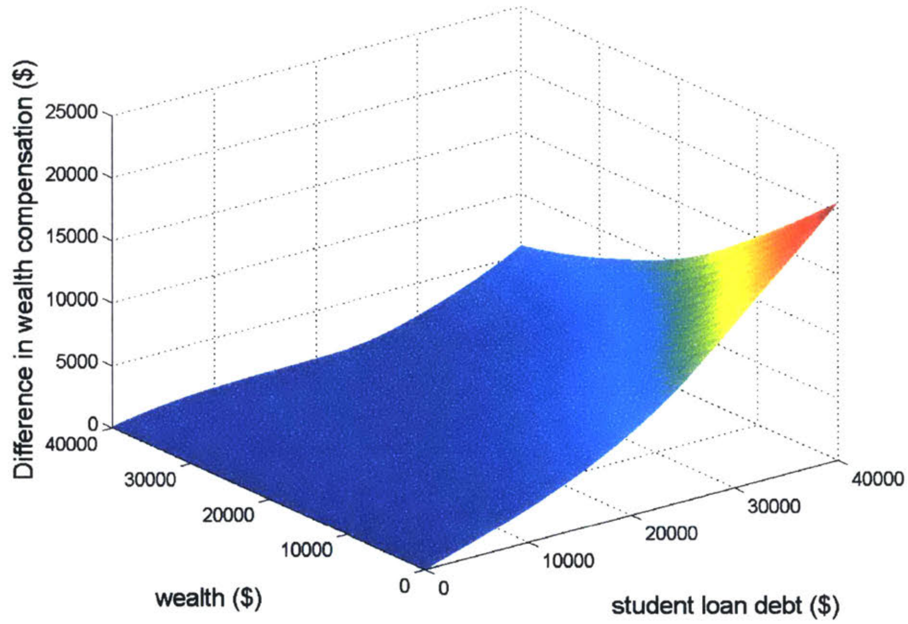
IBR significantly alleviates the distortion of the debt burden on job search decisions. Immediately after college graduation, normalized borrowers under IBR on average spend 22.9 weeks searching for their first jobs, which is 0.5 week below the average of non-borrowers. As a consequence of longer job search, the difference in initial wage income is about $2,000 between non-borrowers and normalized borrowers under IBR.

Distributional Implications on Debt Alleviation

I now evaluate the distributional implications of IBR on alleviating the debt burden. Following Townsend and Ueda [218], I proxy the debt burden using wealth compensation. In particular, for any borrower who just graduated from college, the student debt burden is measured as the amount of wealth that should be transferred to the agent for her to have the same utility as a non-borrower of the same characteristics. In other words, the non-borrower would be indifferent about accepting the debt and the associated wealth compensation at the same time.

I measure the debt alleviation achieved by IBR by calculating the difference in wealth compensation between normalized borrowers under the fixed repayment plan and those under IBR.[24] To assess the distributional effect, I do this calculation for normalized borrowers of the average talent but different levels of wealth and student loan debt. Figure 1-7 illustrates that adopting IBR has significant distributional implications. Borrowers who are poorer and more indebted would benefit more by switching to IBR because they ask for larger wealth compensation. The implication of IBR's distributional benefits coincides with the characteristics of borrowers enrolled in income-driven repayment plans in reality. The Executive Office of the President of the United States [214] documents that undergraduate-only borrowers in income-driven repayment plans have a median outstanding debt of $25,000 compared with $10,000 in the fixed repayment plan in 2015. Moreover, the average family income based on the first application for federal student aid is $45,000 for those in income-driven repayment plans compared with $57,000 in the fixed repayment plan.

[24]Therefore, the numbers reported in Figure 1-7 can be considered as the least amount of wealth compensation that induces the agent to switch from IBR to the fixed repayment plan.



Note: This figure illustrates the distributional effect of IBR. I measure the change in the debt burden caused by IBR using wealth compensation. For any borrower who just graduated from college, I calculate the least amount of wealth transfer that induces the borrower to switch from IBR to the fixed repayment plan. The figure plots the wealth compensation for borrowers of different levels of wealth and student loan debt. It is shown that borrowers who are poorer and more indebted benefit more by switching to IBR.

Figure 1-7: The simulated distributional effect of IBR.

Average Effects on Borrowers

In this subsection, I evaluate the average effects of the fixed repayment plan and IBR on normalized borrowers' labor market outcomes. In particular, I calculate the average debt burden (measured by wealth compensation), unemployment duration, match quality (measured by job productivity), wage income, output, and labor supply for three groups of young agents: non-bororwers, normalized borrowers under the fixed repayment plan, and normalized borrowers under IBR between ages 23-32.

These results are reported in the first three columns of Table 1.6. Normalized borrowers have \$10,370 debt on average and they ask for \$6,274 wealth compensation under the fixed repayment plan and \$3,003 under IBR. This suggests that allowing borrowers to have access to IBR would alleviate their debt burden by about half. Note that although there is debt forgiveness provided by IBR after 25 years, my simulation results indicate that almost the entire debt in the economy is repaid by most borrowers. This implies that the debt alleviation caused by IBR is almost entirely driven by insurance.

Table 1.6 also shows that non-borrowers spend 23.8 weeks on job search on average in their first 10 years. Normalized borrowers on average spend 1.8 weeks fewer when they are under the fixed repayment plan and about 0.4 week fewer under IBR.

Table 1.6: Evaluation of IBR.

	Non-borrowers	Normalized borrowers			Difference	
		FIX	IBR	IBR(w^*_{FIX})	IBR-FIX	IBR(w^*_{FIX})-IBR
Compensation ($)	N/A	6,274	3,003	4,214	-3,271	1,211
Unemp. dur. (week)	23.8	22.0 (-7.6%)	23.4 (-1.7%)	22.4 (-5.9%)	1.4 (5.9%)	-1.0 (-4.2%)
Match quality	0.836	0.812 (-2.9%)	0.826 (-1.2%)	0.813 (-2.8%)	0.014 (1.7%)	-0.013 (-1.6%)
Wage income ($)	47,697	45,689 (-4.2%)	46,586 (-2.3%)	45,121 (-5.4%)	897 (1.9%)	-1,465 (-3.1%)
Output ($)	60,235	57,976 (-3.8%)	58,756 (-2.5%)	56,862 (-5.6%)	780 (1.3%)	-1,894 (-3.1%)
Labor supply (hour)	1,737	1,724 (-0.7%)	1,711 (-1.5%)	1,695 (-2.4%)	-13 (-0.8%)	-16 (-0.9%)

Note: This table compares the aggregate implications of the debt burden in the first 10 years after college graduation under the fixed repayment plan and IBR. Column "FIX" reports outcomes under the fixed repayment plan for normalized borrowers. Column "IBR" reports outcomes under IBR for normalized borrowers. Column "IBR(w^*_{FIX})" reports the effect of IBR when reservation wages are fixed at the values under the fixed repayment plan. Statistics in parentheses report the relative percent change using non-borrowers as the benchmark.

In terms of match quality, normalized borrowers under the fixed repayment plan are on average matched with jobs that are 2.9% less productive relative to jobs associated with non-borrowers. IBR improves match quality by about 1.7% for normalized borrowers. The lower match quality translates to lower output and wage income. On average, normalized borrowers under the fixed repayment plan produce 3.8% and earn 4.2% ($2,008) less annually compared to non-borrowers in the first 10 years after college graduation. Note that at the estimated parameter values, normalized borrowers already need to repay about $1,500 every year under the fixed repayment plan. This suggests that debt repayment imposes a double burden on consumption. The indirect reduction in consumption due to inadequate job search is larger than the direct negative effect from debt repayment, which generates even larger consumption inequality between borrowers and non-borrowers.

IBR makes job search much more affordable, and as a result, output and wage income are increased by about 1.3% and 1.9% for normalized borrowers. Output increases precisely because IBR increases borrowers' reservation wages, implying that the increase in wage income is not entirely caused by a redistribution of profits from firms to workers.

The negative effect on labor supply introduced by IBR is not large. Normalized borrowers work for 1,724 hours on average under the fixed repayment plan, and for 1,711 hours under IBR. The reduction in labor supply after borrowers switch to IBR is 0.8%, which is much smaller than the value suggested by a simple back-of-the-envelope calculation, i.e., 15% (repayment ratio) $\times$ 0.33 (tax-modified elasticity of labor supply) $\approx$ 5%. The small negative effect on labor supply is due to the following reasons. First, because there is not much debt forgiveness in my simulation, the labor supply distortion of IBR is much smaller compared to that of income taxes. With income taxation, people have less incentive to supply labor because a fraction of income is reaped by the government. However, this is not the case for IBR if there is no debt forgiveness in the end. Intuitively, although increasing labor supply increases debt repayment in the current period, it lowers total repayment made in the future.[25] Second, the payment under IBR is capped by the amount under the fixed repayment plan. This implies that if borrowers' earnings are high enough to hit the repayment cap, then IBR would have no distortion on labor supply. Finally, there is a positive substitution effect from having better jobs. IBR improves the job quality of borrowers, which incentivizes them to increase labor supply. This partially offsets the negative substitution effect caused by proportional repayment.

Implication on College Premium My simulation suggests that considering endogenous job search decisions is important for the estimation of college wage premium. In my model, the average wage income for high school graduates is about \$30,505. If college entry is not financed by student debt, entering college increases average wage income to \$47,697, generating a college premium of \$17,192. If instead college entry is financed by student debt, under the fixed repayment plan, the college premium is reduced by about \$2,008, or 11.7%. This suggests that a naive estimate based on the average college premium might overestimate the welfare benefits of higher college entry rates caused by student debt because the debt burden also affects job search decisions after college. Moreover, my analysis implies that there is room for policy intervention. The adoption of IBR could potentially alleviate the debt burden and increase wage income. In the subsection 1.6.3, I conduct a full general equilibrium analysis to shed light on the welfare implication of student debt under the two repayment plans.

Isolate the Reservation Wage Effect

I now use the model to separately quantify the positive reservation wage effect induced by IBR. I conduct one additional experiment, in which borrowers are allowed to make payments according to IBR, but their reservation wages are fixed at the values under the fixed repayment plan. Therefore, in this experiment, IBR provides consumption smoothing but not job search benefits.

The simulation results are reported in column 4 of Table 1.6, "IBR(w^*_{FIX})". The wealth

[25]This intuition can be more clearly illustrated in a two-period model. Suppose that the agent has utility $u(c_1, l_1) + u(c_2, l_2)$. The agent lives hand-to-mouth with no discounting and has to repay debt S under IBR with repayment ratio α. The budget constraints are $c_1 = (1-\alpha)wl_1$ and $c_2 = wl_2 - (S - \alpha wl_1)$. Taking the first order condition at $t = 1$, we obtain

$$wu_1((1-\alpha)wl_1, l_1) - \alpha w[u_1((1-\alpha)wl_1, l_1) - u_1(wl_2 - S + \alpha wl_1, l_2)] = u_2((1-\alpha)wl_1, l_1) \tag{1.6.1}$$

The second term on the LHS of equation (1.6.1) captures the distortion on labor supply. If instead the debt left at $t = 2$ is completely forgiven by the government, then the second term becomes $-\alpha wu_1((1-\alpha)wl_1, l_1)$, which makes IBR equivalent to income taxation at $t = 1$. This example suggests that if there is no debt forgiveness at the end of the IBR repayment period, the labor supply distortion of IBR would depend on the difference in the marginal utility of consumption across periods, which is much smaller than what the back-of-the-envelop calculation implies.

compensation is $4,214 on average for normalized borrowers if reservation wages are fixed. However, if reservation wages are allowed to adjust, the wealth compensation would be $3,003. Therefore, the adjustment in reservation wages caused by IBR on average contributes to a reduction in the wealth compensation by about $1,211. This implies that about one-third of the difference in the wealth compensation between the fixed repayment plan and IBR is attributed to the positive reservation wage effect, and the remaining is due to better consumption smoothing.

Moreover, it is not surprising that almost the entire improvement in match quality is caused by the positive response in reservation wages. When this channel is shut down, borrowers' average wage income and output become even lower than those under the fixed repayment plan. This highlights the standard tradeoff between welfare and output, namely, providing insurance increases welfare but potentially lowers output by distorting the incentive to work.

1.6.3 General Equilibrium Implications of Student Debt

In this subsection, I shed light on the general equilibrium implications of student debt, taking into account the change in college entry, borrowing, firms' job posting decisions, and fiscal tax parameters.

Table 1.7 reports the fraction of college graduates, average debt, wage income, output, match quality, job contact rate, and welfare for all young agents (between ages 23-32) in the economy. Column "FIX" presents the baseline simulation in which all student loan borrowers repay their debt under the fixed repayment plan. Column "IBR"-(1) presents the IBR counterfactual in which all borrowers repay under IBR.

Table 1.7: General Equilibrium Implications of Student Debt.

	FIX	**IBR**		
		(1)	(2)	(3)
Fraction of college graduates	41.4%	47.5%	47.7%	41.4%
Fraction of borrowers	62.2%	67.5%	67.6%	62.2%
Average debt among borrowers ($)	10,370	16,960	17,013	10,370
Job contact rate	0.82	0.88	0.82	0.82
Wage income ($)	37,212	38,452	38,018	37,445
		(3.3%)	(2.2%)	(0.6%)
Output ($)	45,600	46,512	46,317	45,829
		(2.0%)	(1.6%)	(0.5%)
Welfare (%)		2.4%	1.9%	0.8%

Note: This table presents the general equilibrium implications of student debt for all young agents (between ages 22-32). Column "FIX" reports outcomes under the fixed repayment plan. Column "IBR"-(1) reports outcomes under IBR. Column "IBR"-(2) reports outcomes under IBR when the equilibrium job contact rates are set equal to those under the fixed repayment plan. Column "IBR"-(3) reports outcomes under IBR when the equilibrium job contact rates, the college entry, and borrowing decisions are set equal to those under the fixed repayment plan. Statistics in parentheses report the relative percent change using column "FIX" as the benchmark.

In the baseline economy, 41.4% agents choose to attend college, among which about 62.2%

finance their education by borrowing student debt, with an average amount of $10,370. When borrowers are allowed to repay under IBR, the reduction in repayment burden induces more agents to borrow student debt to attend college, increasing the college attendance rate by about 6.1%. Among college graduates, about 67.5% are borrowers with an average amount of $16,960 debt.

The adoption of IBR increases average wage income and output by 3.3% and 2.0%. The equilibrium job contact rate is also higher under IBR. This is because college graduates are more productive compared to high school graduates at any jobs. Thus the increase in college entry rate increases firms' profits, motivating firms to post more vacancies.[26]

Following Abbott et al. [1], I measure the change in welfare by considering the percentage change of lifetime consumption for a newborn economic agent (at age t=0) before drawing her initial conditions (wealth and talent). The last row of Table 1.7 indicates that switching from the fixed repayment plan to IBR increases the welfare of young agents by about 2.4%.

Decomposition It is clear from above analyses that IBR increases social welfare through three channels. First, borrowers conduct more adequate job search because of better insurance in the labor market. Second, debt alleviation achieved through IBR induces a general equilibrium effect that encourages college entry and borrowing. Third, improved education increases match-specific productivity and profits, motivating firms to post more jobs. I now run two additional counterfactual experiments to quantify the importance of these channels.

Column "IBR"-(2) reports outcomes under IBR when the equilibrium job contact rates are set equal to those under the fixed repayment plan. The difference between columns "IBR"-(1) and "IBR"-(2) is thus informative about the importance of more job postings. My simulation suggests that reducing the job contact rate from 0.88 to 0.82 for unemployed agents would reduce the wage income and output by about 1.1% and 0.4%. In terms of welfare, IBR increases consumption by an additional 0.5% by incentivizing firms to post more jobs.

Column "IBR"-(3) reports outcomes under IBR when the equilibrium job contact rates, the college entry, and borrowing decisions are set equal to those under the fixed repayment plan. Thus Column "IBR"-(3) quantifies the importance of better job search and insurance in the labor market, and the difference between columns "IBR"-(2) and "IBR"-(3) is informative about the importance of more college entry and borrowing. My simulation implies that insurance in the labor market increases wage income, output, and welfare by about 0.6%, 0.5%, and 0.8%, while more college entry and borrowing increases these statistics by about 1.6%, 1.1%, and 1.1%.

Connection to Tuition Subsidy Note that my debt burden analysis in Table 1.6 suggests that for an average borrower with $10,370 debt, introducing IBR would cut the debt burden by about half, from $6,274 to $3,003. This implies that if the average borrower under the fixed repayment

[26]There is also a countervailing effect from IBR. When borrowers become pickier under IBR, they set higher reservation wages and reject more wage offers. This dampens firms' profits and their incentive to post vacancies. This effect, however, is dominated by the main effect from a higher college entry rate.

plan chooses to enroll in IBR, loosely speaking, it is equivalent to having $\$10,370 \times \frac{3,003}{6,274} = \$4,963$ debt under the fixed repayment plan. Table 1.7 shows that switching from the fixed repayment plan to IBR would also increase the average debt amount to \$16,960, implying an equivalent debt amount of $\$16,960 \times \frac{3,003}{6,274} = \$8,118$ under the fixed repayment plan. This further implies that the ex-ante effect of introducing IBR is similar to a tuition subsidy by an average amount of $\$10,370 - \$8,118 = \$2,252$ if borrowers finance their education using the fixed repayment plan.

If we treat the estimated monetary college cost, $\mu_k = 12,505$, as mainly reflecting tuition, then my estimation implies a college enrollment elasticity of $\frac{6.1\%/41.4\%}{2,252/12,505} = 0.82$ with respect to college price. In Online Appendix A.7, my sensitivity analyses indicate that the elasticity is around 0.58 if the risk aversion parameter γ is set to be 1.5. If agents are allowed to borrow 18.5% of income after graduation, the elasticity drops to about 0.72. The existing micro estimates surveyed by Leslie and Brinkman [161] and Kane [143] suggests that the elasticity is between 0.52 and 0.83. Therefore, my simple back-of-the-envelop calculation seems to suggest that the model-implied elasticity from providing IBR is in line with the existing estimates from micro studies.

In a related study, Johnson [141] finds that relaxing borrowing constraints does not increase the college entry rate much compared to a tuition subsidy. In my model, agents do not face borrowing constraints when making college entry decisions. My model suggests that the constraint on college entry is more likely due to the repayment burden after college, when agents start searching for their jobs. Enrolling borrowers in IBR alleviates the debt burden, which is essentially similar to a tuition subsidy.

In contrast to a tuition subsidy, providing IBR is less costly from the government's perspective. My simulation suggests that almost every borrower can repay their debt within 25 years, and there is not much debt forgiveness in the end. As a result, the fiscal parameter $\varkappa$ is virtually unchanged during my simulation.[27]

1.7 Conclusion

In this paper, I develop a structural model with college entry, borrowing, and job search to evaluate the implication of student debt on labor market outcomes.

This paper contributes to existing literature in three ways. First, this paper proposes a mechanism through which the debt burden affects individuals' job search decisions and labor market outcomes. I illustrate that borrowers tend to be less patient in job search, and consequently, they are more likely to end up in lower-paid jobs. The exact effect of the debt burden also depends on the repayment schedule.

Second, this paper develops and estimates a quantitative framework that incorporates college entry and borrowing into an equilibrium search model. I estimate the model using NLSY97 data and apply the model to evaluate the aggregate and distributional implication of the mechanism.

[27] In fact, my simulation implies that tax revenue goes up slightly because the increase in college entry rate increases average taxable wage income.

Third, this paper applies the model to shed light on the implication of a realistic income-based repayment plan on welfare, wage income, and output. The counterfactual analyses also separately quantify the potential general equilibrium effects of IBR on inducing more college entry, borrowing, and job postings.

My main quantitative exercise suggests that the distortion of student debt on job search decisions could be large under the fixed repayment plan, and that it is very much relieved by IBR. Debt alleviation achieved by IBR benefits poorer and more indebted borrowers more, and on average, it is equivalent to cutting student debt by half. One-third of debt alleviation is attributed to better job matches. Moreover, adopting IBR also brings two general equilibrium effects that encourage college entry and job postings, effectively offering a tuition subsidy that is much less costly to the government.

Chapter 2

External Financing and Customer Capital: A Financial Theory of Markups

2.1 Introduction

We provide a unified theoretical framework based on financial frictions and imperfect product markets to rationalize the impact of financial slack on markups and offer a set of joint predictions on firms' financing, investment and product price setting behaviors. Our financial theory of markups is motivated by the observation that during the 2007-2009 Great Recession, there was a lack of deflationary pressure on product prices while investment and output suffered large declines in the U.S..[1] This phenomenon is believed to be deeply linked to the controversial debate on the cyclicality of markup dynamics. Despite the extensive debates on cyclicality of markups, there are few formal dynamic models that analyze the effect of firms' financial slack on markup dynamics.[2] We construct an analytically tractable dynamic investment model integrating customer markets with financial constraints to explicitly link markups to firms' financial slack.

In the standard corporate theory of investment and external financing/capital structure, the product market is typically assumed to generate exogenous stochastic cash flows, and firms' financial decisions are usually independent of their decisions in the product market [e.g. 40, 41, 42, 73, 74, 94, 114, 130, 131, 158, 159, 160, 173]. A key contribution of our paper to this literature is that our model features endogenous cash flows which are affected by the optimal choice of product prices. As a result, product market decisions are interlinked to investment and financial decisions.

In our model, the manager is knowledgeable about choosing product prices endogenously in an imperfect product market to balance the tradeoff between current profits and future customer

[1] One recent relevant empirical study is Gilchrist et al. [102] which exploits firm-level data, constructed by merging the U.S. Producer Price Index (PPI) and Compustat datasets, and finds that firms with weak balance sheets increased their product prices significantly relative to industry averages during the height of the crisis in late 2008.

[2] A few exceptions include Chevalier and Scharfstein [60] and Gilchrist et al. [102]. However, the existing theories are not rich enough because endogenous investment and cash hoarding are not allowed. In principle, firms could take precautions ahead of time (e.g., by accumulating cash) so as to circumvent financial problems during a crisis.

base.[3] To set up the frictional environment of product price setting for the manager, on top of the short-term demand effect of product prices (i.e., the intra-temporal demand effect) emphasized in most macroeconomic models, we incorporate the inter-temporal demand effect by introducing the frictional customer market that is originated and formulated in the seminal work of Phelps and Winter [194]. In order to build or keep its customer base, the firm sacrifices its average current profits by strategically reducing its product prices, because a lower product price is more likely to retain existing customers and attract new ones, which could increase its long-term average profits. By contrast, by increasing its product prices, the firm can increase its current profits at the expense of losing customer base, which puts future profits and growth at risk. As emphasized by Rotemberg and Woodford [205], Gilchrist et al. [102] and Gourio and Rudanko [109], the long-term nature of customer base and the upfront-paid costs for customer acquisition render customer base a form of intangible assets held by the firm. Hence, the manager's product price setting decisions are in part investment decisions.[4] Based on this key intuition, we show that product market frictions have nontrivial implications for the firm's dynamics, and in turn, the firm's price setting behavior is also affected by its financial slack. This links to the the "unified q theory" developed by Bolton, Chen and Wang [41]. While the marginal value of liquidity determines the firm's effective marginal cost of investment, we emphasize that the price setting behavior, as a form of investment in intangible assets, is also determined by a q theory with marginal financing costs incorporated.[5] In this sense, our work extends the "unified q theory" to intangible asset investment.

Our model provides a range of novel testable empirical predictions. The first important result concerns how a firm's financial slack affects its product price setting decisions. When setting its product price, the firm in a customer market always balances the tradeoff between current operating revenue and future growth in customer base. When the firm's financial condition is weak, the marginal value of cash is high, which motivates it to set a higher product price in order to mitigate liquidity problems. On the other hand, with abundant internal funds, the firm will stick to a relatively low price in order to build up its customer base. Our model demonstrates

[3]There are numerous examples of imperfect product markets which arise from customers consumption inertia or imperfect information. Chintagunta, Kyriazidou and Perktold [61] reveal that several yogurt brands exist large habit effects. Bhattacharya and Vogt [34] provide both theoretical justification and empirical evidence that the drug industry has consumption inertia. Prices of a new drug are kept low and advertising levels are high early in the life cycle in order to build public knowledge about the drug. As knowledge grows, prices rise and advertising falls. Collado and Browning [67] present evidence that food outside home, alcohol and tobacco are associated with consumption inertia.

[4]The strategy of building up customer base through a lower price is prevalent in the automobile and airline industries. Any airline that recovers from a problem announces discounts to bring back fliers. We have seen Air India and Jet Airways doing it in the past post the pilots' strike. More recently, in February 2015, an India airline company, SpiceJet, announced the launch of a new low fare offer as an attempt to win back the trust of the passengers after recovering from financial distress. In March 2010, Toyota fired the first volley when it announced a variety of discount financing and special lease deals to try to regain market share in the U.S. after a series of recalls. To penetrate in the electric car market, in 2014, Nissan boasts the lowest-price electric car in the U.S., after dropping the base price of a Nissan Leaf by $6,400 earlier this year.

[5]Thus, our paper is also related to the literature on the relevance of intangible assets, including Ai and Kiku [9], Ai, Croce and Li [10], Atkeson and Kehoe [19], Belo, Lin and Vitorino [30], Eisfeldt and Papanikolaou [86], Gourio and Rudanko [109], Hall [118], McGrattan and Prescott [177].

that the time-varying financial slack can generate a strong incentive for firms to manipulate their product prices.[6] This mechanism is reminiscent of the empirical findings in Chevalier and Scharfstein [60], which documents that during regional and macroeconomic recessions, more financially constrained supermarket chains raise their prices relative to less financially constrained chains. In addition, a recent paper by Gilchrist et al. [102] uses confidential product price data and finds that during the "Great Recession" in the United States, firms with "weak" balance sheets increased their prices relative to industry averages, while firms with "strong" balance sheets chose lower product prices relative to industry prices.

A second key result concerns the impact of price stickiness on the firm's cash holdings, investment, financial decisions, and the firm's value. Like the price setting behavior in standard New Keynesian models [e.g. 97], nominal stickiness prevents the firm from setting its product price to the desired markup immediately, and the firm resets to the desired markup once it receives a Calvo price resetting opportunity.[7] Our model predicts that the firm facing larger price stickiness is more precautionary in its financial decisions. In particular, it tends to delay the payment of dividends or equity repurchases and issue less equity, resulting in more cash holdings on its balance sheet. Stickier prices increase the marginal value of cash, as the option of raising product prices to boost up current cash revenue becomes more costly. This implies that the firm will have the incentive to hold more cash, delaying the payment of dividends in order to cushion against negative demand shocks. However, although the marginal value of cash is high for the firm facing larger price stickiness, this does not imply that the firm will issue more equity when it is running out of cash. On the contrary, our model predicts that in most cases, the firm has already set a high product price by the time of pursuing external financing. When the product price is stickier, the firm anticipates that it is less likely/more costly to lower its price in the near future when it is out of liquidity problems. Therefore, it issues less equity since the demand for cash (mainly from investment) is low when the product price is high due to a decreasing customer base. We provide empirical evidence that is consistent with these predictions by exploiting 18 industries within the manufacturing sector. We show that the industries that change prices less frequently[8] issue less equity and conduct less repurchases. Moreover, we provide empirical evidence that the firms, whose product markets feature high price stickiness,

[6]Thus we contribute to the theoretical work that examines markup fluctuations over business cycles. Green and Porter [110] and Haltiwanger and Harrington [121], for example, predict that markups are procyclical with respect to demand shocks using game-theoretic models. But countercyclical markups are predicted by many more papers, featuring either implicit collusion [e.g. 18, 154, 202, 204] or customer markets [e.g. 36, 60, 102, 106, 152, 194]. The empirical evidence on markups so far are still mixed due to the lack of good measures of price-cost margins. For example, Braun and Raddatz [43], Chirinko and Fazzari [62], Domowitz, Hubbard and Petersen [78], Ghosal [101], Gilchrist et al. [102], Hall [120], Machin and Van Reenen [172], Nekarda and Ramey [190] find markups to be procyclical. In contrast to these studies, Bils [35], Chevalier and Scharfstein [60], Galí, Gertler and López-Salido [98], Mazumder [175], Murphy, Shleifer and Vishny [187], Rotemberg and Saloner [202], Rotemberg and Woodford [203] finds markups to be countercyclical.

[7]Under the menu cost framework, it is equivalent to the case that once the gain from resetting the product price to the desired markup is equal to the menu cost incurred as a result of price adjustment (see Appendix B.4).

[8]As shown in Nakamura and Steinsson [188], product price stickiness is a persistent industry characteristic. If an industry has a high price stickiness index, it means that the firms within that industry, on average, face larger price stickiness.

use disproportionately more funds raised from equity issuance to build up cash reserves rather than make investments.

Our model's predictions about the impact of price stickiness on the firm's value is, at first glance, counter-intuitive. We show that the firm facing a stickier price has a larger value in steady state because it endogenously chooses to hold more cash on its balance sheet. However, this does not imply that price stickiness is good in terms of boosting the firm's value. In fact, holding too much cash is costly in our model. Particularly, our model shows that the firm's enterprise value, which is more relevant for operating efficiency and growth [e.g. 33, 58, 156, 222], is lower when its product price is stickier..

Moreover, our theory advances the understanding of the impact of price stickiness on investment. The traditional view is that the firm with a stickier price invests less due to a higher cost of capital [e.g. 220]. Our model suggests that the endogeneity of cash holdings is missing in this argument. In our model, the firm facing larger price stickiness indeed faces a higher cost of capital, but because of this, it has a strong incentive to hold more cash on its balance sheet. This, on the one hand, reduces the cost of capital. On the other hand, it boosts the firm's value and increases the return on investment, motivating the firm to invest more. The force of the endogenous cash holdings channel works oppositely to the cost of capital channel, mitigating the impact of price stickiness on the investment rate.[9]

A third new result is that our model implies that the cost of price stickiness is higher when external financing costs are larger. We call it the "amplification effect" of financial frictions on the cost of price stickiness. In fact, Ball and Romer [24] study the interaction between real price rigidity based on customer markets and nominal price stickiness. They show that significant nominal rigidities can be explained by a combination of real rigidities and small frictions in nominal adjustment, an amplification mechanism via the real rigidity channel. Our model stresses an additional financial channel. Intuitively, the firm loses more value when it is facing more frictions in external financing because a cash-constrained firm tends to rely more on raising the price to boost cash revenue when facing higher costs of external financing. Consider an extreme case when external financing costs are zero, the firm would always prefer to use external financing to replenish cash and the product price will never be distorted. In this case, the cost of price stickiness is zero in our model since the firm would never have the incentive to reset its price. If external financing costs are infinite, the firm would have to raise its price when running out of cash. Since raising prices is more costly for firms with stickier prices, the decrease in the enterprise value is larger.

The remainder of the paper is organized as follows. Section 2.2 sets out the model. Section 2.3 discusses the calibration and quantitative results. Section 2.4 provides brief empirical evidence in support of the model's predictions. Finally, Section 3.7 concludes and provides a brief discussion on the robustness of our main results in a general equilibrium model.

[9]It should be noted that the impact of price stickiness on the firm's enterprise value is exactly consistent with the interesting empirical findings of Weber [220] because hoarding cash is costly for the shareholder in our model (i.e., the deadweight cost of holding cash).

2.2 Model

Our model combines the customer market in the literature of imperfectly competitive product markets [e.g. 194, 195], investment in neoclassical growth models [e.g. 104, 123], and the imperfect capital market in structural corporate models [e.g. 41, 42]. In the following, we first describe the firm's demand dynamics and investment behavior, and then we characterize the evolution of the firm's customer base following traditional customer market models. Next, we introduce the firm's external financing costs, cash holding costs and dynamics of cash holdings. Lastly, we formulate the firm's optimization problem.

Demand and Investment Consider a large and mature firm whose gross incremental operating revenue dY_t during a small interval $[t, t + dt]$ is

$$dY_t = A_t dQ_t, \tag{2.2.1}$$

where A_t represents "effective firm size" and dQ_t is the nominal incremental demand over the interval $[t, t + dt]$ per unit of effective size. Effective firm size A_t can also be interpreted as the firm's average sales. The nominal incremental demand dQ_t is exogenous and following a diffusion process with drift term[10]:

$$dQ_t = (p - \bar{c})\mu\left(\frac{p}{\bar{p}}\right)dt + \sigma dZ_t, \tag{2.2.2}$$

where Z_t is a standard Brownian motion under the risk neutral measure. We assume the shock dZ_t to the incremental nominal demand dQ_t to be exogenous. The variable p is the product price charged by the firm, the term $\bar{c}$ is the marginal cost of production, and $\bar{p}$ is the industry average price of the product. Note that neither the product price p nor the marginal cost $\bar{c}$ loads on the nominal shock. Without loss of generality, we normalize the industry average price to one in our calibration, i.e., $\bar{p} \equiv 1$. The average intratemporal demand (i.e. the demand curve faced by the firm in the short run) is characterized by

$$\mu\left(\frac{p}{\bar{p}}\right) = \mu_A\left(\frac{p}{\bar{p}}\right)^{-\eta}, \quad \text{with } \eta > 1. \tag{2.2.3}$$

This functional form has been widely adopted in the models with monopolistic pricing such as standard New Keynesian models [e.g. 97, 194]. Basically, it means that setting a higher product price p relative to the industry average lowers the average demand from existing customers.[11]

[10]The modeling of demand shocks follows the long history in the literature, such as Caballero [50]. More precisely, the modeling of cash flows dQ_t is very similar to the models of dynamic investment under dynamic optimal incentive contracting [e.g. 73]. In those models, managers can control the drift of cash flows by choosing effort level, and here by choosing the product price level. Moreover, those models assume that managers cannot control the volatility of the cash flow process.

[11]This effect can be rationalized using a search model where customers conduct sequential search for the cheapest

Moreover, in the spirit of Phelps and Winter [194], we assume that the firm's nominal profits in the short-run ($(p-\bar{c})\mu(p)$) are increasing, on average, in the product price p. That is, we require $p \leq p^* \equiv \frac{\eta}{\eta-1}\bar{c}$, where p^* is the optimal static monopolistic product price charged by the firm and $\frac{\eta}{\eta-1}$ is the static monopolistic markup. This assumption is innocuous, as argued by Phelps and Winter [194], which states that firms often charge lower markups relative to the static monopolistic one.

We assume that the firm's effective size depends on two major factors including customer base m_t and capital K_t, i.e.,

$$A_t = a(m_t, K_t). \tag{2.2.4}$$

In particular, we assume the functional form of the aggregator $A(\cdot,\cdot)$ to be Cobb-Douglas:

$$a(m, K) \equiv m^{\alpha} K^{1-\alpha}, \tag{2.2.5}$$

where α is the share of customer base in determining the sales of the firm, conditional on the nominal incremental demand dQ_t. Our model is an extension of the traditional customer market model [e.g. 194, 205], since we incorporate the firm's capital as a factor influencing sales in addition to the firm's customer base. The Cobb-Douglas aggregator is adopted mainly for tractability.

Capital accumulation follows the standard investment model with quadratic adjustment costs. In particular, we assume

$$dK_t = (I_t - \delta K_t)dt, \quad \text{for } t \geq 0, \tag{2.2.6}$$

where I_t is the gross investment rate on $[t, t+dt]$ and δ is the rate of capital depreciation.

With the gross investment adjustment cost, the firm's incremental net profits after paying the investment cost (denoted by dN_t) over the time incremental dt is given by

$$dN_t = A_t dQ_t - \Gamma(I_t, K_t, A_t)dt, \quad \text{for } t \geq 0, \tag{2.2.7}$$

where $\Gamma(I, K, A)$ is the total adjustment cost of investment. We assume that the adjustment cost is homogeneous of degree one in I and K, in line with the neoclassical investment literature [e.g. 123]. That is, we assume $\Gamma(I, K, A) = g(i)A$, where $i \equiv I/K$ is the firm's investment capital ratio and $g(i) = \frac{1}{\varsigma\theta}(1+\theta i)^{\varsigma} - \frac{1}{\varsigma\theta}$ is an increasing and convex function. We use the standard investment adjustment cost function in the neoclassical investment literature [e.g. 192]. Particularly, we take $\varsigma = 2$, and the functional form of $g(i)$ simply becomes

$$g(i) \equiv i + \frac{\theta i^2}{2}, \tag{2.2.8}$$

where θ captures the degree of the adjustment cost.

product and the distribution of search costs is uniform across buyers [e.g. 54].

The Customer Market and Sticky Price Setting In addition to the short-term demand effect of product prices (i.e., the intra-temporal demand), we incorporate the inter-temporal demand based on the customer market. The key idea is that reducing the product price not only reduces current profits, as $p\mu(p)$ is increasing in p. At the same time, the lower product price is more likely to retain existing customers and attract new ones, hence increasing the firm's future profits. Conversely, by increasing the product price, the firm can raise current profits at the cost of losing customer base, hence jeopardizing its future profits and growth.

Following the literature on the customer market [194, 205], we assume that customers gradually learn about prices elsewhere overtime and probably need to overcome some brand switching costs. Therefore, customers drift toward the cheapest sellers slowly. In particular, we postulate that the evolution of the firm's customer base follows the "customer flow equation":

$$dm_t = h\left(\frac{p_t}{\overline{p}}\right) m_t dt, \quad \text{with } h'(\cdot) < 0 \;\; and \;\; h(1) = 1, \tag{2.2.9}$$

where p_t is the product price charged by the firm and $\overline{p}$ is the industry average price. First, note that the customer flow function h in (2.2.9) captures the rate at which customers drift from one firm to others when the firm's product price differs from the industry average. Second, the slow-moving assumption in (2.2.9) captures information frictions faced by customers when searching for the cheapest price[12] or brand switching costs, in line with the search models of product markets, including Gottfries [105], Klemperer [151], Farrell and Shapiro [90], Beggs and Klemperer [29] and Farrell and Klemperer [91], among many others. Third, as in Phelps and Winter [194], the change in customer base is proportional to existing customer base, which implies that a temporal change in relative product prices can bring a permanent effect on the firm's customer base. Fourth, combining (2.2.2), (2.2.3) and (2.2.9), we see that the long-run elasticity of demand, which measures the percentage response of the eventual demand to a permanent increase in the product price, is larger than the short-run elasticity of demand η.

For simplicity, we adopt the following functional form to model the customer flow function h, which is also widely used by other customer market models [e.g. 63, 205]:

$$h\left(\frac{p}{\overline{p}}\right) \equiv \kappa - \kappa\left(\frac{p}{\overline{p}}\right)^{\nu} \quad \text{with } \kappa > 0, \nu > 0. \tag{2.2.10}$$

The relative price $p/\overline{p}$ determines the growth rate of customer base. Given the innocuous normalization $\overline{p} \equiv 1$, the marginal change in customer base is $-\nu\kappa p^{\nu-1}$ when the product price varies. Thus, the quantity $\nu\kappa$ measures how sensitive customers are to changes in the relative price, which can be interpreted as an inverse measure of information frictions or brand switching costs faced by customers.

Price setting follows the continuous-time version of the staggered price-setting model orig-

[12] Phelps and Winter [194] suggest that customers exchange information about the prices charged by different firms through random encounters.

inally developed by Calvo [52]. We assume that the firm's price resetting opportunities arrive randomly following a Poisson process with intensity ξ.[13] Intuitively, within any given period $[t, t+\Delta t]$, the firm can reset its product price with probability $1-e^{-\xi\Delta t}$, independent of the time elapsed since the last adjustment. Thus, the average duration between two consecutive price resetting opportunities is ξ^{-1}. Therefore, the intensity parameter ξ captures the price change frequency which constitutes a natural index of price stickiness. When price resetting opportunities arrive, the firm is free to reset its product price to either p_L or p_H.[14]

Cash Holdings, External Financing and Liquidation The firm has access to an imperfect capital market. For simplicity, we assume that the firm uses outside equity as the only source of external funds for investment [e.g. 41, 42]. The cost of external financing is captured by a fixed cost and a variable cost which is proportional to the amount of issued equity. We assume that the fixed cost is given by ϕA, where ϕ is the fixed cost parameter. The fixed financing cost plays a crucial role in generating an option-exercising type of external financing decisions. This not only produces severe nonlinearity in investment and the marginal value of cash, but also strongly incentivizes the firm to increase its product price when the firm is financially constrained. The fixed cost is proportional to effective firm size since this ensures that the firm does not grow out of its fixed cost of issuing equity. Technically, the proportional fixed cost also helps to keep the model homogeneous. In additional to the fixed cost, the firm needs to pay a variable financing cost γA, for each incremental dollar raised from the capital market.

The firm can also file for bankruptcy, resulting in a liquidation value $L = \ell A$, which is proportional to effective firm size. In the event of bankruptcy, shareholders obtain $L + W$, where W is the amount of cash holdings when the firm files for bankruptcy.

The firm optimally chooses the timing and the amount of external equity financing. When the gain from external financing is smaller than the value of liquidation, the firm will file for bankruptcy when running out of cash. Otherwise, the firm will pursue external financing.

Combining the firm's cash inflows from the incremental operating profits net the investment expenditures (dN_t in (2.2.7)) with cash inflows from financing policies (given by the cumulative payout U_t and the cumulative external financing H_t), the firm's cash inventory W_t evolves

[13] Ball and Romer [24] jointly model the real price rigidity based on customer markets and nominal price stickiness based on small adjustment costs. They emphasize that substantial nominal rigidity can arise from a combination of real rigidities and small nominal frictions. We would like to point out that the main mechanism of the model that the firm increases its price when being financially constrained does not depend on nominal rigidity. Introducing nominal rigidity enables us to analyze its impact on the firm's value, financing, and investment decisions, etc. We adopt the Calvo rule for technical convenience and to follow the convention in the New Keynesian literature. In Appendix B.4, we propose an alternative model based on menu costs, which shows that our main results remain valid even if nominal rigidity is modeled in a different way.

[14] We assume that the firm can only choose two different prices for simplicity and clarity. We are able to solve the model with continuous product prices. However, the problem becomes more complicated, as a PDE with free boundary conditions has to be solved. We find that the qualitative results are unchanged. Therefore, enabling the firm to choose among two prices is sufficient to convey the main idea of our theory and also helps clarify the model's mechanism in a coherent way.

according to the following equation:

$$dW_t = dN_t + (r - \lambda)W_t dt + dH_t - dU_t, \tag{2.2.11}$$

where the term $(r - \lambda)W_t dt$ represents the interest income net of the cash carrying cost; the term dH_t refers to the cash inflows from external financing; and the term dU_t refers to the cash outflows to investors.

We define the cash ratio as $w_t \equiv W_t/A_t$. We show below that the cash ratio w_t plays an important role as an endogenous state variable in characterizing the equilibrium. Using Ito's lemma, the law of motion for w_t within the internal financing region is

$$\begin{aligned} dw = & -w\left[\alpha h(p) + (1-\alpha)(i(w,p) - \delta)\right] dt \\ & + \left[(p - \bar{c})\mu(p) - i(w,p) - \frac{\theta}{2} i(w,p)^2 + (r-\lambda)w\right] dt + \sigma dZ_t. \end{aligned} \tag{2.2.12}$$

It shows that the product price p affects the cash ratio dynamics through three channels. First, a higher p has a positive effect on the cash ratio through the "current profit channel", because the term $(p - \bar{c})\mu(p)$ is increasing in p on the support $(0, p^*]$. Second, a higher p has a positive effect on the cash ratio through the "long-run growth channel" by changing customer base, because the term $-w\alpha h(p)$ is increasing in p. More intuitively, a higher p leads to a smaller growth rate in the firm's effective size, hence making the cash ratio easier to catch up. Third, the product price p affects the cash ratio through the "investment channel" as reflected by the term $-\left[w(1-\alpha) + 1\right] i(w,p) - \frac{\theta}{2} i(w,p)^2$. In fact, the significance of this channel depends on the impact of the product price on investment $i(w,p)$, which is determined both by the current profit channel and the long-run growth channel. As we show in Appendix B.1, without external financing costs, the firm will always focus on the long-run growth channel and investment will not be affected by the current profit channel. However, when there are external financing costs, the firm concerns more about the current profit channel if its financial slack is not sound.

The firm maximizes shareholders' value, as below, by optimally choosing its investment I, its product price p, payout policy U, external equity financing policy H, and liquidation time τ:

$$E_0\left[\int_0^{\tau} e^{-rt}\left(dU_t - dH_t - dX_t\right) + e^{-r\tau}(\ell K_\tau + W_\tau)\right], \tag{2.2.13}$$

where the expectation is taken under the risk-neutral measure. The term $dU_t - dH_t - dX_t$ is the discounted value of net payouts to shareholders. The quantity X_t is the cumulative costs of external financing up to time t, and dX_t is the incremental costs of raising incremental external equity dH_t. The term $\ell K_\tau + W_\tau$ is the liquidation value paid to shareholders at the time of bankruptcy τ.

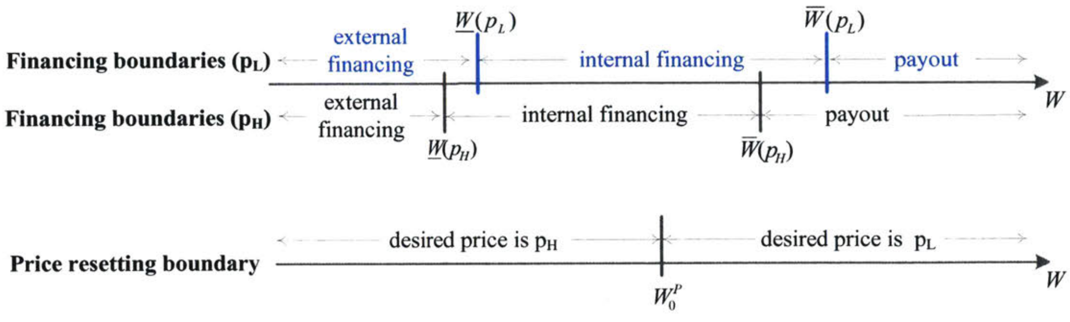


Figure 2-1: Illustrative Graph for the Decision Boundaries and Regions.

2.2.1 Model Solution

Let $U(A, W, p)$ be the value function of the firm. The firm needs to endogenously and simultaneously make three kinds of decisions, namely, investment decisions, financing/liquidation decisions, and price setting decisions. Since both financing/liquidation decisions and price setting decisions are discrete in our model, they can be sufficiently characterized by "decision boundaries". Figure 2-1 elaborates this idea: Basically, the firm's decision-making depends on which of the following four regions the firm finds itself lying in: (1) an external financing/liquidation region within which the firm pursues external financing ($dH > 0$) or liquidation; (2) an internal financing region within which the firm chooses the high product price (p_H) once price resetting opportunities arrive; (3) an internal financing region within which the firm chooses the low product price (p_L) once price resetting opportunities arrive; and (4) a payout region within which the firm chooses to payout dividends ($dU > 0$). More precisely, it is optimal for the firm to hoard up cash to finance future investment as a result of precautionary motives. When exogenous nominal demand shocks drive cash holdings W gradually to some low level $\underline{W}$ (i.e. the "external finance boundary") such that the current financing costs and the discounted future financing costs are equal, the firm will decide to raise outside equity. The product price setting decision essentially depends on the tradeoff between long-run customer base buildup and short-run profits. When cash holdings W are lower than W_0^P (i.e. the "price setting boundary"), the marginal value of cash is large enough so that the marginal value of short-run profits dominates the marginal value of developing future customer base. Thus, the firm desires to raise the product price to increase current profits. Lastly, because holding cash is costly (captured by $\lambda > 0$), the firm chooses to pay out cash when exogenous shocks drive cash holdings W beyond some high level $\overline{W}$ (i.e. the "payout boundary").

Internal Financing Region The equilibrium dynamics within the internal financing region can be described by the following Hamilton-Jacobi-Bellman (HJB) equation:

$$
\begin{aligned}
rU(A,W,p) = \max_{I,p^+\in\{p_L,p_H\}} & [\alpha h(p) + (1-\alpha)(I/K-\delta)]\, AU_A \\
& + [(r-\lambda)W + A(p-\bar{c})\mu(p) - \Gamma(I,K,A)]\, U_W + \frac{1}{2}\sigma^2 A^2 U_{WW} \\
& + \xi\left[U(A,W,p^+) - U(A,W,p)\right], \qquad (2.2.14)
\end{aligned}
$$

where A is the effective firm size defined in (2.2.4). The term U_A represents the marginal effect of increasing effective firm size on the firm's value, while effective firm size is changing due to net investment $(I/K-\delta)$ and the change in customer base $(h(p))$. The term U_W represents the effect of the firm's expected savings and profits on the firm's value, the term U_{WW} represents the effect of the volatility of cash holdings on the firm's value, and the jump term represents the jump in the firm's value caused by the change in the product price.

Price resetting decisions amount to comparing the value functions $U(A,W,p_L)$ and $U(A,W,p_H)$. The optimal price is chosen to be p_L if and only if $U(A,W,p_L) \geq U(A,W,p_H)$. The optimal investment rate $i = I/K$ is pinned down by the following first order condition:

$$
1 + \theta i = (1-\alpha)\frac{U_A(A,W,p)}{U_W(A,W,p)}. \qquad (2.2.15)
$$

A key simplification in our setup is that the firm's three-state optimization problem can be reduced to a two-state problem by exploiting homogeneity. We define the function $u(w,p)$ on $[0,+\infty) \times \{p_L, p_H\}$ such that

$$
U(A,W,p) \equiv Au(w,p), \quad \text{with } w = W/A. \qquad (2.2.16)
$$

Therefore, by taking out the scaling factor A, the HJB equation in (2.2.14) can be rewritten into a system of coupled ordinary differential equations:

$$
\begin{aligned}
ru(w,p_L) = {} & (u(w,p_L) - wu_w(w,p_L))\,[\alpha h(p_L) + (1-\alpha)(i(w,p_L)-\delta)] + \xi \max\{u(w,p_H) - u(w,p_L), \\
& 0\} + u_w(w,p_L)\left[(p_L-\bar{c})\mu(p_L) - i(w,p_L) - \frac{\theta}{2}i(w,p_L)^2 + (r-\lambda)w\right] + u_{ww}(w,p_L)\frac{\sigma^2}{2},
\end{aligned}
$$

and

$$
\begin{aligned}
ru(w,p_H) = {} & (u(w,p_H) - wu_w(w,p_H))\,[\alpha h(p_H) + (1-\alpha)(i(w,p_H)-\delta)] + \xi \max\{u(w,p_L) - u(w,p_H), \\
& 0\} + u_w(w,p_H)\left[(p_H-\bar{c})\mu(p_H) - i(w,p_H) - \frac{\theta}{2}i(w,p_H)^2 + (r-\lambda)w\right] + u_{ww}(w,p_H)\frac{\sigma^2}{2},
\end{aligned}
$$

where

$$
i(w,p) \equiv \left[\frac{u(w,p)}{u_w(w,p)} - w\right]\frac{1-\alpha}{\theta} - \frac{1}{\theta}, \quad \text{for } p \in \{p_L, p_H\}. \qquad (2.2.17)
$$

Payout Region The characterization of the payout boundary is mainly based on the work of Dumas [81]. The firm starts to pay out cash when the marginal value of cash held by the firm is less than the marginal value of cash held by shareholders which is one. Thus, the value matching condition gives the following boundary condition:

$$u_w(\overline{w}(p), p) = 1. \tag{2.2.18}$$

The payout region is characterized by $w \geq \overline{w}(p)$ for each p. Whenever the cash ratio is beyond the boundary, it is optimal for the firm to payout all the extra cash $w - \overline{w}(p)$ in a lump-sum manner and return its cash holdings back to $\overline{w}(p)$. Thus, the firm's value in the payout region has the following form:

$$u(w, p) = u(\overline{w}(p), p) + (w - \overline{w}(p)), \text{ when } w \geq \overline{w}(p). \tag{2.2.19}$$

Lump-sum payouts can occur mainly because payout boundaries are different for different product prices. In our quantitative analysis, we show that $\overline{w}(p_L) > \overline{w}(p_H)$ under the parameter calibration of interest (see Section 2.3.1). Moreover, the first-order condition for maximizing the firm's value over constant payout boundaries leads to the smooth pasting or the super contact condition

$$u_{ww}(\overline{w}(p), p) = 0, \tag{2.2.20}$$

where optimization is achieved at $\overline{w}(p)$.

External Financing/Liquidation Region Although the firm can raise outside equity any time, it is optimal for the firm to raise equity only when it runs out of cash, which means the external financing boundary $\underline{w}(p) \equiv 0$. This is due to the following reasons. First, cash within the firm earns a lower interest rate $r - \lambda$ due to the cash holding cost. Second, the firm's investment is continuous. Third, financing costs have smaller present value when they are paid further in the future.

Once the firm hits the financing boundary and decides to raise external equity, the optimal financing amount is also endogenously determined. The value matching condition for the issuance amount $w^*(p)A$ is

$$u(0, p) = u(w^*(p), p) - \phi - (1 + \gamma)w^*(p). \tag{2.2.21}$$

The left-hand side of equation (2.2.21) is the firm's value per unit of effective size right before the issuance. The right-hand side of equation (2.2.21) is the firm's value per unit of effective size minus both the fixed and variable costs of equity issuance per unit of effective size. The first-order optimality condition for the issue amount leads to the smooth pasting condition

$$u_w(w^*(p), p) = 1 + \gamma. \tag{2.2.22}$$

Since $w^*(p)$ is the optimal equity issuance, the marginal value of the last dollar raised by the firm must equal to one plus the marginal cost of external financing γ.

Now, we characterize the liquidation boundary $\underline{w}^L$ and the decision of liquidation. It is easy to see that $\underline{w}^L \equiv 0$, since as long as $w > 0$, the firm can still invest. In this case, both the marginal value of effective size and the marginal value of cash within the firm are larger than one, thus the firm's value is strictly larger than the value of liquidation. In the model, when the firm uses up all its cash, it needs to decide whether to issue equity or to file for bankruptcy. If it is optimal for the firm to choose filing bankruptcy instead of raising external equity at the boundary $\underline{w}^L = \underline{w}(p) = 0$, the liquidation value per unit of effective size gives

$$u(0, p) = \ell. \tag{2.2.23}$$

2.3 Quantitative Results

2.3.1 Parameter Choices and Calibration

We discipline the model by choosing parameter values based on existing calibration and empirical evidence. The liquidation parameter is set to be $l = 0.9$ following the estimates provided by Hennessy and Whited [126]. We choose the variable cost of financing to be $\gamma = 6\%$ based on the estimates reported by Altinkilic and Hansen [14] and the fixed cost of financing is $\phi = 2\%$. The interest rate is taken to be $r_f = 3\%$, which is within the range of broad empirical evidence in the United States. The volatility of demand shocks is set to be $\sigma = 12\%$, which is consistent with the parameters estimated by Eberly, Rebelo and Vincent [84]. The rate of depreciation is set to be $\delta = 2\%$. The cash holding cost is assumed to be $\lambda = 0.9\%$.[15] The adjustment cost parameter is $\theta = 1.5$ [223]. We set the Calvo price intensity to $\xi = 2.8$ to generate a median price duration of 4.3 months as reported in Bils and Klenow [37]. We set $\eta = 1.5$ as used in Backus, Kehoe and Kydland [21] and Zimmermann [225]. We set $\bar{c} = 0.78$, $p_L = 0.95$, and $p_H = 2.34$, implying that p_H is the price that maximizes cash revenue, namely, $p_H = \frac{\eta}{\eta - 1}\bar{c}$, and the industrial average price is about 1.[16] There is no direct empirical evidence on the growth rate of customer base for different prices, we choose $\kappa = 0.73$ and $\nu = 1.3$ to reflect that the firm sets its price to p_L when cash is abundant, in line with the main implication of the customer market literature [194, 195].

In the end, we are left with two parameters, the expected nominal demand, μ_A, and the capital share in effective firm size, $1 - \alpha$. We calibrate them to match the relevant moments for U.S. public mature large firms during the period of 1998-2012. We interpret effective firm size A as total sales and capital K as total assets. According to the Compustat dataset described in Appendix B.2, the mean cash-sales ratio is 18.45%, and the mean investment-asset ratio is 3.01%.

[15]We follow Bolton, Chen and Wang [42] who interpret the cash holding cost as a result of tax disadvantage or agency frictions. Under the simple tax disadvantage interpretation, compared to borrowing the fund, holding cash as retained earnings bears an additional cost 0.9% because the marginal tax rate is about 30% and the interest rate is 3%.

[16]The simulation results indicate that about 3.6% of the time the firm is setting its price to p_H, implying that the industry average price is equal to 1.

Table 2.1: Summary of key variables and parameters

Variable	Symbol	Parameters	Symbol	Value
Capital stock	K	Risk-free rate	r	3%
Cash holding	W	Rate of depreciation	δ	2%
Effective firm size	A	Mean nominal demand	μ_A	1.02
Customer base	m	Volatility of demand shocks	σ	12%
Investment	I	Adjustment cost parameter	θ	1.5
Cumulative nominal demand	Q	Share of customer base	α	0.145
Cumulative gross operating revenue	Y	Marginal cost of production	$\underline{c}$	0.78
Cumulative external financing	H	Fixed financing cost	ϕ	2%
Cumulative payout	U	Variable financing cost	γ	6%
Price resetting boundary	W_0^P	Demand elasticity	η	1.5
External financing boundary (p_H)	$\underline{W}(p_H)$	Customer base growth parameter	κ	0.73
External financing boundary (p_L)	$\underline{W}(p_L)$	Customer base growth parameter	ν	1.3
Payout boundary (p_H)	$\overline{W}(p_H)$	Proportional cash-carrying cost	λ	0.9%
Payout boundary (p_L)	$\overline{W}(p_L)$	Liquidation parameter	l	0.9
Optimal financing amount (p_H)	$W^*(p_H)$	Calvo parameter	ξ	2.8
Optimal financing amount (p_L)	$W^*(p_L)$	High price	p_H	2.34
		Low price	p_L	0.95

We set $\mu = 1.02$ and $\alpha = 0.145$ to match these two moments. Table 2.1 summarizes the symbols for the key variables of the model and the parameter values in the benchmark case.

2.3.2 Basic Mechanism: Financial Drivers of Markups

In this section, we elaborate on the rich interactions among cash holdings, product prices, investment, and financing decisions. Note that in our model, there are three channels that the firm can raise short-term cash inflows: increasing product prices, disinvesting, or external financing. However, there are costs associated with each channel either directly incurred or indirectly reflected as a loss in the firm's future revenue. The strategy for a liquidity constrained firm is to choose an optimal combination of the three choices to avoid the possibility of liquidation in the short run, while at the same time taking into account long-run growth opportunities.

Enterprise Value The firm's enterprise value is defined as the value of all the firm's marketable claims minus cash, $U(A, W, P) - W$, which can be considered as the value of the firm's total tangible and intangible capital stock. We normalize the enterprise value by effective firm size, and obtain $u(w, p) - w = \dfrac{U(A, W, P) - W}{A}$, where $w = W/A$ denotes the cash-size ratio. This normalized enterprise value $u(w, p) - w$ can be considered as a measure of the firm's average q, thus reflecting operating efficiency and growth [e.g. 33, 58, 156, 222].

Panel A of Figure 2-2 plots the normalized enterprise value as a function of the cash-size ratio for the two product prices, p_L and p_H, respectively. The solid line represents the normalized enterprise value when the product price is set at p_L. It is concave and increasing in the region between zero and the payout boundary $\overline{w}_{p_L} = 0.26$ (the vertical dotted line), and becomes flat (with slope zero) beyond the payout boundary ($w \geq \overline{w}_{p_L}$). The dashed line represents the normalized enterprise value when the product price is set at p_H, which has a similar shape as the solid line but is associated with a lower payout boundary $\overline{w}_{p_H} = 0.22$ (the vertical dotted line).

The two curves capturing the normalized enterprise value intersect with each other at the price resetting boundary, $w_0^P = 0.095$ (the vertical solid line). For $w > w_0^P$, the normalized enterprise value is higher if the product price is set at p_L; while for $w < w_0^P$, the normalized enterprise value is higher for p_H. This implies that when price resetting opportunities arrive (with Poisson intensity ζ), the firm will set its price to p_H if the cash-size ratio is less than w_0^P, and p_L if the cash-size ratio is larger than w_0^P. The fact that the optimal product price varies with the cash-size ratio is generated by two forces underlying our model. The product price not only affects short-term operating revenue (the "current profit channel", captured by $dQ_t = (p_t - \overline{c})\mu(\frac{p_t}{\overline{p}})dt$), but also determines the growth rate in customer base (the "growth channel", captured by $dm_t = h(\frac{p_t}{\overline{p}})m_t dt$). Therefore, there exists a trade-off between p_L and p_H. Setting the price to p_H enables the firm to increase its short-term operating revenue, but customer base will be gradually diminishing. By contrast, the firm builds up its customer base over time by setting the price to p_L, at the cost of lowering short-term operating revenue. The current profit channel is more crucial when the firm is liquidity constrained (i.e. with a low cash-size ratio), as in this case the marginal value of cash is high. Thus the firm is inclined to set its price to p_H for $w < w_0^P$, relying on the current profit channel to accumulate cash. When cash is abundant, the "growth channel" plays a dominating role in determining the firm's product price setting, and p_L would be chosen to build up customer base. This mechanism is reminiscent of the empirical findings in Chevalier and Scharfstein [60] and [102] that firms under weak/strong financial conditions tend to increase/decrease product prices relative to industry average prices during a recession.

As we have elaborated before, the firm issues equity only when its cash holdings hit the zero lower bound because of the proportional cash carrying cost and continuous investment flows [see 41, for more detailed explanations]. At the financing boundaries (i.e. $\underline{w}_{p_L} = \underline{w}_{p_H} = 0$), the firm's normalized enterprise value is strictly higher than its liquidation value. Therefore, external financing is always preferred to liquidation under our model parameterization.

Marginal Value of Cash Panel B plots the marginal value of cash $u_w(w, p)$ for the two product prices p_L and p_H. When the cash-size ratio is beyond the payout boundary, the marginal value of cash is equal to one. The marginal value of cash is higher when the firm becomes more liquidity constrained due to the frictions in external financing. This induces the firm to hoard cash in order to reduce the likelihood of external financing, although holding cash itself is costly, as captured by $\lambda > 0$. The frictions in external financing effectively generate "risk aversion" for the firm, a point emphasized by Bolton, Chen and Wang [41].

To economize on the fixed external financing cost ($\phi = 2\%$), the firm issues equity in lumps.

Conditional on issuing equity and having paid the fixed cost, the amount of equity issued returns the cash-size ratio to the point where the marginal value of cash $u_w(w, p)$ is equal to the marginal cost $1 + \gamma$. The firm's optimal issuance amount is $w^*_{p_L} = 0.103$ (the vertical dashed line) for p_L, and $w^*_{p_H} = 0.066$ (the vertical dashed line) for p_H. To the left of the optimal issuance amount, the marginal value of cash is higher than $1 + \gamma$, reflecting the fact that the fixed external financing cost in raising equity increases the marginal value of cash. If there is no fixed external financing cost ($\phi = 0$), the firm's optimal issuance amount is zero, as the firm raises just sufficient funds to keep positive w and to avoid incurring the cash carrying cost.

Note that within the internal financing region ($\underline{w} < w < \overline{w}$), the marginal value of cash is always higher when the product price is p_L. As a result, the firm with p_L delays the payout ($\overline{w}_{p_L} > \overline{w}_{p_H}$) and replenishes more cash when issuing equity ($w^*_{p_L} > w^*_{p_H}$) compared to the firm with p_H. The firm with p_L faces a higher marginal value of cash because in general it receives less cash inflows. As mentioned above, conditional on the same demand shock, setting a lower product price generates less short-term operating revenue. Since price resetting opportunities arrive in a Poisson fashion, the firm may find itself unable to adjust its product price to p_H immediately after the cash-size ratio w drops below the price resetting boundary w^P_0. This implies that the firm with p_L is more likely to hit the financing boundary after experiencing a sequence of negative demand shocks. The higher likelihood of executing costly external financing drives up the marginal value of cash. As a result, the firm with p_L is motivated to delay the payout of equity and endogenously chooses to hold more cash.

Panel C of Figure 2-2 plots the normalized firm value, which is equal to the enterprise value plus the value of cash.

Investment Panel D of Figure 2-2 plots the firm's optimal investment-capital ratio. The investment-capital ratio is increasing in cash between zero and the payout boundary, and becomes flat beyond the payout boundary. Notably, the firm disinvests when the cash-size ratio is low. This is to move away from the financing boundary to avoid costly external financing. However, disinvesting is costly not only because of its effect on lowering the growth rate of effective firm size, but also due to the convex capital adjustment cost. Since external financing bears a fixed cost, the firm only issues equity when the cash-size ratio hits the zero lower bound. To avoid paying the fixed cost, the firm starts to raise cash through disinvesting and price adjustment before the cash-size ratio hits zero. Once the firm is completely running out of cash, it raises sufficient cash in lump-sum through the external financing channel. This mechanism delivers a pecking-order solution to liquidity problems: the firm holds cash on its balance sheet to cushion against negative demand shocks. When shocks are small or only last for a few periods, the firm can run down cash and partially rely on raising its price or disinvesting to refill its cash reserves. However, if shocks are large or long lasting, the firm will eventually run out of cash, in which case it would fill up the cash reserve through the external financing channel, which is most costly due to the fixed cost.

Moreover, compared to the firm with p_H, the firm with p_L invests more when cash is sufficient;

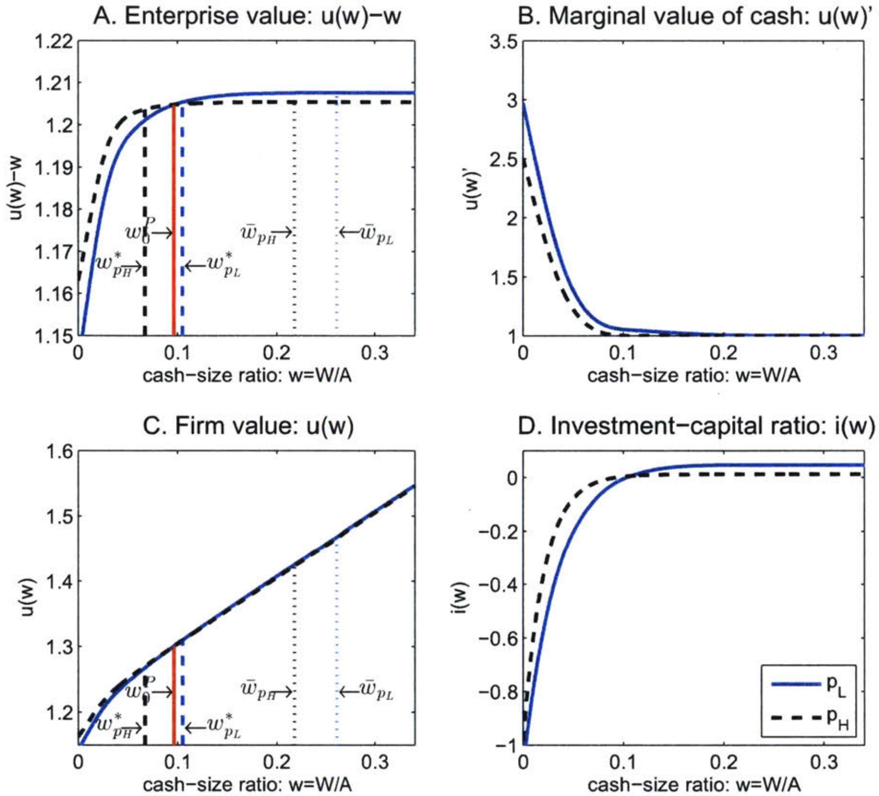


Figure 2-2: The firm's value, marginal value of cash, and optimal investment rates.

however, it invests less (disinvests more) when cash is constrained. This can be explained by the interaction between the current profit channel and the growth channel. When cash is abundant, the marginal value of cash is one, and the firm purely focuses on the growth channel. The firm with p_L enjoys a higher growth rate in customer base. Since investment has a complementary effect in boosting the growth rate of effective firm size, the firm with p_L optimally chooses to invest more.[17] However, if the firm is cash constrained, the current profit channel kicks in and starts to play a more important role in determining the firm's investment-capital ratio. With price stickiness, the firm with p_L cannot adjust its price to p_H immediately after its cash-size ratio drops below the price resetting boundary w_0^P. Therefore, it anticipates less incremental operating revenue in the short term, and an increased likelihood of pursuing costly external financing. Being aware of this anticipation, the firm's investment decision is also more precautionary. In Section 2.3.3, we analyze the impact of price stickiness and show that a higher degree of price stickiness intensifies the precautionary investment motive, leading to less investment.

[17]To see the complementarity between investment and customer base in boosting effective firm size, note that the growth rate of effective firm size, g_A, is determined by both investment, i, and the growth rate of customer base, g_m. The Cobb-Douglas form, $A = m^\alpha K^{1-\alpha}$ indicates that the marginal return of increasing capital stock K increases with the value of customer base m.

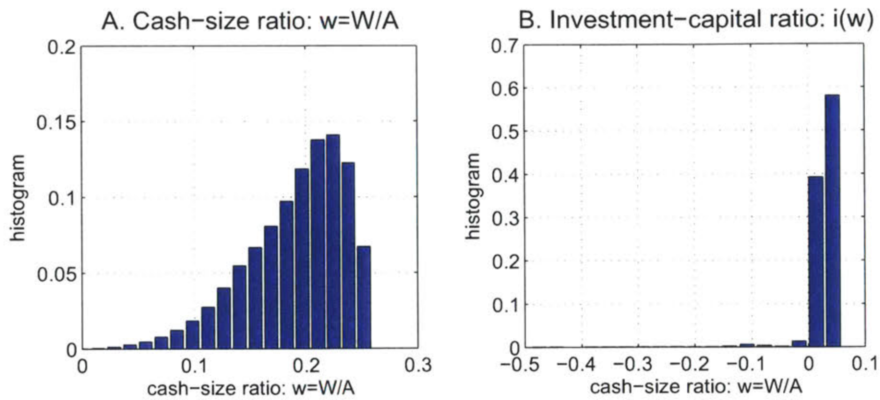


Figure 2-3: The steady-state distribution of cash and investment.

Stationary Distribution Using the optimal policy rules solved from the model, we simulate the evolution of prices and cash holdings of a single firm for 100 years and plot the stationary distribution of normalized cash-size ratios and investment ratios in Figure 2-3. Not surprisingly, as shown in panel A, cash holdings are relatively high during most of the time because using the above mentioned three channels (i.e. raising price, disinvesting, and external financing) to raise cash is costly. As a result, the probability mass of the investment ratio, $i(w)$, is concentrated around the highest value in the relevant support of w (see panel B). However, there are periods where the firm is liquidity constrained, and hence making negative investment. The stationary distribution of prices also reveals (not reported here) that the firm sets p_L during most of the time, to attract customers and build up customer base; while p_H is set for about 3.6% of the entire simulation period to overcome liquidity problems.

2.3.3 The Impact of Price Stickiness

In this section, we investigate the impact of price stickiness. Specifically, we analyze the impact of the Calvo parameter ξ on the firm's enterprise value, financing, payout, and investment decisions.

Price Stickiness on the Enterprise Value Figure 2-4 compares the firm's enterprise value when the Calvo parameter ξ varies. Panel A of Figure 2-4 plots the benchmark case, with $\xi = 2.8$. The case with a more flexible price ($\xi = 40$) is shown in Panel B.

The enterprise value for the firm with either p_L or p_H is higher in panel B, indicating that price stickiness reduces the enterprise value for any cash-size ratio. This is intuitive since being able to adjust the price can be considered as an option for the firm. The firm always prefers to set its price to p_H when the cash-size ratio is low ($w < w_0^P$) to take advantage of the current profit channel, and p_L when the cash-size ratio is high ($w > w_0^P$) to benefit from the growth channel. The likelihood of exercising this option is dependent on the degree of price stickiness, which is captured by the Calvo parameter, ξ. The larger ξ is, the lower the cost of price adjustment, and the higher the option value and the firm's enterprise value.

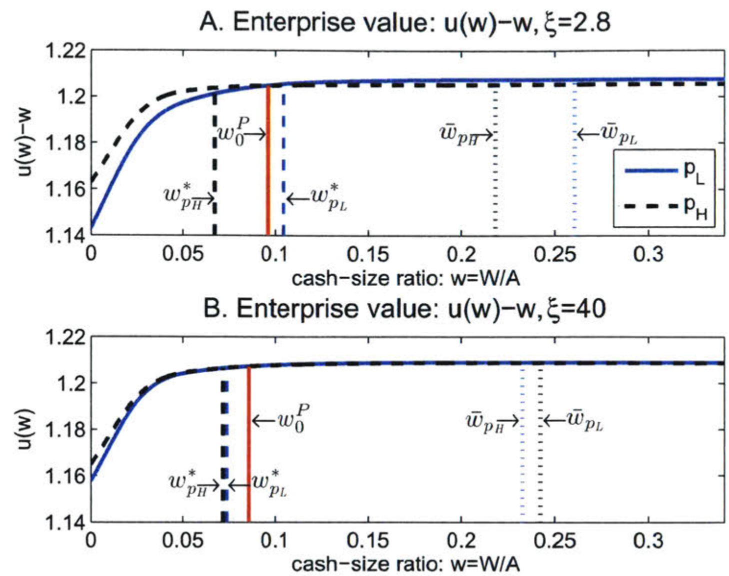


Figure 2-4: The firm's enterprise values for different value of the Calvo parameter. Panel A is plotted for $\xi = 2.8$, and Panel B is plotted for $\xi = 40$.

Moreover, notice that the enterprise value for the firm with p_L and the firm with p_H converges to each other when the price becomes less sticky. This is because when the firm obtains more opportunities to adjust its price, the enterprise value is affected less by the inherited price from the previous instant. On the extreme, when the price is perfectly flexible (with $\xi = \infty$), the two curves coincide with each other, and the price set in the previous instant no longer matters for the enterprise value (i.e., the product price is no longer a state variable).

Price Stickiness on Financing and Payout As shown in Figure 2-4, for the firm with p_L, both the payout boundary (the vertical dotted line) and the issuance amount (the vertical dashed line) shift to the left when the price becomes less sticky, because the firm has a larger chance to adjust its product price to p_H when it is running out of cash. This, as a result, would increase the incremental operating revenue through the current profit channel, and thereby mitigate liquidity problems and decrease the marginal value of cash. Hence, the firm is willing to hold less cash on its balance sheet when facing smaller price stickiness.

On the contrary, for the firm with p_H, both the payout boundary (the vertical dotted line) and the issuance amount (the vertical dashed line) shift to the right when the price becomes less sticky, indicating that the firm is willing to hold more cash. This is because when the price is less sticky, the firm with p_H anticipates that, in the future, it is more likely to get the chance to adjust its price to p_L when its cash-size ratio exceeds the price resetting boundary (w_0^P). As mentioned above, setting the low price increases the investment demand for cash due to the complementarity between capital stock and customer base in determining effective firm size. This motivates the firm with p_H to accumulate more cash, in order to benefit from future investment opportunities through the complementarity channel, when price resetting opportunities arrive.

Price stickiness has diametrically different implications for the firm with p_L and the firm with p_H in their financing and payout decisions. This is essentially because the marginal value of cash for the firm with p_L and the firm with p_H converges to each other when the price becomes less sticky. As shown in panel B of Figure 2-2, for any cash-size ratio, the marginal value of cash for the firm with p_L is higher than that for the firm with p_H. As a result, the convergence in the marginal value of cash triggered by a smaller price stickiness leads to a decrease in the marginal value of cash for the firm with p_L and an increase for the firm with p_H. Since financing and payout decisions are intimately linked to the marginal value of cash, it is not surprising that the firm with p_L tends to payout more and finance less while the firm with p_H tends to do the opposite when the price becomes less sticky.

Our simulation results show that for the benchmark calibration, the firm is setting its price to p_L when repurchasing equity (or issuing dividends) with a probability of 99.5%, thus the model predicts that firms facing larger price stickiness tend to repurchase equity less frequently, as the payout boundary associated with p_L in Figure 2-7 shifts to the right when ξ decreases from 40 to 2.8. On the other hand, during 98.5% of time the firm is setting its price to p_H when pursuing external financing, thus the model predicts that firms facing larger price stickiness issue less equity, as the optimal issuance amount associated with p_H in Figure 2-7 shifts to the left when ξ decreases from 40 to 2.8.

Price Stickiness on Investment Figure 2-5 presents the firm's optimal investment decisions when the Calvo parameter varies.

In panel A, the optimal investment ratio is plotted for the firm with p_L. It is shown that the firm facing smaller price stickiness invests more especially when the cash-size ratio is low. As we noted above, when the cash-size ratio is low, the "current profit channel" constrains the firm with p_L from investing more. Smaller price stickiness dampens the impact of this channel, as it becomes more likely for the firm to adjust its product price to p_H, hence boosting short-term operating revenue. Therefore, the liquidity-constrained firm with p_L increases its investment (or disinvests less) when the price becomes less sticky. For a cash-abundant firm, investment is determined mostly by the "growth channel", which is not affected much by the degree of price stickiness. Therefore, there is no significant difference in investment among the cash-abundant firms with p_L when price stickiness varies.[18]

In panel B, we plot the optimal investment ratio for the firm with p_H. Again, it is shown that the firm facing smaller price stickiness invests more for any cash-size ratio. This is mainly due to the increase in the normalized enterprise value for the firm with p_H when the price becomes less sticky (see Figure 2-4). To see this, note that the firm's value is equal to the sum of the

[18]Note that for the firm with p_L, when cash is abundant, investment ratios are marginally higher when the price is less sticky. This is due to the increase in the enterprise value generated by smaller price stickiness (Figure 2-4).

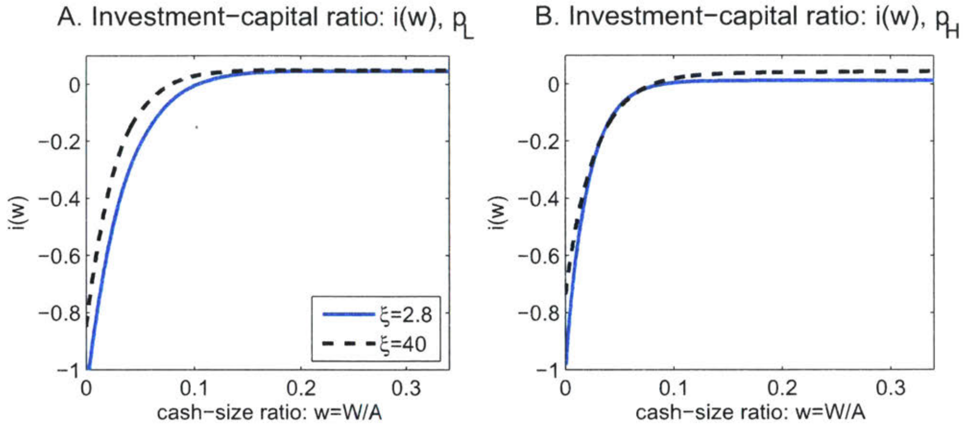


Figure 2-5: The firm's investment for different values of the Calvo parameter. Panel A/B plots investment for the p_L/p_H case. Solid/Dashed lines refer to the $\xi = 2.8/\xi = 40$ case.

enterprise value and cash, which is equal to the sum of the normalized enterprise value and the cash-size ratio multiplied by the firm's effective size. Since investment increases effective firm size, its return is higher when either the normalized enterprise value or the cash-size ratio is larger. Therefore, a higher normalized enterprise value due to smaller price stickiness motivates the firm to make more investments.

To illustrate the impact of price stickiness on investment more clearly, in Figure 2-6, we plot both the optimal investment ratio and the steady-state distribution of the cash-size ratio. For expositional purposes, we also mark the simulated average investment ratio on the figure, and place the up-arrow at the position representing the average cash-size ratio. It shows that the average investment ratio for the baseline calibration ($\xi = 2.8$) is 0.030, whereas it increases to 0.034 if the firm faces a more flexible price ($\xi = 40$). These calculations are subject to numerical errors, but the observed small difference indicates that the impact of price stickiness on investment is not large.

There are three forces underlying our model which affect investment in different directions, and as a result, investment is on average not affected significantly by the degree of price stickiness faced by the firm. As we have said above, investment boosts effective firm size, and thus its marginal return is linked to the firm's value per unit of effective size, which is equal to the sum of the normalized enterprise value and the cash size ratio. The first force reduces investment as the enterprise value is lower when the price is stickier, which decreases the marginal return on investment (see Figure 2-5). However, there is a countervailing force transmitted through endogenous cash holdings which pushes a stickier firm to make more investments. In fact, a stickier firm is more precautionary and holds more cash during most of the time, which increases its average cash-size ratio. Since the marginal return on investment also increases with the cash-size ratio, the firm with a stickier price tends to invest more due to more cash holdings. This can be directly seen from Figure 2-6, although investment for the firm with a stickier price ($\xi = 2.8$) is lower for any cash-size ratio, the distribution of cash holdings is more right skewed.

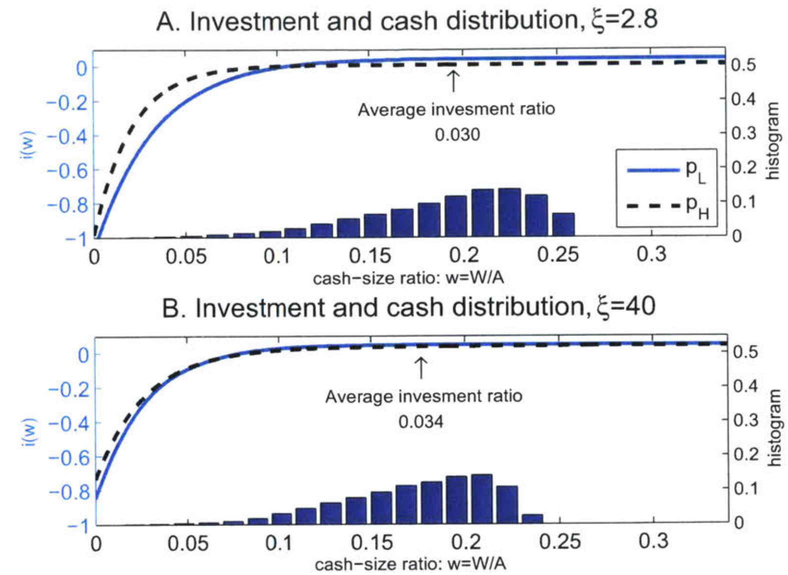


Figure 2-6: The firm's investment and cash-size ratio distribution for different values of the Calvo parameter. Panel A is plotted for $\xi = 2.8$, and Panel B is plotted for $\xi = 40$. The left y-axis is for the firm's investment and the right y-axis is for the steady-state distribution of the cash-size ratio. The up-arrow in each panel is positioned at the average cash-size ratio.

Third, as shown in Figure 2-5, the difference in investment ratios when price stickiness varies is quantitatively large only for the firm with p_L when the cash-size ratio is low and for the firm with p_H when the cash-size ratio is high. However, during most of the time the firm with a low cash-size ratio is setting its price to p_H and the firm with a high cash-size ratio is setting p_L. Therefore, the difference in the average investment ratio may not be significant. The latter two results are related to the steady-state distribution of the firm's cash holdings, which has been largely ignored in the existing literature. Our result complements the traditional view that firms facing larger price stickiness invest less due to the higher cost of capital [e.g. 220]. In fact, the negative impact of price stickiness on investment could be largely mitigated due to the channel of endogenous cash holdings.

Quantifying the Cost of Price Stickiness To shed light on the impact of price stickiness on the firm's value, we simulate the model for 100 years and compute the average normalized firm value, cash-size ratio, and normalized enterprise value, respectively, over the whole simulation period. We focus on the steady-state outcomes and discard the simulated path of the first 10 years as burn in. Panel A of Figure 2-7 shows that the firm's value increases with the degree of price stickiness (decreases with the Calvo parameter). The firm facing larger price stickiness bears a higher marginal value of cash, and is more precautionary in its payout decisions. As a result, it endogenously chooses to hold more cash on its balance sheet, boosting up the firm's

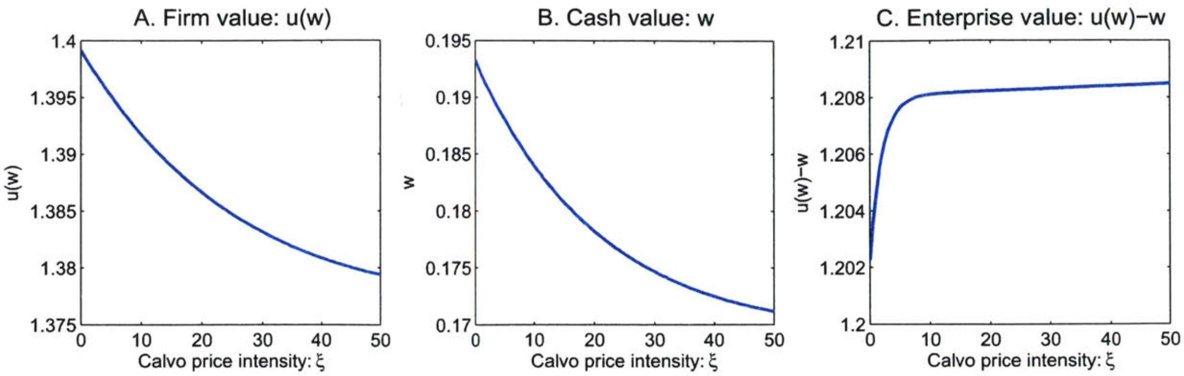


Figure 2-7: The average steady-state normalized firm value, cash holdings and normalized enterprise value for different values of the Calvo parameter.

value.[19] This is confirmed in panel B of Figure 2-7, which shows that cash holdings increase when the price becomes stickier. However, note that the firm facing larger price stickiness has a lower enterprise value (panel C of Figure 2-7). Since the enterprise value is equivalent to the average Tobin's q in our model, this implies that the firm with a stickier price is less efficiently operated [e.g. 33, 58, 156, 222] or riskier [e.g. 115, 163]. This result is consistent with Weber [220], which finds that firms facing larger price stickiness are riskier and demand a higher risk premium.

Quantitatively, our simulation results shown in Figure 2-7 indicate that the firm facing a completely sticky price ($\xi = 0$) on average holds 14% more cash than the firm facing a perfectly flexible price ($\xi = \infty$), while the enterprise value is lowered by 0.5%. A sticky price is costly not only through its direct effect on reducing the enterprise value, but also by indirectly inducing the firm to carry more cash, which is costly as captured by $\lambda > 0$. Moreover, note that although the reduction in the enterprise value due to price stickiness has a smaller magnitude on average, it is magnified especially when the firm is liquidity constrained. As shown in Figure 2-4, when the cash-size ratio is near zero, the enterprise value of the firm with p_L is decreased by about 1.5% when the value of the Calvo parameter decreases from $\xi = 40$ to $\xi = 2.8$.

2.3.4 The Interaction Between Price Stickiness and Financial Frictions

In this section, we seek to understand the interaction between price stickiness and financing costs and their joint impact on the firm's value.[20]

Panel A of Figure 2-8 presents the firm's normalized enterprise value when the fixed financing cost varies from $\phi = 0$ to $\phi = 3\%$ for the firm with $\xi = 2.8$ and $\xi = 40$, respectively. Confirming our previous results, the enterprise value is lower as the price becomes stickier. This is true regardless of the value of the fixed financing cost. As shown in panel A, the firm with $\xi = 2.8$ has uniformly a lower enterprise value (solid line) than the firm with $\xi = 40$ (dashed line).

[19]Note that the results presented in Panel A of Figure 2-7 are about the average firm's value when the firm is in steady state. A stickier price indeed reduces the firm's value at the impact of the change in price stickiness.

[20]The impact of financing costs is similar to that of Bolton, Chen and Wang [41] (see Appendix B.3.1).

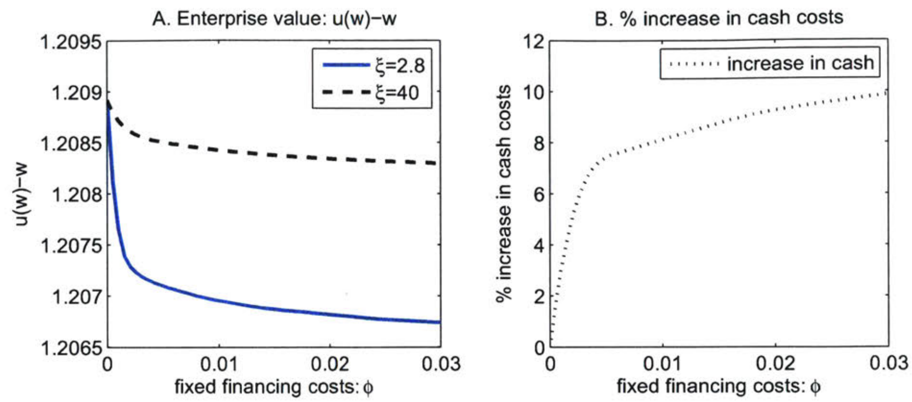


Figure 2-8: The interaction effect between price stickiness and financing costs on the enterprise value.

Notably, the decrease in the enterprise value due to larger price stickiness increases with the value of the fixed financing cost. This implies that the firm loses more value when it is facing more frictions in external financing. This is because cash-constrained firms tend to rely more on raising prices to boost cash revenue when facing higher costs of external financing. Consider an extreme case when external financing costs are zero, the firm will always prefer to use external financing to replenish cash and stick to the optimal price forever. As shown in panel A, the two curves converge to each other as the fixed financing cost approaches zero, implying that the degree of price stickiness has no impact on the enterprise value. If external financing costs are infinite, the firm will have to raise its price when running out of cash since external financing is not feasible. The firm facing a stickier price is more likely to hit the liquidation boundary, because it is less likely to find a chance to reset its price when the cash-size ratio is low. While the firm facing a more flexible price can timely increase its price and boost cash revenue to avoid liquidation. Hence, the difference in the enterprise value between the two firms, or the cost of price stickiness, is particularly large when external financing is not allowed.

In our model, the cost of holding cash is captured by parameter $\lambda > 0$. Panel B of Figure 2-8 presents the increase in the cash carrying cost when the Calvo parameter decreases from $\xi = 40$ to $\xi = 2.8$. When external financing costs are zero, the firm's cash management policy is no longer affected by price stickiness, thus there is no change in the cash carrying cost when the price becomes stickier. However, for a relatively large fixed external financing cost, $\phi = 3\%$, the cost of holding cash increases by about 10% when ξ is reduced from 40 to 2.8.

In sum, the impact of price stickiness varies significantly with financing costs. Price stickiness becomes more costly for the firm when the frictions in external financing are more severe.

2.4 Empirical Evidence

The model makes a number of testable predictions on the interaction between price stickiness and financial frictions, and how they jointly affect the firm's financial and investment decisions.

The first set of predictions is about how financial frictions and firms' financial slack affect their product price setting behavior. Our model predicts that financially constrained firms tend to raise their prices in order to increase cash flows. Moreover, the incentive to manipulate prices is larger when prices are less sticky or external financing costs are high. This prediction is supported by the empirical evidence from Chevalier and Scharfstein [60], which find that during regional and macroeconomic recessions, more financially constrained supermarket chains raise their prices relative to less financially constrained chains. In addition, a recent paper by Gilchrist et al. [102] uses confidential product-level price data and finds that during the "Great Recession" in the United States, firms with "weak" balance sheets increased their prices relative to industry averages, while firms with "strong" balance sheets lowered prices.

The second set of predictions is about how price stickiness affects firms' cash holdings and financial decisions. Our model predicts that firms facing larger price stickiness are more precautionary in their financial decisions. In particular, they tend to delay the payment of dividends or equity repurchases (see Figure 2-4 for the shift of the payout boundary associated with p_L) and issue less equity (see Figure 2-4 for the shift of the optimal issuance amount associated with p_H). Below, we provide the empirical evidence that are consistent with these predictions. Due to the lack of firm-level price data, our analysis focuses on the 18 industries within the manufacturing sector. Firms in the same industry produce similar products, and the price stickiness facing them is a persistent industry characteristic, as shown in Nakamura and Steinsson [188]. Hence, our analysis at the industry level is also informative about large firms' behavior.

The industries are defined to be consistent with the categories used by the Bureau of Economic Analysis (BEA), which include 10 industries producing durable goods and 8 non-durable goods industries. The measure of price stickiness is obtained following the empirical literature, which proxies the degree of stickiness by measuring how frequently firms change their prices [e.g. 23, 37, 57, 188]. Firms facing larger price stickiness adjust their prices less frequently. In this sense, the average frequency of price change can be regarded as an inverse measure of price stickiness. Among the industries we consider, 14 industries are broadly consistent with the major groups defined in Nakamura and Steinsson [188]. Therefore, we use the median frequency of industry-level price change for these 14 industries from Table VI of Nakamura and Steinsson [188], and the frequency of price change for the rest 4 industries are estimated based on the regressions elaborated in Appendix B.2.3. The industry-level financing, investment, and other variables used in our analysis are constructed from the Compustat and CRSP quarterly dataset over the period 1998 – 2012. Our analysis starts with year 1998, to be consistent with the price stickiness measure obtained from Nakamura and Steinsson [188] and the industry categories

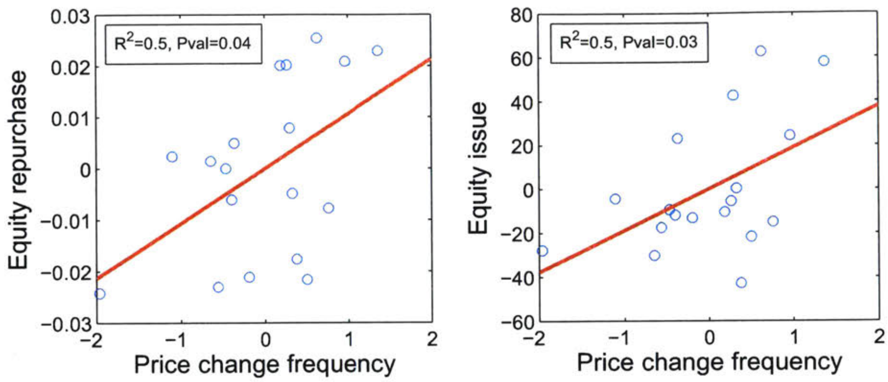


Figure 2-9: Equity financing, repurchases, and price stickiness. Each dot in the figure characterizes one industry. The solid line is the linear curve fitting of all the dots. The y-value of the dots in the left/right panel are obtained as the residuals from regressing the industry-level average equity repurchase/issuance ratio on the change in average debt ratios, the average industry stock return, and the average abnormal return. The x-value of the dots are obtained as the residuals from regressing the log industry-level price change frequency on the same set of controls. The slope of the fitted lines in the left/right panel is equivalent to the value of coefficient $\beta_1^{rep}/\beta_1^{ei}$ in regression specification (2.4.1)/(2.4.2).

defined by BEA.[21] We focus on large and mature firms only and thereby exclude the bottom 30% firms sorted by asset value in each industry. All the details are provided in Appendix B.2.2.

We construct industry-level average equity repurchase ratios and equity issuance ratios from the Compustat and CRSP datasets. For each firm, the equity repurchase ratio in period t is measured as the increase in the firm's treasury stock divided by the value of total assets. The equity issuance ratio is measured as the number of common shares issued multiplied by the stock price at the end of the quarter over the value of total assets. The industry-level ratios of equity repurchases and issuances are computed as the average firm-level repurchase and issuance ratios weighted by sales (see Appendix B.2.2 for more details).

The empirical specifications being employed are (2.4.1) and (2.4.2):

$$REPR_i = \beta_0^{rep} + \beta_1^{rep} log(FREQ_i) + \beta_2^{rep} X_i + \epsilon_i^{rep}, \tag{2.4.1}$$

$$\Delta EIR_i = \beta_0^{ei} + \beta_1^{ei} log(FREQ_i) + \beta_2^{ei} X_i + \epsilon_i^{ei}. \tag{2.4.2}$$

We regress industry-level average equity repurchase ($REPR_i$) and issuance ratio (ΔEIR_i) on the log industry-level price change frequency ($FREQ_i$) controlling for a set of control variables (X_i), including the change in average debt ratios ($\Delta DEBTR_i$), average industry stock return (R_i), and average abnormal return (AR_i).

Figure 2-9 shows that industries that change prices more frequently (thus face smaller price

[21] BEA adopted new industry categories after 1998.

stickiness) issue more equity and conduct more repurchases during the sample period, consistent with the model's predictions. The results are significant at the 5% level.

Next, we zoom in the consequence of the linkage between equity financing, repurchases, and price stickiness. Based on a simple regression analysis, we show that for firms facing stickier prices, a larger chunk of the free cash flows obtained from equity financing is held on their balance sheets rather than being invested. This is consistent with our model's predictions, as firms facing stickier prices have the tendency to build up cash reserves when markets are favorable to cushion against the deterioration of their balance sheet conditions during recession periods. The motive of "save for a rainy day" is inherently generated by the firms with stickier prices, because they are more restricted from using the "raising price channel" to boost up short-term operating revenue.

To show this, we focus on normal periods, and employ firm-level data from Compustat and CRSP.[22] For each firm f in industry i, we define the change in cash in period t as the increase in the firm's cash holdings from period t to period $t+1$. We associate the firm's price change frequency with the resided industry's price change frequency. Then, we run a regression according to specification (2.4.3):

$$\Delta CASH^f_{i,t} = \alpha_0 + \alpha_1 EI^f_{i,t} + \alpha_2 log(FREQ_i) + \alpha_3 EI^f_{i,t} \times log(FREQ_i) + \alpha_4 X^f_{i,t} + \varepsilon^f_{i,t}. \tag{2.4.3}$$

We regress the change in cash ($\Delta CASH^f_{i,t}$) on the amount of equity issuance ($EI^f_{i,t}$), the firm's log price change frequency ($log(FREQ_i)$), their interaction term ($EI^f_{i,t} \times log(FREQ_i)$), and a set of control variables ($X^f_{i,t}$) including the firm's average q ($q^f_{i,t}$) and the industry stock return ($R_{i,t}$).

The coefficient on the interaction term, α_3 is to our interest. Table 2.2 shows that it is negative regardless whether the average q and/or time fixed effects are controlled or not. This indicates that firms with stickier prices are in general using disproportionately more externally financed funds to build up cash reserves.

2.5 Conclusion and General Equilibrium Discussion

We propose a tractable model which demonstrates price setting decisions when a firm is operated in an environment featuring both customer markets and financial frictions, and the impact of price stickiness on a firm's investment and financing.

Our model offers several testable predictions on the interaction between price stickiness and financial frictions, and how they jointly affect a firm's financial and investment decisions. In

[22] Over our sample period 1998-2012, the U.S. economy was in a financial crisis or recession for 21 quarters. The first period was from the third quarter of 1998 to the end of 2001, which was started with the Russian financial crisis, continued with the Argentine economic crisis, and ended up with a recession due to the collapse of the speculative dot-com bubble [200, 201]. The second was between the fourth quarter of 2007 and the second quarter of 2009, known as the Great Recession.

Table 2.2: Firms with sticky prices build up cash reserves through equity financing

	(1)	(2)	(3)	(4)
Amount of issued equity	0.252***	0.251***	0.251***	0.251***
	[0.007]	[0.007]	[0.007]	[0.007]
log(price change frequency)	5.536***	5.278**	5.659***	5.370**
	[1.260]	[1.271]	[1.262]	[1.275]
Amount of issued equity	-0.096***	-0.096***	-0.096***	-0.096***
× log(price change frequency)	[0.004]	[0.004]	[0.004]	[0.004]
Tobin's q			-26.015	-24.956
			[18.553]	[18.881]
Industry stock return			-11.571**	-8.098
			[5.676]	[7.987]
Time fixed effects	No	Yes	No	Yes
Adj. R^2	0.564	0.565	0.566	0.567
Observations	1225	1225	1225	1225

***p<0.01, **p<0.05, *p<0.1. The reported adjusted R^2 for the regressions with time fixed effects is the overall R^2. Note that the positive and significant coefficient on the log amount of issued equity implies that the increase in cash is larger when more equity is issued. The positive and significant coefficient on the log price change frequency indicates that firms facing stickier prices are building up cash slowly. This is also consistent with our theory as stickier firms have less power to manipulate prices and increase cash revenue. Most importantly, the negative and significant coefficient on the interaction term coincides with the model's predictions, namely, for the same amount of equity financing, firms with stickier prices are building cash reserves more in normal periods.

particular, our model predicts that financially constrained firms have a tendency to set a relatively higher markup, and such tendency is larger when prices are less sticky or external financing costs are high. Moreover, our model predicts that firms facing larger price stickiness are more precautionary in their financial decisions—they tend to delay the payment of dividends and issue less equity. Existing literature [e.g. 60, 102] and our industry-level analysis provide empirical evidence in line with these predictions. Empirical tests based on more detailed firm-level price data set a research agenda of markup dynamics and the impact of price stickiness.

The interaction between price stickiness and financial frictions highlighted in our model can strengthen countercyclical markups. We analyze markup dynamics in Appendix B.3.2, which is only intended to be illustrative, since our model is a partial equilibrium model with both the industry average price and the interest rate being exogenously given. Yet, we believe its implications on the cyclicality of markups are robust even after taking into account the feedback effect from firms' optimal decisions on the industry average price. During recessions, financially weak firms have a tendency to increase their product prices to boost revenue. In a symmetric Nash equilibrium, the financial distress can reinforce the motivation for "implicit collusion" during recessions [see 202], since there is little loss in each firm's customer base when all firms keep high prices or even raise prices all together at the same time. Even without "implicit collusion", under a pure Walrasian equilibrium framework, the higher product prices charged by financially weak firms increase the industry average price. This may further induce financially strong firms to increase their product prices since now they are facing a less elastic demand curve due to a higher industry average price.

Chapter 3

Distinguishing Constraints on Financial Inclusion and Their Impact on GDP, TFP, and Inequality

3.1 Introduction

Financial deepening has accelerated in emerging market and low-income countries over the past two decades. The record on financial inclusion, however, has not kept apace. Large amounts of credit do not always correspond to broad use of financial services, as credit is often concentrated among the largest firms. Moreover, firms in developing countries continue to face barriers in accessing financial services. For instance, 95 percent of firms in advanced economies have access to a bank loan or line of credit as compared with 58 percent in developing countries, and 20 percent in low-income countries (Figure 3-1). Collateral requirements for loans, which impose borrowing constraints on firms, are also two to three times higher in developing countries as compared to advanced economies. Similarly, interest rate spreads (the difference between lending and deposit rates) tend to be much higher than in advanced economies. Firms also differ in terms of their own identification of access to finance as a major obstacle to their operations and growth: in developing countries, 35 percent of small firms report that access to finance is a major obstacle to their operations, compared with 25 percent of large firms, and 8 percent of large firms in advanced economies (Figure 3-2).[1]

[1]This problem is more acute for firms in the informal sector. This paper focuses primarily on firms in the formal sector.

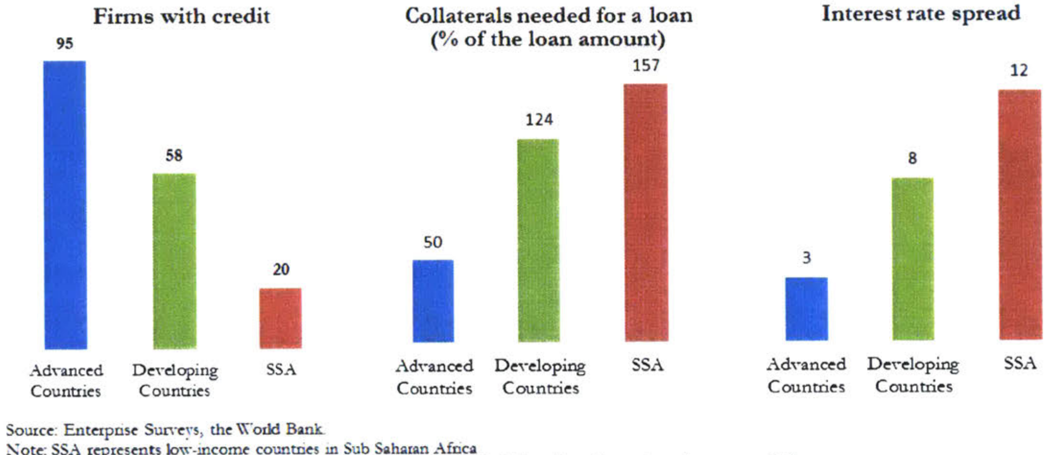


Figure 3-1: Financial inclusion in the world.

These considerations warrant a tractable framework that allows for a systematic examination of the linkages between financial inclusion, GDP, and inequality. Given that financial inclusion is multi-dimensional, involving both participation barriers and financial frictions that constrain credit availability, policies to foster financial inclusion are likely to vary across countries. In this paper, we develop a micro-founded general equilibrium model to highlight, distinguish, and evaluate the differential impacts of different financial constraints on GDP, TFP, and inequality and examine how these constraints interact both theoretically and numerically.

In the model, agents are heterogeneous—distinguished from each other by wealth and talent. Agents choose in each period whether to become entrepreneurs or to supply labor for a wage. Workers supply labor to entrepreneurs and are paid the equilibrium wage. Entrepreneurs have access to a technology that uses capital and labor for production. In equilibrium, only talented agents with a certain level of wealth choose to become entrepreneurs. Untalented agents, or those who are talented but wealth constrained, are unable to start a profitable business, choosing instead to become wage earners. Thus, occupational choices determine how agents can save and also what risks they can bear, with long-run implications for growth and the distribution of income.

The model features an economy with two "financial regimes", one with credit and one with savings only. Agents in the savings regime can save (i.e., make a deposit in banks to transfer wealth over time) but cannot borrow. Participation in the savings regime is free, but agents must pay a participation cost to borrow. The size of this participation cost is one of the determinants of financial inclusion, capturing the fixed transactions costs and high annual fees, documentation requirements, and other access barriers facing entrepreneurs in developing countries.

Once in the credit regime, agents can obtain credit, but its size is constrained by two additional types of financial frictions—limited commitment and asymmetric information. These distort the allocation of capital and entrepreneurial talent in the economy, lowering TFP. The first financial

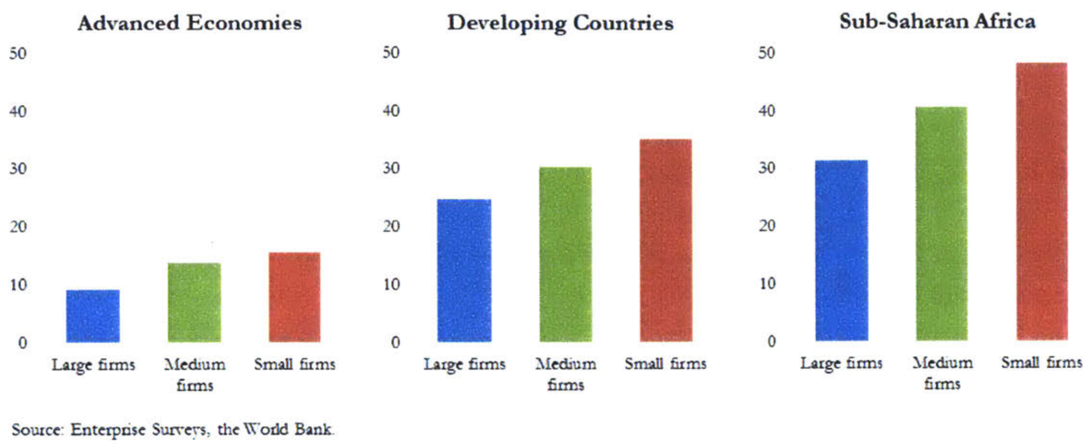


Figure 3-2: Percent of firms identifying access to finance as a major constraint.

friction is modeled as a borrowing constraint, which arises from imperfect enforceability of contracts. Entrepreneurs have to post collateral in order to borrow. The value of collateral is thus another determinant of financial inclusion, affecting the amount of credit available. The second financial friction arises from asymmetric information between banks and borrowers. In this environment, interest rates charged on loans must cover the cost of monitoring of highly-leveraged entrepreneurs. Because more productive and poorer agents are more likely to be highly leveraged, the ensuing higher intermediation cost is another source of inefficiency and financial exclusion. As only highly-leveraged entrepreneurs are monitored, entrepreneurs face differential costs of capital and may choose not to borrow even when credit is available.

We distinguish the effect of financial inclusion on the extensive and intensive margins. On the one hand, relaxing financial constraints can increase GDP through the extensive margin by increasing the credit access ratio (i.e., moving entrepreneurs from the savings regime to the credit regime). On the other hand, it enables entrepreneurs in the credit regime to produce more output, which boosts up GDP. This is the effect on the intensive margin.

In a partial equilibrium analysis with fixed interest rates and wages, we show that relaxing the ex-ante friction captured by the credit participation cost and the two ex-post frictions within the credit regime can increase GDP through both the extensive and intensive margins. We obtain closed-form solutions which indicate that relaxing different financial constraints has differential quantitative impacts, depending upon country-specific (the primitive model parameters calibrated from data) and individual-specific (wealth and talent) characteristics. We find that there are non-trivial interactions among the three financial constraints. The credit participation cost, the borrowing constraint, and the intermediation cost have complementary effects on the intensive margin, but are substitutes on the extensive margin. Intuitively, this is because a lower credit participation cost increases the credit access ratio, such that relaxing the borrowing constraint and reducing the intermediation cost have less of an impact. In other words, when the credit participation cost is low, the credit access ratio is already high, so that there is little room for increasing this ratio further through the other two channels. Essentially, the substitution effect on the extensive margin is due to the natural bound on the maximum credit access ratio (100%). On the intensive margin, relaxing one constraint amplifies the effects of relaxing other constraints. This is because, when the credit participation cost is low, entrepreneurs are left with more wealth after entering the credit regime. Since the amount of credit and the total intermediation cost are proportional to wealth, relaxing the borrowing constraint and reducing the intermediation cost increases business profits more.

The general equilibrium effect of financial inclusion does not allow for deriving analytical solutions, since it involves the endogenous distribution of wealth and talent and equilibrium interest rates and wages. To better understand the differential impacts of relaxing the various financial constraints, and in particular, how they interact in general equilibrium, we calibrate the model using data from the World Bank Enterprise Surveys and World Development Indicators. We jointly choose the model's key parameters to match the simulated moments, such as the percent of firms with credit and the firm employment distribution, as well as the economy-wide non-performing loans (NPL) ratio, the interest rate spread, and the bank overhead costs to assets ratio. We calibrate the model separately for six developing countries at varying degrees of economic development: three low-income countries (Uganda in 2005, Kenya in 2006, and Mozambique in 2006), and three emerging market economies (Malaysia in 2006, the Philippines in 2007, and Egypt in 2007).

The model simulations confirm our partial equilibrium analysis, suggesting that the impact of financial inclusion policies depends upon country-specific characteristics. For example, Uganda's GDP is most responsive to a relaxation of the borrowing constraint. This is because entrepreneurs in Uganda are severely constrained by high collateral requirements, so that reducing the intermediation cost only benefits a small number of highly-leveraged entrepreneurs. By contrast, a high fixed participation cost is a major obstacle to financial inclusion in Malaysia. These results suggest that understanding the specific factors constraining financial inclusion in an economy is critical for tailoring policy advice.

The model simulations also indicate that different dimensions of financial inclusion unambiguously increase the economy's GDP and TFP as talented entrepreneurs, who desire to operate firms at a larger scale, benefit disproportionately. However, they have a differential impact on inequality and there can be trade-offs. For example, a decline in the intermediation cost increases income inequality as it raises the profits of entrepreneurs living in the credit regime (whose income is already higher than others). Relaxing the borrowing constraint, on the other hand, can have an ambiguous impact on inequality, with inequality initially increasing and then declining. In other words, a Kuznets-type response can be generated. In fact, different dimensions of financial inclusion can result in different distributional consequences. In partial equilibrium, everyone can benefit from a more inclusive financial system, albeit to varying degrees. However, in general equilibrium, the resulting changes in interest rates and wages can lead to losses for some agents. For example, a policy that is most effective in increasing access (reducing the participation cost) benefits the poor and talented agents primarily, while wealthy agents lose due to higher interest rates and wages. By contrast, policies that target financial depth (relaxing the borrowing constraint) benefit wealthy and talented agents but can impose losses on wealthy but less-talented agents.

Finally, a GDP decomposition shows that relaxing the credit participation cost increases GDP mainly through the extensive margin by enabling more entrepreneurs to obtain credit from banks. By contrast, relaxing the borrowing constraint or reducing the intermediation cost raises GDP mostly through the intensive margin by allowing entrepreneurs who are already in the credit regime to expand their businesses. Our TFP decomposition shows that there are large losses in TFP in the savings regime as talented entrepreneurs leave the savings regime when financial constraints are relaxed. More importantly, a large proportion of the increase in TFP generated by financial inclusion is due to a between-regime shifting effect, namely, talented but relatively poor entrepreneurs move from the savings to the credit regime and expand their businesses.

The remainder of the paper is organized as follows. The next section provides a brief overview of the related literature. Section 3.3 sets out the structure of the model. Section 3.4 highlights the differential impacts of relaxing different financial constraints and their interactions. Section 3.5 presents the data and the model calibration. Section 3.6 discusses the quantitative results. Finally, Section 3.7 provides concluding remarks.

3.2 Literature Review

A growing theoretical literature has emphasized the aggregate and distributional impacts of financial intermediation in models of occupational choice and financial frictions. Banerjee and Newman [25] develop a framework with occupation choiceto capture the process of economic development; Lloyd-Ellis and Bernhardt [169] extend the model to explain income inequality and the existence of a Kuznets curve. Cagetti and Nardi [51] build on the framework to show that the introduction of a bequest motive generates lifetime savings profiles more consistent with data. In these studies, improved financial intermediation leads to greater entry into entrepreneurship, higher productivity and investment, and a general equilibrium effect that increases wages. Moreover, the models suggest that the distribution of wealth or the joint distribution of wealth and productivity is critical.

A related literature has found sizable impacts of improved financial intermediation on aggregate productivity and income [15, 47, 103, 112, 138, 139]. Buera, Kaboski and Shin [47] incorporate forward-looking agents in an occupational choice framework, and show that financial frictions account for a substantial part of the observed cross-country differences in output per worker and aggregate TFP. Moreover, Buera, Kaboski and Shin [48] focus on the general equilibrium effects of micro finance. They find that the impact of scaling-up micro finance on per-capita income is small, because of the ensuing redistribution of income from high-savers to low-savers, but the vast majority of the population benefits from higher wages. Moll [184] shows that the impact of financial frictions on GDP and TFP depends on the persistence of idiosyncratic shocks, and that the short-run effects of financial frictions tend to be larger than their long-run impacts.

Our model builds on this occupational choice framework, but with novel features. We focus on several dimensions of financial inclusion within an economy. Although these dimensions have typically been considered separately in the previous literature, our paper provides a unified framework for examining them individually as well as jointly. Our model features three types of financial frictions: fixed costs of credit entry, limited commitment, and asymmetric information. Unlike previous studies, our model allows us to also uncover how different frictions interact with each other. In this sense, our paper is related to studies in which multiple financial frictions co-exist and are compared. Clementi and Hopenhayn [66] and Albuquerque and Hopenhayn [11] argue that moral hazard and limited commitment have different implications for firm dynamics. Abraham and Pavoni [2] and Doepke and Townsend [77] discuss how consumption allocations differ under moral hazard with and without hidden savings versus full information. Martin and Taddei [174] study the implications of adverse selection on macroeconomic aggregates and contrast them with those under limited commitment. Karaivanov and Townsend [146] estimate the financial/information regime in place for households (including those running businesses) in Thailand and find that a moral hazard constrained financial regime fits the data best in urban areas, while a more limited savings regime is more applicable for rural areas. Similarly, Paulson, Townsend and Karaivanov [193] argue that moral hazard best fits the data in the more urban Central region of Thailand but not in the more rural Northeast. Kinnan [150] uses a different metric based on the first-order conditions characterizing optimal insurance under moral hazard, limited commitment, and hidden income to distinguish between these regimes in Thai data. Finally, Moll, Townsend and Zhorin [185] use a general equilibrium framework that encompasses different types of frictions, and examine the equilibrium interactions among various frictions. Our paper is related to these studies, but we emphasize the rich interactions among financial constraints, which in partial equilibrium can be complements on the intensive margin and substitutes on the extensive margin. Our paper also constitutes a normative policy analysis. By developing a quantitative macroeconomic framework and disciplining it with micro data, we shed light on a number of policy issues. For instance, what financial frictions are most relevant for the economy's GDP and income inequality? And what is the impact of alleviating these financial frictions individually or jointly?

Our paper is also related to a large empirical literature on the real effects of credit. The view that financial inclusion spurs economic growth is supported by empirical evidence [149, 162]. Regression-based analyses at the aggregate level reveal a strong correlation between broad measures of financial depth (such as M2 or credit to GDP) and economic growth. For firms, access to finance is positively associated with innovation, job creation, and growth [20, 28]. There is also evidence that aggregate financial depth is positively associated with poverty reduction and income inequality [27, 65]. Cross-sectional regression analysis, however, can be problematic as causality cannot easily be established, causal mechanisms are difficult to pin down, and policy evaluation is more challenging. Moreover, the implicit assumptions of stationarity and linearity in regression analysis could be incorrect, even after taking logs and including lags, if these variables lie on complex transitional growth paths [217]. The advantage of using a structural framework such as ours lies in capturing salient features of the economy and the pertinent financial sector frictions.

Our paper is also broadly related to the literature on misallocation [55, 132, 179, 184] and inequality [7, 56, 70, 134, 189]. Our contribution is to show that policy options that target different financial sector frictions have different impacts on resource allocation and inequality. More importantly, even for the same policy, the impacts on inequality can differ due to country-specific characteristics.

3.3 The Model

The economy is populated by a continuum of agents of measure one. Agents are heterogeneous in terms of initial wealth b and talent z.

Agents live for two periods. In the first period, agents make credit participation, occupational choice, and investment decisions, taking the optimal consumption and bequest decisions made in the second period as given. In the second period, agents realize income as wages or business profits, depending on their occupations, and make consumption and bequest decisions to maximize utility. Each agent has an offspring, whose wealth is equal to the bequest, and talent is drawn from a stochastic process.[2] The time subscript t is omitted unless necessary.

3.3.1 Agents

Agents generate utility only in the second period through consumption and a bequest to their offspring. The utility function is Cobb-Douglas, given by

$$u(c, b') = c^{1-\omega} b'^{\omega}, \tag{3.3.1}$$

where c is consumption, and b' is bequest. The bequest motive transfers wealth across periods, which endogenously determines the economy's wealth distribution. The assumption that utility is generated by bequest rather than the offspring's utility simplifies the analysis and captures the idea of a tradition for bequest giving following Andreoni [16].[3]

[2] The successor of an agent can be interpreted as the reincarnation of the original agent with potentially new talent.

[3] This is equivalent to assuming a myopic savings rate for the same agent. In Appendix C.2, we consider robustness checks and explore the implications of myopic savings rate by contrasting the simulation results in the baseline model with the results obtained from a model with forward-looking agents.

In the second period, agents maximize (3.3.1) by choosing c and b' subject to the budget constraint $c + b' = W$, where W denotes the second-period wealth, and it depends on the initial wealth and the realized first-period income.

The Cobb-Douglas form implies that the optimal bequest rate is ω.[4] Hence, the utility function $u(c, b')$ is a linear function of the end-of-period wealth (W), i.e., agents are risk neutral. This implies that maximizing expected utility is equivalent to maximizing expected second-period wealth. Therefore, in the first period, agents make credit participation decisions, occupational choices, and investment decisions to maximize expected income.

In the first period, agents need to make an occupational choice between being workers or entrepreneurs.[5] Each worker supplies one unit of labor, and the income realized in the first period is equal to the equilibrium wage, w. Entrepreneurs employ capital and labor, and obtain income through business profits.

Talent is drawn from a Pareto distribution $\mu(z)$ with a tail parameter θ. The offspring inherits the talent of her parents (or former self) with probability γ, otherwise, a new talent is drawn from $\mu(z)$.[6]

Entrepreneurs have access to a production technology, the productivity of which depends on agents' talent. The production function is given by

$$f(k, l) = z(k^{\alpha} l^{1-\alpha})^{1-\nu}, \tag{3.3.2}$$

where $1 - \nu$ is the Lucas span-of-control parameter, representing the share of output accruing to the variable factors. Out of this, a fraction α goes to capital, and $1 - \alpha$ goes to labor. Production exhibits diminishing returns to scale, with $\nu > 0$. Capital depreciates by δ after use.

Production fails with probability p, in which case output is zero and agents are able to recover only a fraction $\eta < 1$ of installed capital, net of depreciation in the second period. To simplify the calculation, we assume workers get paid only when production is successful. Therefore, each worker earns a wage with probability $1 - p$.

All agents can make a deposit in banks so as to transfer income and initial wealth across periods for consumption and bequest. However, following Greenwood and Jovanovic [111] and Townsend and Ueda [217], agents need to pay a fixed credit participation cost ψ to obtain a borrowing contract from banks. We assume that an agent lives in a "credit regime", if the agent pays the cost ψ and can borrow; that an agent lives in a "savings regime", if the agent does not pay ψ and can thereby only save. This cost can be considered as a contractual fee or a bargaining cost with banks. Intuitively, since workers do not invest, they never demand external credit. Entrepreneurs may want to borrow in order to expand their business scale and profits. In equilibrium, the fixed entry cost ψ is more likely to exclude poor entrepreneurs from financial

[4] The value of ω affects the amount of wealth transferred from the current period to the next period. Therefore, ceteris paribus a higher ω implies that the economy would have a higher level of wealth.

[5] In our framework, farmers can be considered as entrepreneurs, who operate their own farming businesses.

[6] The shock to talent is interpreted as changes in market conditions that affect the profitability of individual skills as in Buera, Kaboski and Shin [47].

markets, because this amounts to a larger fraction of their initial wealth. The next subsection illustrates the structure of the borrowing contract in detail.

Note that both the wage and the deposit rate are potentially time-varying and determined endogenously by the labor and capital market clearing conditions. Given the equilibrium wage w and deposit rate r^d, agents of type (b,z) make credit participation and occupational choice decisions to maximize expected income.

We solve the problem in two steps: first, agents choose their occupations conditional on the regime they are living in; second, agents choose the underlying regime by comparing the expected income that can be obtained in each regime. Next, we present the occupational choice problem in the savings and credit regimes, respectively.

Savings Regime

Agents living in the savings regime cannot borrow from banks—they have to finance the production exclusively using their own resources.

In the first period, the goal of agents is to maximize expected income. Given a certain initial wealth, maximizing expected income is equivalent to maximizing expected end-of-period wealth, W. Let $\pi(b,z)$ be the expected end-of-period wealth function for entrepreneurs of type (b,z). Denoting variables in the savings regime with superscript S, one can write

$$W^S = \begin{cases} (1+r^d)b + (1-p)w & \text{for workers,} \\ \pi^S(b,z) & \text{for entrepreneurs,} \end{cases} \tag{3.3.3}$$

where workers are paid only if production is successful, with probability $(1-p)$. Since agents are risk neutral, they choose to be workers if $(1+r^d)b + (1-p)w > \pi^S(b,z)$, and entrepreneurs otherwise. Therefore, the end-of-period wealth can be simply written as $W^S = \max\{(1+r^d)b + (1-p)w, \pi^S(b,z)\}$.

The wealth function $\pi^S(b,z)$ for entrepreneurs is obtained from the following maximization problem

$$\begin{aligned} \pi^S(b,z) = \max_{k,l} \quad & (1-p)[z(k^\alpha l^{1-\alpha})^{1-\nu} - wl + (1-\delta)k] + p\eta(1-\delta)k + (1+r^d)(b-k), \\ \text{subject to} \quad & k \leq b. \end{aligned} \tag{3.3.4}$$

With probability $1-p$, production succeeds, and entrepreneurs get revenue, $z(k^\alpha l^{1-\alpha})^{1-\nu} - wl$, plus the undepreciated working capital, $(1-\delta)k$. With probability p, production fails, and entrepreneurs can only get a fraction η of the undepreciated working capital. The last term in the maximization problem accounts for the wealth that is not used in production, which earns the equilibrium interest rate r^d. The constraint reflects the fact that entrepreneurs need to finance capital through their own initial wealth. The optimal choice of capital and labor is characterized in Proposition 4.

Proposition 4. *In the savings regime, the optimal amount of capital invested by entrepreneurs of type* (b,z) *is given by*

$$\begin{aligned} k^*(b,z) &= \min(b, \tilde{k}^S(z)), \\ l^*(b,z) &= [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^*(b,z)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}, \end{aligned}$$

where $\tilde{k}^S(z) = [\frac{\alpha w(1-p)}{(1-\alpha)(r^d+(1-p)\delta - p\eta(1-\delta)+p)}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)z}{w})^{\frac{1}{\nu}}$ *is the unconstrained level of capital (scale of business) in the savings regime.*

Note that $\tilde{k}^S(z)$ is the desired amount of capital that entrepreneurs living in the savings regime would like to invest when facing no wealth constraints. The value of $\tilde{k}^S(z)$ is finite because production has diminishing returns to scale. For entrepreneurs whose wealth is lower than $\tilde{k}^S(z)$, capital investment is constrained by wealth, i.e., $k^*(b,z) = b$.

Credit Regime

By paying an up-front credit participation cost ψ, agents enter the credit regime and obtain access to external credit. As workers do not need credit, they never pay ψ. Therefore, we only consider the entrepreneurs' problem in the credit regime.

We assume that the banking sector is perfectly competitive, driving the profit of intermediation to zero. This assumption can be easily relaxed by adding a profit margin for intermediation to capture noncompetitive banking sectors in many developing countries. This serves to increase the lending rate facing entrepreneurs, but the model's quantitative predictions would not change much.

In order to borrow, agents need to sign a contract with banks. A financial contract is characterized by three variables, (Φ, Δ, Ω), where Φ is the amount of borrowing, Δ is the value of collateral, and Ω is the face value of the contract. The face value Ω is the amount of money that needs to be repaid by the borrower if there is no default, which is determined by banks' zero profit condition. For simplicity, we assume that collateral is interest bearing, that is, agents earn the deposit rate r^d on the value of collateral.

Although the financial contract does not specify the lending rate, we can define the implied interest rate in the following way

$$r^l = \frac{\Omega}{\Phi} - 1. \tag{3.3.5}$$

Note that r^l would be potentially different for different entrepreneurs, depending on the terms of the contract.

Similarly, the leverage ratio (the amount of loans relative to the size of collateral) is defined as

$$\tilde{\lambda} = \frac{\Phi}{\Delta}. \tag{3.3.6}$$

If production fails, entrepreneurs may not be able to repay the loan's face value Ω. If this happens, entrepreneurs default and banks seize the interest-bearing collateral, $(1+r^d)\Delta$, and the recovered value of undepreciated working capital, $\eta(1-\delta)k$. In equilibrium, since highly-leveraged entrepreneurs default in the case of a production failure, they are charged with a higher lending rate in the event of success (to compensate for losses in the event of failure).

Limited commitment In order to borrow, entrepreneurs need to post collateral at banks. Suppose that entrepreneurs can borrow Φ if amounts of collateral Δ is posted. Suppose further that contract enforcement is imperfect, therefore, entrepreneurs can immediately abscond with a fraction $1/\lambda$ of the rented capital. The only punishment is that they lose their collateral Δ. In equilibrium, entrepreneurs do not abscond only if $\Phi/\lambda < \Delta$.[7] Therefore, banks are only willing to lend $\lambda\Delta$ to entrepreneurs if Δ units of collateral are posted. This single parameter $\lambda \geq 1$ parsimoniously captures the degree of financial friction resulting from limited commitment. A special case of $\lambda = 1$ implies that entrepreneurs cannot borrow.

Asymmetric information There is asymmetric information between entrepreneurs and banks (i.e. whether the production of a particular entrepreneur fails or not is only known to the entrepreneur). Due to limited liability, entrepreneurs have a default option when production fails. This implies that they could repay less if a production failure is reported and the lie is not discovered by banks. Banks have a monitoring technology through which they get information on the success of production at a cost proportional to the scale of the production (denoted by χ). If entrepreneurs are caught cheating, banks can legally enforce the full repayment of the loan's face value. As banks make zero profit in equilibrium, the monitoring cost is borne by entrepreneurs when the financial contract is designed. In sum, all agents are truth-telling. However, this comes at a cost.

The banks' optimal verification strategy follows Townsend [216], which occurs if and only if entrepreneurs cannot repay the face value of the loan. This happens when entrepreneurs are highly leveraged and also experience a production failure.[8] To be more specific, when production

[7]See Banerjee and Newman [26], Buera and Shin [46], and Moll [184] for a similar motivation of this type of constraint. The borrowing constraint is derived based on the assumption that entrepreneurs can immediately walk away with the rented capital. Another possibility is that entrepreneurs may want to put this capital into production and walk away after output is realized. In this case, the condition that regulates diversion is $\Phi(R^k - R^l) + \Delta R^l \geq \frac{\Phi}{\lambda} R^k$, where R^k and R^d are the average gross return on capital and the gross lending rate, respectively. The implied borrowing constraint by this condition is $\Phi/\lambda < \frac{\Delta}{\lambda-(\lambda-1)R^k/R^d}$, which is more relaxed than the one in the main text. In fact, the two are equivalent only when $R^k = R^d$, which is the capital return obtained by the least talented entrepreneur. Since it is realistic to believe that banks cannot observe entrepreneurs' talent (an assumption we make later when discussing the optimal contract), it is reasonable to assume that banks would impose the most stringent borrowing constraint. As a result, the borrowing constraint derived from ex-post diversions is consistent with the borrowing constraint specified in the main text.

[8]Implicit here is the assumption that entrepreneurs would not decline the repayment of the loan if they have sufficient funds because banks monitor and seize the face value of the loan when default happens.

succeeds, entrepreneurs can repay the face value of the loan.[9] Therefore, there is no incentive for banks to monitor. However, if a production failure is reported, banks monitor only if the loan contract is highly leveraged. This is because a low-leveraged loan contract implies that entrepreneurs are not borrowing much. Therefore, the required repayment is small, and can be covered by the value of interest-bearing collateral, $(1 + r^d)\Delta$, plus the value of recovered working capital, $\eta(1 - \delta)k$, even if production fails. In this case, entrepreneurs have no incentive to lie because regardless of the production outcome, they can and have to repay the face value of the loan. For the same reason, banks have no incentive to monitor.

On the other hand, if the loan contract is highly leveraged[10], and if production fails, the amount that entrepreneurs can repay is not sufficient to cover the face value of the loan. As a result, default happens. Finally, note that in this case entrepreneurs have an incentive to lie when production is successful because they know that with high leverage, they would repay less if a production failure is reported. Therefore, to motivate truth-telling, banks verify all highly-leveraged loan contracts if a production failure is reported. We formalize the optimal verification strategy in Proposition 5.

Proposition 5. *Banks' optimal verification strategy is pinned down ex-ante and determined by the contract* (Φ, Δ, Ω), *parameter* η *and* δ, *and the deposit rate* r^d:

i. For a low-leveraged loan, i.e. $\eta(1 - \delta)\Phi + (1 + r^d)\Delta \geq \Omega$, *no verification occurs.*

ii. For a highly-leveraged loan, i.e. $\eta(1 - \delta)\Phi + (1 + r^d)\Delta < \Omega$, *verification occurs iff production fails.*

In the credit regime, the end-of-period wealth is denoted by

$$W^C = \pi^C(b, z),$$

where the superscript C refers to the credit regime. Agents choose to pay the credit participation cost when $W^C > W^S$.

We assume that banks cannot observe entrepreneurs' type (b, z), and therefore have to provide a menu of contracts. Entrepreneurs choose their optimal contracts from the menu. Notice that the schedule of contracts is designed to be incentive compatible, namely, entrepreneurs of type (b, z) would have no incentive to imitate type (b', z') and choose the optimal contract of other entrepreneurs. Moreover, all loan contracts make zero profit given that financial intermediation is perfectly competitive. Below, we first elaborate the optimal contract for entrepreneurs of type (b, z). We then discuss why the contract is incentive compatible.

[9] To see this, notice that entrepreneurs would borrow to produce only if they can make profits. Therefore, when production succeeds, gross output should be at least higher than the capital input. On the other hand, if entrepreneurs default, banks monitor output and seize the face value of the loan anyway. Thus, entrepreneurs have no incentive to default.

[10] The threshold between low and high leverage ratios is derived by considering whether the value of interest-bearing collateral plus the recovered working capital is sufficient to repay the face value of the loan. In particular, as we discuss later, the loan contract is highly leveraged if $\eta(1 - \delta)\Phi + (1 + r^d)\Delta < \Omega$.

To solve the optimal loan contract (Φ, Δ, Ω) for entrepreneurs of type (b, z), we use the following steps:

First, since collateral is interest-bearing, entrepreneurs are willing to post all of their wealth net of credit participation cost, $b - \psi$, as collateral instead of depositing a fraction of it in a savings account. Hence, the collateral term, $\Delta = b - \psi$, belongs to the set of optimal loan contracts.[11]

Second, entrepreneurs borrow to increase production scale and make higher profits. Therefore, there is no reason to borrow more funds from banks and not use them in production, since this would only increase the leverage ratio, which, in turn, potentially increases the cost of capital. Hence, the amount of loan Φ is equal to the amount of capital $k(b, z)$, if the loan contract is optimal.

The above arguments suggest that the optimal loan contract chosen by entrepreneurs of type (b, z) should be of the form $(k(b, z), b - \psi, \Omega)$. Hence, Ω remains the only element to be determined.

The face value of the loan Ω in the optimal contract is set such that banks make zero profit knowing that only entrepreneurs of type (b, z) will choose it. From banks' perspective, the expected payoff of this loan contract is $(1 - p)\Omega + p \min(\Omega, \eta(1 - \delta)k + (1 + r^d)(b - \psi))$. The first term refers to the payoff when production succeeds, which happens with probability $1 - p$. In this case, banks receive the full face value of the loan, Ω. The second term refers to the payoff when production fails. When production fails, before repaying debt, entrepreneurs' "net value" is equal to the recovered working capital, $\eta(1 - \delta)k$, plus the after-interest value of collateral, $(1 + r^d)(b - \psi)$. Banks receive the full face value of the loan, Ω, if entrepreneurs' "net value" is sufficient to repay it. Otherwise, banks only receive the "net value" due to limited liability, and entrepreneurs would end up with nothing. In sum, when production fails, banks receive either Ω or $\eta(1 - \delta)k + (1 + r^d)(b - \psi)$, whichever is smaller.

On the other hand, the cost of creating the loan contract is equal to the after-interest value of the loan, $(1 + r^d)k$, plus the expected cost of monitoring. Note that monitoring occurs only if entrepreneurs cannot repay the loan, namely, when production fails and the net value, $\eta(1 - \delta)k + (1 + r^d)(b - \psi)$, is smaller than the loan's face value, Ω. In this case, a monitoring cost, χk, is incurred. Therefore, the expected cost of monitoring is equal to the monitoring cost, χk, multiplied by the monitoring rate. The monitoring rate is equal to the production failure rate, p, when entrepreneurs are highly leveraged, i.e. $\eta(1 - \delta)k + (1 + r^d)(b - \psi) < \Omega$, and zero otherwise. Thus the expected cost of monitoring can be expressed as $p\chi k \cdot \mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}$, where $\mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}$ is an indicator function, which equals to 1 if $\eta(1 - \delta)k + (1 + r^d)(b - \psi) < \Omega$ and 0 otherwise. Hence, the cost of creating the loan contract is $(1 + r^d)k + p\chi k \cdot \mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}$.

[11]Note that there might exist multiple optimal contracts for wealthy entrepreneurs since they do not demand much credit. But all these contracts would result in an identical net outcome for both entrepreneurs and banks. The optimal contract we consider here is the one with the lowest leverage ratio, i.e., all wealth b is posted as collateral.

The zero profit function is obtained when the expected payoff of the loan is equal to its cost:

$$(1-p)\Omega + p\min(\Omega, \eta(1-\delta)k + (1+r^d)(b-\psi)) = (1+r^d)k + p\chi k \cdot \mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}. \tag{3.3.7}$$

Equation (3.3.7) pins down Ω, and implies that in the optimal contract we consider, Ω is a function of k and b only. The optimal contract chosen by entrepreneurs of type (b,z) can be written as $(k^*(b,z), b-\psi, \Omega(k^*(b,z),b))$, where $k^*(b,z)$ is the optimal amount of capital invested in production, and $\Omega(k^*(b,z),b)$ is determined by equation (3.3.7). This implies that to exactly characterize the optimal contract as a function of initial variables b and z, we only need to know $k^*(b,z)$, which solves the following problem:

$$\begin{aligned} \pi^C(b,z) = \max_{k,l} \quad & (1-p)[z(k^\alpha l^{1-\alpha})^{1-\nu} - wl + (1-\delta)k - \Omega + (1+r^d)(b-\psi) \\ & +p\max(0, \eta(1-\delta)k + (1+r^d)(b-\psi) - \Omega), \\ \text{subject to} \quad & k \leq \lambda(b-\psi), \end{aligned} \tag{3.3.8}$$

where the term Ω in problem (3.3.8) is the solution to banks' zero profit condition (3.3.7). The solution to (3.3.7) and (3.3.8) determines the optimal capital k as a function of b and z, and pins down the optimal contract.

In (3.3.8), the first term refers to the end-of-period wealth when production succeeds. The second term refers to the case of production failure. Entrepreneurs have something left only if $\eta(1-\delta)k + (1+r^d)(b-\psi) > \Omega$, that is when the recovered undepreciated working capital plus the after-interest value of collateral is sufficient to repay the loan. Otherwise, entrepreneurs end up with zero end-of-period wealth.

Below we restrict ourselves to the case where default occurs, with the endogenously determined interest rate satisfying, $r^d > \frac{\eta(1-\delta)\lambda}{\lambda-1} - 1$.[12] Note that this condition is satisfied for all the six countries in our quantitative analysis. We first illustrate the default boundary (Lemma 2) and the associated cost of capital for different cases (Lemma 3), and then we characterize the optimal amount of capital in Proposition 6.

Lemma 2. *In the credit regime, default occurs for highly-leveraged entrepreneurs. In particular, there is a default boundary,* $\bar{\lambda} = \frac{1+r^d}{1+r^d-\eta(1-\delta)}$, *depending on parameters* η *and* δ *and the endogenous deposit rate* r^d. *For an entrepreneur who operates a business with leverage ratio* $\tilde{\lambda}$:

- *i. If* $\tilde{\lambda} \leq \bar{\lambda}$ *(low-leverage region), default never occurs, and the implied lending rate is* $r^l = r^d$.
- *ii. If* $\tilde{\lambda} > \bar{\lambda}$ *(high-leverage region), default occurs when production fails, and the implied lending rate is increasing in* $\tilde{\lambda}$, *i.e.* $r^l = \frac{1+r^d+p\chi-p\eta(1-\delta)-p(1+r^d)/\tilde{\lambda}}{1-p} - 1$.

[12] If $r^d \leq \frac{\eta(1-\delta)\lambda}{\lambda-1} - 1$, there is no default in the economy. This is because in our model, whether an entrepreneur defaults or not depends on the leverage ratio. As shown in Lemma 2, only entrepreneurs whose leverage ratios are larger than $\bar{\lambda}$ default when production fails. Notice that $\bar{\lambda}$ is decreasing in r^d. Therefore, $\bar{\lambda}$ could be higher than λ (the highest possible leverage ratio imposed by limited commitment) for small r^d. In this case, even entrepreneurs with fully leveraged loans do not default.

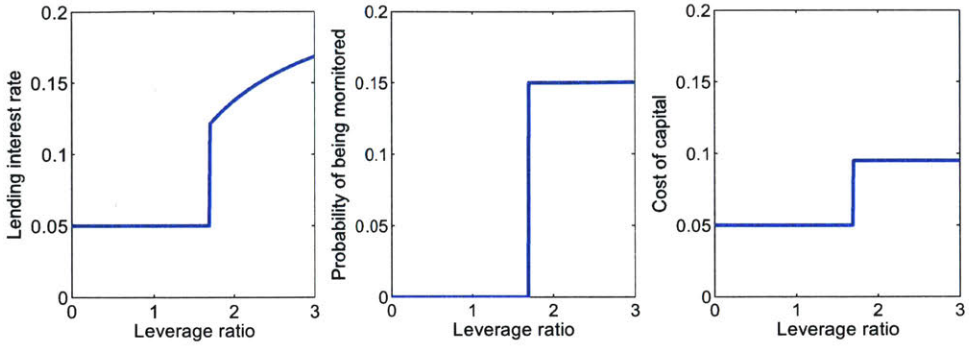
Lemma 2 states that default happens only for highly-leveraged entrepreneurs whose production fails. Moreover, for entrepreneurs with no default risk (i.e., $\tilde{\lambda} \leq \overline{\lambda}$), banks can always get repaid the face value of the loan, and the implied lending rate r^l is equal to the deposit rate r^d. For entrepreneurs facing a risk of default (i.e., $\tilde{\lambda} > \overline{\lambda}$), the implied lending rate is increasing in the leverage ratio to compensate for losses from default. In general, for highly-leveraged entrepreneurs, the lending rate includes a risk premium which depends on the leverage ratio and the fixed intermediation cost from bank monitoring.

Note that the implied lending rate is not equal to the cost of capital facing entrepreneurs. The lending rate should be considered as the interest rate entrepreneurs need to pay when production is successful. But if production fails, entrepreneurs have the option to default and pay less. The cost of capital includes this default option. Therefore, it is a weighted average of the lending rate and the repayment rate during default. This is characterized in Lemma 3.

Lemma 3. *In the credit regime, for an entrepreneur who operates a business with leverage ratio $\tilde{\lambda}$:*

i. If $\tilde{\lambda} \leq \overline{\lambda}$, the cost of capital is r^d.
ii. If $\tilde{\lambda} > \overline{\lambda}$, the cost of capital is $r^d + p\chi$.

In Figure 3-3, we show how the lending rate, the probability of being monitored, and the cost of capital change when the leverage ratio varies. As noted in Proposition 5, only highly-leveraged entrepreneurs are monitored. In particular, there is a default boundary, $\overline{\lambda} = 1.69$, below which the probability of being monitored is zero, and thus both the lending rate and the cost of capital are equal to the deposit rate. If entrepreneurs increase leverage beyond this boundary, they cannot repay the face value of the loan when production fails. Therefore, the probability of being monitored is exactly equal to the production failure rate, p. Since banks are making zero profit, the monitoring cost is completely borne by entrepreneurs, generating a higher cost of capital. Note that the cost of capital in this case is $r^d + p\chi$, which is constant regardless of the leverage ratio (see Lemma 3). This is due to our assumption that the monitoring cost is proportional to the scale of production but not the value of the loan. Moreover, the implied lending rate characterized in Lemma 2 is strictly increasing in the leverage ratio when the leverage ratio is higher than the default boundary. This is because banks have to be repaid more (as reflected by a higher face value Ω) when production succeeds to compensate for larger losses during production failure arising from higher leverage.



Note: The left panel plots the implied lending rate against the leverage ratio; the middle panel plots the monitoring frequency against the leverage ratio; the right panel plots the implied cost of capital against the leverage ratio. All panels are plotted using the following parameter values: $r^d = 0.05$, $\eta = 0.35$, $\delta = 0.06$, $p = 0.15$, $\chi = 0.3$.

Figure 3-3: The lending rate, the monitoring frequency, and the cost of capital for different leverage ratios.

Next we characterize the optimal amount of capital invested by entrepreneurs of type (b, z).

Proposition 6. *In the credit regime, for entrepreneurs of type (b, z), denote the optimal leverage ratio by $\lambda^*(b, z)$ and optimal capital by $k^*(b, z)$. There is a threshold level of wealth $\overline{b}(z)$, such that:*

i. If wealth b is between the participation cost and the threshold level, $\psi \leq b < \overline{b}(z)$, the optimal leverage ratio lies between the default boundary and the inverse of the absconding rate,

$$\begin{aligned} \overline{\lambda} &< \lambda^*(b,z) \leq \lambda, \\ k^*(b,z) &= \min(\lambda(b-\psi), \tilde{k}^h(z)), \end{aligned}$$

where $\tilde{k}^h(z)$ is defined in (iii) below.

ii. If wealth b is above the threshold level, $b \geq \overline{b}(z)$, the optimal leverage ratio is below the default boundary,

$$\begin{aligned} \lambda^*(b,z) &\leq \overline{\lambda}, \\ k^*(b,z) &= \min(\overline{\lambda}(b-\psi), \tilde{k}^l(z)), \end{aligned}$$

where $\tilde{k}^l(z)$ is defined in (iii) below.

iii. $\tilde{k}^h(z)$ is the unconstrained level of capital in the high-leverage region,

$$\tilde{k}^h(z) = [\frac{(1-p)\alpha w}{(r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p)(1-\alpha)}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)z}{w})^{\frac{1}{\nu}}.$$

$\tilde{k}^l(z)$ is the unconstrained level of capital in the low-leverage region,

$$\tilde{k}^l(z) = [\frac{(1-p)\alpha w}{(r^d + (1-p)\delta - p\eta(1-\delta) + p)(1-\alpha)}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)z}{w})^{\frac{1}{\nu}}.$$

Note that $k^h(z) < k^l(z)$ for all z. This is because in the high-leverage region, banks monitor when production fails, which increases the cost of capital. When entrepreneurs are constrained by wealth, increasing the leverage ratio can generate higher revenue, but this may also push them into the "default region", increasing their cost of capital. Entrepreneurs want to maximize profits, but are always facing this trade-off when making investment decisions. For entrepreneurs with low wealth, the marginal return on capital is high. The extra revenue generated by increasing leverage beyond $\bar{\lambda}$ outweighs the increase in the cost of capital, hence they choose higher leverage ($\tilde{\lambda} > \overline{\lambda}$). By contrast, for relatively wealthy entrepreneurs, the marginal return on capital is low. As a result, they choose to borrow less and stay in the low-leverage region to avoid paying the monitoring cost.

Our model features both limited commitment and asymmetric information. In a model with only limited commitment, the supply of credit is rationed exogenously by the parameter λ. When asymmetric information is introduced, since monitoring is costly, in equilibrium there are some entrepreneurs who restrain themselves from borrowing more. For these entrepreneurs, the borrowing constraint imposed by limited commitment is not binding. In fact, they are restricting themselves from using up the credit line precisely because obtaining more credit brings them into the high-leverage region and increases their cost of capital. In this sense, credit rationing is endogenously imposed by entrepreneurs themselves.

Intuitively, the return on production is higher for talented entrepreneurs, which induces them to leverage more. This leads to Proposition 7.

Proposition 7. *The threshold level of wealth $\overline{b}(z)$ is increasing in z.*

Finally, all contracts offered by banks are incentive compatible, although talent is not observable. This implies that entrepreneurs with low talent have no incentive to pretend to be highly talented and ask for a different contract, or vice versa. To see this, divide both sides of equation (3.3.7) by k,

$$(1-p)\frac{\Omega}{k} + p\min(\frac{\Omega}{k}, \eta(1-\delta) + (1+r^d)\frac{b-\psi}{k}) = (1+r^d) + p\chi \cdot \mathbb{1}_{\{\eta(1-\delta)+(1+r^d)\frac{b-\psi}{k} < \frac{\Omega}{k}\}}. \quad (3.3.9)$$

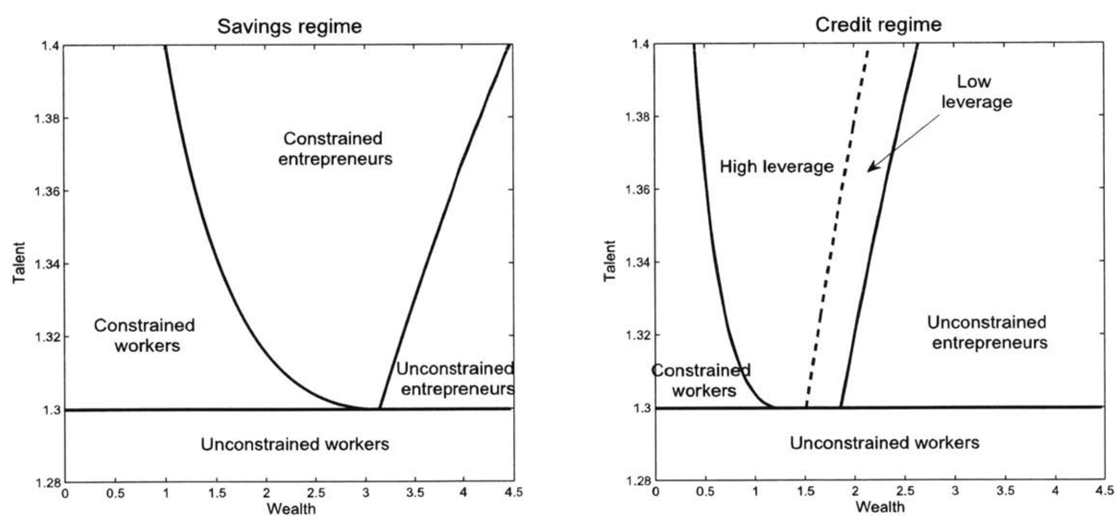
Equation (3.3.9) suggests that the implied gross lending rate, $\frac{\Omega}{k}$, depends only on the inverse of the leverage ratio $\frac{b-\psi}{k}$[13], but not directly on entrepreneurs' talent. That is, capital k and talent z enter equation (3.3.9) only through the leverage ratio, which is observable. Therefore, for all entrepreneurs, given the amount of capital they want to invest (or demand for credit) and the amount of wealth they own (or collateral value), it is impossible to receive a lower interest rate from banks by cheating on talent. This result is obtained because it is assumed that the recovered value of undepreciated working capital does not depend on entrepreneurs' talent.

[13]According to (3.3.6), the inverse of leverage ratio is defined as $\frac{\Delta}{\Phi}$. In the optimal contract illustrated above, $\Delta = b - \psi$, and $\Phi = k$.

Occupational Choice

The occupation map is plotted based on the choice of occupation for agents with different talent z and wealth b, and whether this choice is constrained by wealth. We identify four categories of agents in the savings regime, separated by the solid lines in the left panel of Figure 3-4: unconstrained workers, constrained workers, constrained entrepreneurs, and unconstrained entrepreneurs.

As shown in the figure, there is a certain threshold level of talent (1.3) below which agents always find working for a wage better than operating a business. These agents are identified as unconstrained workers, suggesting that their talent is so low that they never find it optimal to become entrepreneurs. Above this talent level, the figure is further segmented into three regions. In the left region, agents are talented, but do not have sufficient wealth, so they cannot operate businesses at a profitable scale. Hence, they choose to be workers. These are constrained workers. The middle region represents agents with sufficient wealth to operate profitable businesses but scale is still constrained by wealth ($k^*(b,z) < k^s(z)$). These agents are constrained entrepreneurs. Agents in the right region of the figure choose to be entrepreneurs, operating businesses at the unconstrained scale ($k^*(b,z) = k^s(z)$), with the marginal return on capital equal to the deposit rate. Thus, they are identified as unconstrained entrepreneurs.



Note: The left panel plots the occupation choice map in the savings regime, which is partitioned into four regions depending on talent and wealth: unconstrained/constrained workers, unconstrained/constrained entrepreneurs. The right panel plots the occupation choice map in the credit regime. The region of constrained entrepreneurs is further partitioned into entrepreneurs with high leverage ratios and entrepreneurs with low leverage ratios. All panels are plotted using the following parameter values: $r^d = 0.05$, $w = 0.6$, $\eta = 0.35$, $\delta = 0.06$, $\nu = 0.21$, $p = 0.15$, $\alpha = 0.33$, $\lambda = 2.5$, $\psi = 0$, $\chi = 0$.

Figure 3-4: The occupation choice map in the two regimes.

When agents obtain external credit and enter the credit regime, the occupation map changes to the one in the right panel of Figure 3-4. The occupation map for the credit regime is plotted with the same parameter values, and under the assumption that there is no credit participation cost, $\psi = 0$, or monitoring cost, $\chi = 0$.[14] This is to highlight the effect of external credit. Clearly, the area of constrained workers shrinks and that of unconstrained entrepreneurs increases. This implies that agents are more likely to become entrepreneurs and operate their businesses at a larger scale once credit is obtained from banks. Note that the region of constrained entrepreneurs is further partitioned by the dotted line into two sub-regions depending on their leverage ratios. Agents in the low-leverage region are not borrowing much in the sense that the face value of the loan can be repaid even if production fails. Thus, banks do not monitor them, and the lending rate is equal to the deposit rate, as shown in Figure 3-3. By contrast, agents in the high-leverage region default when production fails, in which case banks monitor and seize the recovered undepreciated working capital and after-interest collateral. In accordance with Proposition 6, the high-leverage region is to the left of the low-leverage region, implying that entrepreneurs prefer to leverage more when wealth is low given the high marginal return on capital.

The policy options we consider in Section 3.6, move the lines in the occupation map (Figure 3-4), and alter the relative income received by different agents. This kind of micro-level adjustment for each agent impacts the aggregate economy and generates a movement in GDP and income inequality.

3.3.2 Competitive Equilibrium

Given an initial joint probability density distribution of wealth and talent $h_0(b,z)$, a competitive equilibrium consists of allocations $\{c_t(b,z), k_t(b,z), l_t(b,z)\}_{t=0}^{\infty}$, sequences of joint distributions of wealth and talent $\{h_t(b,z)\}_{t=1}^{\infty}$ and prices $\{r^d(t), w(t)\}_t$, such that:

(1). Agents of type (b,z) optimally choose the underlying regime, occupations, consumption $c_t(b,z)$, capital $k_t(b,z)$, and labor $l_t(b,z)$ to maximize utility at $t \geq 0$.

(2). The capital market clears at all $t \geq 0$,

$$\iint_{(b,z)\in E(t)} k_t(b,z)h_t(b,z)dbdz = \iint_{(b,z)} bh_t(b,z)dbdz - \psi \iint_{(b,z)\in Fin(t)} h_t(b,z)dbdz,$$

where $E(t)$ is the set of agents, who choose to be entrepreneurs at time t; $Fin(t)$ is the set of agents, who are in the credit regime.

(3). The labor market clears at all $t \geq 0$,

$$\iint_{(b,z)\in E(t)} l_t(b,z)h_t(b,z)dbdz = \iint_{(b,z)\notin E(t)} h_t(b,z)dbdz.$$

[14] Note that we also use the same wage and interest rate while plotting the occupation choice map for the credit regime. This is to highlight the partial equilibrium result of moving an agent from the savings regime to the credit regime. When financial inclusion allows more agents to get credit, the wage and interest rate would also change in general equilibrium.

(4). $\{h_t(b,z)\}_{t=1}^{\infty}$ evolves according to the equilibrium mapping.

$$h_{t+1}(\bar{b},\bar{z})db = \gamma\mu(\bar{z})\int_z \mathbb{1}_{\{b'=\bar{b}\}}h_t(b,z)dbdz + (1-\gamma)\int_b \mathbb{1}_{\{b'=\bar{b}\}}h_t(b,\bar{z})db,$$

where b' is the bequest for agents of type (b,z), and $\mathbb{1}_{\{b'=\bar{b}\}}$ is an indicator function which equals 1 if $b'=\bar{b}$, and equals 0 otherwise.

The steady state of the economy is defined as the invariant joint distribution of wealth and talent $h(b,z)$,

$$h(b,z) = \lim_{t\to\infty} h_t(b,z).$$

3.4 Distinguishing the Impact of Financial Constraints

In this section, we explore the impact of different financial constraints on GDP when these constraints are relaxed separately and in combination. We first provide a partial equilibrium analysis focusing on individuals' output and credit access conditions. This enables us to uncover the underlying mechanisms of the model and distinguish the impact of different financial constraints. We then decompose GDP and TFP to shed light on the macroeconomic effects of financial inclusion.

Financial inclusion is reflected by three parameters in our model. The credit participation cost ψ directly measures the difficulty of obtaining credit. A decrease in its value therefore reflects greater financial access. The borrowing constraint parameter λ coincides directly with the maximum leverage ratio, an increase in which reflects lower collateral requirements. Finally, a decrease in the monitoring cost χ indicates an increase in the efficiency of financial intermediation.

Because financial inclusion is multidimensional, it is difficult to precisely identify these three parameters from an empirical standpoint. However, one can find evidence of policies that address one dimension or the other. For example, Assuncao, Mityakov and Townsend [17] and Alem and Townsend [12] find that the distance to a bank branch matters for credit access, which suggests that policies that promote branch openings in rural unbanked locations could help reduce the credit participation cost ψ in our model.[15] Moreover, during the recent financial crisis, many countries widened the range of securities that could be accepted as collateral with the aim of boosting lending to firms and households. This reflects an increase in λ in our model. Finally, financial liberalization and the resultant competition between financial institutions could accelerate investment in computerization, thereby improving intermediation efficiency (as reflected by a decrease in χ in our model). For example, from 1985 to 1994, the Thai banking sector had become a more capital-intensive industry, substituting physical capital for labor. The

[15] Many developing countries have conducted such kind of policies. For example, after a bank nationalization in 1969, the Indian government launched an ambitious social banking program which sought to improve the access of the rural poor to formal credit and savings opportunities [49].

average cost of raising funds decreased from 14.40% in 1985 to 5.61% in 1994 for large-sized banks [191].

We distinguish the effect of financial inclusion on the extensive margin and the intensive margin. On the one hand, relaxing financial constraints can increase GDP through the extensive margin by increasing the credit access ratio (i.e., moving entrepreneurs from the savings regime to the credit regime). On the other hand, relaxing financial constraints enable entrepreneurs in the credit regime to produce more output, which boosts GDP. This effect operates on the intensive margin.

3.4.1 The Impact at the Individual Level

We focus on the partial equilibrium with interest rates and wages fixed. We consider constant returns to scale production function (i.e., $\nu = 0$) to simplify the algebra. All the results also hold in the general case with decreasing returns to scale production function, but at the expense of losing closed-form solutions.

When $\nu = 0$, the threshold level of wealth, $\overline{b}(z)$, which separates "high-leverage" and "low-leverage" regions in Proposition 6 only takes two values, 0 or ∞, as shown in Lemma 4.

Lemma 4. *In the credit regime with $\nu = 0$, for an entrepreneur of type (b, z), there exists a threshold of talent $\overline{z}$, such that:*

i. If $z > \overline{z}$, then $\overline{b}(z) = \infty$ and the optimal leverage ratio is $\lambda^*(b, z) = \lambda$.
ii. If $z \leq \overline{z}$, then $\overline{b}(z) = 0$ and the optimal leverage ratio is $\lambda^*(b, z) = \overline{\lambda}$.
iii. The threshold of talent is $\overline{z} = \frac{w}{1-\alpha}[(\frac{p\chi\lambda}{(\lambda - \overline{\lambda})} + (1 + r^d) + p\eta(1-\delta) - (1-p)(1-\delta))\frac{1-\alpha}{\alpha w(1-p)}]^{\alpha}$.

In fact, Lemma 4 can be considered as a corollary of Proposition 7 for the case where $\nu = 0$. Talented entrepreneurs face a steeper profit function, therefore they would like to choose higher leverage ratios. When the production function exhibits constant returns to scale, we obtain a "bang-bang" solution for the wealth threshold $\overline{b}(z)$ since the unconstrained level of capital is infinite.

Denote $\underline{b}(\psi, \lambda, \chi; z)$ as the threshold of wealth above which entrepreneurs of type (b, z) choose to enter the credit regime. A lower $\underline{b}(\psi, \lambda, \chi; z)$ implies that, all else equal, entrepreneurs with talent z are more likely to enter the credit regime.

In the following theorem, we show that relaxing financial constraints (decreasing ψ, increasing λ, or decreasing χ) can have differential quantitative impacts on reducing agents' wealth thresholds $\underline{b}(\psi, \lambda, \chi; z)$, reflecting their effect on the extensive margin.

Theorem 1. ***The Impact of Financial Constraints on the Extensive Margin***
In the credit regime with fixed interest rates and wages, and when $\nu = 0$:

i. Relaxing each financial constraint improves credit access:

$$-\frac{\partial \underline{b}}{\partial \psi} \leq 0, \quad \frac{\partial \underline{b}}{\partial \lambda} \leq 0, \quad -\frac{\partial \underline{b}}{\partial \chi} \leq 0.^{16} \tag{3.4.1}$$

ii. Financial constraints are substitutes on the extensive margin:

$$-\frac{\partial^2 \underline{b}}{\partial \psi \partial \lambda} \geq 0, \quad -\frac{\partial^2 \underline{b}}{\partial \lambda \partial \chi} \geq 0, \quad \frac{\partial^2 \underline{b}}{\partial \chi \partial \psi} \geq 0. \tag{3.4.2}$$

Theorem 1.i indicates that relaxing any constraint can unambiguously reduce the wealth threshold $\underline{b}(\psi, \lambda, \chi; z)$, which facilitates credit access (as captured by the first derivatives). The exact quantitative impacts, however, depend on individual characteristics (b, z) and country-specific parameters $(p, \eta, \delta, \alpha)$ as presented in Appendix C.1.7.[17] Note that the underlying mechanisms for relaxing different constraints are not identical. Lowering the credit participation cost ψ induces entrepreneurs to enter the credit regime by decreasing the ex-ante cost of borrowing, while a lower intermediation cost χ reduces the ex-post cost of borrowing. Relaxing the borrowing constraint λ motivates entrepreneurs to obtain credit by increasing their profits in the credit regime.

Importantly, Theorem 1.ii indicates that the three financial constraints are pair-wise substitutes on the extensive margin. For example, a lower credit participation cost dampens the effect of relaxing the borrowing constraint or reducing the intermediation cost. This is because a lower credit participation cost results in a lower wealth threshold $\underline{b}(\psi, \lambda, \chi; z)$, thus relaxing the borrowing constraint and reducing the intermediation cost have less of an impact on further reducing this threshold. In other words, when the credit participation cost is low, the credit access ratio is already high, so with little room for increasing this ratio further through the other two channels. Essentially, the substitution effect arises due to the natural bound on the maximum credit access ratio, which is 100%.

Financial inclusion increases agents' well being not only through its impact on promoting credit access (the extensive margin), but also by increasing the net output of entrepreneurs living in the credit regime (the intensive margin).

We define entrepreneurs' net output as the expected output net of direct costs arising from financial frictions, if any.[18] Thus in the savings regime, the net output $y^S(\psi, \lambda, \chi; b, z)$ is equal to the difference between the end-of-period wealth and the beginning-of-period wealth plus the user

---

[16] There is a negative sign in front of $\frac{\partial \underline{b}}{\partial \psi}$ and $\frac{\partial \underline{b}}{\partial \chi}$ since relaxing the two constraints implies reducing the credit participation cost and the intermediation cost.

[17] For the case with $z \leq \underline{z}$, the impact of relaxing the borrowing constraint λ has no impact on the wealth threshold (i.e., $\frac{\partial \underline{b}}{\partial \lambda} = 0$). This is because entrepreneurs with $z \leq \underline{z}$ choose the leverage ratio $\overline{\lambda} = \frac{1+r^d}{1+r^d-\eta(1-\delta)}$ (see Lemma 4). Therefore, parameter λ can only affect $\overline{\lambda}$ through its impact on the equilibrium interest rate r^d, which is ruled out in our partial equilibrium analysis. Similar arguments also apply to Theorem 2.

[18] The net output defined in this way enables us to calculate GDP as the sum of all entrepreneurs' net output.

cost of capital and the labor cost:

$$y^S(\psi, \lambda, \chi; b, z) = \pi^S(b, z) + (r^d + \delta)k^*(b, z) + (1 - p)wl^*(b, z) - (1 + r^d)b, \tag{3.4.3}$$

where $\pi^S(b,z)$, $k^*(b,z)$, and $l^*(b,z)$ are solutions to problem (3.3.4).

In the credit regime, the net output $y^C(\psi, \lambda, \chi; b, z)$ is

$$y^C(\psi, \lambda, \chi; b, z) = \pi^C(b, z) + (r^d + \delta)k^*(b, z) + (1 - p)wl^*(b, z) - (1 + r^d)b, \tag{3.4.4}$$

where $\pi^C(b,z)$, $k^*(b,z)$, and $l^*(b,z)$ are solutions to problem (3.3.8).

The next theorem presents the impact of relaxing financial constraints on agents' net output in the credit regime.

Theorem 2. ***The Impact of Financial Constraints on the Intensive Margin***
In the credit regime with fixed interest rates and wages, and when $v = 0$:

i. Relaxing each financial constraint raises net output:

$$-\frac{\partial y^C}{\partial \psi} \geq 0, \quad \frac{\partial y^C}{\partial \lambda} \geq 0, \quad -\frac{\partial y^C}{\partial \chi} \geq 0. \tag{3.4.5}$$

ii. Financial constraints are complements on the intensive margin:

$$-\frac{\partial^2 y^C}{\partial \psi \partial \lambda} \geq 0, \quad -\frac{\partial^2 y^C}{\partial \lambda \partial \chi} \geq 0, \quad \frac{\partial^2 y^C}{\partial \chi \partial \psi} \geq 0. \tag{3.4.6}$$

Theorem 2.i indicates that relaxing any constraint increases entrepreneurs' net output in the credit regime. However, it is important to emphasize that the exact quantitative impacts are different (see Appendix C.1.8). Theorem 2.ii says that relaxing any two constraints has complementary effects in boosting output. For example, when the credit participation cost is lower, entrepreneurs are left with more wealth after entering the credit regime. Since both the amount of credit and the total intermediation cost are proportional to wealth, relaxing the borrowing constraint and reducing the intermediation cost increases business profits by more.

In sum, the above partial equilibrium discussions provide important policy implications. First, financial inclusion policies matter since the quantitative effects differ when different financial constraints are relaxed. In general, the quantitative effects depend on the joint distribution of agents' wealth and talent and country-specific characteristics. If policy makers are constrained in using a single policy, then choosing the right policy to address the "bottleneck" constraint is tempting. Second, different constraints are complements on the intensive margin and substitutes on the extensive margin. Therefore, an optimal combination of policies is necessary in order to boost GDP.

3.4.2 The Impact on the Aggregate Economy

We now discuss the general equilibrium impact of financial constraints on the aggregate economy.

GDP Decomposition

The economy's GDP can be written as the sum of net output produced by entrepreneurs in the savings regime and those in the credit regime,

$$\mathrm{GDP} = \underbrace{\int_z \int_0^{\bar{b}(\psi,\lambda,\chi;z)} y^S(\psi,\lambda,\chi;b,z)h(b,z)dbdz}_{\text{Total output in savings regime}} + \underbrace{\int_z \int_{\bar{b}(\psi,\lambda,\chi;z)}^{\infty} y^C(\psi,\lambda,\chi;b,z)h(b,z)dbdz}_{\text{Total output in credit regime}}. \quad (3.4.7)$$

Note that entrepreneurs' output in the savings regime $y^S(\psi,\lambda,\chi;b,z)$ is indirectly affected by financial parameters (ψ,λ,χ) through changes in equilibrium wages and interest rates. When financial constraints are relaxed from (ψ,λ,χ) to (ψ',λ',χ') with $\psi' \leq \psi$, $\lambda' \geq \lambda$, and $\chi' \leq \chi$, the increase in GDP can be decomposed into three margins as follows:

$$\begin{aligned}
\Delta\mathrm{GDP}_{(\psi,\lambda,\chi)\text{ to }(\psi',\lambda',\chi')} &= \underbrace{\int_z \int_{\bar{b}(\psi',\lambda',\chi';z)}^{\bar{b}(\psi,\lambda,\chi;z)} [y^C(\psi',\lambda',\chi';b,z) - y^S(\psi,\lambda,\chi;b,z)]h(b,z)dbdz}_{\text{Gains on the extensive margin}} \\
&\quad + \underbrace{\int_z \int_{\bar{b}(\psi,\lambda,\chi;z)}^{\infty} [y^C(\psi',\lambda',\chi';b,z) - y^C(\psi,\lambda,\chi;b,z)]h(b,z)dbdz}_{\text{Gains on the intensive margin}} \quad (3.4.8) \\
&\quad - \underbrace{\int_z \int_0^{\bar{b}(\psi',\lambda',\chi';z)} [y^S(\psi,\lambda,\chi;b,z) - y^S(\psi',\lambda',\chi';b,z)]h(b,z)dbdz}_{\text{General equilibrium effects in the savings regime}}.
\end{aligned}$$

First, more entrepreneurs enter the credit regime as implied by a reduction in $\bar{b}$. Gains on the extensive margin arise since external credit enables entrepreneurs to expand their businesses and produce more output. Moreover, gains on the intensive margin (within the credit regime) accrue since relaxing financial constraints limits the losses from financial contracts (lower ψ) and inefficient monitoring (lower χ) and improves the provision of credit (higher λ). When general equilibrium effects are considered, gains on both margins are likely to be smaller due to the increase in the equilibrium wage and interest rate. In fact, entrepreneurs living in the savings regime shrink the size of their businesses and produce less output after financial inclusion due to the general equilibrium effect. This implies that not everyone is better off with greater financial inclusion. In particular, entrepreneurs remaining in the savings regime would incur losses due to higher equilibrium factor prices. Some entrepreneurs in the credit regime can also lose if gains on the intensive margin are lower than the losses from the general equilibrium effects. We discuss the heterogeneous welfare effects in Subsection 3.6.4.

An important variation which is not captured in the GDP decomposition is the endogeneity of the steady-state distribution of wealth and talent. Equation (3.4.8) holds only for analyzing the immediate impact of financial inclusion as the distribution of wealth and talent $h(b,z)$ is assumed to be unchanged. However, in the new steady state after relaxing financial constraints, the endogenously determined joint distribution would also have a new shape. Therefore, the long-run effects of financial inclusion cannot be easily evaluated based on equation (3.4.8). We provide a numerical analysis in Subsection 3.6.3.

TFP Decomposition

In this subsection, we provide a TFP decomposition in the spirit of Jeong and Townsend [139] by exploiting the equivalence between growth accounting by factors and growth accounting by regimes. Again, we consider relaxing the financial constraints from (ψ, λ, χ) to (ψ', λ', χ') with $\psi' \leq \psi$, $\lambda' \geq \lambda$, and $\chi' \leq \chi$. Our decomposition generalizes that of Jeong and Townsend [139] since in our model the credit access ratio is endogenous and depends on the three micro-founded financial frictions. All the variables presented in this subsection are functions of (ψ, λ, χ), and the explicit dependence is omitted unless necessary.

We follow Buera and Shin [46] and define the model-implied TFP as

$$TFP = \frac{Y}{K^{\alpha} L^{1-\alpha}}, \tag{3.4.9}$$

where Y is aggregate output, K is aggregate capital, and L is aggregate labor.

Growth Accounting by Factors By taking an approximation of the first difference of equation (3.4.9), we obtain

$$g_{TFP} = g_Y - \alpha g_K - (1-\alpha) g_L, \tag{3.4.10}$$

where g_x is the growth rate of variable x, i.e., $g_x = \dfrac{X(\psi', \lambda', \chi')}{X(\psi, \lambda, \chi)} - 1$.

Since the economy consists of two regimes, aggregate capital and labor are equal to the sum of capital and labor employed by entrepreneurs living in the two regimes separately. Denote $\overline{K}^s / \overline{L}^s$, and $\overline{K}^c / \overline{L}^c$ as the average capital/labor employed by entrepreneurs in the savings regime, and the credit regime, respectively. Denote p_c as the percent of entrepreneurs living in the credit regime and E as the total number of entrepreneurs. Therefore, aggregate capital and labor can be written as,

$$\begin{aligned} K &= E(1-p_c)\overline{K}^s + E p_c \overline{K}^c, \\ L &= E(1-p_c)\overline{L}^s + E p_c \overline{L}^c. \end{aligned} \tag{3.4.11}$$

From (3.4.11), the growth rates of aggregate factors g_K and g_L can be further decomposed into,

$$\begin{aligned} g_K &= g_E + (s_K^c - s_K^s)g_{p_c} + s_K^s g_{\overline{K}^s} + s_K^c g_{\overline{K}^c}, \\ g_L &= g_E + (s_L^c - s_L^s)g_{p_c} + s_L^s g_{\overline{L}^s} + s_L^c g_{\overline{L}^c}, \end{aligned} \tag{3.4.12}$$

where $s_K^c = \frac{E(1-p_c)\overline{K}^s}{2K}(\psi,\lambda,\chi) + \frac{E(1-p_c)\overline{K}^s}{2K}(\psi',\lambda',\chi')$ is the average fraction of capital employed by entrepreneurs in the credit regime before and after relaxing the financial constraints. The variables s_K^s, s_L^c, and s_L^s are defined in a similar way. Note that $s_K^s + s_K^c = 1$, and $s_L^s + s_L^c = 1$.

Substituting (3.4.12) into (3.4.10), we obtain

$$g_Y = g_{TFP} + g_E + \alpha(s_K^c - s_K^s)g_{p_c} + \alpha s_K^s g_{\overline{K}^s} + \alpha s_K^c g_{\overline{K}^c} + (1-\alpha)(s_L^c - s_L^s)g_{p_c} + (1-\alpha)s_L^s g_{\overline{L}^s} + (1-\alpha)s_L^c g_{\overline{L}^c}. \tag{3.4.13}$$

Growth Accounting by Regimes The economy's output Y is equal to the sum of output in each regime,

$$Y = E(1-p_c)\overline{Y}^s + Ep_c\overline{Y}^c, \tag{3.4.14}$$

where $\overline{Y}^s$ and $\overline{Y}^c$ are the average output produced by entrepreneurs in the savings and credit regimes, respectively.

Thus, the growth rate of output can be expressed as

$$g_Y = g_E + (s_Y^c - s_Y^s)g_{p_c} + s_Y^s g_{\overline{Y}^s} + s_Y^c g_{\overline{Y}^c}, \tag{3.4.15}$$

where $s_{Y^s} = \frac{Y^s}{2Y}(\psi,\lambda,\chi) + \frac{Y^s}{2Y}(\psi',\lambda',\chi')$ and $s_{Y^c} = \frac{Y^c}{2Y}(\psi,\lambda,\chi) + \frac{Y^c}{2Y}(\psi',\lambda',\chi')$ are the average fraction of output produced in the savings regime and credit regime, respectively.

Decomposition Formula Equating the two growth accounting identities (3.4.13) and (3.4.15), we obtain,

$$\begin{aligned} g_{TFP_{(\psi,\lambda,\chi)}\text{ to }(\psi',\lambda',\chi')} &= \underbrace{(s_Y^c - s_Y^s - \alpha(s_K^c - s_K^s) - (1-\alpha)(s_L^c - s_L^s))g_{p_c}}_{\text{Between-regime shifting}} \\ &\quad + \underbrace{s_Y^s g_{\overline{Y}^s} - \alpha s_K^s g_{\overline{K}^s} - (1-\alpha)s_L^s g_{\overline{L}^s}}_{\text{Growth within savings regime}} \\ &\quad + \underbrace{s_Y^c g_{\overline{Y}^c} - \alpha s_K^c g_{\overline{K}^c} - (1-\alpha)s_L^c g_{\overline{L}^c}}_{\text{Growth within credit regime}} \end{aligned} \tag{3.4.16}$$

Therefore, the growth rate of TFP is decomposed into three terms. The first term captures growth generated by entrepreneurs who shift from the savings regime to the credit regime. In fact, the between-regime shifting effect can be considered as TFP gains on the extensive margin. The second and third term captures TFP growth within the savings regime and the credit regime, respectively. TFP growth within the credit regime can be considered as TFP gains on the intensive margin, and growth within the savings regime is caused by general equilibrium effects.

To evaluate and identify the macro, general equilibrium, and long-term impact of financial inclusion, a calibration and quantitative analysis is examined in the following sections.

3.5 Data and Calibration

We calibrate the model for six countries at various stages of economic development: three low-income countries (Uganda in 2005, Kenya in 2006, and Mozambique in 2006), and three emerging market economies (Malaysia in 2007, the Philippines in 2008 and Egypt in 2007). We use the data from World Bank Enterprise Surveys and World Development Indicators (WDI).[19]

In general, financial inclusion in low-income countries is more constrained compared with emerging market economies across different dimensions, as indicated by high collateral requirements, low share of firms with credit, and high borrowing costs (see Table 3.1). However, there is significant heterogeneity within country groups across these different dimensions. For example, access to the financial system, as measured by the share of firms with credit, is lower in Mozambique than in Uganda and Kenya, despite relatively lower collateral requirements and interest rate spreads. In the Philippines, collateral requirements are very high, while interest rate spreads are comparable to other emerging market economies in the sample.

Table 3.1: Overview of the Data

	Low-income countries			Emerging market economies		
	Uganda	Kenya	Mozambique	Malaysia	The Philippines	Egypt
Savings (% of GDP)	8	15.4	7.1	39	25.7	24.5
Collateral (% of loan)	173	120.8	92	64.6	238.4	85.5
Firms with credit (%)	17.2	25.4	14.2	60.4	33.2	17.4
Non-perfor. loan (%)	2.3	10.6	3.1	8.5	4.5	19.3
Interest rate spread	10.9	8.5	8.2	3.3	4.3	6.1
Overhead costs/assets	6.9	6.6	7.4	1.5	3.2	1.5
Top 5% emp. share	53.8	54.1	41.3	29.5	52.7	58.4
Top 10% emp. share	64.2	66.	55.8	46.3	65.7	72.7
Top 20% emp. share	74.6	81	71.9	63.5	79	85.9
Top 40% emp. share	86.4	93.2	87.2	84.1	90.8	95

We use standard values from the literature for some of the parameters. The one-year depreciation rate δ is set to be 0.06. Following Buera and Shin (2013), we choose the share of output going to the variable factors in the production function $1 - v$ to be 0.79, of which the share of capital α is 0.33. The probability that the offspring inherits the talent of her parents γ is assumed to be 0.894. The other parameters are estimated by matching the simulated moments to real data.

Each generation is interpreted as one year as in Gine and Townsend [103] and Jeong and

[19]The selection of countries is mainly driven by data availability. First and foremost, we need sufficient cross-section units to run our framework. The numbers of firms in our sample are 563 for Uganda, 781 for Kenya, 599 for Mozambique, 1115 for Malaysia, 1326 for the Philippines, and 996 for Egypt. Second, we consider relatively recent cases but exclude countries with financial turbulence around the year of the survey.

Table 3.2: Data, model and calibrated parameters

	Uganda			Kenya			Mozambique		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	8	8	$\omega = 0.08$	15.4	15.4	$\omega = 0.15$	7.1	7.1	$\omega = 0.07$
Collateral (% of loan)	173	173	$\lambda = 1.58$	120.8	120.8	$\lambda = 1.83$	92	92	$\lambda = 2.09$
Firms with credit (%)	17.2	17.1	$\psi = 0.03$	25.4	25.1	$\psi = 0.07$	14.2	14.2	$\psi = 0.03$
Non-perfor. loan (%)	2.3	2.4	$p = 0.15$	10.6	5.8	$p = 0.17$	3.1	3.1	$p = 0.14$
Interest rate spread	10.9	8.9	$\chi = 0.90$	8.5	11.3	$\chi = 0.61$	8.2	11.2	$\chi = 0.95$
Overhead costs/assets	6.9	6.6	$\eta = 0.37$	6.6	6.5	$\eta = 0.45$	7.4	7.3	$\eta = 0.54$
Top 5% emp. share	53.8	52.9	$\theta = 4.80$	54.1	58.1	$\theta = 4.40$	41.3	46.9	$\theta = 6.00$
Top 10% emp. share	64.2	64.4		66.9	70.1		55.8	58.9	
Top 20% emp. share	74.6	74.7		81	80.5		71.9	69.1	
Top 40% emp. share	86.4	84.7		93.2	88.7		87.2	80.5	
	Malaysia			The Philippines			Egypt		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	39	39	$\omega = 0.39$	25.7	25.7	$\omega = 0.26$	24.5	24.5	$\omega = 0.25$
Collateral (% of loan)	64.6	64.6	$\lambda = 2.56$	238.4	238.4	$\lambda = 1.42$	85.5	85.5	$\lambda = 2.17$
Firms with credit (%)	60.4	60.5	$\psi = 0.13$	33.2	33.0	$\psi = 0.07$	17.4	17.5	$\psi = 0.23$
Non-perfor. loan (%)	8.5	7.6	$p = 0.12$	4.5	3.8	$p = 0.11$	19.3	15.7	$p = 0.28$
Interest rate spread	3.3	5.8	$\chi = 0.16$	4.3	6.2	$\chi = 0.6$	6.1	8.0	$\chi = 0.08$
Overhead costs/assets	1.5	1.5	$\eta = 0.37$	3.2	3.1	$\eta = 0.27$	1.5	1.4	$\eta = 0.44$
Top 5% emp. share	29.5	34.5	$\theta = 6.80$	52.7	54.5	$\theta = 4.30$	58.4	62.0	$\theta = 4.25$
Top 10% emp. share	46.3	46.9		65.7	66.0		72.7	74.2	
Top 20% emp. share	63.5	61.5		79	77.0		85.9	83.5	
Top 40% emp. share	84.1	78.5		90.8	87.0		95	90.3	

Townsend [138].[20] We match the gross savings rate, which measures the overall funds available for financial intermediation in a closed economy, in the data and the model to calibrate the optimal bequest rate, ω. We use the average value of collateral as a percentage of the loan to calibrate parameter λ, which captures the degree of financial frictions caused by limited commitment.[21]

The credit participation cost ψ, the intermediation cost χ, the probability of production failure p, the project recovery rate η, and the parameter governing the talent distribution θ are jointly calibrated to match the moments of the percent of firms with a line of credit, non-performing loans (NPLs) as a percentage of total loans, the interest rate spread, the bank overhead costs to assets ratio[22], and the employment share distribution (using four brackets of employment shares—top 5% / 10% /20% / 40%). Even though parameters ψ, χ, p, η, and θ affect the value of all these moments, and are jointly calibrated, each moment is primarily affected by particular parameters. Specifically, the moment of percent of firms with credit is mostly determined by the credit participation cost ψ. Increasing the value of ψ decreases the percent of firms with credit. The NPL ratio is mainly determined by parameter p. However, the relationship is non-monotonic for some parameter values. For example, when p increases, if entrepreneurs' leverage ratios

[20] In Appendix C.2, we consider a model with forward-looking agents as a robustness check.

[21] Note that the definition of the loan-to-collateral ratio in the model is slightly different from that in the data. In our model, we assume that entrepreneurs deposit their wealth b in banks as collateral and borrow k^*. Hence, the loan-to-collateral ratio is $\lambda_{model} = \frac{k^*}{b}$. In the data, the loan-to-collateral ratio is measured as $\lambda_{data} = \frac{k^* - b}{b}$ (i.e., entrepreneurs put their own wealth into the project, and the wealth is used as collateral to borrow the difference $k^* - b$). The relationship between the two is $\lambda_{data} = \lambda_{model} - 1$.

[22] The bank overhead costs to assets ratio is obtained from World Bank Global Financial Development Database (GFDD). We compute its average value over the period 2000 – 2011. Greenwood, Sanchez and Wang [112, Figure 6] show that this variable reflects monitoring efficiency in a cross-country analysis.

were unchanged, NPLs should increase. However, a higher p may reduce leverage ratios as higher riskiness induce entrepreneurs to take less loans, which results in a lower NPL ratio. The employment share distribution is matched primarily by adjusting the value of parameter θ, which governs the shape of the entrepreneurial talent distribution. Parameter η is jointly identified with parameter χ to simultaneously match the bank overhead costs to assets ratio and the interest rate spread.

From Table 3.2, it is clear that the model performs well in terms of matching the macroeconomic moments. The percent of firms with credit and the bank overhead costs to assets ratio generated by the model are almost exactly matched with those in the data for all six countries. Both NPLs and interest rate spreads are matched well, although some countries have high NPLs but relatively low interest rate spreads (e.g. Maylasia and Egypt) while other countries have low NPLs and high interest rate spreads (e.g. Uganda and Mozambique). The employment share distribution is also captured, but in general the model tends to generate more large firms compared to the data.

The linkages between different characteristics of an economy and financial inclusion are complex. For example, it might seem surprising that the calibrated credit participation cost ψ, in general, is lower in low-income countries despite their lower credit access ratios. This is because ψ is not the only factor determining credit access. In fact, both λ and χ affect the credit access ratio in the model—higher λ and lower χ increase credit access in emerging market economies, whose partial equilibrium effects are summarized in Theorem 1. Moreover, these countries have higher savings rates (higher ω), which implies that agents transfer more wealth to the next generation. In this case, the credit participation cost is a relatively smaller proportion of agents' wealth in emerging market economies. Therefore, it is less binding, as reflected in the high credit access ratio. In the next section, we analyze the macroeconomic implications of financial inclusion and identify the role that country characteristics play in the process.

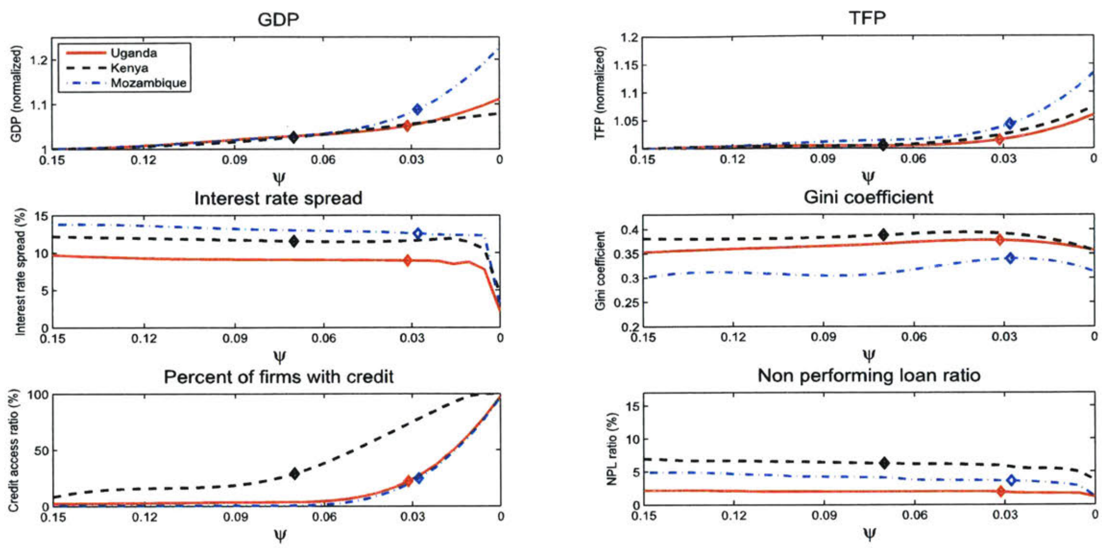
3.6 Quantitative Analysis

In our quantitative analysis, we first evaluate the impact of relaxing different financial constraints individually for the countries in our sample. Then we take the Phillipines as an example to evaluate the interaction effect among the three financial constraints. Finally, we decompose GDP and TFP and provide a welfare analysis.

3.6.1 Evaluation of Policy Options

This subsection analyzes the policy implications of promoting financial inclusion across these three dimensions for the countries in our sample. Specifically, we focus on changes across the steady states of the economy when these parameters change.[23] Figures 3-5 – 3-10 below present

[23]Note that it takes time for the economy to transition from one steady state to another when these parameters change. The models' transitional dynamics are also computable. However, we only report the simulation results in



Note: The solid line represents Uganda, the dashed line represents Kenya, and the dash-dotted line represents Mozambique. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 3-5: Comparative statics: Credit participation cost—low-income countries.

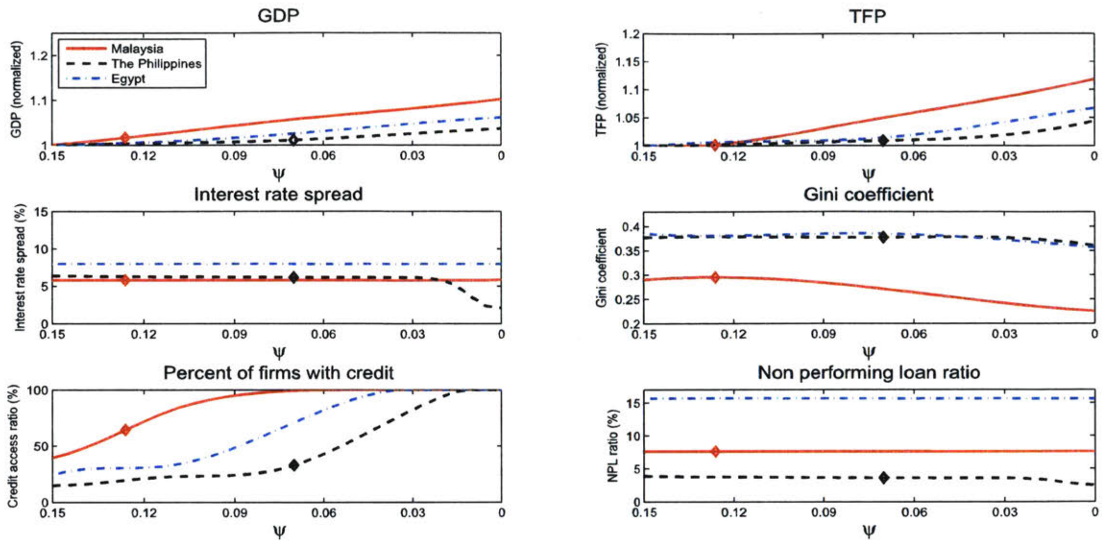
the simulation results when each of the three financial parameters changes separately (on the horizontal axis).

Reducing the Participation Cost

Figures 3-5 – 3-6 present the impact of a decline in the credit participation cost ψ from 0.15 to 0 (moving from left to right). A decrease in the participation cost pushes up GDP through its positive impact on investment for two reasons. First, a lower credit participation cost enables more entrepreneurs to have access to credit. Second, less funds are wasted in unproductive contract negotiation. Both tend to increase entrepreneurs' investment in production. TFP increases as capital is more efficiently allocated among entrepreneurs.

The average interest rate spread is stable when ψ is high, but eventually decreases in some countries (Uganda, Mozambique, Mozambique and the Philippines). A smaller ψ enables some of the constrained workers to become entrepreneurs. These entrepreneurs are severely wealth constrained, and therefore choose high leverage ratios, driving the average interest rate spread

steady states because focusing on the transitional dynamics could be misleading for at least two reasons: (1) The transition is rapid at the beginning but becomes slower when the economy is approaching the steady state. This is inconsistent with reality, where the impact of financial reforms happen gradually, or at least the immediate impact is not significant; and (2) Numerical errors are large relative to those in the steady state, possibly leading to overshooting in some variables if parameters are adjusted a lot. These two problems associated with computing transitional dynamics exist for all quantitative macroeconomic models, although the first problem could be mitigated to some extent if agents were modeled as forward-looking [e.g. 46, and our robustness check].



Note: The solid line represents Malaysia, the dashed line represents the Philippines, and the dash-dotted line represents Egypt. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 3-6: Comparative statics: Credit participation cost—emerging market economies.

up.[24]

As financial inclusion increases, income inequality (the Gini coefficient in our simulation) first increases and then decreases in low-income countries, consistent with the Kuznets' hypothesis. This is because when ψ decreases from a particularly high value, it only enables a very small number of constrained workers to become entrepreneurs. As shown in Figure 3-5, the percent of firms with credit is almost unchanged for high values of ψ. However, the effect on the incumbent entrepreneurs is large since it reduces their contracting costs, thus allowing them to invest more capital in production. These entrepreneurs make higher profits, pushing up income inequality. If ψ decreases further (all the way to zero), it becomes disproportionately more beneficial for constrained workers and entrepreneurs without access to credit. This enables relatively poorer agents to earn higher income, driving down the Gini coefficient.

By contrast, in emerging market economies, this Kuznets' pattern is not observed. The reason is that at $\psi = 0.15$, financial systems in these economies are already highly developed compared to low-income countries. In other words, emerging market economies are already in the "second stage" of development.[25] A decrease in ψ unambiguously leads to a lower Gini coefficient in emerging market economies, such as Malaysia. Since ψ is a fixed cost, a decrease in ψ benefits

[24] A decrease in ψ also has a countervailing effect on interest rate spreads in our model. As entrepreneurs need to pay less to get credit, they become richer and tend to deleverage, resulting in a lower interest rate spread. However, this effect is secondary.

[25] As reflected in Figures 3-5 – 3-6, at $\psi = 0.15$, the percent of firms with credit is about 50% in Malaysia while it is close to zero in Uganda. Identified by the circle on the blue solid line in Figure 3-5, Uganda in 2005 was about to move from the initial stage of development (in the Kuznets' sense).

poor entrepreneurs disproportionately as this constitutes a larger proportion of their wealth. In the Philippines and Egypt, the decline in inequality is less noticeable, reflecting other binding constraints to financial inclusion.

Relaxing the Borrowing Constraint

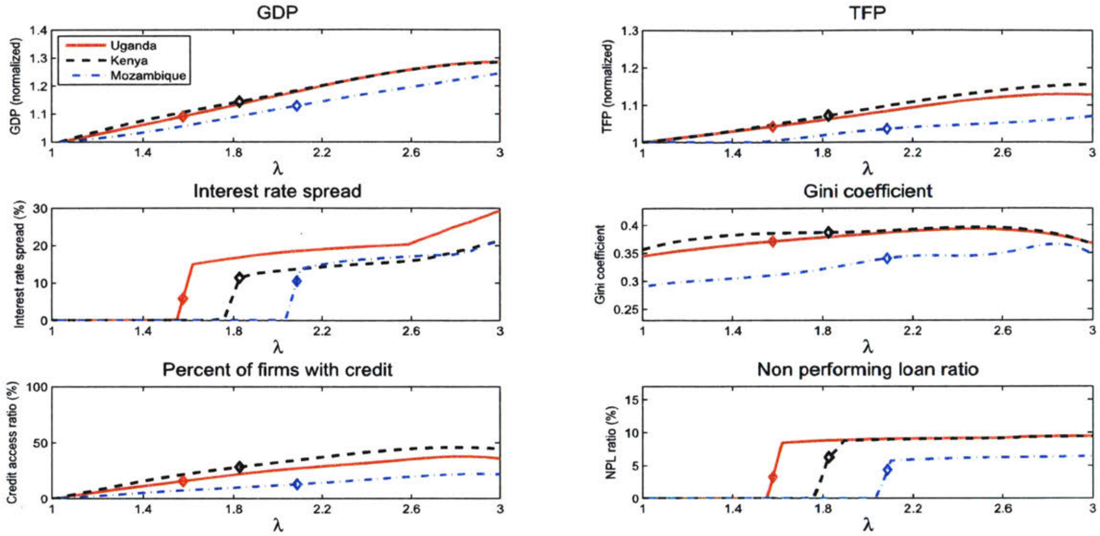
In Figures 3-7 – 3-8, we vary the borrowing constraint parameter λ from 1 to 3. Following the relaxation of the borrowing constraint, GDP increases in all countries. However, the responsiveness of output is highly dependent on the economy's savings rate. In low-income countries, GDP is typically more responsive as agents' production relies heavily on external financing due to small transfers across periods (low savings rates). This suggests that the borrowing constraint is one of the major obstacles to economic development for low-income countries in our sample. In the Philippines, GDP also responds well to the relaxation of the borrowing constraint; however, the reason for this is different from that for low-income countries. Financial access is moderate in the Philippines, but interest spreads are low and savings rates are high. Therefore, the relaxation of the borrowing constraint unlocks financial resources, leading to a significant increase in GDP.

As λ increases, TFP increases, implying a more efficient resource allocation across entrepreneurs. A relaxation of the borrowing constraint benefits talented entrepreneurs disproportionately as they often desire to operate businesses at a larger scale than untalented entrepreneurs (i.e., both $\tilde{k}^l(z)$ and $\tilde{k}^h(z)$ increase in z). Relaxing the borrowing constraint allows all entrepreneurs to borrow more, but on average untalented ones do not borrow as much because their smaller (maximum) business scale may have already been achieved. As a result, more talented entrepreneurs enter the credit regime and expand their scale of operations, driving up TFP.

The interest rate spread increases in this scenario. The spread is zero when λ is low, because entrepreneurs' leverage is low—no default happens even when production fails. As λ increases above a threshold, agents leverage more, the share of NPLs increases, and the interest rate spread starts increasing. Note that, in general, low-income countries have higher interest rate spreads relative to emerging market economies due to higher intermediation costs.

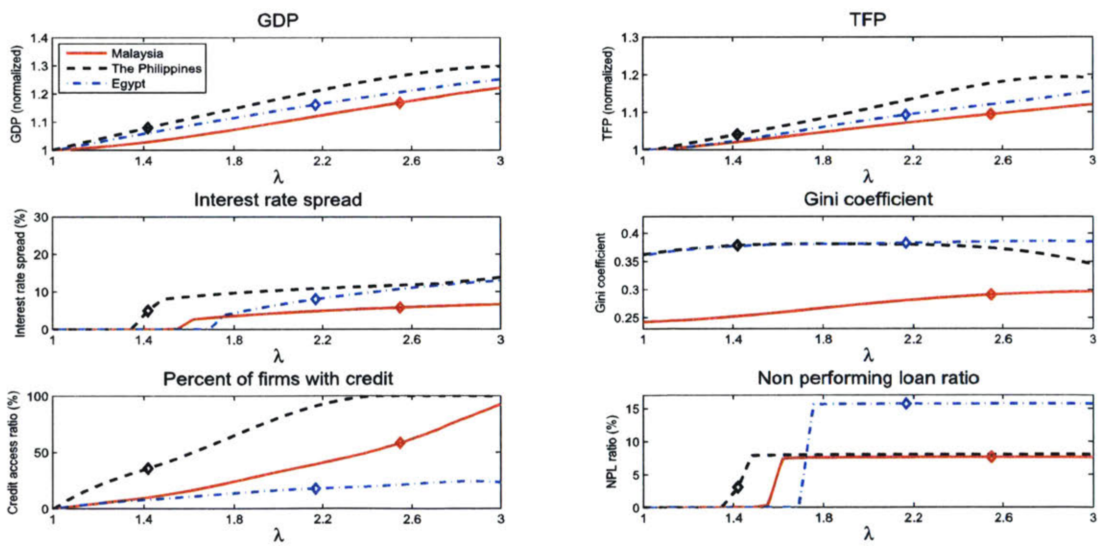
In terms of inequality, the Kuznets pattern is again observed for low-income countries. As λ increases, talented entrepreneurs can leverage more and increase their profits, which drives up the Gini coefficient. In low-income countries, the savings rate is low. As a result, external credit is limited and the interest rate increases by more, the easier the borrowing constraint is. As λ becomes larger, the sharp increase in the interest rate shrinks entrepreneurs' profits, leading to a lower Gini coefficient.

Relaxing the borrowing constraint provides more external credit to entrepreneurs once they pay the participation cost. This induces more entrepreneurs to join the credit regime. However, NPLs also increase. This occurs as a relaxation of the borrowing constraint opens up the doors for small new entrants who tend to be more leveraged.



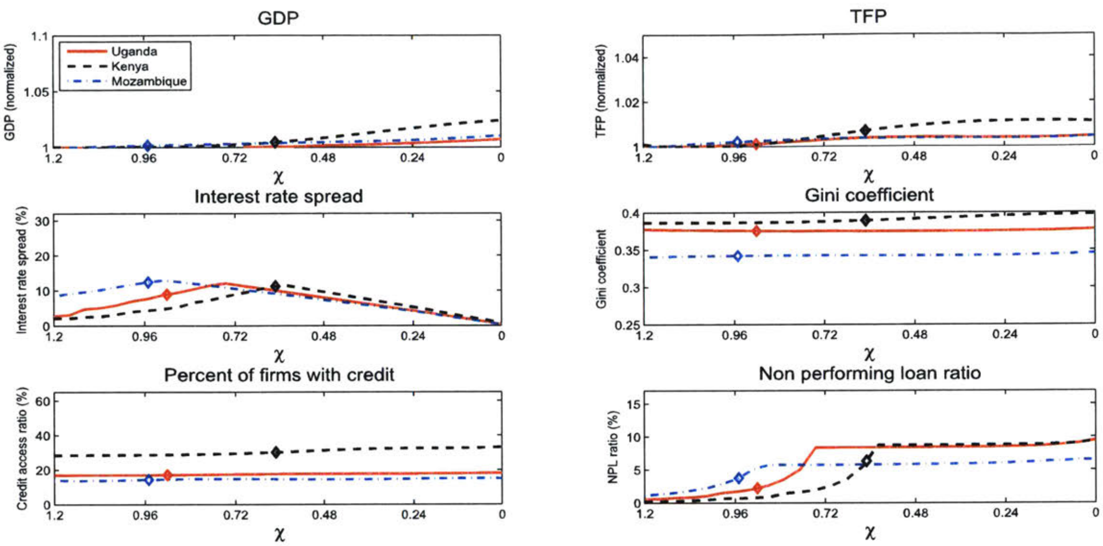
Note: The solid line represents Uganda, the dashed line represents Kenya, and the dash-dotted line represents Mozambique. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 3-7: Comparative statics: Borrowing constraint—low-income countries.



Note: The solid line represents Malaysia, the dashed line represents the Philippines, and the dash-dotted line represents Egypt. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 3-8: Comparative statics: Borrowing constraint—emerging market economies.



Note: The solid line represents Uganda, the dashed line represents Kenya, and the dash-dotted line represents Mozambique. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 3-9: Comparative statics: Intermediation cost—low-income countries.

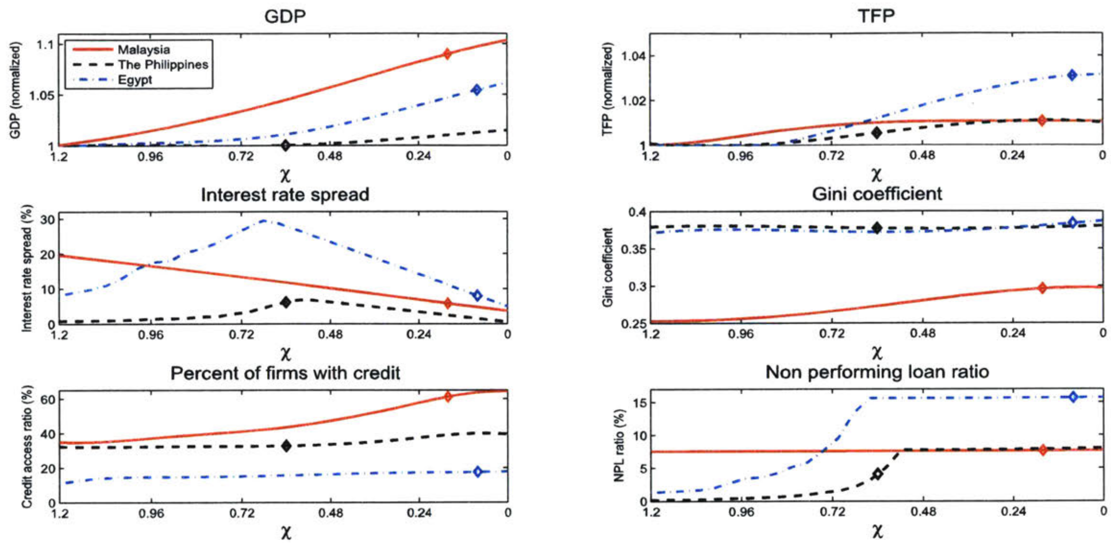
Increasing Intermediation Efficiency

In Figure 3-9 – Figure 3-10, we vary the monitoring cost χ from 1.2 to 0 to reflect financial inclusion from an intermediation efficiency angle. When χ decreases, the response of GDP varies across countries. In some countries (Uganda, Mozambique and the Philippines), GDP is not responsive as a lower intermediation cost only benefits highly-leveraged entrepreneurs, which are few due to the low credit access ratio.

TFP increases (but only slightly) as a lower intermediation cost facilitates the allocation of capital to efficient entrepreneurs. The interest rate spread monotonically declines in Malaysia, but displays an inverted V-shape in other countries. Two opposing forces are in effect here. First, the decline in the net lending rate induces entrepreneurs to increase leverage because it reduces the cost of capital for risky firms, pushing up the share of NPLs. This tends to increase the endogenous interest rate spread. Second, a lower intermediation cost decreases the interest spread through its pass-through effect. Whether the interest rate spread increases or decreases depends on which effect dominates.

The Gini coefficient increases as efficient intermediation disproportionately benefits highly-leveraged entrepreneurs (who already have higher income than others).[26] Moreover, a lower intermediation cost induces more agents to borrow, hence increasing the percent of firms with credit.

[26]There is only a slight increase (almost invisible on the figure) in the Gini coefficient of Uganda, Mozambique, and the Philippines, as the model suggests that the intermediation cost is not a binding constraint in these countries.



Note: The solid line represents Malaysia, the dashed line represents the Philippines, and the dash-dotted line represents Egypt. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 3-10: Comparative statics: Intermediation cost—emerging market economies.

Impact on GDP and Inequality: A Numerical Comparison

Figures 3-5 – 3-10 suggest that the economic implications of financial inclusion policies depend on the source of frictions. In this subsection, we zoom-in on a numerical comparison of the marginal responses of GDP and inequality. The numbers in Table 3.3 are calculated as the difference between the current state of the country (shown with circles in Figures 3-5 – 3-10) and the eventual steady state when the economy's credit to investment ratio is increased by one percentage point.

As before, although financial inclusion brings an increase in GDP and TFP in all cases, its impact on inequality varies. The impact on the Gini coefficient can be positive or negative for a reduction in the credit participation cost, depending on country-specific characteristics.

Moreover, in line with the discussion above, the numbers highlight that the form of financial inclusion and country characteristics matter in how the economies respond. For example, Uganda's GDP responds more if the increase in credit to investment ratio comes from a lower participation cost. However, Egypt's GDP responds more to relaxing the borrowing constraint; while the other countries are more responsive to a lower intermediation cost.[27]

How far are these countries from the world best financial sector technology in terms of these three financial parameters? Which country is most underdeveloped along which dimension? To

[27] Using the credit to investment ratio might bias the results on the effectiveness of different sources of financial inclusion since the credit to investment ratio itself is more responsive to some factors (e.g. λ), and significantly less responsive to other factors (e.g. χ). Therefore, the impact of λ is likely to be underestimated, while the impact of χ is likely to be overestimated.

Table 3.3: The impact of financial inclusion of various forms on GDP per capita, TFP and income inequality.

	Participation cost ψ			Borrowing constraint λ			Intermediation cost χ		
	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini
Uganda	0.40	0.28	-0.0007	0.35	0.20	0.0007	0.03	0.14	-0.0006
Kenya	0.67	0.40	0.0033	0.28	0.22	0.0001	0.07	0.10	0.0004
Mozambique	0.38	0.28	0.0002	0.29	0.13	0.0011	0.38	0.10	-0.0001
Malaysia	0.38	0.37	-0.0005	0.52	0.32	0.0010	1.10	0.00	0.0015
The Philippines	0.28	0.16	0.0006	0.20	0.15	0.0002	0.04	0.08	-0.0003
Egypt	0.26	1.26	-0.0093	0.46	0.35	0.0003	0.69*	0.02*	0.0033*

Note: We consider the change of parameters that increases the credit to investment ratio by 1%. In cases marked with *, we report the change in GDP, TFP, and Gini when parameter χ is reduced to zero. This is because in these cases, even if parameter χ is reduced to zero, the increase in the credit to investment ratio is still less than 1%.

shed light on these questions, we show a numerical comparison for changes in GDP, TFP and the Gini coefficient when the six countries adopt the best-possible intermediation technology. Obviously, the best possible value for the credit participation cost and monitoring cost are zero ($\psi = \chi = 0$). Among the 127 countries in enterprise surveys, we consider countries that require the lowest amount of collateral (Germany, Spain, and Portugal). The average amount of collateral required as a percent of loans in these countries is about 50% ($\lambda = 3$), which is regarded as the best possible borrowing constraint.

Table 3.4 shows the simulation results when one of the financial parameters is equal to the world frontier value. The increase in GDP is largest when the borrowing constraint is relaxed in Uganda, Kenya, the Philippines and Egypt, implying that the financial sector in these countries is facing disproportionately higher collateral requirements. By contrast, Mozambique and Malaysia's GDP are more responsive to a decrease in the credit participation cost, indicating that limited credit availability or low financial access is the major obstacle. Moreover, reducing the credit participation cost leads to a uniform increase in TFP and decrease in the Gini coefficient in all countries for reasons discussed above, while relaxing the borrowing constraint increases TFP, but has an ambiguous impact on income inequality. Not surprisingly, adopting the most efficient intermediation technology ($\chi = 0$) does not boost GDP significantly. However, this does not imply that the intermediation cost is not crucial in terms of financial inclusion. As we show in Theorem (1), (2) and Subsection 3.6.2, there exist rich interactions among these parameters: inefficient intermediation will dampen the responsiveness of GDP to a lower credit participation cost and a relaxed borrowing constraint, or even block these channels.

3.6.2 Interactions among the Three Financial Constraints

In Sections 3.6.1 – 3.6.1, we have shown that policies targeting different financial parameters have differential effects on macroeconomic aggregates. Moreover, the effects vary across countries depending on how country-specific economic characteristics interact with financial sector characteristics. In this subsection, we shed light on how the three financial sector parameters interact with each other and examine its implication on the macroeconomy and financial policies.

Table 3.4: The impact of financial inclusion of various forms on GDP per capita, TFP and income inequality.

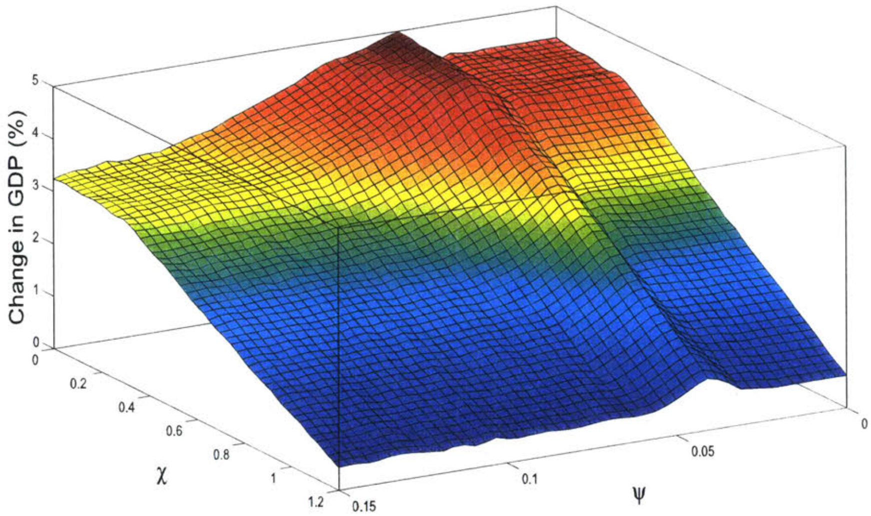
	Participation cost ψ			Borrowing constraint λ			Intermediation cost χ		
	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini
Uganda	5.77	5.67	-0.0210	17.94	10.41	-0.0034	0.74	0.42	0.0018
Kenya	5.16	6.50	-0.0314	12.28	9.30	-0.0203	1.93	0.74	0.0082
Mozambique	12.72	10.16	-0.0267	10.30	4.83	0.0217	0.88	0.32	0.0033
Malaysia	8.44	10.94	-0.0696	4.52	2.85	0.0059	1.26	0.00	0.0013
The Philippines	2.56	3.40	-0.0165	20.21	16.45	-0.0336	1.48	0.58	0.0033
Egypt	7.04	11.31	-0.0590	7.78	6.61	0.0026	0.69	0.02	0.0033

Note: In all cases, we consider financial inclusion that moves the country to world financial sector frontier in one of the three parameters.

We take a specific country—the Philippines—and study the change in GDP per capita following a relaxation of the borrowing constraint (i.e. an increase in parameter λ).[28] In particular, we relax the borrowing constraint by 20%, and compare the increase in GDP relative to the previous state (i.e. before relaxing the borrowing constraint) for different levels of the credit participation cost ψ and the intermediation cost χ. Figure 3-11 shows that the relative change in GDP following an increase in λ depends on the two costs, ψ and χ. When χ increases, the increase in GDP becomes smaller for all ψ. This is because relaxing the borrowing constraint increases GDP by providing more credit to entrepreneurs. However, this channel is partially blocked if the intermediation cost is very high, due to the complementarity on the intensive margin as shown in Theorem 2. A higher intermediation cost restricts entrepreneurs from borrowing more as they want to keep low leverage ratios to avoid being monitored. This dampens the GDP-boosting effect that arises from a relaxation of the borrowing constraint. If the intermediation cost is too high, relaxing the borrowing constraint would be futile as all entrepreneurs prefer to stay with low leverage ratios to avoid paying the monitoring cost.

However, the change in GDP is non-monotonic when ψ increases. The change in GDP stays almost constant for low values of ψ ($\psi < 0.03$); it is increasing in ψ when ψ lies between 0.03 and 0.04, and is decreasing for large values of ψ ($\psi > 0.04$). This non-monotonic pattern results from the two channels through which relaxing the borrowing constraint impacts GDP (as emphasized in Theorem 1 and 2). On the one hand, it enables agents in the credit regime to borrow more (the intensive margin). On the other hand, it induces more agents to join the credit regime, as a lower borrowing constraint increases the benefit of obtaining a credit contract (the extensive margin). Gains on both the intensive and extensive margins depend on the fraction of agents in the credit regime. A decrease in ψ promotes financial inclusion, increasing gains on the intensive margin (i.e. there is complementarity on the intensive margin between ψ and λ as shown in Theorem 2). However, it decreases gains on the extensive margin as relaxing the borrowing constraint has less of an impact on increasing the credit access ratio when this ratio is already high (i.e. there is a substitution effect on the extensive margin between ψ and λ as shown in Theorem 1). Therefore, as ψ decreases, change in GDP first increases and then decreases. The change in GDP, however,

[28]Using other countries' calibrated parameters does not change the qualitative results we emphasize.



Note: The horizontal axes refer to the intermediation cost χ and the credit participation cost ψ; The vertical axis refers to the relative change in GDP when the borrowing constraint λ is relaxed by 20%. The calibrated parameters for the Philippines are used for this study.

Figure 3-11: The increase in relative GDP per capita when the borrowing constraint is relaxed by 20% for different credit participation costs and intermediation costs.

stays almost constant for low values of ψ. This is because the credit access ratio is about 100% when $\psi < 0.03$ (see Figure 3-6), so that further reducing ψ has no impact on gains accruing on both margins.

This exercise suggests that the effectiveness of financial inclusion policies depends crucially on the underlying financial sector characteristics within an economy. Relaxing the borrowing constraint is less effective if the intermediation cost is high, which is partially reflected in a high interest rate spread. The impact of relaxing the borrowing constraint also depends on the credit access ratio, although the relationship is not as clear-cut because of the coexistence of the two margins. The exercise also suggests that financial inclusion policies can be used in a complementary way in order to be more effective. For example, reducing the intermediation cost not only directly boosts GDP, but also amplifies the effect of relaxing the borrowing constraint. However, simultaneously reducing the participation cost and relaxing the borrowing constraint may be partially substitutable, as both policies increase GDP by promoting credit access. The optimal mix of policies thus depends on the underlying financial sector parameters and country-specific characteristics.

3.6.3 Decomposition of GDP and TFP

Still taking the Philippines as an example, we conduct a GDP decomposition using equation (3.4.8) and the endogenous steady-state distributions of wealth and talent. Table 3.5 reveals that relaxing different financial constraints promote GDP through different channels. In partial equilibrium, about two third of the increase in GDP from reducing the credit participation cost arises from gains on the extensive margin. However, most of the increase in GDP due to either relaxing the borrowing constraint or reducing the intermediation cost arises from gains on the intensive margin. This is intuitive since the credit participation cost can be considered as an ex-ante cost for credit while the borrowing constraint and the intermediation cost are both ex-post costs. The increase in GDP in general equilibrium is much smaller than that in partial equilibrium due to higher wages and interest rates. Notably, in general equilibrium, when the credit participation cost is lowered, there are losses on the intensive margin. This is because entrepreneurs who are already in the credit regime do not benefit much but the increase in wages and interest rates depresses their profits substantially. Finally, entrepreneurs in the savings regime also incur losses in general equilibrium due to the spill-over effects of equilibrium factor prices.

Table 3.5: GDP decomposition.

		General Equilibrium		Partial Equilibrium	
		GDP(%)	Contribution(%)	GDP(%)	Contribution(%)
ψ	Extensive margin	8.94	348.96	3.69	68.08
	Intensive margin	-6.32	-246.73	1.73	31.92
	Savings regime	-0.06	-2.23	0	0
	Total	2.56	100	5.42	100
λ	Extensive margin	2.75	13.59	9.68	19.61
	Intensive margin	24.68	122.13	39.7	80.39
	Savings regime	-7.22	-35.72	0	0
	Total	20.21	100	49.38	100
χ	Extensive margin	0.09	5.76	0.09	4.27
	Intensive margin	1.79	120.70	2.04	95.73
	Savings regime	-0.39	-26.46	0	0
	Total	1.48	100	2.13	100

Note: In all cases, we consider financial inclusion that moves the Philippines to the world financial sector frontier in one of the three parameters.

Next we conduct a TFP decomposition for the Philippines using equation (3.4.16). As shown in Table 3.6, most of the increase in TFP arises from the between-regime shifting effect regardless of which financial constraint is relaxed. Moreover, there are large TFP losses in the savings regime as talented entrepreneurs enter the credit regime when financial constraints are relaxed. In partial equilibrium, the qualitative results do not change significantly, but the magnitude of each TFP component is amplified while the increase in the economy's aggregate TFP is smaller. In fact, in general equilibrium, the increase in the equilibrium interest rate and wage drives out less talented entrepreneurs, hence increasing average productivity. Therefore, since prices are fixed in partial equilibrium, the increase in TFP tends to be smaller.[29]

3.6.4 Welfare Analysis

Financial inclusion engenders growth in GDP, however, not all agents are necessarily better off. In this subsection, we investigate the heterogeneous welfare redistribution effects following different

[29]Quantitatively, the difference in TFP between partial equilibrium and general equilibrium is smaller compared with that in Buera, Kaboski and Shin [48], which is due to myopic savings rate assumed here.

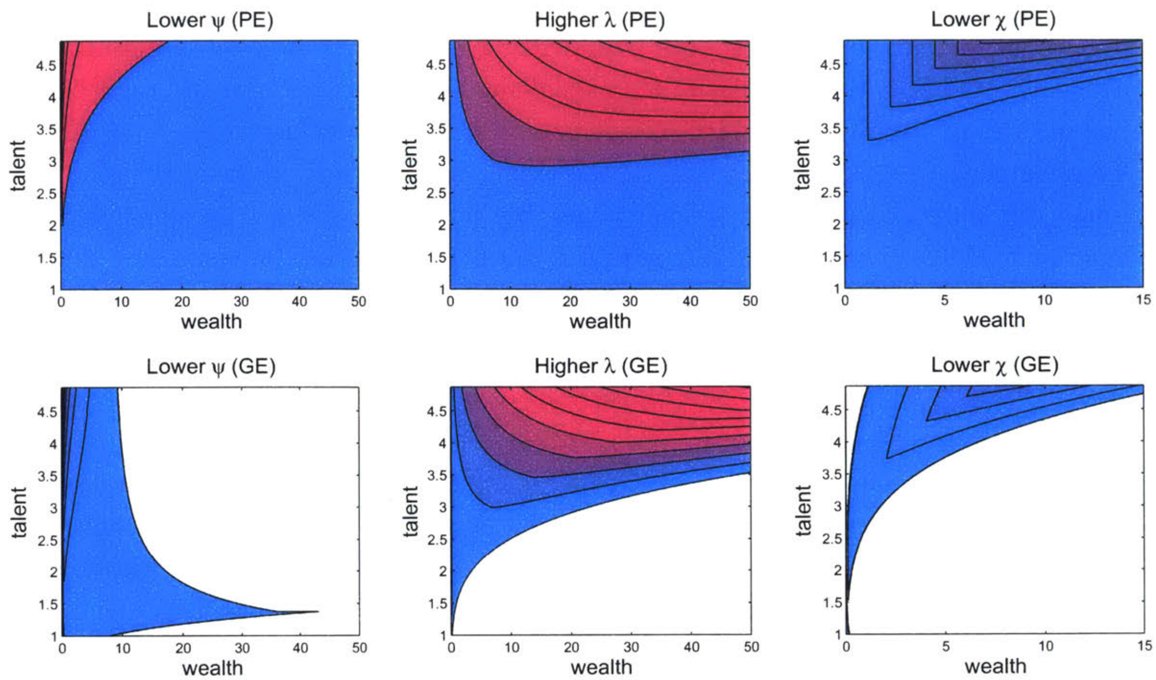
Table 3.6: TFP decomposition.

		General Equilibrium		Partial Equilibrium	
		TFP(%)	Contribution(%)	TFP(%)	Contribution(%)
ψ	Between-regime shifting	7.68	226.02	15.32	471.36
	Credit regime	-0.85	-25.05	-1.69	-51.95
	Savings regime	-3.43	-100.97	-10.38	-319.41
	Total	3.40	100	3.25	100
λ	Between-regime shifting	22.53	136.95	29.61	296.11
	Credit regime	4.16	25.27	-1.04	-104.05
	Savings regime	-10.24	-62.22	-12.56	-125.60
	Total	16.45	100	16.01	100
χ	Between-regime shifting	0.78	133.98	0.81	142.76
	Credit regime	-0.19	-32.36	-0.22	-38.80
	Savings regime	-0.01	-1.62	-0.02	-3.96
	Total	0.58	100	0.57	100

Note: In all cases, we consider financial inclusion that moves the Philippines to the world financial sector frontier in one of the three parameters.

financial inclusion policies. In particular, we quantify the amount of consumption change for different agents (endowed with different wealth and talent) when one of the financial sector parameters (ψ, λ, χ) changes. In Figure 3-12, we present partial equilibrium (top three panels) and general equilibrium (bottom three panels) results separately to highlight their differences. A comparison suggests that the changes in equilibrium interest rates and wages are sources of welfare losses for some agents. That is, if interest rates and wages were fixed, all agents would gain following financial inclusion.

The left-most panels show the change in consumption when the credit participation cost ψ declines. Agents in the white areas experience a reduction in utility after financial inclusion. This is because a reduction in ψ enables more entrepreneurs to borrow, driving up equilibrium wages and interest rates. Wealthier agents lose as they benefit less from a lower credit participation cost and suffer more from the ensuing increase in wages and interest rates. Interestingly, the boundary line is not monotonic. As talent increases, the threshold level of wealth beyond which agents lose first increases then decreases. To understand this pattern we compare agents around the boundary line. We find that the lower part of the boundary line (talent < 1.5) separates agents who use external funds from those who do not. Agents to the right of the boundary line are sufficiently wealthy to self finance production. These agents do not demand external credit, so a reduction in the participation cost does not benefit them. However, because of the increase in wages and interest rates, these agents make lower profits when the participation cost declines. The threshold wealth level is increasing in talent when talent is below 1.5 because talented agents have a higher demand for capital, and therefore need to have higher wealth in order to self finance production. By contrast, when talent is above 1.5 and increases further, the threshold level of wealth decreases. This is because in our model, talented entrepreneurs disproportionately demand more labor than capital (see the optimal labor decision in Proposition 4), therefore they suffer more from the increase in wages. Since labor demand is also increasing in wealth, as talent increases, the marginal gainer should have lower wealth to mitigate the wage effect. Notice that the biggest winner after a reduction in the credit participation cost lies in the upper left corner.



Note: The horizontal and vertical axes refer to wealth and talent, respectively. Consumption gains are reflected by the difference in shades of color—gains are low for light areas (white areas incur losses). Panels in the first row are partial equilibrium results when interest rates and wages are fixed; panels in the second row are general equilibrium results. The left, middle, and right columns represent the change of ψ, λ, and χ, respectively.

Figure 3-12: The impact of financial inclusion of various forms on welfare redistribution.

These agents are poor but very talented. The reduction in the participation cost enables them to have access to external credit, allowing them to expand businesses and increase their profits.

The middle two panels present the consumption change following a relaxation of the borrowing constraint. In this case, untalented entrepreneurs whose demand for credit is low lose. They incur consumption losses because they do not benefit as much from the relaxation of the borrowing constraint due to their low credit demand. Instead, they suffer from the increase in wages and interest rates. The biggest winners are talented and wealthy agents. As credit is proportional to wealth, relaxing the borrowing constraint enables wealthier agents to receive more funds, increasing their profits. Note that if talented and wealthy agents are not financially constrained, these agents will actually lose due to the general equilibrium effect. However, in our calibration, a severe credit constraint is observed for almost all agents due to low savings rates and the finite horizon framework.

The right-most panels show the consumption change following a decrease in the intermediation cost χ. The biggest winners are the most talented agents with moderate amounts of wealth. Intuitively, talented agents employ more capital, hence a reduction in the intermediation cost reduces their costs of production by more. Note that the biggest winners are not the wealthiest agents, because they already have sufficient internal funds, and have low demand for credit. Agents in the two white areas both experience a decrease in their income, but for different reasons. Agents in the upper-left area are talented but poor, and operate their businesses at the maximum leverage ratio. Hence, they benefit from the decrease in the intermediation cost. However, because their demand for capital is low, the benefit from the lower cost of capital is smaller than the increased cost of labor wages.[30] Agents in the lower-right area lose because they operate businesses with low leverage ratios (not being monitored), hence they do not receive benefits from a lower intermediation cost but suffer from the increase in labor costs.

3.7 Conclusion

We develop a tractable micro-founded general equilibrium model with heterogeneous agents to analyze the implications of financial inclusion policies on GDP and inequality in developing countries. In particular, we focus on three specific dimensions of greater financial inclusion: access (as measured by the size of credit participation costs), depth (as measured by the size of borrowing constraints resulting from limited commitment), and intermediation efficiency (as measured by the size of interest rate spreads, reflecting default risk and asymmetric information).

We show theoretically that relaxing each constraint can have a differential quantitative impact on individuals' output and their credit access conditions. More importantly, there exist rich interactions among these financial constraints. In partial equilibrium, we prove that relaxing all constraints simultaneously can have complementary effects of increasing GDP on the intensive margin. However, these constraints are substitutes on the extensive margin. Such interaction effects in general equilibrium are confirmed in our quantitative analysis. Moreover, we find that the sources of GDP gains vary across financial policies. When the participation cost is reduced, GDP gains arise from the extensive margin. However, most of the GDP gains are from the intensive margin when the borrowing constraint and the intermediation cost are reduced. Our TFP decomposition indicates that irrespective of the sources of financial inclusion, most of the TFP gains are captured by a between-regime shifting effect, namely, talented entrepreneurs enter the credit regime and expand their businesses.

Using analytical and numerical methods, we calibrate the model for six countries—Uganda, Kenya, Mozambique, Malaysia, the Philippines, and Egypt. While our simulation results are intended to be illustrative, they indicate that relaxing various financial sector frictions may affect GDP and inequality in different ways. Our findings suggest that country-specific characteristics play a central role in determining the impacts, interactions, and trade-offs among policies. Thus, understanding the specific constraints generating the lack of financial inclusion in an economy is critical for tailoring policy recommendations.

A defining feature of our model is having three different types of financial frictions, limited

[30]According to the labor demand function in Proposition 4, the capital/labor ratio is increasing in wealth. Therefore, poor agents benefit less from a reduction in the cost of intermediation.

participation, limited commitment, and asymmetric information (or costly state verification) all embedded in a unified framework. Most of the recent literature only considers one. We emphasize that the micro-foundations for each of the three financial frictions included in this paper are very different. Moreover, even within a given economy, there exists evidence that individuals face different types of financial frictions depending on location [8, 146, 193]. Thus a quantitative framework that incorporates multiple types of financial frictions is needed to capture the multifaceted nature of financial systems. Even at the aggregate level, alleviating different sources of financial frictions can have differential impacts, either qualitatively or quantitatively, on key macro variables. Incorporating these frictions altogether in a unified framework enables us to identify the most binding constraints that hinder financial inclusion within an economy and uncover their interactions. Moreover, multiple frictions are necessary ingredients to match the data for a wide set of countries. If the only financial friction is limited commitment, relaxing it increases the credit access ratio, interest rate spread, and non-performing loans in tandem. However, in the data, the three moments are not perfectly positively correlated across countries. For example, Uganda has a high interest rate spread but low NPLs, while Egypt has low NPLs and a low interest rate spread. It is not possible to match the two moments in both countries without allowing for both limited commitment and costly state verification. Our bottom-line contribution is to allow and quantify the impact of different types of constraints on GDP, TFP, and inequality in an economy where all these frictions are potentially present.

Appendix A

Appendix for Job Search under Debt: Aggregate Implications of Student Loans

A.1 Data

My empirical analysis uses panel data from the National Longitudinal Survey of Youth 1997 (NLSY97). This is a nationally representative survey conducted by the Bureau of Labor Statistics. In round 1, 8,984 youths were initially interviewed in 1997. Follow-up surveys were conducted annually. Almost 83% (7,423) of the round 1 sample were interviewed in round 15 (2011-2012). Youths were born between 1980 and 1984. Their ages ranged from 12 to 18 in round 1 and were 26 to 32 in round 15.

My analysis focuses on high school and college graduates. I do not include college dropouts because it is not clear when they enter the labor market. I drop youths who have ever served in the military or attended graduate schools because they are not in the same position as the other youths in my sample when it comes to making labor market decisions. I also drop youths who received the bachelor's degree before 1997 due to the lack of labor market information upon college graduation. This leaves me with a sample of 1,721 high school graduates and 1261 college graduates.

All dollar-denominated variables are converted to 2009 dollars using GDP deflator. The details of the data construction follow.

A.1.1 Variables Used for Sample Selection

Highest degree In each year, NLSY97 collects the highest degree received to the start of the interview year. The cumulative variable CVC_HIGHEST_DEGREE_EVER documents the highest degree received ever according to the most recent survey. I only keep the youths with a bachelor's degree (CVC_HIGHEST_DEGREE_EVER=4) or a high school degree (CVC_HIGHEST_DEGREE_EVER=2).

Military service I check two variables for military services. The variable YCPS_2400, available in years 1997, 2000, 2006, documents whether the youth is now in the active Armed Forces. I drop the the youths who answered yes in any of these surveys. The variable YEMP_59000, available in years 1998-2012, documents whether the youth is/was in the regular, the Reserves, or the National Guard. I drop youths who ever had these statuses.

Enrollment in grad schools Some youths choose to continue a graduate program after college graduation. I drop these students because their labor market experience is likely to be different. The variable CV_ENROLLSTAT, available in each year since 1997, documents the enrollment status as of the survey year. I drop youths who ever enrolled in a graduate program (CV_ENROLLSTAT=11).

Degree receiving date The variable CVC_BA_DEGREE documents the date on which the youth received a bachelor' s degree in a continuous month scheme. I drop youths who received the bachelor's degree before 1997 due to the lack of labor market information upon college graduation.

A.1.2 Variables Used for Model Estimation and Regression Analyses

Variables Used for Both Model Estimation and Regression Analyses

Student loan debt I construct the student loan debt variable following Addo [6]. The variable YSCH_25600 documents the amount of loans borrowed in government-subsidized loans or other types of loans while the youth attended schools/institutions in each term and each college. Together with the records on enrollment information, I construct the amount of student loans taken out in each year and the total amount of student loans borrowed before college graduation. Unfortunately, there is no information on repayment in the data. Because students rarely repay student loan debt during college, I consider the total amount of student loans borrowed as the amount of outstanding student loan debt upon college graduation. To prevent the skewness of the debt distribution having a large effect on the estimated means, the total amount of student loan debt is top coded at 99 percentile ($49,280).

Last date enrolled I construct a "last-enrolled" variable to record the last date on which the youth is in school. I consider the youth as in the labor market after this date is passed. For college graduates, the variable SCH_COLLEGE_STATUS documents the youth's college enrollment status in each month since 1997. Based on this information, I set the value of "last-enrolled" to be the latest month that the youth was enrolled in college (SCH_COLLEGE_STATUS=3). Then, I check whether the value of "last-enrolled" variable is consistent with the date that the youth receives her bachelor's degree, documented by the variable CVC_BA_DEGREE. Among the 1261 college graduates in my sample, 83 youths have the last date enrolled being inconsistent with the degree date for at least 1 year. These youths are not considered when constructing the labor market moments below. For high school graduates, I use the high school degree receiving date as the last date in school.

Duration of unemployment spells I construct the duration of unemployment spells by tracking the period until an unemployed (or out of the labor force) youth finds a job. In my sample, there are 7,969 unemployment spells with an average duration of 27.2 weeks and a standard deviation of 59.4 weeks.

Wage income The variable YINC_1700 documents income that the youth received from wages, salary, commissions, or tips from all jobs in past year, before deductions for taxes or anything else. This is the variable I use to construct annual wage income. An alternative method to construct annual wage income is to use the information on hours and hourly wage rate. The two methods usually provide different numbers due to measurement errors. I prefer to use the variable YINC_1700 to construct annual wage income because the value of this variable is directly obtained from the questionnaire but the second method uses data constructed by BLS staff based on several discretionary assumptions. To be consistent, I construct an average hourly wage rate by dividing deflated values of YINC_1700 by the total number of hours worked in that year. When constructing annual wage income for each youth, I follow Rubinstein and Weiss [207] by excluding the youths whose hourly wage rates are below $4 or higher than $2,000 and who worked less than 35 weeks or less than 1,000 annual hours.

Hours The variable EMP_HOURS documents the total number of hours worked by a youth at any job in each week. Hours per week worked at each job are assumed constant except during a reported gap, when the hours for that job are assumed to be zero. Weekly hours are top coded at 140 hours.

Variables Used Only for Model Estimation

Net liquid wealth I construct the net liquid wealth variable using financial assets. Loans received from family members and friends to help pay for college are not subtracted in the measure of net liquid wealth. This is because, as argued by Johnson [141], it is not clear whether or when these youths would need to repay the loans from family members and friends for educational purposes. I do not include non-financial assets, e.g., housing and property values, farm operation, etc., because these assets are not as liquid, and accounting for their values downplays the marginal propensity to consume. As I show in section 1.3, the repayment of student loans affects job search strategy through the liquidity channel, which depends on the marginal propensity to consume.

The variable CVC_ASSETS_FINANCIAL documents the value of financial assets when the youth reaches ages 18, 20, and 25. The financial assets include savings and checking accounts, money market funds, retirement accounts, stocks, bonds, and life insurance, etc. I use the value of financial assets at age 18 to proxy the net liquid wealth right before making the college entry decision. To prevent the skewness of the asset distribution having a large effect on the estimated means, the net liquid wealth values are top coded at 99 percentile ($69,695).

One concern is that money in retirement accounts is not as liquid. The adjustment is made using the variable YAST_4292, which documents the amount of savings in pension/retirement plans. Making this adjustment has almost no effect on the distribution of liquid wealth because only 50 youths reported to have positive balance in these plans with an average amount of $39.7.

Work status I construct the youth's work status using the variable EMP_STATUS, which documents the youth's weekly employment status since 1997. This variable documents whether the youth is employed, unemployed, or out of the labor force. Because my model does not distinguish between unemployed and out of the labor force, I consider the youths who are out of the labor force as unemployed. For employed youths, the associated employer number is also documented.

Duration of employment spells For each youth, I construct the duration of her employment spells by tracking the period between the date of moving from unemployment status to employment status and the date of moving from employment status to unemployment status. I drop employment spells whose duration is less than five weeks, because these are likely to be temporary or insecure jobs. In my sample, there are 8,130 employment spells with an average duration of 113.2 weeks and a standard deviation of 136.2 weeks.

Job tenure For each youth, I construct her tenure at each job (employer) by tracking the period between the date of moving to the job and the date of leaving the job. In my sample, there are 12,086 job spells with an average duration of 76.3 weeks and a standard deviation of 106.9 weeks.

Hourly wage rate The variable CV_HRLY_PAY documents the hourly rate of pay as of either the job's stop date or the interview date for on-going jobs. This variable is used to construct the wage increase upon job-to-job transitions (not wage income; see above).

Wage increase upon job-to-job transitions I construct the log wage increase upon job-to-job transitions by calculating the change in log hourly wage rate between consecutive job spells.

Government benefits The monthly take-up status and benefit amount of AFDC, food stamps, and WIC between 1997-2009 are documented in variables, AFDC_AMT, AFDC_STATUS, FDSTMPS_AMT, FDSTMPS_STATUS, WIC_AMT, WIC_STATUS.

Others The remaining moments are constructed using other data sources. The vacancy to unemployment ratio is constructed using job openings information since December 2000 from

JOLTS. The life-cycle earnings profile between ages 23-60 is constructed using March CPS 1997-2008 from Acemoglu and Autor [3] (available on David Autor's website).

Variables Used Only for Regression Analyses

Parental wealth and education The variable CV_HH_NET_WORTH_P documents household net worth from parent interview in 1997. I use this variable to proxy parental wealth. The variable, CV_HGC_BIO_DAD and CV_HGC_BIO_MOM, document the highest grade completed by each youth's biological father and mother. I use the mean of the two variables to proxy parental education.

Gender, race, age, and AFQT score can be found from variables, KEY!SEX, KEY!RACE_ETHNICITY, KEY!BDATE, ASVAB_MATH_VERBAL_SCORE_PCT.

County of residence is available from NLSY restricted geocode CD. The variable GEO01 documents the youth's residence in each survey year.

Job industry The variable YEMP_INDCODE_2002 documents the 4-digit business or industry code based 2002 Census Industry Codes for each youth between 1997-2013. Industry codes between 6870-6990 are classified as finance and banking jobs and those between 7270-7460 are classified as consulting jobs.

Length of college study The length of college study is constructed by taking the difference between the first date enrolled in college, available from variable SCH_COLLEGE_STATUS, and the BA degree receiving date, documented by variable CVC_BA_DEGREE.

Sector The variable YEMP_58500 documents whether the worker is employed by government, a private company, a nonprofit organization , or is working without pay in a family business or farm since 1997. I consider the respondent as working in the public sector if she is employed by government or by a nonprofit organization. There is only one respondent working without pay in a family business or farm. This data point is dropped when running regressions.

College major Respondents in rounds 1-13 (1997-2009) indicated their college majors from a pick list. The variable YSCH_21300 documents the youth's major field in each college each term since the date of last interview. Beginning in round 14 (2014), respondents' majors were collected in a verbatim format and then coded using the CIP (Classification of Instructional Programs) 2010 codes under the variable YSCH_21300_COD. In my sample, only 7 youths received the BA degree after 2010 (the most recent graduate received his degree in September 2011). For these youths, I use the majors recorded before round 14 to be consistent with the old coding system. Among the rest 1254 youths, 1234 youths' majors are documented in at least one of the survey between 1997-2009. For the 104 youths who changed majors during college study, I use the most recently reported major before the degree receiving date to represent the major associated with the BA degree. The old coding system has a very fine category with 45 different majors, which generates a collinearity problem (with the county fixed effect) in my wage regressions because of the small sample size. Therefore, I reclassify the recorded majors into four broader category, including physical science, social science, engineering, and others.

A.1.3 Adjusting the Higher-Order Moments for Unmodeled Variation

In the model, the exogenous sources of variation among agents come from differences in initial wealth, talent, student loan debt, and histories of shocks to job offers. By contrast, the data contain unmodeled variation due to heterogeneity in personal characteristics, family background, occupation, and industry fixed effects. Ignoring these sources of variation would not be problematic if the moments used in identification only include sample averages. However, because the talent and vacancies' productivity distribution are identified using the second (variation) and third (skewness) moments of the cross-sectional log wage income distribution and the variance of log wage increase upon job-to-job transitions, ignoring these sources of variation would bias the estimation result. Intuitively, failure to account for the unmodeled variation in the data would result in a more dispersed estimated productivity distribution, which will in turn exaggerate the option value of staying unemployment and overestimate the effect of the debt burden on job search decisions.

I adjust the data by purging the unmodeled sources of variation from the data following the approach of Gourinchas and Parker [108] and Kaboski and Townsend [142]. In particular, I run linear regressions of log wage income. The estimated equation is:

$$\log Wage_{i,t} = \beta_w X_{i,t} + \epsilon_{w,i,t}, \tag{A.1.1}$$

where $X_{i,t}$ is a vector of controls including race, gender, parental net worth and education, occupation, and year fixed effects. I then construct the adjusted data for individuals with mean values of the explanatory variables $(\overline{X})$ using the estimated coefficients and residuals:

$$\widetilde{\log Wage}_{i,t} = \hat{\beta}_w \overline{X} + \hat{\epsilon}_{w,i,t}.$$

Finally, I construct the variance and skewness moments of the cross-sectional log wage income distribution using the adjusted log wage income $\widetilde{\log Wage}_{i,t}$.

A.1.4 Suggestive Evidence

In this subsection, I present the full regression table for the validation test conducted in section 1.5.3.

Table A.1: The duration of the first unemployment spell after college graduation.

	Duration of the first unemployment spell		
	(1)	(2)	(3)
Loan amount	-1.54**	-2.08***	-1.92***
(in $10,000)	(0.66)	(0.68)	(0.63)
Parental wealth	-0.02	-0.00	0.03
(in $10,000)	(0.06)	(0.07)	(0.08)
Parental education	0.36	0.68	0.57
	(0.41)	(0.53)	(0.53)
Female		3.37	1.91
		(2.23)	(2.27)
AFQT		-0.01	-0.03
		(0.06)	(0.06)
Race: Black		-0.23	-2.10
		(5.24)	(4.09)
Hispanic		2.62	2.92
		(9.49)	(9.18)
Mixed Race		1.56	3.51
		(4.00)	(3.60)
Married		1.00	-0.81
		(3.41)	(3.29)
age		-28	-148
		(271)	(227)
age^2		1.38	6.17
		(10.91)	(9.04)
age^3		-0.02	-0.08
		(0.15)	(0.12)
Major: Physical Science			6.55
			(4.31)
Social Science			4.35
			(2.70)
Others			5.71*
			(3.28)
Industry: finance, banking,			-6.78***
and consulting			(2.02)
Length of college study			0.42
			(0.58)
Observations	884	771	728
County fixed effect	√	√	√
R^2	0.0057	0.0183	0.0291

Note: This table examines the impact of student loan debt on the duration of the first unemployment spell after college graduation. A $10,000 increase in the amount of student loans reduces the duration of the first unemployment spell by about 2 weeks. Each observation is at the individual level. The dependent variable is the number of weeks elapsed from the college graduation date to the date of starting the first full-time job (i.e., work more than 35 hours per week for at least two consecutive weeks). The dependent variable is regressed on the total amount of student loan debt borrowed during college study, recorded in units of $10,000. All regressions control for parental wealth, parental education, and the county of residence in the graduation year. Column (2) adds additional controls for gender, race, AFQT score, marital status, and the cubic age polynomials. Column (3) adds additional controls for college major, job industry, and the length of college study. Standard errors are clustered at the county level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Table A.2: The impact of student loan debt on post-graduation wage income.

	First year			Second year			Third year		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Loan amount	-1,830**	-2,067**	-2274**	-1,812**	-2,152**	-2,232**	-2,009*	-2,619**	-2,821**
(in $10,000)	(770)	(890)	(920)	(789)	(865)	(882)	(1,117)	(1,309)	(1,372)
Parental wealth	100*	94*	77	91	106	95	53	33	56
(in $10,000)	(56)	(55)	(56)	(70)	(84)	(69)	(85)	(83)	(90)
Parental education	19	-376	-146	290	-364	-130	611	-29	320
	(305)	(380)	(405)	(389)	(523)	(516)	(538)	(623)	(565)
Female		-6,140***	-3,585*		-6,347***	-3,135		-8,154***	-4,738*
		(1,969)	(1,864)		(2,142)	(2,155)		(2,765)	(2,513)
AFQT		80.7	55.4		112.0	94		117	108
		52.6	(51.8)		(69.5)	(68)		(78)	(74)
Race: Black		1,491	52		-835	-142		992	1,613
		(3,679)	(3,741)		(4,986)	(4,825)		(5,340)	(5,931)
Hispanic		-730	-696		-8,496	-5,825		-12,583	-6,366
		(8,473)	(8,049)		(8,113)	(8,049)		(13,008)	(11,574)
Mixed Race		2,051	513		-1,323	-2,841		1,326	-446
		(2,850)	(2,820)		(3,515)	(3,335)		(4,102)	(4,129)
Married		-1,153	-2,415		-2,337	-2,081		-4,563	-4,860
		(2,457)	(2,469)		(3,349)	(3,166)		(3,616)	(3,871)
age		-9.3e4	1.5e4		2.4e5	-2.3e5		9.9e4	5.9e5
		(3.0e5)	(3.0e5)		(4.2e5)	(4.8e5)		1.1e6	1.6e6
age^2		3.4e3	-1.0e3		1.0e4	9.8e3		-3.3e3	-2.4e4
		(1.2e4)	(1.2e4)		(1.8e4)	(2.0e4)		(4.7e4)	6.6e4
age^3		-42	18		-145	-138		33	323
		(163)	(163)		(244)	(276)		(662)	929
Major: Physical Science			-20,189***			-19,244***			-20,969***
			(4,988)			(4,631)			(6,697)
Social Science			-20,370***			-21,147***			-23,233***
			(4,627)			(4,512)			(6,453)
Others			-24,729***			-26,608***			-28,201***
			(4,532)			(5,184)			(6,708)
Industry: finance, banking,			5,632***			5,498**			4,358
and consulting			(2,158)			(2,615)			(3,088)
Length of college study			495			-536			-164
			(563)			(647)			(863)
Observations	671	596	582	588	518	507	483	427	415
County fixed effect	√	√	√	√	√	√	√	√	√
R^2	0.0175	0.0651	0.1455	0.0221	0.0733	0.1361	0.0185	0.0713	0.1311

Note: This table examines the impact of student loan debt on wage income in the first three years after college graduation. A $10,000 increase in the amount of student loans reduces the annual wage income by about $2,000. The dependent variable is wage income in the t-th year ($t = 1, 2, 3$) after college graduation. The dependent variable is regressed on the total amount of student loan debt borrowed during college study, recorded in units of $10,000. All regressions control for parental wealth, parental education, and the county of residence in the graduation year. Column (2) adds additional controls for gender, race, AFQT score, marital status, and the cubic age polynomials. Column (3) adds additional controls for college major, job industry, and the length of college study. Standard errors are clustered at the county level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Table A.3: The impact of student loan debt on first jobs' industry, sector, and labor supply.

	High-paid industry			Private sector			Labor supply		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Loan amount	-0.005	-0.032	-0.033	0.026	0.037	0.031	13.0	24.9	20.2
(in \$10,000)	(0.047	(0.051)	(0.052)	(0.061)	(0.069)	(0.070)	(23.8)	(29.4)	(29.0)
Parental wealth	0.005*	0.005*	0.005	-0.002	-0.002	-0.002	1.55	1.00	1.25
(in \$10,000)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(1.69)	(1.76)	(1.79)
Parental education	-0.008	-0.041*	-0.042*	0.067**	0.041	0.041	-9.4	-30.1**	-29.5**
	(0.019)	(0.022)	(0.022)	(0.029)	(0.034)	(0.034)	(12.5)	(13.2)	(13.7)
Female		-0.28***	-0.24**		-0.38**	-0.35**		-235***	-219***
		(0.11)	(0.11)		(0.16)	(0.17)		(60)	(60)
AFQT		0.006**	0.006**		-0.001	-0.001		1.68	1.40
		(0.003)	(0.003)		(0.003)	(0.003)		(1.30)	(1.32)
Race: Black		0.015	-0.005		0.103	0.107		3.8	-16.0
		(0.231)	(0.231)		(0.319)	(0.322)		(159.2)	(154.5)
Hispanic		-0.270	-0.170		0.277	0.419		12.3	-3.8
		(0.579)	(0.578)		(0.711)	(0.702)		(184.0)	(193.2)
Mixed Race		0.011	-0.004		0.288	0.298		59.1	42.2
		(0.193)	(0.195)		(0.244)	(0.245)		(99.4)	(101.4)
Married		0.118	0.132		-0.372*	-0.374*		-130.5	-158.6*
		(0.152)	(0.153)		(0.224)	(0.227)		(91.4)	(82.4)
age		-32.9*	-33.9*		-44.0***	-44.0**		748	3,825
		(18.6)	(18.7)		(17.0)	(17.2)		(7,066)	(7,085)
age^2		1.37*	1.42*		1.76***	1.77**		-35.8	-162.1
		(0.76)	(0.77)		(0.68)	(0.69)		(286.3)	(287.3)
age^3		-0.019*	-0.020*		-0.023***	-0.024**		0.545	2.242
		(0.010)	(0.011)		(0.009)	(0.009)		(3.851)	(3.863)
Major: Physical Science			0.148			0.030			-222.3
			(0.246)			(0.415)			(147.1)
Social Science			-0.040			-0.019			-242.9*
			(0.209)			(0.354)			(131.7)
Others			-0.306			-0.278			-167.1
			(0.232)			(0.374)			(135.8)
Length of college study			-0.027			0.021			42.3
			(0.021)			(0.035)			(33.5)
Observations	884	775	773	365	319	317	812	705	705
County fixed effect							√	√	√
R^2	0.0037	0.0417	0.0506	0.0142	0.0638	0.0694	0.0029	0.0383	0.0521

Note: This table examines the impact of student loan debt on the industry and sector of first jobs and the number of working hours in the first year after college graduation. There is no significant finding on these margins. The first three columns estimate a Probit model using whether the respondent's first job is in finance, banking, and consulting industry as the dependent variable. The next three columns estimate a Probit model using whether the respondent's first job is in private sector as the dependent variable. The last three columns estimate an OLS regression using the number of workings hours in the first year after college graduation as the dependent variable. The treatment variable is the total amount of student loan debt borrowed during college study, recorded in units of \$10,000. All regressions control for parental wealth, parental education, and the county of residence in the graduation year. Column (2) adds additional controls for gender, race, AFQT score, marital status, and the cubic age polynomials. Column (3) adds additional controls for college major, job industry, and the length of college study. Standard errors in the last three columns are clustered at the county level. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

A.2 Proofs

A.2.1 Proof of Proposition 1

Proof. Rearranging equation (1.3.3), the reservation wage is implicitly determined by

$$1 = \frac{\beta}{1-\beta} \int_{w^*_{FIX}}^{\overline{w}} \frac{u(w-s) - u(w^*_{FIX} - s)}{u(w^*_{FIX} - s) - u(\theta - s)} dF(w). \tag{A.2.1}$$

Consider increasing debt by Δs, and denote the reservation wage corresponding to $s + \Delta s$ as $\hat{w}^*_{FIX}$, thus according to (A.2.1),

$$1 = \frac{\beta}{1-\beta} \int_{\hat{w}^*_{FIX}}^{\overline{w}} \frac{u(w-s-\Delta s) - u(\hat{w}^*_{FIX} - s - \Delta s)}{u(\hat{w}^*_{FIX} - s - \Delta s) - u(\theta - s - \Delta s)} dF(w). \tag{A.2.2}$$

Define $u_2(x) = u(x - \Delta s)$, we can rewrite (A.2.2) as

$$1 = \frac{\beta}{1-\beta} \int_{\hat{w}^*_{FIX}}^{\overline{w}} \frac{u_2(w-s) - u_2(\hat{w}^*_{FIX} - s)}{u_2(\hat{w}^*_{FIX} - s) - u_2(\theta - s)} dF(w). \tag{A.2.3}$$

Let $r(x)$ and $r_2(x)$ be the local absolute risk aversion for $u(x)$ and $u_2(x)$. Thus

$$\begin{aligned} r(x) &> r_2(x) \quad \text{If } u(\cdot) \text{ has IARA;} \\ r(x) &= r_2(x) \quad \text{If } u(\cdot) \text{ has CARA;} \\ r(x) &< r_2(x) \quad \text{If } u(\cdot) \text{ has DARA.} \end{aligned} \tag{A.2.4}$$

Taking DARA as an example, note that $\theta - s < w^*_{FIX} - s < w - s$ for all $w \in (w^*_{FIX}, \overline{w}]$, thus according to Pratt [199, Theorem 1],

$$\begin{aligned} 1 &= \frac{\beta}{1-\beta} \int_{w^*_{FIX}}^{\overline{w}} \frac{u(w-s) - u(w^*_{FIX} - s)}{u(w^*_{FIX} - s) - u(\theta - s)} dF(w) \\ &> \frac{\beta}{1-\beta} \int_{w^*_{FIX}}^{\overline{w}} \frac{u_2(w-s) - u_2(w^*_{FIX} - s)}{u_2(w^*_{FIX} - s) - u_2(\theta - s)} dF(w). \end{aligned} \tag{A.2.5}$$

Then (A.2.3) and (A.2.5) imply

$$\int_{\hat{w}^*_{FIX}}^{\overline{w}} \frac{u_2(w-s) - u_2(\hat{w}^*_{FIX} - s)}{u_2(\hat{w}^*_{FIX} - s) - u_2(\theta - s)} dF(w) > \int_{w^*_{FIX}}^{\overline{w}} \frac{u_2(w-s) - u_2(w^*_{FIX} - s)}{u_2(w^*_{FIX} - s) - u_2(\theta - s)} dF(w). \tag{A.2.6}$$

Because $\int_{w^*_{FIX}}^{\overline{w}} \frac{u_2(w-s) - u_2(w^*_{FIX} - s)}{u_2(w^*_{FIX} - s) - u_2(\theta - s)} dF(w)$ is decreasing in w^*_{FIX}, this implies $\hat{w}^*_{FIX} < w^*_{FIX}$.

Note that Danforth [68] extends the result of Pratt [199] to multi-dimensional lotteries. By applying Danforth [68, Theorem 2], we can obtain a more general result, which indicates that higher debt reduces the agent's reservation wage even in a perfect credit market.

As an extension, if we assume that borrowers are protected from limited liability, i.e., they do not need to make repayment during unemployment, then equation (A.2.1) can be written as

$$
\begin{aligned}
1 &= \frac{\beta}{1-\beta}\int_{w^*}^{\overline{w}} \frac{u(w-s)-u(w^*-s)}{u(w^*-s)-u(\theta)}dF(w) \\
&= \frac{\beta}{1-\beta}\int_{w^*}^{\overline{w}} \left[\frac{u(w-s)-u(\theta)}{u(w^*-s)-u(\theta)}-1\right]dF(w).
\end{aligned} \tag{A.2.7}
$$

Equation (A.2.7) implies that an increase in s increases the reservation wage w^*. This is the risk-shifting effect of debt proposed by Donaldson, Piacentino and Thakor [79] (a related discussion is in footnote 7).

□

A.2.2 Proof of Proposition 2

Proof. The mileage that CRRA utility buys me is that it is a homogeneous utility function with multiplicative scaling behavior. With CRRA utility, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, equation (1.3.5) becomes

$$
(w^*_{IBR})^{1-\gamma} = \theta^{1-\gamma} + \frac{\beta}{1-\beta}\int_{w^*_{IBR}}^{\overline{w}} [w^{1-\gamma} - (w^*_{IBR})^{1-\gamma}]dF(w). \tag{A.2.8}
$$

Clearly, w^*_{IBR} does not depend on α. Therefore, under the IBR, when the utility has CRRA, the agent's reservation wage is equal to the reservation wage of the agent who has no debt. This suggests that

$$
w^*_{IBR} = w^*|_{s=0} > w^*_{FIX}, \tag{A.2.9}
$$

where the last inequality is from Proposition 1 because CRRA utility has decreasing absolute risk aversion. Note that another way to see that the reservation wage does not depend on α when utility has CRRA is to calculate the absolute risk aversion for utility $u((1-\alpha)x)$, which is γ/x, not a function of α. Then, according to the proof of Proposition 1, the reservation wage stays the same because the local absolute risk aversion does not change for any x when α changes.

In fact, we can further show that the disposable reservation wage also satisfies $\overline{w}^*_{IBR} > \overline{w}^*_{FIX}$. This indicates that the liquidity channel plus the risk channel strictly dominates the optionality channel according to Proposition 3. This result is obtained by applying the following lemma to equation (1.3.7).

Lemma 5. *The reservation wage under IBR satisfies:*

$$
w^*_{IBR} < \frac{s}{\alpha}, \tag{A.2.10}
$$

where α solves equation (1.3.4).

Proof. According to equation (1.3.5), w^*_{IBR} is determined by

$$u((1-\alpha)w^*_{IBR}) = \frac{(1-\beta)u((1-\alpha)\theta) + \beta\int_{w^*_{IBR}}^{\overline{w}} u((1-\alpha)w)dF(w)}{1-\beta F(w^*_{IBR})}. \tag{A.2.11}$$

Thus

$$(1-\alpha)w^*_{IBR} = u^{-1}\left[\frac{(1-\beta)u((1-\alpha)\theta) + \beta\int_{w^*_{IBR}}^{\overline{w}} u((1-\alpha)w)dF(w)}{1-\beta F(w^*_{IBR})}\right]. \tag{A.2.12}$$

Notice that

$$\frac{(1-\beta) + \beta\int_{w^*_{IBR}}^{\overline{w}} dF(w)}{1-\beta F(w^*_{IBR})} = 1, \tag{A.2.13}$$

thus, we can think of the LHS of equation (A.2.13) as probability weights, which are imposed on $u((1-\alpha)x)$ to generate the RHS of equation (A.2.12).

By Jensen's inequality, equation (A.2.12) can be written as

$$w^*_{IBR} < \frac{(1-\beta)\theta + \beta\int_{w^*_{IBR}}^{\overline{w}} wdF(w)}{1-\beta F(w^*_{IBR})}. \tag{A.2.14}$$

According to equation (A.4.17) and (1.3.4), α is determined by

$$\frac{s}{\alpha} = \frac{(1-\beta)\theta F(w^*_{IBR}) + \int_{w^*_{IBR}}^{\overline{w}} wdF(w)}{1-\beta F(w^*_{IBR})}. \tag{A.2.15}$$

Therefore,

$$\begin{aligned} \frac{s}{\alpha} - w^*_{IBR} &> \frac{(1-\beta)\int_{w^*_{IBR}}^{\overline{w}} wdF(w) - (1-\beta)\theta[1-F(w^*_{IBR})]}{1-\beta F(w^*_{IBR})} \\ &> \frac{(1-\beta)[1-F(w^*_{IBR})](w^*_{IBR}-\theta)}{1-\beta F(w^*_{IBR})} \\ &> 0. \end{aligned} \tag{A.2.16}$$

□

Using Lemma 5, the disposable reservation wage satisfies

$$\begin{aligned}
\overline{w}^*_{IBR} - \overline{w}^*_{FIX} &= (1-\alpha)w^*_{IBR} - w^*_{FIX} + s \\
&> (1-\alpha)w^*_{IBR} - w^*_{IBR} + s \\
&= \alpha\left(\frac{s}{\alpha} - w^*_{IBR}\right) \\
&> 0. \qquad \text{(A.2.17)}
\end{aligned}$$

□

A.2.3 Proof of Lemma 1

Proof. Proposition 9 indicates that $I(x)$ is increasing in x when $x < \hat{w}$ (note: $\hat{w}$ is the reservation wage chosen by a risk-neutral agent). Therefore, equation (1.3.4) implies

$$\alpha = \frac{S}{\beta I(w^*_{IBR})} < \frac{S}{\beta I(\theta)} = \frac{s}{\int_\theta^{\overline{w}} w dF(w)}. \qquad \text{(A.2.18)}$$

The expected disposable wage offer under the two contracts are

$$E_{IBR} = \int_{(1-\alpha)\theta}^{(1-\alpha)\overline{w}} w dF_{IBR}(w) = \int_\theta^{\overline{w}} (1-\alpha) w dF(w); \qquad \text{(A.2.19)}$$

$$E_{FIX} = \int_{\theta-s}^{\overline{w}-s} w dF_{FIX}(w) = \int_\theta^{\overline{w}} (w-s) dF(w). \qquad \text{(A.2.20)}$$

Taking the difference,

$$E_{IBR} - E_{FIX} = s - \int_\theta^{\overline{w}} \alpha w dF(w) > 0, \qquad \text{(A.2.21)}$$

according to equation (A.2.18). Moreover, because s/α is the unique solution to $F_{IBR}(w) = F_{FIX}(w)$ and $(1-\alpha)\theta > \theta - s$, $F_{IBR}(w)$ single crosses $F_{FIX}(w)$ from below, i.e.,

$$\begin{aligned}
F_{IBR}(w) &< F_{FIX}(w) \quad \text{for } w < s/\alpha \\
F_{IBR}(w) &> F_{FIX}(w) \quad \text{for } w > s/\alpha.
\end{aligned} \qquad \text{(A.2.22)}$$

For $z \in [0, s/\alpha]$, the single-crossing property implies

$$\int_0^z F_{IBR}(w) dw < \int_0^z F_{FIX}(w) dw. \qquad \text{(A.2.23)}$$

For $z \in (s/\alpha, \overline{w}]$,

$$
\begin{aligned}
0 &< E_{IBR} - E_{FIX} \\
&= \int_0^\infty [1 - F_{IBR}(w)]dw - \int_0^\infty [1 - F_{FIX}(w)]dw \\
&= \int_0^z [1 - F_{IBR}(w)]dw - \int_0^z [1 - F_{FIX}(w)]dw + \int_z^\infty [1 - F_{IBR}(w)]dw - \int_z^\infty [1 - F_{FIX}(w)]dw \\
&< \int_0^z [1 - F_{IBR}(w)]dw - \int_0^z [1 - F_{FIX}(w)]dw \\
&= \int_0^z F_{FIX}(w)dw - \int_0^z F_{IBR}(w)dw. \qquad \text{(A.2.24)}
\end{aligned}
$$

Note that the second inequality uses the single-crossing property, and the second equality uses an expectation formula derived below. For a continuous random variable x taking only non-negative values,

$$
\begin{aligned}
E(x) &= \int_0^\infty x f(x)dx \\
&= \int_0^\infty (-x)d(1 - F(x)) \\
&= [-x(1 - F(x))]_0^\infty + \int_0^\infty [1 - F(x)]dx. \qquad \text{(A.2.25)}
\end{aligned}
$$

The first term in bracket vanishes because

$$
1 - F(x) = o(\frac{1}{x}) \quad \text{as } x \to \infty. \qquad \text{(A.2.26)}
$$

□

A.2.4 Proof of Proposition 3

Proof. Consider the fixed repayment contract. The disposable reservation wage, $\overline{w}_{FIX}^*$, is determined by

$$
\begin{aligned}
u(\tilde{w}_{FIX}^*) &= u(\theta - s) + \frac{\beta}{1-\beta} \int_{\tilde{w}_{FIX}^*}^{w-s} [u(w) - u(\tilde{w}_{FIX}^*)] dF_{FIX}(w) \\
&= u(\theta - s) + \frac{\beta}{1-\beta} \int_{\tilde{w}_{FIX}^*}^{\infty} [u(w) - u(\tilde{w}_{FIX}^*)] dF_{FIX}(w) \\
&= u(\theta - s) + \frac{\beta}{1-\beta} \left[M - u(\tilde{w}_{FIX}^*) - \int_{\tilde{w}_{FIX}^*}^{\infty} u'(w) d\left(\int_0^w F_{FIX}(x)dx \right) \right] \\
&= u(\theta - s) + \frac{\beta}{1-\beta} \left[M - u(\tilde{w}_{FIX}^*) - \lim_{x\to\infty} u'(x) \int_0^x F_{FIX}(w)dw + u'(\tilde{w}_{FIX}^*) \int_0^{\tilde{w}_{FIX}^*} F_{FIX}(w)dw \right. \\
&\quad \left. + \int_{\tilde{w}_{FIX}^*}^{\infty} \left(\int_0^w F_{FIX}(x)dx \right) u''(w)dw \right], \qquad \text{(A.2.27)}
\end{aligned}
$$

where $M = \lim_{x\to\infty} u(x)$. The last two equalities are derived by doing integration by parts.

Rearranging the above equation,

$$u(\tilde{w}^*_{FIX}) = (1-\beta)u(\theta - s) + \beta \left[M - \lim_{x\to\infty} u'(x) \int_0^x F_{FIX}(w)dw + u'(\tilde{w}^*_{FIX}) \int_0^{\tilde{w}^*_{FIX}} F_{FIX}(w)dw \right.$$
$$\left. + \int_{\tilde{w}^*_{FIX}}^{\infty} \left(\int_0^w F_{FIX}(x)dx \right) u''(w)dw \right]. \tag{A.2.28}$$

Similarly, for IBR, we have

$$u(\tilde{w}^*_{IBR}) = (1-\beta)u((1-\alpha)\theta) + \beta \left[M - \lim_{x\to\infty} u'(x) \int_0^x F_{IBR}(w)dw + u'(\tilde{w}^*_{IBR}) \int_0^{\tilde{w}^*_{IBR}} F_{IBR}(w)dw \right.$$
$$\left. + \int_{\tilde{w}^*_{IBR}}^{\infty} \left(\int_0^w F_{IBR}(x)dx \right) u''(w)dw \right]. \tag{A.2.29}$$

Taking the difference between (A.2.28) and (A.2.29):

$$u(\tilde{w}^*_{IBR}) - u(\tilde{w}^*_{FIX}) = (1-\beta)\left[u((1-\alpha)\theta) - u(\theta - s)\right] - \beta \lim_{x\to\infty} u'(x) \int_0^x \left[F_{IBR}(w) - F_{FIX}(w)\right] dw$$
$$+ \beta \left[\int_{\tilde{w}^*_{IBR}}^{\infty} \left(\int_0^w F_{IBR}(x)dx \right) u''(w)dw - \int_{\tilde{w}^*_{FIX}}^{\infty} \left(\int_0^w F_{FIX}(x)dx \right) u''(w)dw \right].$$
$$+ \beta \left[u'(\tilde{w}^*_{IBR}) \int_0^{\tilde{w}^*_{IBR}} F_{IBR}(w)dw - u'(\tilde{w}^*_{FIX}) \int_0^{\tilde{w}^*_{FIX}} F_{FIX}(w)dw \right] \tag{A.2.30}$$

Because $\lim_{x\to\infty} u'(x) = 0$ and $\lim_{x\to\infty} \int_0^x \left[F_{FIX}(w) - F_{IBR}(w)\right] dw$ is finite,

$$\lim_{x\to\infty} u'(x) \int_0^x \left[F_{FIX}(w) - F_{IBR}(w)\right] dw = 0. \tag{A.2.31}$$

Thus,

$$u(\tilde{w}^*_{IBR}) - u(\tilde{w}^*_{FIX}) = (1-\beta)\left[u((1-\alpha)\theta) - u(\theta - s)\right]$$
$$+ \beta \left[\int_{\tilde{w}^*_{IBR}}^{\infty} \left(\int_0^w F_{IBR}(x)dx \right) u''(w)dw - \int_{\tilde{w}^*_{FIX}}^{\infty} \left(\int_0^w F_{FIX}(x)dx \right) u''(w)dw \right]$$
$$+ \beta \left[u'(\tilde{w}^*_{IBR}) \int_0^{\tilde{w}^*_{IBR}} F_{IBR}(w)dw - u'(\tilde{w}^*_{FIX}) \int_0^{\tilde{w}^*_{FIX}} F_{FIX}(w)dw \right]. \tag{A.2.32}$$

In equation (A.2.32), increasing $\tilde{w}^*_{IBR}$ increases the LHS by $u'(\tilde{w}^*_{IBR})$, more than the increase in the RHS, $\beta F_{IBR}(\tilde{w}^*_{IBR})u'(\tilde{w}^*_{IBR})$. Thus, given $\tilde{w}^*_{FIX}$, there is a unique $\tilde{w}^*_{IBR}$, and whether it is greater or less than $\tilde{w}^*_{FIX}$ depends on the sign of the RHS conditional on $\tilde{w}^*_{FIX} = \tilde{w}^*_{IBR}$.

The first term is positive because $(1-\alpha)\theta > \theta - s$ according to Lemma 1. When $\tilde{w}^*_{FIX} = \tilde{w}^*_{IBR}$, the second term is

$$\beta\left[\int_{\tilde{w}^*_{FIX}}^{\infty}\left(\int_0^w F_{IBR}(x)dx - \int_0^w F_{FIX}(x)dx\right)u''(w)dw\right], \tag{A.2.33}$$

which is positive because $u''(w) < 0$ and $\int_0^w F_{IBR}(x)dx - \int_0^w F_{FIX}(x)dx < 0$ for all $w > \tilde{w}^*_{FIX}$ according to Lemma 1.

When $\tilde{w}^*_{FIX} = \tilde{w}^*_{IBR}$, the third term is

$$\beta u'(\tilde{w}^*_{FIX})\left[\int_0^{\tilde{w}^*_{FIX}} F_{IBR}(w)dw - \int_0^{\tilde{w}^*_{FIX}} F_{FIX}(w)dw\right], \tag{A.2.34}$$

which is negative according to Lemma 1.

□

A.3 Estimation and Numerical Methods

In this appendix section, I discuss the estimation and numerical method for the quantitative model in section 1.4. Different from existing search-theoretic models, the quantitative model is developed to allow most of the parameters being estimated in partial equilibrium without iterating on the equilibrium job contact rates. This largely simplifies the computation and makes the estimation of the full general equilibrium model tractable. Below I first discuss the estimation method and its limitations. Then I discuss the numerical algorithm that solves the model.

A.3.1 Estimation Method

The standard way to estimate an equilibrium search model is to iterate on the set of parameters Ξ in order to minimize the objective function (1.5.5). However, this method is not sufficiently tractable due to the large number of parameters in Ξ and the model complexity. The computation burden mainly comes from numerically searching for the equilibrium job contact rates, which are endogenously determined by the firms' job posting decisions and the workers' search decisions. Although searching for the equilibrium objects is not difficult in a standard search model, it is enormously time consuming in my model due to the many features introduced. If there are ways to estimate a subset of parameters without searching for the equilibrium, then the total estimation time would be possibly reduced. This is the logic that underlies my estimation method.

In particular, I estimate the model in two steps: first, I treat the equilibrium job contact rates λ_u and λ_e as parameters and estimate a subset of parameters

$$\Xi_p = \{A_0, A_1, \kappa, \xi, f_1^a, f_2^a, f_1^p, f_2^p, \phi, \eta, \mu_e, \sigma_e, \vartheta, \mu_k, \sigma_k, \mu_0, \mu_1, \mu_2, \mu_3\}$$

along with λ_u and λ_e to match the moments in Table 1.2 except for the vacancy to unemployment ratio. Second, I fix the values of Ξ_p and estimate the rest parameters $\Xi_q = \Xi / \Xi_p = \{h_e, h_u, \chi, \nu\}$.

I normalize s_e to be 1, and the other three parameters are estimated to match the vacancy to unemployment ratio and the job contact rates λ_u and λ_e, which are estimated in the first step. This is straightforward, because equation (1.4.11) indicates that $h_u = \frac{\lambda_u}{\lambda_e}$. Therefore, the second step only needs to estimate two parameters χ and ν to match two moments, λ_u and the vacancy to unemployment ratio.

Essentially, in the first step, I estimate a partial equilibrium search model with exogenous job contact rates. In the second step, I estimate a general equilibrium search model with only two parameters. This estimation method is much faster because most parameters are estimated in the first step without searching for the equilibrium objects when parameters are optimized. This is because the only equilibrium objects are job contact rates, which are treated as parameters. The estimation in the second step needs to search for the equilibrium objects, but it is much easier now because only two parameters are left to be optimized.

Discussions and Limitations

This two-step estimation method obtains the same result as the standard way of estimating all the parameters together because my quantitative model satisfies three conditions: first, the only equilibrium objects are job contact rates, which are estimated in the first step. Second, all the parameters estimated in the second step affect the model outcomes only through their impacts on job contact rates. Third, all the moments used in the second-step estimation can be exactly matched.

The first condition is satisfied because I assume that the productivity of vacancies is randomly drawn from an exogenous distribution $F(\rho)$. This ensures that the equilibrium vacancy distribution $V(\rho)$ is the same as $F(\rho)$. This condition would be violated if the productivity is not randomly drawn. For example, if different firms can post vacancies of different productivity as in Lise and Robin [165], then the vacancy distribution is also endogenous. As a result, there is no way to execute the first step due to the unknown vacancy distribution.[1] The limitation of assuming that vacancies' productivity is randomly drawn is that the model cannot capture the potential change in the distribution of vacancies' productivity when repayment policy changes. That is, my model does not capture the possibility that firms would create more productive jobs because IBR motivates borrowers to search for these jobs, the general equilibrium effect proposed by Acemoglu and Shimer [4, 5].

The second condition is satisfied because the search intensity parameters and the vacancy posting cost do not affect either agents' or firms' decisions once the job contact rates are given. It is straightforward to prove that the third condition is also satisfied.[2] If the third condition is

[1] This could be solved if in the first step we treat both the job contact rates and the vacancy distribution as parameters. But then the third condition would be violated because it is almost impossible to fit exactly the distribution of vacancy by selecting the vacancy posting cost.

[2] To see this, note that given λ^u and λ^e, the equilibrium distributions $\phi^u(\Omega)$ and $\phi^e(\Omega, \rho)$ are unique in the stationary equilibrium. The unemployment rate $\bar{u}$ is determined by equation (1.4.30). Substituting equations (1.4.10-1.4.11) into equation (1.4.29), then N_v is uniquely determined as a function of ν, λ^u, λ^e, and the equilibrium distributions. Thus,

not satisfied, this two-step estimation is guaranteed to be inconsistent with the standard way of estimating all parameters together. This is because if we cannot adjust the parameters in the second step to perfectly match the job contact rates, it means that we are over-fitting the model in the first step by selecting those contact rates that could never be achieved in equilibrium. Moreover, if we cannot perfectly match the vacancy to unemployment ratio in the second step, then the estimation result could also be different depending on the weighting matrix. This is because when all the parameters are estimated together, we may want to sacrifice the matched moments in the first step in order to better match the moment in the second step, namely, the vacancy to unemployment ratio.

Estimating Standard Errors

Once the two-step estimation is finished, standard errors of parameters can be constructed in the standard way.

First, I estimate the variance-covariance matrix $\widehat{COV}$ for all moments. Because the vector of moments in the data can be computed without knowing parameter values, $\widehat{COV}$ can be computed by bootstrapping the data directly without doing iterated MSM. Specifically, I calculate the moments $N = 200$ times by bootstrapping, then use these N observations of moments to construct the variance-covariance matrix. There are two issues in estimating $\widehat{COV}$. First, moments are constructed using different data sources. The life-cycle moments are constructed using March CPS, the vacancy to unemployment ratio is constructed using JOLTS, the default rate is constructed using NSLDS, and the remaining moments are constructed using NLSY97. The covariance between moments constructed in different data sources is set to be zero. Second, the moments in NLSY97 are constructed using different number of observations due to missing values. The covariance between any pair of moments is constructed by bootstrapping non-missing-value observations for both moments. Thus the assumption here is that values are missing randomly, though it is not likely to be true in reality.

In my estimation, I use a diagonal weighting matrix, $\hat{\Theta} = [\text{diag}(\widehat{COV})]^{-1}$, because covariance is not precisely estimated and may bias the estimated parameter values. The asymptotic variance-covariance matrix for MSM estimators $\hat{\Xi}_2$ is given by:

$$Q(\hat{\Theta}) = (\nabla^T \hat{\Theta} \nabla)^{-1} \nabla^T \hat{\Theta} \widehat{COV} \hat{\Theta}^T \nabla (\nabla^T \hat{\Theta}^T \nabla)^{-1}, \tag{A.3.1}$$

where $\nabla = \frac{\partial \hat{m}_S(\Xi_2)}{\partial \Xi_2}|_{\Xi_2 = \hat{\Xi}_2}$ is the Jacobian matrix of the simulated moments evaluated at the estimated parameters.[3] The first derivatives are calculated numerically by varying each parameter's

there is a unique v to match the vacancy to unemployment ratio. Because the number of matches M is a function of N_v and χ in equation (1.4.9), given N_v, χ is uniquely solved to match the job contact rates.

[3] In general, the formula should also incorporate simulation errors, thus the variance-covariance matrix for MSM estimators also depends on the number of simulated agents [107]. The formula I use does not consider simulation errors because instead of simulating a number of agents, I adopt the histogram method by simulating the distribution of characteristics. Therefore, as long as I focus on the stationary equilibrium, the simulation outcomes are not dependent on randomly drawn shocks.

value by 1%. The standard errors of $\hat{\Xi}_2$ are given by the square root of the diagonal elements of $Q(\hat{\Theta})$.

A.3.2 Numerical Method

I solve the model numerically. The computational complexity of this model is extremely large because this is an equilibrium model with five state variables (wealth, debt, talent, job productivity, and the negotiation benchmark's productivity).[4]

In addition to the complexity introduced by five state variables, the model is hard to solve due to the violation of the linear sharing rule in the Nash bargaining problem. Therefore, for each possible worker-job combination, the algorithm needs to solve a maximization problem whose objective function does not have an analytical solution and is determined endogenously. In the following, I first present the numerical algorithm. Then I describe the initialization of value functions in the final period. Finally, I discuss the implementation of this algorithm.

Algorithm

The model is solved by backward induction using the following algorithm:

(1). Guess the equilibrium job contact rates λ_u for unemployed workers, and $\lambda_e = \frac{s^e}{s^u}\lambda_u$ for employed workers.

(2). Solve the value functions $U(\Omega)$, $W(\Omega,\rho,\rho')$, and $J(\Omega,\rho,\rho')$ in the following steps:

- **(2.1).** Guess wage functions $w(\Omega,\rho,\rho')$ for all Ω, ρ, and ρ'.
- **(2.2).** Solve problems (1.4.25-1.4.27) by backward induction from $t = T$ to $t = 1$ to obtain $U(\Omega)$, $W(\Omega,\rho,\rho')$, $J(\Omega,\rho,\rho')$, and the corresponding policy functions.
- **(2.3).** Solve the Nash bargaining problems (1.4.15-1.4.17) to obtain wage $w'(\Omega,\rho,\rho')$.
- **(2.4).** If $w'(\Omega,\rho,\rho') = w(\Omega,\rho,\rho')$ for all Ω, ρ, and ρ', go to step (3); otherwise, go to step (2.1).

(3). Given initial distributions $\mho(a,b_0)$ and the computed value functions, solve the optimal college entry decisions. Then given the policy functions, forward simulate the model from $t = 1$ to $t = T$ to obtain distributions $\phi^u(\Omega)$ and $\phi^e(\Omega,\rho)$.

(4). Compute the equilibrium unemployment rate $\overline{u}$ using equation (1.4.30) and the aggregate level of search intensity S using equation (1.4.8). Compute the probability of contacting a worker q using the free entry condition (1.4.29).

[4]Loosely speaking, solving the model is as difficult as solving the quantitative models of Krusell, Mukoyama and Sahin [155] and Lise and Robin [165]. Krusell, Mukoyama and Sahin [155] do not model on-the-job search and Lise and Robin [165] consider risk-neutral agents. But Krusell, Mukoyama and Sahin [155] and Lise and Robin [165] also consider aggregate shocks in their models, which I do not have.

(5). Substituting S and q into equations (1.4.9-1.4.11) to obtain the number of meetings M, the number of vacancies N_v, and the equilibrium job contact rates λ_u'.

(6). Check if $\lambda_u' = \lambda_u$. If not, go to step (1).

Because I focus on the stationary equilibrium, the value functions and policy functions across different generations are identical. The final period represents age T. When the model is solved in partial equilibrium, the job contact rates λ_u and λ_e are given as parameters. Thus only steps (2) and (3) are executed.

Initialization of Value Functions

The value functions at age T are initialized by assuming that the agent consumes all wealth in the end. In the simulation, all agents should have paid off the outstanding debt before reaching age T. To have a well-defined problem, I also need to specify what happens off the equilibrium, i.e., if there is outstanding debt left at age T. I assume that the agent needs to pay off all the outstanding debt if wealth at age T is sufficient to make the payment. If wealth is not sufficient, I punish the agent to keep the level of consumption at the floor value $\underline{c}$, and the rest wealth is used to repay the debt.

Formally, the value for unemployed workers at age T is:

$$U(\Omega) = \begin{cases} \dfrac{[(1+r)b + \varkappa\theta^{1-\tau} - (1+r_s)s]^{1-\gamma}}{1-\gamma} & \text{if } (1+r)b + \varkappa\theta^{1-\tau} \geq (1+r_s)s + \underline{c} \\ \dfrac{\underline{c}^{1-\gamma}}{1-\gamma} & \text{otherwise} \end{cases} \tag{A.3.2}$$

The agent dies after age T, and the worker-job match separates as a consequence. Therefore, the value of a filled job at age T is

$$J(\Omega, \rho, \rho') = [z\rho - w(\Omega, \rho, \rho')]l, \tag{A.3.3}$$

where $l = \left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1}{\sigma+\tau}} w(\Omega, \rho, \rho')^{\frac{1-\tau}{\sigma+\tau}}$.

To calculate the value for employed workers at age T, I need to solve the Nash bargaining problem to obtain the wage functions at age T. This can be solved directly from a root-finding problem.[5] Depending on whether the agent's wealth is sufficient to repay the debt, there are two cases:

Case 1: Insolvency If the agent is employed at job ρ, the highest wage rate that job ρ can offer is its marginal product of labor $z\rho$. If the agent could not repay the debt when being offered this wage rate, then she is insolvent and would consume $\underline{c}$. Note that because the current job's

[5]Note that the wage functions at the final period T can be solved directly from a root-finding problem because the agent consumes all wealth in the final period. In other periods $t < T$, due to the endogenous consumption and savings decisions, multiple iterations are needed to obtain the wage functions as fixed points.

productivity is always higher than the negotiation benchmark's productivity, the agent would also be insolvent and consume $\underline{c}$ at the negotiation benchmark. In this scenario, the agent is indifferent between being employed at the current job ρ or at the negotiation benchmark, thus the match surplus for the agent is zero. Moreover, the agent has no incentive to supply labor as this only increases repayment but not consumption. Thus the firm would also obtain zero match surplus. This implies that the value for employed workers at age T is $\underline{c}^{1-\gamma}/(1-\gamma)$ in the case of insolvency.

To pin down the condition for insolvency, consider the highest wage rate $z\rho$ being offered by job ρ to agent Ω. The after-tax income is $\varkappa(z\rho)^{1-\tau}$. Substituting the optimal labor supply, $l = \left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1}{\sigma+\tau}} (z\rho)^{\frac{1-\tau}{\sigma+\tau}}$, the maximum wealth that the agent can obtain is

$$\begin{aligned} \bar{b} &= (1+r)b + \varkappa(z\rho l)^{1-\tau} \\ &= (1+r)b + \varkappa \left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1-\tau}{\sigma+\tau}} (z\rho)^{\frac{(1+\sigma)(1-\tau)}{\sigma+\tau}}. \end{aligned} \tag{A.3.4}$$

The agent's utility is

$$\begin{aligned} u(\bar{b}-(1+r_s)s, l) &= \frac{1}{1-\gamma}\left[\bar{b}-(1+r_s)s - \phi\frac{l^{1+\sigma}}{1+\sigma}\right]^{1-\gamma} \\ &= \frac{1}{1-\gamma}\left[\bar{b}-(1+r_s)s - \frac{\phi}{1+\sigma}\left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1+\sigma}{\sigma+\tau}} (z\rho)^{\frac{(1-\tau)(1+\sigma)}{\sigma+\tau}}\right]^{1-\gamma}. \end{aligned} \tag{A.3.5}$$

For the agent to be insolvent, it should hold that $u(\bar{b}-(1+r_s)s, l) \leq u(\underline{c}, 0)$, which requires

$$\bar{b} \leq (1+r_s)s + \frac{\phi}{1+\sigma}\left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1+\sigma}{\sigma+\tau}} (z\rho)^{\frac{(1-\tau)(1+\sigma)}{\sigma+\tau}} + \underline{c}. \tag{A.3.6}$$

Case 2: Solvency When condition (A.3.6) is not satisfied, the agent is solvent if the highest wage rate is offered by job ρ. Therefore, the actual wage rate $w(\Omega,\rho,\rho')$ that solves the Nash bargaining problem should also satisfy the solvency condition, i.e.,

$$\begin{aligned} &(1+r)b + \varkappa\left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1-\tau}{\sigma+\tau}} w(\Omega,\rho,\rho')^{\frac{(1+\sigma)(1-\tau)}{\sigma+\tau}} \\ &> (1+r_s)s + \frac{\phi}{1+\sigma}\left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1+\sigma}{\sigma+\tau}} w(\Omega,\rho,\rho')^{\frac{(1-\tau)(1+\sigma)}{\sigma+\tau}} + \underline{c}. \end{aligned} \tag{A.3.7}$$

Otherwise, the agent would obtain a zero match surplus and choose not to supply labor, which also results in a zero match surplus for the firm. Thus both sides could be better off if the firm increases the wage rate to satisfy the solvency condition (A.3.7). I now derive the wage function $w(\Omega,\rho,\rho')$ under condition (A.3.7).

The value for employed workers at age T is:

$$W(\Omega,\rho,\rho') = \frac{1}{1-\gamma}\left[(1+r)b + \varkappa[w(\Omega,\rho,\rho')l]^{1-\tau} - (1+r_s)s - \phi\frac{l^{1+\sigma}}{1+\sigma}\right]^{1-\gamma}, \tag{A.3.8}$$

where $l = \left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1}{\sigma+\tau}} w(\Omega,\rho,\rho')^{\frac{1-\tau}{\sigma+\tau}}$.

The outside option value for employed workers with negotiation benchmark ρ' at age T is

$$\overline{W}(\Omega,\rho') = \max\left\{\frac{1}{1-\gamma}\left[(1+r)b + \varkappa(z\rho'l')^{1-\tau} - (1+r_s)s - \phi\frac{l'^{1+\sigma}}{1+\sigma}\right]^{1-\gamma}, \frac{\underline{c}^{1-\gamma}}{1-\gamma}\right\}, \tag{A.3.9}$$

where $l' = \left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1}{\sigma+\tau}} (z\rho')^{\frac{1-\tau}{\sigma+\tau}}$. The max operator considers the solvency/insolvency case at the negotiation benchmark ρ'.

The $w(\Omega,\rho,\rho')$ is chosen to maximize the bargaining objective function:

$$w(\Omega,\rho,\rho') = \underset{w(\Omega,\rho,\rho')}{\operatorname{argmax}} \left[W(\Omega,\rho,\rho') - \overline{W}(\Omega,\rho')\right]^{\xi} J(\Omega,\rho,\rho')^{1-\xi}. \tag{A.3.10}$$

Substituting equations (A.3.3), (A.3.8) and (A.3.9) into problem (A.3.10) and taking the first order condition, we obtain $w(\Omega,\rho,\rho')$ by solving the following root-finding problem:

$$\begin{aligned}
&\frac{\xi(1-\gamma)\left[B + Kw(\Omega,\rho,\rho')^{\frac{(1+\sigma)(1-\tau)}{\sigma+\tau}}\right]^{-\gamma} \varkappa(1-\tau)\left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1-\tau}{\sigma+\tau}} w(\Omega,\rho,\rho')^{\frac{1-2\tau-\sigma\tau}{\sigma+\tau}}}{\left[B + Kw(\Omega,\rho,\rho')^{\frac{(1+\sigma)(1-\tau)}{\sigma+\tau}}\right]^{1-\gamma} - \left[\max\{B + K(z\rho')^{\frac{(1+\sigma)(1-\tau)}{\sigma+\tau}}, \underline{c}\}\right]^{1-\gamma}} \\
&= (1-\xi)\left[\frac{1}{z\rho - w(\Omega,\rho,\rho')} - \frac{1-\tau}{(\sigma+\tau)w(\Omega,\rho,\rho')}\right],
\end{aligned} \tag{A.3.11}$$

where $B = (1+r)b - (1+r_s)s$ and $K = \varkappa\frac{\sigma+\tau}{1+\sigma}\left[\frac{\varkappa(1-\tau)}{\phi}\right]^{\frac{1-\tau}{\sigma+\tau}}$.

I use bisection method to solve equation (A.3.11) with initial lower bound,

$$LB = \left[\frac{1+\sigma}{\varkappa(\sigma+\tau)}\left[\frac{\phi}{\varkappa(1-\tau)}\right]^{\frac{1-\tau}{\sigma+\tau}} \left[[(1-\gamma)\overline{W}(\Omega,\rho')]^{\frac{1}{1-\gamma}} - (1+r)b + (1+r_s)s\right]\right]^{\frac{\sigma+\tau}{(1+\sigma)(1-\tau)}}, \tag{A.3.12}$$

and upper bound,

$$UB = z\rho. \tag{A.3.13}$$

Finally, substituting the solution of $w(\Omega,\rho,\rho')$ into equations (A.3.3) and (A.3.8), we obtain $J(\Omega,\rho,\rho')$ and $W(\Omega,\rho,\rho')$.

Table A.4: Discretization of state space.

Parameters	Value	Description
n_b	400	Number of wealth grids
Δ_b	\$500	Length of wealth grids
$[\underline{b} \quad \bar{b}]$	[\$0 \$200,000]	Range of wealth
n_s	100	Number of student loan debt grids
Δ_s	\$500	Length of student loan debt grids
$[\underline{s} \quad \bar{s}]$	[\$0 \$50,000]	Range of student loan debt
n_ρ	20	Number of productivity grids
Δ_ρ	0.05	Length of productivity grids
$[\underline{\rho} \quad \bar{\rho}]$	[0 1]	Range of productivity

Implementation

To ensure accuracy, I choose relatively fine grids (see Table A.4), and the values between grids are approximated by linear interpolation. I use the golden section search method to find the optimal decision rules. The advantage of the golden section search method is that it is robust to the choice of initial values because convergence is guaranteed. However, convergence to the global optimum is not ensured if there are many local optima. Therefore, I further divide the whole decision space into multiple sub-space and select the largest local optimum. I do a robustness check after the estimation using a sequential grid search, and the results are identical. When solving the Nash bargaining problem, I need to invoke the calculation for utility from consumption and utility from the future multiple times. I save the computation time by calculating these values in advance and store them in memory.

The numerical algorithm is implemented using C++. The program is run on the server of MIT Economics Department, *supply.mit.edu*, which is built on Dell PowerEdge R910 running RedHat 6.7 (64-core processor, Intel(R) Xeon(R) CPU E7-4870, 2.4GHz). I use OpenMP for parallelization when iterating value functions and simulating the model. My baseline model requires 200GB of RAM to store the large number of decision rules and value functions.

A.4 Additional Theoretical Results

A.4.1 Restructuring the Fixed Repayment Contract

The existence of the liquidity channel suggests that the lender can restructure the schedule of repayment to mitigate the debt burden. In reality, the federal student loan system has such features. For example, under the Direct Loan Program and the FFEL Program, borrowers have a 6-month grace period after graduation before payments are due. Moreover, the graduated repayment plan allows borrowers to make smaller payments at first and then increase their payments over time. The extended repayment plan allows qualified borrowers to extend the repayment period up to 25 years.

To formalize the intuition behind these realistic repayment plans, consider a particular contract that requires the agent to repay s_1 at $t = 1$, and s_2 at $t \geq 2$, such that the outstanding debt S is recovered:

$$\frac{s_1}{1+r} + \sum_{t=2}^{\infty} \frac{s_2}{(1+r)^t} = S. \tag{A.4.1}$$

Proposition 8 shows that back-loading debt payments increases the reservation wage at $t = 1$ through the liquidity channel.

Proposition 8. *Reducing s_1 and increasing s_2 subject to the constraint (A.4.1) strictly increases the reservation wage at $t = 1$ when the borrowing constraint is binding.*

Proof. If the wage offer is accepted at $t = 1$, then the wage income becomes flat in the future. Therefore, the agent would perfectly smooth consumption by saving $s - s_1$ at $t = 1$, and consuming $w - s$ in every period. The value function is

$$W_1(w) = \frac{u(w-s)}{1-\beta}. \tag{A.4.2}$$

Under the twisted repayment schedule, suppose that the agent's borrowing constraint is binding when unemployed, i.e., the agent does not save at $t = 1$ if the wage offer is rejected. Then the value function is

$$U_1 = u(\theta - s_1) + \beta \int_{\theta}^{w_2^*} U_2 dF(w) + \beta \int_{w_2^*}^{\overline{w}} W_2(w) dF(w), \tag{A.4.3}$$

where U_2 and $W_2(w)$ are the value functions of rejecting and accepting the wage offer at $t = 2$ conditional on the wage offer being rejected at $t = 1$. w_2^* is the reservation wage at $t = 2$; It is also the reservation wage for all $t > 2$ because the job search problem is stationary in later periods due to constant debt repayment and zero initial wealth. Therefore, we can write U_2 and $W_2(w)$ as

$$W_2(w) = \frac{u(w-s_2)}{1-\beta}. \tag{A.4.4}$$

$$U_2 = \frac{u(\theta - s_2)}{1-\beta} + \frac{\beta}{1-\beta} \int_{w_2^*}^{\overline{w}} [W_2(w) - U_2] dF(w). \tag{A.4.5}$$

The reservation wage at $t = 1$, w_1^*, is determined by

$$U_1 = W_1(w_1^*) \tag{A.4.6}$$

Substituting equations (A.4.2) and (A.4.3) into equation (A.4.6), we obtain

$$\frac{u(w_1^* - s)}{1-\beta} = u(\theta - s_1) + \beta \int_{\theta}^{w_2^*} U_2 dF(w) + \beta \int_{w_2^*}^{\overline{w}} W_2(w) dF(w). \tag{A.4.7}$$

Substituting equation (A.4.4) and $U_2 = W_2(w_2^*)$ into equation (A.4.7), we obtain

$$\frac{u(w_1^* - s)}{1-\beta} = u(\theta - s_1) + \frac{\beta}{1-\beta}\int_\theta^{w_2^*} u(w_2^* - s_2)dF(w) + \frac{\beta}{1-\beta}\int_{w_2^*}^{\overline{w}} u(w - s_2)dF(w). \tag{A.4.8}$$

Consider small changes of payments, $\Delta s_1 < 0$, equation (A.4.1) and assumption $\beta(1+r) = 1$ imply

$$\Delta s_2 = -r\Delta s_1 = -\frac{1-\beta}{\beta}\Delta s_1 > 0. \tag{A.4.9}$$

Differentiating equation (A.4.8):

$$\begin{aligned}\Delta w_1^* = & -\frac{1}{Q}u'(\theta - s_1)\Delta s_1 + \frac{\beta}{Q(1-\beta)}u'(w_2^* - s_2)F(w_2^*)\Delta w_2^* \\ & + \frac{\beta}{Q(1-\beta)}\left[-u'(w_2^* - s_2) + \int_{w_2^*}^{\overline{w}}[u'(w_2^* - s_2) - u'(w - s_2)]dF(w)\right]\Delta s_2,\end{aligned} \tag{A.4.10}$$

where $Q = \frac{u'(w_1^* - s)}{1-\beta} > 0$.

The reservation wage at $t = 2$, w_2^*, is determined by $U_2 = W(w_2^*)$,

$$u(w_2^* - s_2) = u(\theta - s_2) + \frac{\beta}{1-\beta}\int_{w_2^*}^{\overline{w}}[u(w - s_2) - u(w_2^* - s_2)]dF(w). \tag{A.4.11}$$

Differentiating equation (A.4.11):

$$\Delta w_2^* = \frac{u'(w_2^* - s_2)\dfrac{1-\beta F(w_2^*)}{1-\beta} - u'(\theta - s_2) - \dfrac{\beta}{1-\beta}\displaystyle\int_{w_2^*}^{\overline{w}} u'(w - s_2)dF(w)}{u'(w_2^* - s_2)\dfrac{1-\beta F(w_2^*)}{1-\beta}}\Delta s_2. \tag{A.4.12}$$

Substituting (A.4.9) and (A.4.12) into (A.4.10), I obtain

$$\Delta w_1^* = -\frac{1}{Q}[u'(\theta - s_1) - \frac{(1-\beta)F(w_2^*)}{1-\beta F(w_2^*)}u'(\theta - s_2) - \frac{1}{1-\beta F(w_2^*)}\int_{w_2^*}^{\overline{w}} u'(w - s_2)dF(w)]\Delta s_1. \tag{A.4.13}$$

When the wage offer at $t = 1$ is rejected, the marginal utility of one unit of consumption at $t = 1$ is $u'(\theta - s_1)$, and the marginal utility of one unit of savings is

$$\beta\left[(1+r)u'(\theta - s_2)F(w_2^*) + \frac{r}{1-\beta}\int_{w_2^*}^{\overline{w}} u'(w - s_2)dF(w)\right]. \tag{A.4.14}$$

In (A.4.14), the first term captures that the agent would consume $(1+r)$ at marginal utility $u'(\theta - s_2)$ if the wage offer is below w_2^* at $t = 2$ and rejected. The agent does not save in this case because debt payment is flat during $t \geq 2$ and expected income is higher. The second term captures that the agent would consume r at marginal utility $u'(w - s_2)$ in every future period, $t \geq 2$, if the wage offer w is above w_2^* at $t = 2$ and accepted. This is because both wage income and debt payment are flat in every future period, $t \geq 2$. Thus the agent would only consume the interest of her one unit of wealth to perfectly smooth consumption.

The binding borrowing constraint implies that the marginal utility of one unit of consumption at $t = 1$ is larger than the marginal utility of one unit of savings, i.e.,

$$u'(\theta - s_1) \geq F(w_2^*)u'(\theta - s_2) + \int_{w_2^*}^{\overline{w}} u'(w - s_2)dF(w). \tag{A.4.15}$$

Substituting (A.4.15) into (A.4.13),

$$\begin{aligned}
\Delta w_1^* &\geq -\frac{1}{Q}\left[\frac{\beta F(w_2^*)[1 - F(w_2^*)]}{1 - \beta F(w_2^*)}u'(\theta - s_2) - \frac{\beta F(w_2^*)}{1 - \beta F(w_2^*)}\int_{w_2^*}^{\overline{w}} u'(w - s_2)dF(w)\right]\Delta s_1 \\
&= -\frac{\beta F(w_2^*)}{Q[1 - \beta F(w_2^*)]}\left[[1 - F(w_2^*)]u'(\theta - s_2) - \int_{w_2^*}^{\overline{w}} u'(w - s_2)dF(w)\right]\Delta s_1 \\
&= -\frac{\beta F(w_2^*)\Delta s_1}{Q[1 - \beta F(w_2^*)]}\int_{w_2^*}^{\overline{w}} [u'(\theta - s_2) - u'(w - s_2)]dF(w) > 0
\end{aligned} \tag{A.4.16}$$

□

In contrast to Proposition 1, Proposition 8 holds for any risk-averse agent, but it requires an imperfect credit market. When the borrowing constraint is not binding, the liquidity channel is absent because any change in the repayment schedule only results in a change in savings rather than affecting the job search decisions. When the borrowing constraint is binding, back-loading debt payments affects the reservation wage through two effects. First, reducing s_1 has a direct positive effect on the reservation wage at $t = 1$, because it provides liquidity for continued search. Second, reducing s_1 induces a higher s_2, resulting in a lower reservation wage at $t \geq 2$. The lower future reservation wages reduce the value of continued job search, which in turn indirectly imposes a negative effect on the reservation wage at $t = 1$. When the borrowing constraint is binding, the direct effect dominates the indirect effect. Intuitively, this is because the agent faces a higher marginal utility of consumption in the current period, thus she has the incentive to transfer wealth from future periods by setting a lower reservation wage. Requiring a smaller payment in the current period dampens this incentive by reducing the intertemporal gap in the marginal utility of consumption. As a result, the agent would increase her reservation wage to pursue a higher expected future return.

A.4.2 Implication on Expected Income

A lower reservation wage implies that the agent would have less expected income when she is indebted under the fixed repayment contract. To see this, let $I(w_{FIX}^*)$ denote the present value of expected income as a function of the reservation wage w_{FIX}^*, and then it can be solved recursively:

$$I(w_{FIX}^*) = F(w_{FIX}^*)[\theta + \beta I(w_{FIX}^*)] + \int_{w_{FIX}^*}^{\overline{w}} \frac{w}{1 - \beta}dF(w). \tag{A.4.17}$$

Equation (A.4.17) states that when the agent draws an offer below w^*_{FIX} with probability $F(w^*_{FIX})$, she rejects it and receives UI benefits θ in the current period and the same present value of expected income $I(w^*_{FIX})$ in the next period. When the wage offer is above w^*, she accepts it and gets paid perpetually. The compensation for search risks implies a monotonic relationship between w^*_{FIX} and $I(w^*_{FIX})$:

Proposition 9. *There exists a unique income-maximizing reservation wage $\hat{w}$, determined by*

$$\hat{w} - \frac{\beta}{1-\beta}\int_{\hat{w}}^{\overline{w}} (w - \hat{w})dF(w) = \theta. \tag{A.4.18}$$

The present value of expected income is strictly increasing in w^*_{FIX} when $w^*_{FIX} < \hat{w}$, and strictly decreasing in w^*_{FIX} when $w^*_{FIX} > \hat{w}$. Moreover, the optimal reservation wage for any risk-averse agent satisfies $w^*_{FIX} < \hat{w}$.

Proof. Rearranging equation (A.4.17),

$$I(w^*_{FIX}) = \frac{\theta F(w^*_{FIX}) + \frac{1}{1-\beta}\int_{w^*_{FIX}}^{\overline{w}} w dF(w)}{1 - \beta F(w^*_{FIX})}. \tag{A.4.19}$$

Take the first derivative,

$$I'(w^*_{FIX}) = \frac{f(w^*_{FIX})}{[1-\beta F(w^*_{FIX})]^2}\left[\theta - w^*_{FIX} + \frac{\beta}{1-\beta}\int_{w^*_{FIX}}^{\overline{w}} (w - w^*_{FIX})dF(w)\right]. \tag{A.4.20}$$

Denote

$$h(x) = \theta - x + \frac{\beta}{1-\beta}\int_{x}^{\overline{w}} (w - x)dF(w). \tag{A.4.21}$$

It is straightforward to show that $h(\theta) > 0$, $h(\overline{w}) < 0$, and $h(x)' < 0$. Thus there exists a unique $w^*_{FIX} \in (\theta, \overline{w})$, denoted as $\hat{w}$, such that $I'(\hat{w}) = 0$. When $w^* < \hat{w}$, $I'(w^*_{FIX}) > 0$ and expected income is strictly increasing in w^*_{FIX}; when $w^*_{FIX} > \hat{w}$, $I'(w^*_{FIX}) < 0$ and expected income is strictly decreasing in w^*_{FIX}. Therefore, $\hat{w}$ maximizes expected income and is determined by

$$\hat{w} - \frac{\beta}{1-\beta}\int_{\hat{w}}^{\overline{w}} (w - \hat{w})dF(w) = \theta. \tag{A.4.22}$$

Now, I prove that a risk-neutral agent sets her reservation wage to be $\hat{w}$. Because the interest rate is assumed to satisfy $\beta(1+r) = 1$, the risk-neutral agent is indifferent about savings. Without loss of generality, I assume that the risk-neutral agent also behaves hand-to-mouth, like a risk-averse agent. Therefore, her reservation wage is determined by equation (1.3.3).

The utility function of the risk-neutral agent has a linear form, i.e., $u(x) = ax + b$. Substituting this into equation (1.3.3), I obtain

$$w_{FIX}^{*}-\frac{\beta}{1-\beta}\int_{w_{FIX}^{*}}^{\overline{w}}(w-w_{FIX}^{*})dF(w)=\theta. \tag{A.4.23}$$

There is a unique solution to equation (A.4.23), thus $w_{FIX}^{*}=\hat{w}$ for the risk-neutral agent. □

In fact, the income-maximizing reservation wage $\hat{w}$ is the reservation wage set by risk-neutral agents. In an incomplete market, the existence of uninsured search risks incentivizes risk-averse agents to set a strictly lower reservation wage in order to smooth consumption.

A.4.3 Tradeoff Between Insurance and Incentive to Work

Because IBR provides insurance, it is not surprising that it increases welfare relative to the fixed repayment contract. I prove this result under CRRA utility:

Proposition 10. *With CRRA utility (and inelastic labor supply), IBR improves the agent's welfare relative to the fixed repayment contract.*

Proof. The welfare of the agent under the two repayment contracts is given by:

$$\begin{aligned}\text{Welfare}_{IBR}&=\frac{F(w_{IBR}^{*})u((1-\alpha)\theta)}{1-\beta F(w_{IBR}^{*})}+\int_{w_{IBR}^{*}}^{\overline{w}}\frac{u((1-\alpha)w)}{(1-\beta)[1-\beta F(w_{IBR}^{*})]}dF(w);\\ \text{Welfare}_{FIX}&=\frac{F(w_{FIX}^{*})u(\theta-s)}{1-\beta F(w_{FIX}^{*})}+\int_{w_{FIX}^{*}}^{\overline{w}}\frac{u(w-s)}{(1-\beta)[1-\beta F(w_{FIX}^{*})]}dF(w).\end{aligned} \tag{A.4.24}$$

Notice that

$$\frac{(1-\beta)F(w^{*})+\int_{w^{*}}^{\overline{w}}dF(w)}{1-\beta F(w^{*})}=1, \tag{A.4.25}$$

which allows us to define a CDF $G(w;w^{*})$ with parameter w^{*} as follows:

$$G(w;w^{*})=\begin{cases}\dfrac{(1-\beta)F(w^{*})}{1-\beta F(w^{*})} & \text{if } w\in[\theta,w^{*}],\\ \dfrac{F(w)-\beta F(w^{*})}{1-\beta F(w^{*})} & \text{if } w\in(w^{*},\overline{w}].\end{cases} \tag{A.4.26}$$

Then the welfare equations (A.4.24) can be written as:

$$\text{Welfare}_{IBR}=\frac{1}{1-\beta}\int_{\theta}^{\overline{w}}u((1-\alpha)w)dG(w;w_{IBR}^{*}); \tag{A.4.27}$$

$$\text{Welfare}_{FIX}=\frac{1}{1-\beta}\int_{\theta}^{\overline{w}}u(w-s)dG(w;w_{FIX}^{*}). \tag{A.4.28}$$

To prove that Welfare$_{IBR}$ > Welfare$_{FIX}$, it is sufficient to show that the lottery with value $(1-\alpha)w$ and CDF $G(w; w^*_{IBR})$ is second-order stochastically dominant over the lottery with value $w - s$ and CDF $G(w; w^*_{FIX})$.

Following the proof of Lemma 1, the single-crossing condition is satisfied because $(1-\alpha)\theta > \theta - s$. Thus I only need to show that the mean of lottery $G(w; w^*_{IBR})$ is larger than the mean of lottery $G(w; w^*_{FIX})$:

$$\int_\theta^{\overline{w}} (1-\alpha) w dG(w; w^*_{IBR}) > \int_\theta^{\overline{w}} (w - s) dG(w; w^*_{FIX}). \tag{A.4.29}$$

Define $L(w^*)$ as follows:

$$L(w^*) = \int_\theta^{\overline{w}} w dG(w; w^*) = \frac{(1-\beta)\theta F(w^*) + \int_{w^*}^{\overline{w}} w dF(w)}{1 - \beta F(w^*)}. \tag{A.4.30}$$

Taking the first derivative w.r.t. w^*:

$$L(w^*)' = \frac{(1-\beta) f(w^*)}{[1-\beta F(w^*)]^2} \left[\theta - w^* + \frac{\beta}{1-\beta} \int_{w^*}^{\overline{w}} (w - w^*) dF(w)\right]. \tag{A.4.31}$$

According to the proof of Proposition 9, $L(w^*)' > 0$ as long as $w^* < \hat{w}$, which is always the case because $\hat{w}$ is the reservation wage chosen by a risk-neutral agent.

Proposition 2 shows that with CRRA utility $w^*_{IBR} > w^*_{FIX}$. Then $L(w^*)' > 0$ implies that

$$\int_\theta^{\overline{w}} w dG(w; w^*_{IBR}) > \int_\theta^{\overline{w}} w dG(w; w^*_{FIX}). \tag{A.4.32}$$

The repayment ratio α is determined by equations (A.4.17) and (1.3.4), thus

$$s = \frac{(1-\beta) F(w^*_{IBR}) \alpha\theta + \int_{w^*_{IBR}}^{\overline{w}} \alpha w dF(w)}{1 - \beta F(w^*_{IBR})} = \int_\theta^{\overline{w}} \alpha w dG(w; w^*_{IBR}). \tag{A.4.33}$$

This implies

$$\int_\theta^{\overline{w}} \alpha w dG(w; w^*_{IBR}) = \int_\theta^{\overline{w}} s dG(w; w^*_{FIX}). \tag{A.4.34}$$

Equations (A.4.32) and (A.4.34) lead to (A.4.29). □

However, this proposition may not hold if labor supply is sufficiently elastic, because IBR also naturally introduces an income-tax-like distortion, resulting in an efficiency loss. Moreover, the proof of Proposition 10 also hinges on Proposition 2.[6] This implies that the reservation wage effect of income contingency plays a role in determining the agent's welfare.

[6]For a general DARA utility, the proof is not obtained because it is not clear whether IBR raises the reservation wage.

In the following, I introduce elastic labor supply to elucidate the tradeoff between the two contracts and the implication of the reservation wage effect on welfare. To this end, I consider a simple mix of the two contracts by assuming that the lender is restricted to using a linear combination of IBR and the fixed repayment contract.[7]

In particular, the lender makes a fraction of debt mS income contingent, and the rest $(1-m)S$ is repaid under the fixed repayment contract, where $m \in [0,1]$. Under this linear contract, in each period the agent repays

$$s = \begin{cases} \alpha\theta + r(1-m)S & \text{if unemployed,} \\ \alpha z + r(1-m)S & \text{if employed with earnings } z = wl(w,\alpha), \end{cases} \tag{A.4.35}$$

where labor supply $l(w,\alpha)$ is a function of the wage rate w and the repayment ratio α.

For the lender to break even, the repayment ratio α is chosen to satisfy the recoverability constraint[8],

$$\frac{mD}{\beta} = \frac{F(w^*)}{1-\beta F(w^*)}\alpha\theta + \frac{\alpha}{(1-\beta)[1-\beta F(w^*)]}\int_{w^*}^{\overline{w}} wl(w,\alpha)dF(w). \tag{A.4.36}$$

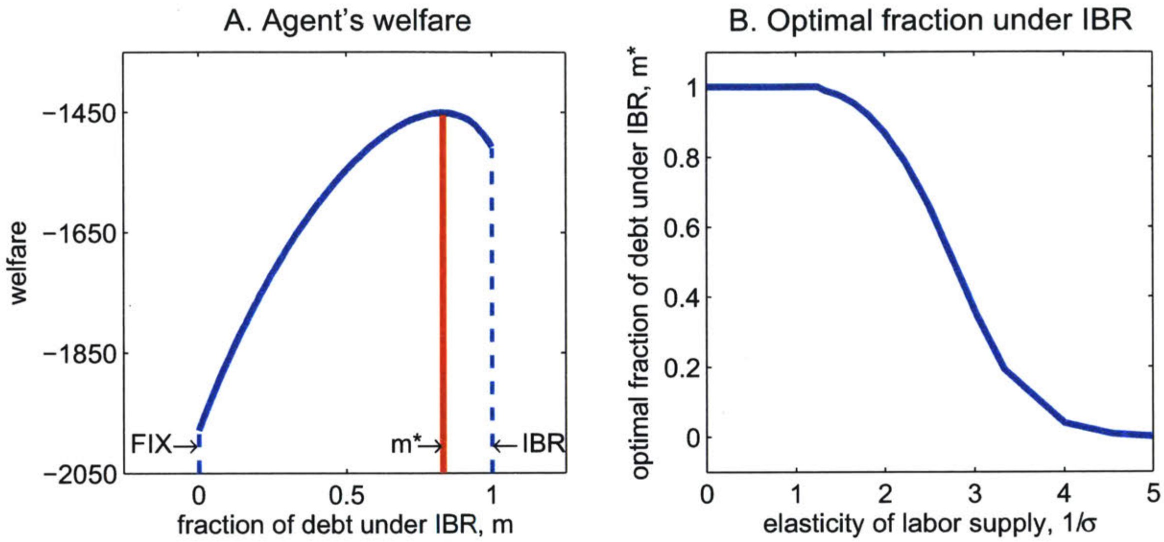
I use GHH utility [113], $u(c,l) = \frac{1}{1-\gamma}\left(c - \phi\frac{l^{1+\sigma}}{1+\sigma}\right)^{1-\gamma}$, to provide several numerical examples. Panel A of Figure A.1 shows that depending on parameter values, increasing m may increase or decrease the agent's welfare due to the the tradeoff in insurance and the incentive to work. The optimal fraction of debt repaid under IBR m^* that maximizes the agent's welfare could be an interior point. Intuitively, there are diminishing returns in providing insurance through IBR due to the decreasing marginal utility of consumption. On the other hand, the distortion on labor supply increases as a higher m increases the repayment ratio α. The optimal value of m^* is achieved when the marginal benefit from providing insurance is equal to the marginal cost of labor supply distortion.

In general, m^* could also be a corner solution, in which case the full IBR is strictly better than the fixed repayment contract or vice versa. Panel B of Figure A.1 indicates that whether IBR results in a higher welfare crucially depends on the elasticity of labor supply. This is because the elasticity of labor supply determines how responsive labor supply would be when a fraction of income is extracted by the lender. When labor supply is completely inelastic, IBR is strictly better as shown in Proposition 10. However, when labor supply is very elastic, the distortion on labor supply is large; so the fixed repayment contract results in a higher welfare.

In Figure A.2, I illustrate that the insurance provided by IBR is more valuable due to the

[7]I consider this linear contract for its simplicity and transparency to illustrate the idea. It is also partially motivated by the numerical examples of Mirrlees [182] that the optimal tax schedule is hardly different from an affine function with a constant marginal tax rate. However, numerical simulations from later research show that optimal tax schedules are very sensitive to the utility functions and income distributions.

[8]Because earnings depend on labor supply, which is a function of α, there is a Laffer curve for expected debt repayment, and there may not exist a solution to equation (A.4.36) when the debt level is high. My following numerical analyses consider the case in which there exist solutions to equation (A.4.36) and the smaller α is always selected.



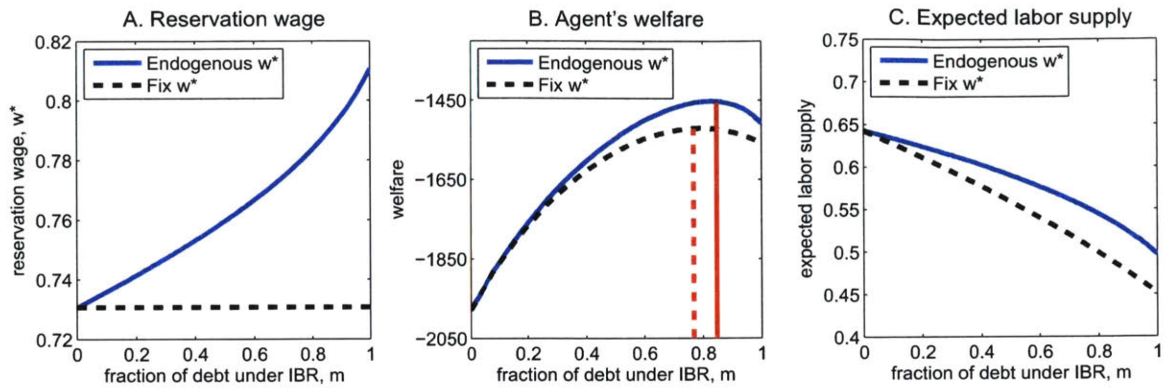
Note: This figure illustrates the tradeoff between insurance and the incentive to work. I consider a simple mix of the two contracts by assuming that the lender is restricted to using a linear combination of IBR and the fixed repayment contract. Panel A plots the agent's welfare when the fraction of debt repaid under IBR varies from zero (corresponding to the pure fixed repayment contract) to one (corresponding to the pure IBR). It shows that the agent's welfare first increases then decreases due to the benefit from insurance and the distortion on labor supply. The optimal fraction under this parametrization is given by an interior point m^*. Panel B plots the optimal fraction of debt under IBR when the elasticity of labor supply varies. A more elastic labor supply increases the distortion on labor supply, thus making IBR less desirable. The figure is plotted using the GHH utility, $u(c,l) = [c - \phi l^{1+\sigma}/(1+\sigma)]^{1-\gamma}/(1-\gamma)$, and beta distribution of wage offers, $Beta(a,b)$, with parameter values: $a = 2, b = 4, \gamma = 3, \theta = 0.1, \overline{w} = 1.1, \beta = 0.96, S = 1, \phi = 1, \sigma = 0.47$.

Figure A.1: A numerical illustration of the agent's welfare and the optimal fraction of debt repaid under IBR.

positive response in the reservation wage. In particular, I gradually increase m from 0 to 1 and compare the change in welfare and expected labor supply in two scenarios. In one scenario, I allow the agent to endogenously choose the reservation wage; in the other scenario, I fix the reservation wage at the beginning. Because IBR raises the reservation wage, the reservation wage in the first scenario is increasing as m increases (see panel A of Figure A.2).

Panel B of Figure A.2 indicates that the welfare is significantly higher in the first scenario. This illustrates that IBR increases welfare not only by directly providing insurance, but also by indirectly increasing the reservation wage. In other words, the insurance provided by IBR is more desirable when there are search risks because the agent would choose a higher reservation wage when search risks are partially insured. As a result, the optimal fraction of debt, m^*, is also higher in the first scenario.

Note that this result is not general and would be violated if the elasticity of labor supply is very large. In fact, when labor supply is elastic, IBR raises the reservation wage through an additional channel. This is because repaying debt as a fraction of income disproportionately reduces income more during employment relative to during unemployment because of the negative response in labor supply. This generates a "debt overhang" effect. The "debt overhang" channel not only reduces labor supply, but also further incentivizes the agent to set a higher reservation



Note: This figure illustrates the reservation wage effect. The blue solid line plots the agent's reservation wage, welfare, and expected labor supply when the reservation wage is allowed to increase as a larger fraction of debt is made income contingent. The black dashed line plots the agent's reservation wage, welfare, and expected labor supply when the reservation wage is fixed at the initial value under the pure fixed repayment contract (see panel A). Panel B shows that the agent's welfare is higher if the reservation wage is allowed to increase; as a result, the optimal fraction of debt repaid under IBR is also larger. Panel C shows that the reduction in labor supply also becomes smaller due to the higher reservation wage. The figure is plotted using the GHH utility, $u(c,l) = [c - \phi l^{1+\sigma}/(1+\sigma)]^{1-\gamma}/(1-\gamma)$, and the beta distribution of wage offers, $Beta(a,b)$, with parameter values: $a = 2$, $b = 4$, $\gamma = 3$, $\theta = 0.1$, $\overline{w} = 1.1$, $\beta = 0.96$, $S = 1$, $\phi = 1$, $\sigma = 0.47$.

Figure A.2: A numerical illustration of the reservation wage effect.

wage in order to stay unemployed.[9] When the elasticity of labor supply is sufficiently large, the reservation wage could be higher than the efficient one; as a result, fixing the reservation wage at some lower level could be welfare improving. See Appendix A.4.4 for a detailed discussion.

Panel C compares the expected labor supply in the two scenarios. It shows that the negative effect on labor supply is smaller when the reservation wage is endogenous. This is due to two channels: first, there is a direct positive substitution effect on labor supply as IBR increases the average wage rate by raising the reservation wage. Second, there is an indirect effect due to a lower repayment ratio. This is because a higher reservation wage increases expected repayment conditional on any repayment ratio. Therefore, when the reservation wage increases, the lender would set a lower repayment ratio according to the recoverability constraint (1.3.4). This in turn alleviates the distortion on labor supply.

These numerical examples suggest that the positive response in the reservation wage under IBR offers a channel that not only increases the agent's welfare but also alleviates the distortionary effect on labor supply. These results highlight that despite the canonical tradeoff between insurance and the incentive to work, IBR is in fact more valuable compared to the fixed repayment contract because of uninsured search risks.

[9] I would like to highlight the distinction between the three channels: the risk channel, the liquidity channel, and the debt overhang channel. Although all three channels raise the reservation wage under IBR, they have divergent welfare implications. The increase in the reservation wage through the risk channel and the liquidity channel is a beneficial response to the correction of the credit and insurance market failures. However, the increase in the reservation wage through the debt overhang channel is a sub-optimal response to the distortion in the relative price of employment and unemployment.

A.4.4 Understanding the Reservation Wage Effect

In this appendix section, I discuss the conditions under which the reservation wage effect of IBR raises the agent's welfare.

In Figure A.2, I provided a numerical example showing that IBR also indirectly increases welfare by increasing the reservation wage (i.e., the reservation wage effect). While this result holds for a wide range of empirically reasonable parameter values, it is not generally true. I now elucidate the economic intuitions.

In subsection A.4.4, I characterize the efficient IBR under the assumption that the reservation wage is observable and contractible. I define the reservation wage set by this contract as the efficient reservation wage. In subsection A.4.4, I show that when labor supply is inelastic, the reservation wage under IBR is below the efficient reservation wage. This explains why the reservation wage effect increases welfare. In subsection A.4.4, I show that when labor supply is sufficiently elastic, the reservation wage under IBR could be above the efficient reservation wage. This is because there is an additional debt overhang channel under IBR that further increases the reservation wage. The implication of this is that the reservation wage effect could reduce welfare. Finally, in subsection A.4.4, I provide several numerical examples and discuss that this counter-intuitive result is not likely to happen in reality. Therefore, I argue that IBR indirectly increases welfare by increasing the reservation wage.

Efficient Reservation Wage

For a certain reservation wage w^*, the agent's welfare under IBR can be expressed recursively:

$$\text{Welfare}_{IBR}(w^*) = F(w^*)\left[u((1-\alpha)\theta,0) + \beta\text{Welfare}_{IBR}(w^*)\right] + \int_{w^*}^{\overline{w}} \frac{u((1-\alpha)wl,l)}{1-\beta}dF(w). \tag{A.4.37}$$

Thus, the agent's welfare is

$$\text{Welfare}_{IBR}(w^*) = \frac{F(w^*)u((1-\alpha)\theta,0)}{1-\beta F(w^*)} + \int_{w^*}^{\overline{w}} \frac{u((1-\alpha)wl,l)}{(1-\beta)[1-\beta F(w^*)]}dF(w). \tag{A.4.38}$$

The agent determines the reservation wage w^*_{IBR} to maximize welfare under IBR:

$$\begin{array}{rlcl} \max\limits_{w^*} & \text{Welfare}_{IBR}(w^*) & & , \\ \text{subject to} & (1-\alpha)wu_1((1-\alpha)wl,l) + u_2((1-\alpha)wl,l) & = & 0, \forall w \in [w^*,\overline{w}], \end{array} \tag{A.4.39}$$

where the constraint is the intra-temporal Euler equation on labor supply, $l(w,\alpha)$. If labor supply is inelastic, the solution to problem (A.4.39) gives the indifference equation (1.3.5). Conditional on the reservation wage that solves problem (A.4.39), the lender sets the repayment ratio α according to the recoverability constraint:

$$\frac{F(w^*_{IBR})\alpha\theta}{1-\beta F(w^*_{IBR})} + \int_{w^*_{IBR}}^{\overline{w}} \frac{\alpha w l(w,\alpha)}{(1-\beta)[1-\beta F(w^*_{IBR})]} dF(w) = \frac{S}{\beta}. \tag{A.4.40}$$

The reservation wage w^*_{IBR} is inefficient because the agent's reservation wage generates an externality on the lender's revenue. The agent would be better off if she can internalize this effect when choosing the reservation wage. For the discussion of the reservation wage effect, it is useful to introduce the efficient reservation wage as a benchmark.

Definition 6. *The efficient reservation wage, w^*_{EFI}, is the reservation wage that the lender would set under IBR if the reservation wage is observable and contractible, i.e., w^*_{EFI} solves:*

$$\begin{aligned} \max_{w^*} \quad & \text{Welfare}_{IBR}(w^*) \\ \text{subject to} \quad & (1-\alpha)wu_1((1-\alpha)wl,l) + u_2((1-\alpha)wl,l) &=& \; 0, \forall w \in [w^*,\overline{w}], \\ \frac{F(w^*)\alpha\theta}{1-\beta F(w^*)} + & \int_{w^*}^{\overline{w}} \frac{\alpha w l(w,\alpha)}{(1-\beta)[1-\beta F(w^*)]} dF(w) &=& \; \frac{S}{\beta}, \end{aligned} \tag{A.4.41}$$

where the first constraint is the intra-temporal Euler equation on labor supply, and the second constraint is the lender's recoverability constraint.

Clearly, w^*_{EFI} is different from w^*_{IBR} as the agent takes into account the lender's recoverability constraint when setting the reservation wage.

Inelastic Labor Supply

To provide some intuitions, I begin by discussing thė reservation wage effect when the agent has inelastic labor supply.

Suppose that the agent has CRRA utility, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$. Denote λ as the Lagrangian multiplier for the recoverability constraint in problem (A.4.41). The shadow price λ is negative as the agent's welfare decreases when debt S marginally increases. The first order condition that determines the efficient reservation wage is:

$$\begin{aligned} &\frac{[(1-\alpha)\theta]^{1-\gamma}}{[1-\beta F(w^*_{EFI})](1-\gamma)} + \int_{w^*_{EFI}}^{\overline{w}} \frac{\beta[(1-\alpha)w]^{1-\gamma}}{(1-\beta)[1-\beta F(w^*_{EFI})](1-\gamma)} dF(w) - \frac{[(1-\alpha)w^*_{EFI}]^{1-\gamma}}{(1-\beta)(1-\gamma)} \\ &= \lambda \frac{1-\beta F(w^*_{EFI})}{f(w^*_{EFI})} \alpha I'(w^*_{EFI}), \end{aligned} \tag{A.4.42}$$

where $I'(w^*_{EFI})$ is the first derivative of expected income with respect to the reservation wage, characterized by equation (A.4.20). The RHS of equation (A.4.42) captures the effect of the reservation wage on expected repayment.

Define

$$g(x) = \frac{[(1-\alpha)\theta]^{1-\gamma}}{1-\gamma} + \int_{x}^{\overline{w}} \frac{\beta[(1-\alpha)w]^{1-\gamma}}{(1-\beta)(1-\gamma)} dF(w) - \frac{[1-\beta F(x)][(1-\alpha)x]^{1-\gamma}}{(1-\beta)(1-\gamma)}, \tag{A.4.43}$$

$$h(x) = \lambda \frac{[1-\beta F(x)]^2}{f(x)} \alpha I'(x). \tag{A.4.44}$$

Equation (A.4.42) can be rewritten as

$$g(w^*_{EFI}) - h(w^*_{EFI}) = 0. \tag{A.4.45}$$

In fact, $g(x) = 0$ coincides with the indifference equation (1.3.5), thus the solution to $g(x) = 0$ gives the reservation wage under IBR, i.e., $g(w^*_{IBR}) = 0$.

The proof of Proposition 2 indicates that with CRRA utility $w^*_{IBR} = w^*|_{s=0}$. When the agent is risk averse, according to Proposition 9, $w^*|_{s=0} < \hat{w}$ and $I'(w^*_{IBR}) > 0$. With $\lambda < 0$, we have $h(w^*_{IBR}) < 0$. Thus

$$g(w^*_{IBR}) - h(w^*_{IBR}) > 0. \tag{A.4.46}$$

Take the first derivative for $g(x)$ and $h(x)$, we obtain:

$$g'(x) = -\frac{(1-\alpha)[1-\beta F(x)]}{1-\beta}[(1-\alpha)x]^{-\gamma} < 0, \tag{A.4.47}$$

$$h'(x) = -\frac{\lambda\alpha}{1-\beta}[1-\beta F(x)] > 0. \tag{A.4.48}$$

Thus

$$g'(x) - h'(x) < 0. \tag{A.4.49}$$

Equations (A.4.45-A.4.49) imply $w^*_{IBR} < w^*_{EFI}$. Therefore, the agent's efficient reservation wage is higher than the reservation wage under IBR when labor supply is inelastic. Intuitively, this is because the efficient reservation wage internalizes the choice of the reservation wage on expected repayment. By increasing the reservation wage, the agent could increase the lender's revenue, motivating the lender to set a smaller repayment ratio α given the recoverability constraint, which in turn increases welfare. The efficient reservation wage is not incentive compatible because facing a lower repayment ratio ex-post, the agent would have the incentive to reduce the reservation wage in order to take fewer risks and increase her utility. As a result, the lender would take a loss.

What this implies is that IBR indirectly raises the agent's welfare by increasing the reservation wage. If we restrict the agent from choosing a higher reservation wage, as in the experiment of Figure A.2, the agent's welfare would be lowered because the reservation wage is further away from the efficient one.

Elastic Labor Supply

Now I turn to the discussion of the reservation wage effect when the agent has elastic labor supply. I show that with elastic labor supply, there is an additional channel that increases the reservation wage under IBR. As a result, the agent could possibly choose a reservation wage higher than the efficient reservation wage.

To illustrate the economic channel, I begin my analysis with risk-neutral agents. Suppose that the agent has quasi-linear utility $u(c,l) = c - \frac{l^{1+\sigma}}{1+\sigma}$. Using equations (1.3.3) and (1.3.5), w^*_{FIX} and w^*_{IBR} can be derived from:

$$\frac{\sigma}{1+\sigma}\left(w^*_{FIX}\right)^{\frac{1+\sigma}{\sigma}} = \theta + \frac{\sigma}{1+\sigma}\frac{\beta}{1-\beta}\int_{w^*_{FIX}}^{\overline{w}}\left[w^{\frac{1+\sigma}{\sigma}} - \left(w^*_{FIX}\right)^{\frac{1+\sigma}{\sigma}}\right]dF(w), \tag{A.4.50}$$

$$\frac{\sigma}{1+\sigma}\left(w^*_{IBR}\right)^{\frac{1+\sigma}{\sigma}} = \underbrace{(1-\alpha)^{-\frac{1}{\sigma}}}_{\text{debt overhang channel}} \theta + \frac{\sigma}{1+\sigma}\frac{\beta}{1-\beta}\int_{w^*_{IBR}}^{\overline{w}}\left[w^{\frac{1+\sigma}{\sigma}} - \left(w^*_{IBR}\right)^{\frac{1+\sigma}{\sigma}}\right]dF(w). \tag{A.4.51}$$

The only difference between the two equations lies in the term $(1-\alpha)^{-\frac{1}{\sigma}} > 1$, due to the response in labor supply when the agent is employed and repaying debt under IBR. As a result, the reservation wage under IBR is higher than that under the fixed repayment contract when $\sigma < \infty$. Note that Proposition 9 implies that under the fixed repayment contract, the risk-neutral agent sets the reservation wage equal to $\hat{w}$, which already maximizes expected income. However, IBR further raises the reservation wage, which reduces expected income (before repayment). Intuitively, the agent chooses to set a higher reservation wage to avoid employment because supplying labor is costly. Therefore, elastic labor supply generates an additional force that increases the reservation wage under IBR. This channel is exposed starkly when the agent is risk neutral, because with inelastic labor supply ($\sigma = \infty$), the two reservation wages are equalized, $w^*_{FIX} = w^*_{IBR} = \hat{w}$, due to the absence of the risk channel and the liquidity channel discussed in subsection 1.3.2.

I name the effect on the reservation wage introduced by the elastic labor supply as the *debt overhang channel* of IBR.[10] I would like to highlight the distinction between the three channels: the debt overhang channel, the risk channel, and the liquidity channel. Although all three channels raise the reservation wage under IBR, they have divergent welfare implications. The increase in the reservation wage through the risk channel and the liquidity channel is a beneficial response to the correction of the credit and insurance market failures. However, the increase in the reservation wage through the debt overhang channel is a sub-optimal response to the distortion in the relative price of employment and unemployment.[11] Because the reservation wage controls the extensive

[10]This channel is related to the moral hazard problem in the labor market associated with debt collection policies [186].

[11]Due to the response in labor supply, IBR essentially subsidizes unemployment by reducing income during employment by a proportion, $1-(1-\alpha)^{\frac{1+\sigma}{\sigma}}$, larger than the proportional reduction during unemployment, α.

participation margin of labor supply, we can interpret this result in an alternative way: IBR generates a moral hazard problem that reduces labor supply on the intensive margin. This in turn generates a moral hazard problem that reduces labor supply on the extensive margin, i.e., increasing the reservation wage.

The discussion above suggests that the reservation wage under IBR could be larger than the efficient reservation wage when the risk-neutral agent has elastic labor supply. To see this, I substitute the utility function into equation (A.4.38) and obtain the agent's welfare:

$$(1-\alpha)\left[\frac{F(w^*)}{1-\beta F(w^*)}\theta + \frac{\sigma}{1+\sigma}\frac{(1-\alpha)^{\frac{1}{\sigma}}}{(1-\beta)[1-\beta F(w^*)]}\int_{w^*}^{\overline{w}} w^{\frac{1+\sigma}{\sigma}} dF(w)\right]. \tag{A.4.52}$$

By substituting the expression for labor supply, $l = [(1-\alpha)w]^{1/\sigma}$, into equation (A.4.40), we obtain the recoverability constraint:

$$\alpha\left[\frac{F(w^*)}{1-\beta F(w^*)}\theta + \frac{(1-\alpha)^{\frac{1}{\sigma}}}{(1-\beta)[1-\beta F(w^*)]}\int_{w^*}^{\overline{w}} w^{\frac{1+\sigma}{\sigma}} dF(w)\right] = \frac{S}{\beta}. \tag{A.4.53}$$

The reservation wage w^*_{IBR} is chosen to maximize the objective function (A.4.52) with the repayment ratio α set separately according to equation (A.4.53). The efficient reservation wage w^*_{EFI} is chosen to maximize the objective function (A.4.52) subject to the constraint (A.4.53). It is clear that when $\sigma = \infty$, the reservation wage that maximizes the objective function (A.4.52) also simultaneously maximizes expected repayment, i.e., the LHS of equation (A.4.53). This implies that the first-order derivative of equation (A.4.53) with respect to the reservation wage is equal to zero. Therefore, the unconstrained maximization problem yields the same solution as the constrained maximization problem, i.e., $w^*_{IBR} = w^*_{EFI}$. Intuitively, this is saying that the risk-neutral agent would choose the efficient reservation wage that maximizes expected repayment when labor supply is inelastic.

However, when $\sigma < \infty$, the terms inside the bracket of (A.4.52) differ from those of (A.4.53) as less weight is given for the value of employment ($\frac{\sigma}{1+\sigma} < 1$).[12] This suggests that, compared with the efficient reservation wage w^*_{EFI} that solves the constrained maximization problem, the unconstrained maximization would set a relatively higher reservation wage w^*_{IBR} to avoid employment.

The analysis of a risk-neutral agent presents the stark result that the reservation wage under IBR is always higher than the efficient reservation wage as long as labor supply is elastic. When the agent is risk averse, the risk and liquidity channels of debt repayment would reduce the reservation wage. Therefore, whether the reservation wage under IBR is higher than the efficient

[12]Intuitively, the agent puts less weight on the value of employment in the objective function because supplying labor generates a dis-utility equaling to $\frac{1}{1+\sigma}$ of the agent's wage income. Mathematically, the efficient reservation wage that solves the constrained maximization problem can be thought of as the average of the reservation wage maximizing (A.4.52) and the one maximizing (A.4.52) weighted by the Lagrangian multiplier. Due to the existence of the term $\frac{\sigma}{1+\sigma}$ in (A.4.52), the reservation wage that maximizes (A.4.52) is higher than the one maximizing (A.4.53).

one depends on which channel dominates. Intuitively, the strength of the debt overhang channel increases with the elasticity of labor supply. Therefore, when labor supply is sufficiently elastic, the debt-overhang channel would dominate and the reservation wage under IBR would be inefficiently high.[13]

The implication of the debt-overhang channel is that the agent could be better off if the reservation wage is restricted at some lower value when being provided with IBR. Therefore, it is not generally true that IBR also indirectly raises welfare by increasing the reservation wage.

Numerical Examples and Discussions

In this subsection, I provide numerical examples by setting different values for the elasticity of labor supply. The goal of this simple exercise is to show that for empirically reasonable values of risk aversion and the elasticity of labor supply, IBR increases welfare by raising the reservation wage.

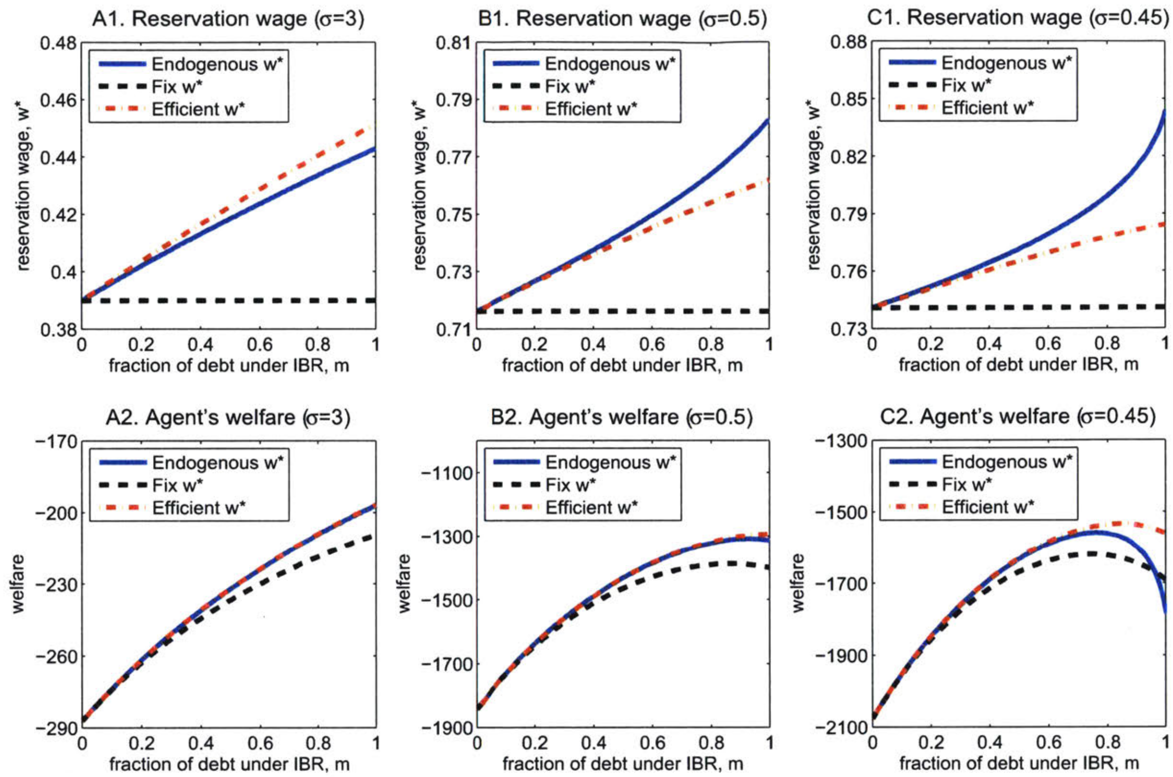
In Figure A.3, I report the agent's reservation wage and welfare for different values of parameter σ. In each panel, I vary the fraction of debt under IBR and plot the outcome of interest when the reservation wage is endogenous, fixed at its value under the fixed repayment contract (i.e., $m = 0$), or efficient.

In panels A, I set $\sigma = 3$ to consider an empirically reasonable elasticity of labor supply, 0.33, according to Keane [147]. Panel A2 shows that welfare increases when a larger fraction of debt is made income contingent. It is clear that the inefficiency due to reservation wages is minimal as the welfare with endogenous reservation wages (blue solid line) is almost on top of that under the efficient contract (red dash-dotted line). Importantly, allowing the reservation wage to respond increases the agent's welfare relative to fixing the reservation wage at the beginning (black dashed line). This is because the reservation wage under the fixed repayment contract is too low compared to the efficient reservation wage. Increasing the fraction of income contingency raises the reservation wage, closing the gap to the efficient one (see panel A1) and lowering the repayment ratio.

In panels B, I dramatically increase the elasticity of labor supply to 2 by setting $\sigma = 0.5$. Similar to the result of Figure A.2, welfare first increases and then decreases due to the increasing distortion of income contingency on labor supply (see panel B2). The welfare with endogenous reservation wages is still higher than that with fixed reservation wages, but by contrast, the endogenous reservation wage is above the efficient one (see panel B1).

In panels C, I further increase the elasticity of labor supply to 2.22 by setting $\sigma = 0.45$. I obtain the result in which the debt-overhang channel dominates, and increasing the fraction of income contingency indirectly reduces welfare by increasing the reservation wage. Panel C2 shows that the agent's welfare would be higher if the reservation wage is fixed at the beginning. As shown

[13] For example, in the extreme case with $\sigma = 0$, the second term in equation (A.4.52) vanishes to zero, and thus $w^*_{IBR} = \overline{w} > w^*_{EFI}$.



Note: This figure illustrates the reservation wage effect for different elasticities of labor supply. In panel A1, B1, and C1, the blue solid line plots the agent's reservation wage when the reservation wage is allowed to increase as a larger fraction of debt is made income contingent. The black dashed line plots the agent's reservation wage when the reservation wage is fixed at the initial value under the pure fixed repayment contract. The red dash-dotted line plots the agent's efficient reservation wage. The corresponding welfare is plotted in panel A2, B2, and C2. The elasticity of labor supply is 0.33 ($\sigma = 3$), 2 ($\sigma = 0.5$), and 2.22 ($\sigma = 0.45$) in panels A, B, and C. The figure is plotted using the GHH utility, $u(c,l) = [c - \phi l^{1+\sigma}/(1+\sigma)]^{1-\gamma}/(1-\gamma)$ and the beta distribution of wage offers, $Beta(a,b)$, with parameter values: $a = 2$, $b = 4$, $\gamma = 3$, $\theta = 0.1$, $\overline{w} = 1.1$, $\beta = 0.96$, $S = 1$, $\phi = 1$.

Figure A.3: A numerical illustration of the reservation wage effect for different elasticities of labor supply.

in panel C1, this is essentially caused by the sharp increase in the reservation wage relative to the efficient one when a larger fraction of debt is made income contingent.

In sum, IBR increases the agent's welfare by directly providing insurance. The insurance leads to a higher reservation wage, which may or may not increase the agent's welfare. The key parameters governing whether a higher reservation wage is beneficial are the degree of risk aversion and the elasticity of labor supply. All else equal, a more risk-averse agent sets a lower reservation wage relative to the efficient one under the fixed repayment contract. Thus increasing the reservation wage by providing insurance increases welfare. A larger elasticity of labor supply intensifies the debt overhang channel. Thus when the incentive to work is distorted by IBR, it is more likely to result in a reservation wage too high compared to the efficient one. In this case, by committing to a lower reservation wage, the agent could increase her expected repayment, inducing the lender to set a lower repayment ratio, which consequently increases welfare. However, such commitment is not incentive compatible because ex-post a lower repayment ratio generates a steeper wage offer distribution due to the elastic labor supply. This motivates the agent to stay unemployed longer by setting a higher reservation wage, and the lender would take a loss on debt collection.

Despite the theoretical possibility, in reality, it is plausible that IBR indirectly increases welfare by increasing the reservation wage. This is due to two reasons. First, as suggested by the numerical examples in Figure A.3, a higher reservation wage reduces welfare only when the elasticity of labor supply is about two, while a consensus empirical estimate is usually below one. Second, the theoretical possibility roots from the inefficiency in IBR, which is designed to allow the lender to collect all debt in expectation. In other words, if the repayment ratio is fixed, instead of being varied with the endogenous reservation wage, then the inefficiency would disappear by construction. This is the case in reality, as the government is willing to take a loss by offering debt forgiveness for federal student loans.

A.4.5 Optimal Repayment Contract

In theory, IBR is not the most efficient way to provide insurance because the repayment ratio is constant regardless of the level of income. In this subsection, I characterize the optimal repayment contract under the assumption that the reservation wage is not contractible. I show that the existence of search risks sets up the optimal contract that also considers the level of reservation wages. The implication is that the lender should provide more insurance in an economy with search risks, because this would increase the reservation wage. Therefore, IBR although not constrained efficient, is designed in the spirit of the optimal repayment contract as it both provides insurance and increases the reservation wage.

To gain some insight, let us begin with the first-best contract. The first-best contract not only provides full insurance against search risks but also sets the reservation wage to $\hat{w}$ to maximize expected income. When labor supply is inelastic, the first-best contract is also incentive compatible because perfect insurance makes the agent indifferent about the reservation wage. This suggests that in contrast to a model without search risks, insurance is more desirable in my model because income risks are controlled by the agent's endogenous job search decisions. The full insurance provided by the first-best contract not only directly increases welfare through consumption smoothing; but also indirectly increases welfare by making a higher reservation wage incentive compatible.

When labor supply is elastic, the first-best contract is not incentive compatible because supplying labor generates disutility. The second-best contract solves the problem in which the lender chooses a nonlinear repayment schedule $\alpha(z)$ conditional on earnings $z = wl$ subject to the recoverability constraint and the agent's incentive compatibility constraints on labor supply and the reservation wage. This problem is more complicated compared to the optimal income taxation problem solved by Mirrlees [182] as there is an additional incentive compatibility constraint on the reservation wage.

Below I first show that when there is no job search (i.e., the reservation wage is fixed at $w^* = 0$), the mathematical problem is exactly the same as Mirrlees [182]'s problem with a utilitarian social welfare function. I then show that my problem is different due to the introduction of endogenous job search decisions. I formulate the optimal contracting problem and use the perturbation approach inspired by Saez [208] to elucidate the economic channels.

Without Job Search

When the reservation wage w^* is set to be 0, the agent accepts all wage offers drawn from $F(w)$ in the first period. Therefore, the agent's life-time utility conditional on receiving a wage offer w is

$$W(w) = \frac{u(w,l)}{1-\beta}, \tag{A.4.54}$$

where l is the labor supply that satisfies the first-order condition.

To maximize the agent's expected life-time utility, the lender chooses an optimal nonlinear repayment schedule $\alpha^{SB}(z)$, as a function of the agent's earnings $z = wl$ to collect debt S/β. The nonlinear repayment schedule is not written on wage rates because wage rates are not observable or contractible.[14] The intercept $\alpha^{SB}(0)$ can be thought of as a lump-sum repayment or subsidy that is applied to any realization of earnings. The marginal repayment rate is $\alpha^{SB}(z)'$.

This problem is exactly the same as Mirrlees [182] if we interpret it in the following way. There is a continuum of agents with different skills w and homogeneous utility functions $\frac{u(c,l)}{1-\beta}$. They work in a static economy and optimally choose their labor supply l in the tax system. The government values a utilitarian social welfare function and optimally designs a nonlinear tax schedule $\alpha^{SB}(z)$ in terms of earnings z to maximize social welfare conditional on collecting S/β revenue.

The problem is solved by Mirrlees [182] by applying an optimal control approach on direct truth-telling mechanisms. The advantage of this approach comes from its rigorousness to obtain the technical conditions.[15] However, the derived formula is not useful to elucidate the economic intuitions underlying the optimal contract.

With Job Search

Now I consider the optimal contracting problem with endogenous job search decisions as specified in section 1.3. The only departure from the problem of Mirrlees [182] is that the agent chooses a reservation wage below which the wage offer is rejected. Therefore, in this problem, the types

[14] If wage rates are contractible, then the first-best allocation is attainable because labor supply would not be distorted by repayment contracts. It is reasonable to assume that wage rates are unobservable because if they are observable, then labor supply is also observable from wage income. But this contradicts with the assumption made in the optimal income taxation literature.

[15] For example, in order to have the local incentive-compatibility constraint being sufficient, the problem is required to satisfy the Spence-Mirrlees single crossing condition and the monotonicity condition.

of agents in the problem of Mirrlees [182] are restricted to a mass point with earnings θ with probability $F(w^*)$ and a continuum of types in $[w^*, \overline{w}]$ with density $\frac{f(w)}{1-F(w^*)}$, where w^* is chosen by the agent to maximize her welfare.

Facing any nonlinear repayment contract $\alpha(z)$ in terms of earnings z, the agent makes two decisions to maximize her welfare. First, the agent chooses a reservation wage w^*. Second, conditional on accepting the wage offer w, the agent chooses her labor supply l. Therefore, the resulting distribution of earnings $H(z)$ depends both on the exogenous wage offer distribution $F(w)$ and the repayment schedule $\alpha(z)$.

Below, I use a perturbation approach inspired by Saez [208] to characterize the shape of the optimal repayment contract $\alpha^{SB}(z)$. For tractability, I make the following assumptions.

Assumption 1. *Earnings z and utility $u(z - \alpha^{SB}(z), l^{SB}(z))$ weakly increase with wage rates w under the optimal repayment contract $\alpha^{SB}(z)$.*

Assumption 2. *The optimal repayment contract $\alpha^{SB}(z)$ is twice differentiable for all z.*

Assumption 1 is saying that the agent earns more and enjoys higher welfare at jobs with higher wage rates. This is intuitively reasonable given that the monotonicity condition in the mechanism design problem of Mirrlees [182] requires net earnings $z - \alpha^{SB}(z)$ to be weakly increasing in w. This assumption ensures that there is an injective function under the optimal contract $\alpha^{SB}(z)$, $w \mapsto z = q(w)$. Thus I denote z^* as the earnings corresponding to the reservation wage offer w^*, i.e., $z^* = q(w^*)$.

Assumption 2 comes from Saez [208]. This assumption has additional meaning in the problem I solve because it also restricts the specification of contract off the equilibrium, i.e., for $z \in (\theta, z^*)$. In general, because the agent rejects the wage offer whenever the resulting earnings are below z^*, there exist infinite numbers of optimal repayment contracts in my problem, and some of them could have a discontinuous jump at z^*.[16] This assumption ensures that the reservation wage is derived from a first-order condition instead of being a corner solution. That is, when the reservation wage is slightly changed, the change in the agent's welfare is of second order.

Denote $\lambda < 0$ as the Lagrangian multiplier associated with the lender's recoverability constraint,

$$\frac{H(z^*)}{1-\beta H(z^*)}\alpha^{SB}(\theta) + \frac{1}{(1-\beta)[1-\beta H(z^*)]}\int_{z^*}^{\infty} \alpha^{SB}(z)dH(z) = \frac{S}{\beta}. \tag{A.4.55}$$

The multiplier λ is also the shadow value measuring the change in the agent's welfare when the amount of debt marginally increases.[17] Denote $g(z) > 0$ as the marginal value of consumption for the agent with earnings z under the optimal repayment contract, expressed in terms of the

[16] For example, given the optimal contract $\alpha^{SB}(z)$. We can specify $\tilde{\alpha}^{SB}(z)$ such that $\tilde{\alpha}^{SB}(z) = \alpha^{SB}(z)$ *for* $z \geq z^*$ and $z - \tilde{\alpha}^{SB}(z) = \theta - \alpha^{SB}(\theta)$ *for* $z < z^*$. Under $\tilde{\alpha}^{SB}(z)$, the net earnings are flat up to the reservation earnings z^*, and there is a discontinuous jump in net earnings at z^*. The contract $\tilde{\alpha}^{SB}(z)$ is incentive compatible because the agent has no incentive to change her reservation earnings z^* as reducing this lowers her utility more than what would be under $\alpha^{SB}(z)$. Moreover, $\tilde{\alpha}^{SB}(z)$ also satisfies the lender's recoverability constraint so it is an optimal contract.

[17] The negative of λ corresponds to the social value of public funds defined by Saez [208].

shadow cost of debt $(-\lambda)$, i.e.,

$$g(z) = \frac{u_1(z - \alpha^{SB}(z), l^{SB}(z))}{-\lambda}, \tag{A.4.56}$$

where $l^{SB}(z)$ corresponds to the labor supply at earnings z under the optimal contract $\alpha^{SB}(z)$.

I follow Saez [208] and consider a small perturbation around the optimal repayment schedule $\alpha^{SB}(z)$. Suppose that the marginal repayment rate is increased by $d\alpha$ for earnings between z and $z + dz$, where $z \geq z^*$. This would generate the following effects on expected repayment R, defined as:

$$R = \frac{H(z^*)}{1 - \beta H(z^*)}\alpha^{SB}(\theta) + \frac{1}{(1-\beta)[1 - \beta H(z^*)]}\int_{z^*}^{\infty} \alpha^{SB}(z) dH(z). \tag{A.4.57}$$

Mechanical effect The agent pays $d\alpha dz$ more when her earnings are above z, with probability $1 - H(z)$. Thus expected repayment increases by

$$M = \frac{1 - H(z)}{(1-\beta)[1 - \beta H(z^*)]} d\alpha dz. \tag{A.4.58}$$

Elasticity effect The increase in the marginal repayment rate distorts labor supply when the agent's earnings are between z and $z + dz$, which consequently affects expected repayment. The change in earnings is caused by two effects. First, there is a direct effect due to the increase in $d\alpha$. Second, there is an indirect effect as the agent would face a different marginal repayment rate when her earnings are changed by the direct effect.

As noted by Saez [208], the direct effect can be decomposed into two parts: an overall uncompensated increase in the marginal rate and an overall increase in virtual income. Therefore, the relevant one that determines the behavioral response is the Hicksian (compensated) elasticity of earnings, which is defined as

$$\zeta^c(z) = \frac{1 - \alpha^{SB}(z)'}{z} \frac{\partial z}{\partial(1 - \alpha^{SB}(z)')}\bigg|_u. \tag{A.4.59}$$

Suppose that the two effects result in an earnings change by Δ, then the direct effect is $-\zeta^c(z) z \frac{d\alpha}{1-\alpha^{SB}(z)'}$, and the indirect effect is $-\zeta^c(z) z \frac{\Delta \alpha^{SB}(z)''}{1-\alpha^{SB}(z)'}$. Hence,

$$\Delta = -\zeta^c(z) z \frac{d\alpha}{1 - \alpha^{SB}(z)'} - \zeta^c(z) z \frac{\Delta \alpha^{SB}(z)''}{1 - \alpha^{SB}(z)'}. \tag{A.4.60}$$

This implies that

$$\Delta = -\zeta^c(z) z \frac{d\alpha}{1 - \alpha^{SB}(z)' + \zeta^c(z) z \alpha^{SB}(z)''}. \tag{A.4.61}$$

Following Saez [208], I assume that $1 - \alpha^{SB}(z)' + \zeta^c(z) z \alpha^{SB}(z)'' > 0$ so that bunching of types does not occur. The elasticity effect on expected repayment is

$$E = \frac{\Delta\alpha^{SB}(z)'h(z)dz}{(1-\beta)[1-\beta H(z^*)]}$$
$$= -\frac{\zeta^c(z)z\alpha^{SB}(z)'}{1-\alpha^{SB}(z)'+\zeta^c(z)z\alpha^{SB}(z)''}\frac{h(z)}{(1-\beta)[1-\beta H(z^*)]}d\alpha dz. \tag{A.4.62}$$

Income effect If the agent accepts a wage offer generating earnings above $z+dz$, her earnings are reduced by $d\alpha dz$ due to the higher marginal rate between z and $z+dz$. This would generate an income effect that induces the agent to work more. As a result, for any $x > z+dz$, earnings increase by $\Delta(x)$, which in turn increases expected repayment. The earnings response $\Delta(x)$ is due to two effects. First, there is a direct effect due to the increase in marginal rate $d\alpha$ between z and $z+dz$. Second, there is an indirect effect due to the change in marginal rates caused by the shift in earnings.

Let $\eta(z) \leq 0$ denote the income effect and $\zeta^u(z)$ denote the Marshallian (uncompensated) elasticity of earnings at earnings z, thus the income effect is derived by the Slutsky equation,

$$\zeta^u(z) = \frac{1-\alpha^{SB}(z)'}{z}\frac{\partial z}{\partial(1-\alpha^{SB}(z)')}; \tag{A.4.63}$$
$$\eta(z) = \zeta^u(z) - \zeta^c(z). \tag{A.4.64}$$

Therefore, the direct effect is $-\frac{\eta(x)d\alpha dz}{1-\alpha^{SB}(x)'}$ and the indirect effect is $-\zeta^c(x)x\frac{\alpha^{SB}(x)''\Delta(x)}{1-\alpha^{SB}(x)'}$, and the change in earnings is

$$\Delta(x) = -\frac{\eta(x)d\alpha dz}{1-\alpha^{SB}(x)'} - \zeta^c(x)x\frac{\alpha^{SB}(x)''\Delta(x)}{1-\alpha^{SB}(x)'}, \tag{A.4.65}$$

which implies

$$\Delta(x) = -\eta(x)\frac{d\alpha dz}{1-\alpha^{SB}(x)'+x\zeta^c(x)\alpha^{SB}(x)''}. \tag{A.4.66}$$

The total income effect on expected repayment is

$$I = -\frac{d\alpha dz}{(1-\beta)[1-\beta H(z^*)]}\int_z^\infty \eta(x)\frac{\alpha^{SB}(x)'}{1-\alpha^{SB}(x)'+x\zeta^c(x)\alpha^{SB}(x)''}h(x)dx. \tag{A.4.67}$$

Reservation wage effect There is a fourth effect on expected repayment due to the change in reservation earnings, which is not in the problem of Mirrlees [182]. The reservation earnings are determined by the following indifference equation:

$$\frac{u(z^*-\alpha^{SB}(z^*),l^{SB}(z^*))}{1-\beta} = \frac{u(\theta-\alpha^{SB}(\theta),0)}{1-\beta H(z^*)} + \frac{\beta}{(1-\beta)[1-\beta H(z^*)]}\int_{z^*}^\infty u(x-\alpha^{SB}(x),l^{SB}(x))dH(x), \tag{A.4.68}$$

where the LHS of this equation represents the value of being employed at the reservation earnings z^*, and the RHS represents the value of staying unemployed. Assumption 2 ensures that the reservation earnings also satisfy the first-order condition. Rearranging it:

$$1 = \frac{\beta}{1-\beta}\int_{z^*}^{\infty}\frac{u(x-\alpha^{SB}(x),l^{SB}(x)) - u(z^*-\alpha^{SB}(z^*),l^{SB}(z^*))}{u(z^*-\alpha^{SB}(z^*),l^{SB}(z^*)) - u(\theta-\alpha^{SB}(\theta),0)}dH(x). \tag{A.4.69}$$

Assumption 1 ensures that the integrand is non-negative and decreasing in z^*. The integration is executed from z^* to infinity, thus the RHS of equation (A.4.69) decreases with z^*. The increase in the marginal repayment rate $d\alpha$ between z and $z+dz$ reduces $u(x-\alpha^{SB}(x),l^{SB}(x))$ for all $x>z$, thus lowering the RHS of equation (A.4.69). This implies that the reservation earnings z^* would decrease.

For $x>z$, the change $d\alpha$ would change $u(x-\alpha^{SB}(x),l^{SB}(x))$ by

$$\begin{aligned} du(x) &= -u_1(x-\alpha^{SB}(x),l^{SB}(x))d\alpha dz \\ &= g(x)\lambda d\alpha dz. \end{aligned} \tag{A.4.70}$$

Note that the elasticity effect and the income effect discussed above indicate that labor supply $l^{SB}(x)$ would also change due to the change $d\alpha$, but the Envelope Theorem implies that such a change does not have a first-order effect on utility. Differentiating equation (A.4.68) and substituting (A.4.70), we obtain

$$dz^* = d\alpha dz\frac{\beta\lambda}{[1-\beta H(z^*)]u_z(z^*)}\int_z^{\infty} g(x)dH(x), \tag{A.4.71}$$

where $u_z(z) = \frac{du(z-\alpha^{SB}(z),l^{SB}(z))}{dz}$ denotes the marginal change in utility due to a marginal change in earnings at z under the optimal contract $\alpha^{SB}(z)$.

The change in reservation earnings dz^* does not affect the agent's welfare due to the envelope condition from Assumption 2. However, it affects expected repayment R determined by equation (A.4.57). Define ζ^{z^*} as the elasticity of expected repayment with respect to the reservation earnings,

$$\zeta^{z^*} = \frac{\partial R/R}{\partial z^*/z^*}. \tag{A.4.72}$$

Differentiating (A.4.57), we obtain

$$\begin{aligned} \zeta^{z^*} &= \frac{\beta R + \alpha^{SB}(\theta) - \dfrac{\alpha^{SB}(z^*)}{1-\beta}}{[1-\beta H(z^*)]R}z^*h(z^*) \\ &= \frac{\beta S + \beta\alpha^{SB}(\theta) - \dfrac{\beta\alpha^{SB}(z^*)}{1-\beta}}{[1-\beta H(z^*)]S}z^*h(z^*), \end{aligned} \tag{A.4.73}$$

where the second equation is obtained by substituting $R = S/\beta$.

In general, ζ^{z^*} could be positive or negative. The discussion in Appendix A.4.4 suggests that $\zeta^{z^*} > 0$ for empirically reasonable elasticities of labor supply. Therefore, higher reservation earnings increase expected repayment. Using equations (A.4.71) and (A.4.73), we obtain the reservation wage effect on expected repayment:

$$\begin{aligned} RW &= \frac{dz^*}{z^*}\zeta^{z^*}R \\ &= d\alpha dz \frac{S\lambda\zeta^{z^*}}{[1-\beta H(z^*)]u_z(z^*)z^*}\int_z^\infty g(x)dH(x). \end{aligned} \tag{A.4.74}$$

Deriving the Optimal Contract During Employment The small perturbation around the optimal contract should have no first-order effect on welfare. Therefore, the sum of the four effects, M, E, I, and RW, multiplied by the shadow cost of debt $(-\lambda)$ should be equal to the agent's expected welfare loss when earnings are above z. The agent's welfare under $\alpha^{SB}(z)$ is

$$\text{Welfare}_{SB} = \frac{H(z^*)u(\theta-\alpha^{SB}(\theta),0)}{1-\beta H(z^*)} + \int_{z^*}^\infty \frac{u(x-\alpha^{SB}(x), l^{SB}(x))}{(1-\beta)[1-\beta H(z^*)]}dH(x) \tag{A.4.75}$$

The expected welfare loss is

$$\begin{aligned} WL &= d\alpha dz \int_z^\infty \frac{u_1(x-\alpha^{SB}(x), l^{SB}(x))}{(1-\beta)[1-\beta H(z^*)]}dH(x) \\ &= d\alpha dz \int_z^\infty \frac{-\lambda g(x)}{(1-\beta)[1-\beta H(z^*)]}dH(x). \end{aligned} \tag{A.4.76}$$

Again, the Envelope Theorem implies that the change in labor supply has a second-order effect on welfare. At the optimum,

$$WL = -\lambda(M+E+I+RW), \tag{A.4.77}$$

which implies

$$\begin{aligned} \underbrace{\int_z^\infty g(x)dH(x)}_{\text{direct welfare loss}} = \underbrace{1-H(z)}_{\text{mechanical effect}} & \underbrace{-\frac{z\zeta^c(z)\alpha^{SB}(z)'}{1-\alpha^{SB}(z)'+z\zeta^c(z)\alpha^{SB}(z)''}h(z)}_{\text{elasticity effect}} \\ & \underbrace{-\int_z^\infty \eta(x)\frac{\alpha^{SB}(x)'}{1-\alpha^{SB}(x)'+x\zeta^c(x)\alpha^{SB}(x)''}dH(x)}_{\text{income effect}} \\ & \underbrace{+\frac{S(1-\beta)\lambda\zeta^{z^*}}{u_z(z^*)z^*}\int_z^\infty g(x)dH(x)}_{\text{reservation wage effect}}. \end{aligned} \tag{A.4.78}$$

This equation implicitly determines the optimal contract $\alpha^{SB}(z)$. It is different from the one derived by Saez [208] due to the existence of the reservation wage effect. As a result, it does not admit an explicit solution for $\alpha^{SB}(z)$ because the elasticity of earnings with respect to the reservation earnings, ζ^{z^*}, is a function of $\alpha^{SB}(z)$.

To gain some intuitions, consider the case with inelastic labor supply, which implies that there is no elasticity effect or income effect in equation (A.4.78). If there are no endogenous search decisions, the reservation wage effect is also absent. Then the optimal contract requires $\int_z^\infty g(x)dH(x) = 1 - H(z)$ for all $z > z^*$. This happens only when $g(z) = 1, \forall z > z^*$, suggesting perfect insurance against earnings risks.

When there are search risks, the direct welfare loss is equal to the sum of the mechanical effect and the reservation wage effect. If the agent is provided with perfect insurance, $g(z) = 1$, then the marginal utility does not change when different earnings offers are accepted. This implies that the term $u_z(z^*)$ in the reservation wage effect is equal to zero. In this case, for the reservation wage effect to be well defined, it is required that $\zeta^{z^*} = 0$, which happens when the reservation earnings z^* is set to maximize expected repayment.

Note that the lender can set the reservation wage to maximize expected repayment precisely because the agent with inelastic labor supply is indifferent among different reservation wages when being perfectly insured. Hence, any reservation wage is incentive compatible. This simple discussion with inelastic labor supply highlights the role of reservation wages in optimal contract design: in the context of elastic labor supply, the optimal contract not only cares about the tradeoff between efficiency (incentive to work) and insurance, but also to some extent, uses the reservation wage to increase expected repayment in order to have a smaller distortion on efficiency.

Equation (A.4.78) characterizes the formula that implicitly determines the optimal marginal repayment rate during employment. In the following, I derive the optimal repayment during unemployment.

Deriving the Optimal Contract During Unemployment Suppose that repayment is increased by $d\alpha$ during unemployment, which is achieved by smoothly perturbing the repayment schedule below z^* so that Assumption 2 is still satisfied. This is going to have a mechanical effect and a reservation wage effect on expected repayment.

The mechanical effect is given by

$$M = \frac{H(z^*)}{1 - \beta H(z^*)} d\alpha, \tag{A.4.79}$$

which captures the fact that the agent repays more during unemployment. Similar to equation (A.4.70), for earnings θ, the increase in repayment reduces utility during unemployment by

$$du(\theta) = -u_1(\theta - \alpha^{SB}(\theta), 0)d\alpha = g(\theta)\lambda d\alpha. \tag{A.4.80}$$

The reservation earnings are determined by equation (A.4.68). Differentiating this equation and substituting (A.4.80) yields:

$$dz^* = d\alpha \frac{(1-\beta)\lambda g(\theta)}{[1-\beta H(z^*)]u_z(z^*)}. \tag{A.4.81}$$

Thus the reservation wage effect is

$$\begin{aligned} RW &= \frac{dz^*}{z^*}\zeta^{z^*}R \\ &= d\alpha \frac{S(1-\beta)\lambda g(\theta)\zeta^{z^*}}{\beta[1-\beta H(z^*)]u_z(z^*)z^*}. \end{aligned} \tag{A.4.82}$$

According to equation (A.4.75), this perturbation generates a direct welfare loss:

$$\begin{aligned} WL &= \frac{H(z^*)}{1-\beta H(z^*)}u_1(\theta - \alpha^{SB}(\theta), 0)d\alpha \\ &= -\frac{H(z^*)}{1-\beta H(z^*)}g(\theta)\lambda d\alpha. \end{aligned} \tag{A.4.83}$$

At the optimum,

$$WL = -\lambda(M + RW), \tag{A.4.84}$$

which yields

$$\underbrace{H(z^*)g(\theta)}_{\text{direct welfare loss}} = \underbrace{H(z^*)}_{\text{mechanical effect}} + \underbrace{\frac{S(1-\beta)\lambda\zeta^{z^*}}{\beta u_z(z^*)z^*}g(\theta)}_{\text{reservation wage effect}}. \tag{A.4.85}$$

If the reservation earnings are fixed, then the reservation wage effect is absent in equation (A.4.85). In this case, the optimal contract subsidizes unemployment such that $g(\theta) = 1$, i.e., to the point where the marginal utility of consumption during unemployment is equal to the shadow cost of debt. This is because there is no behavioral response during unemployment, thus it is always optimal to equalize the cost of fund to the marginal utility of consumption when the agent is unemployed. When there is a negative reservation wage effect, the optimal contract sets $g(\theta) < 1$, indicating that the lender subsidizes the agent more during unemployment. Intuitively, this is because providing more liquidity to unemployment incentivizes the agent to increase her reservation wage and search longer, which would raise expected repayment.

A.5 Quantitative Model Details

In this appendix section, I fill in the details for my quantitative model in section 1.4. I first present the value functions for non-defaulted agents and jobs matched with these agents. Then I illustrate the value functions under the fixed repayment plan and IBR. Finally, I illustrate the wage function with respect to student debt and show that the amount of student debt does not affect the wage rate much through Nash bargaining.

A.5.1 Value Functions for Non-Defaulted Agents

Unemployed Workers

An unemployed worker who has not defaulted yet has the option to default, thus her value is

$$U(\Omega_t) = \max\left\{ \underbrace{\hat{U}(\Omega_t)}_{\text{value of default}}, \underbrace{\tilde{U}(\Omega_t)}_{\text{value of non-default}} \right\}, \tag{A.5.1}$$

where $\hat{U}(\Omega_t)$ and $\tilde{U}(\Omega_t)$ represent the value of default and not default for unemployed workers:

$$\begin{aligned}
\hat{U}(\Omega_t) = \max_{c_t, l_t} \quad & u(c_t, l_t) \quad \underbrace{-\eta}_{\text{default cost}} \quad +\beta\left[\underbrace{\lambda^u \int_{x \geq \rho_u^d} W^d(\Omega_{t+1}, x, \rho_u^d) dV(x)}_{\text{accept the job}} + \underbrace{[1 - \lambda^u + \lambda^u V(\rho_u^d)] U^d(\Omega_{t+1})}_{\text{not accept the job}} \right] \\
\text{subject to} \quad & b_{t+1} = (1+r)b_t + \varkappa\theta^{1-\tau} - c_t, \\
& s_{t+1} = (1+r^s)s_t, \\
& c_t \geq \underline{c}, \\
& b_{t+1} \geq 0,
\end{aligned} \tag{A.5.2}$$

$$\begin{aligned}
\tilde{U}(\Omega_t) = \max_{c_t, l_t} \quad & u(c_t, l_t) + \beta\left[\underbrace{\lambda^u \int_{x \geq \rho_u} W(\Omega_{t+1}, x, \rho_u) dV(x)}_{\text{accept the job}} + \underbrace{[1 - \lambda^u + \lambda^u V(\rho_u)] U(\Omega_{t+1})}_{\text{not accept the job}} \right] \\
\text{subject to} \quad & b_{t+1} = (1+r)(b_t - y_t^{fix}) + \varkappa\theta^{1-\tau} - c_t, \\
& s_{t+1} = (1+r^s)(s_t - y_t^{fix}), \\
& c_t \geq \underline{c}, \\
& b_{t+1} \geq 0,
\end{aligned} \tag{A.5.3}$$

The difference between the recursive formula of $\hat{U}(\Omega_t)$ and $\tilde{U}(\Omega_t)$ lies in the default cost and debt repayment. The agent incurs a default cost η if she chooses to default in period t, and the benefit of this is that no payment is elicited in that period.

Denote $d(\Omega_t)$ as the default decision for unemployed workers,

$$d(\Omega_t) = \begin{cases} 1, & \text{if } \hat{U}(\Omega_t) > \tilde{U}(\Omega_t) \\ 0, & \text{if } \hat{U}(\Omega_t) \leq \tilde{U}(\Omega_t) \end{cases} \tag{A.5.4}$$

Employed Workers

If the agent has not defaulted, her value function is:

$$W(\Omega_t,\rho,\rho') = \max\left\{ \underbrace{\hat{W}(\Omega_t,\rho,\rho')}_{\text{value of default}}, \underbrace{\tilde{W}(\Omega_t,\rho,\rho')}_{\text{value of non-default}} \right\}. \tag{A.5.5}$$

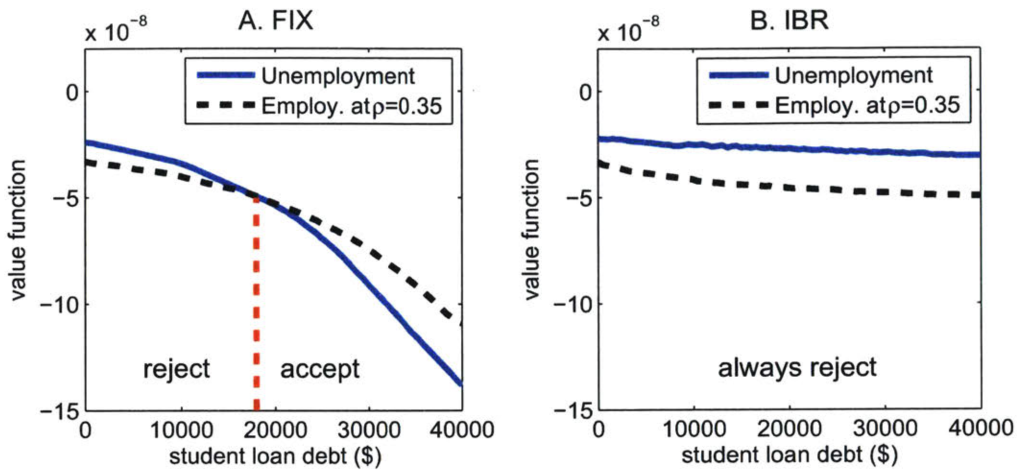
where $\hat{W}(\Omega_t,\rho,\rho')$ and $\tilde{W}(\Omega_t,\rho,\rho')$ represent the value of default and not default for employed workers:

$$\begin{aligned}
\hat{W}(\Omega_t,\rho,\rho') = \max_{c_t,l_t} \quad & u(c_t,l_t) \quad \underbrace{-\eta}_{\text{default cost}} \quad +\beta\Bigg\{ \underbrace{\kappa U^d(\Omega_{t+1})}_{\text{job separation}} + (1-\kappa)\Bigg[\underbrace{[1-\lambda^e+\lambda^e V(\rho')]W^d(\Omega_{t+1},\rho,\rho')}_{\text{not poached or poached by a low vacancy}} \\
& +\lambda^e\Bigg(\underbrace{\int_{x\geq\rho} W^d(\Omega_{t+1},x,\rho)dV(x)}_{\text{transition to a new vacancy}} + \underbrace{\int_{\rho'<x<\rho} W^d(\Omega_{t+1},\rho,x)dV(x)}_{\text{negotiation for a wage rise}} \Bigg)\Bigg]\Bigg\}, \\
\text{subject to} \qquad b_{t+1} &= (1+r)b_t + \varkappa[w^e(\Omega_t,\rho,\rho')l_t]^{1-\tau} - c_t, \\
s_{t+1} &= (1+r^s)s_t. \\
c_t &\geq \underline{c}, \\
b_{t+1} &\geq 0.
\end{aligned} \tag{A.5.6}$$

$$\begin{aligned}
\tilde{W}(\Omega_t,\rho,\rho') = \max_{c_t,l_t} \quad & u(c_t,l_t) + \beta\Bigg\{ \underbrace{\kappa U(\Omega_{t+1})}_{\text{job separation}} + (1-\kappa)\Bigg[\underbrace{[1-\lambda^e+\lambda^e V(\rho')]W(\Omega_{t+1},\rho,\rho')}_{\text{not poached or poached by a low vacancy}} \\
& +\lambda_e\Bigg(\underbrace{\int_{x\geq\rho} W(\Omega_{t+1},x,\rho)dV(x)}_{\text{transition to a new vacancy}} + \underbrace{\int_{\rho'<x<\rho} W(\Omega_{t+1},\rho,x)dV(x)}_{\text{negotiation for a wage rise}} \Bigg)\Bigg]\Bigg\} \\
\text{subject to} \qquad b_{t+1} &= (1+r)(b_t - y_t^{fix}) + \varkappa[w^e(\Omega_t,\rho,\rho')l_t]^{1-\tau} - c_t, \\
s_{t+1} &= (1+r^s)(s_t - y_t^{fix}), \\
c_t &\geq \underline{c}, \\
b_{t+1} &\geq 0.
\end{aligned} \tag{A.5.7}$$

Denote $d(\Omega_t,\rho,\rho')$ as the default decision for employed workers,

$$d(\Omega_t,\rho,\rho') = \begin{cases} 1, & \text{if } \hat{W}(\Omega_t,\rho,\rho') > \tilde{W}(\Omega_t,\rho,\rho'), \\ 0, & \text{if } \hat{W}(\Omega_t,\rho,\rho') \leq \tilde{W}(\Omega_t,\rho,\rho'). \end{cases} \tag{A.5.8}$$



Note: This figure plots the value function with respect to the level of student loan debt for the agent with zero wealth in the first year after college graduation ($t = 1$). The left panel plots the value functions under the fixed repayment plan. The solid line represents the value function of an unemployed agent and the dashed line represents the value function of being employed at a job with productivity $\rho = 0.35$ and negotiation benchmark $\rho' = 0.35$. The right panel plots the value functions under IBR. In this example, when the level of student loan debt is below \$18,000, the agent rejects the job offer and stays unemployed if she is under the fixed repayment plan. However, when the level of debt is above \$18,000, the agent takes the job. This is in contrast to IBR, under which the agent always rejects the job offer.

Figure A.4: An illustration of the value functions under the fixed repayment plan and IBR.

A.5.2 Value Functions for Jobs Filled with Non-Defaulted Agents

The value of a job filled by a worker who has not defaulted yet is

$$\begin{aligned} J(\Omega_t,\rho,\rho') &= \underbrace{[Az_t\rho - w^e(\Omega_t,\rho,\rho')]l(\Omega_t,\rho,\rho')}_{\text{production profit in current period}} \\ &+\beta(1-\kappa)\Bigg\{ \underbrace{d(\Omega_t,\rho,\rho')}_{\text{default}} \Bigg[\underbrace{\lambda^e \int\limits_{\rho'<x<\rho} J^d(\Omega_{t+1},\rho,x)dV(x)}_{\text{negotiation for a wage rise}} + \underbrace{[1-\lambda^e+\lambda^e V(\rho')]J^d(\Omega_{t+1},\rho,\rho')}_{\text{not poached or poached by a low vacancy}} \Bigg] \\ &+ \underbrace{(1-d(\Omega_t,\rho,\rho'))}_{\text{not default}} \Bigg[\underbrace{\lambda^e \int\limits_{\rho'<x<\rho} J(\Omega_{t+1},\rho,x)dV(x)}_{\text{negotiation for a wage rise}} + \underbrace{[1-\lambda^e+\lambda^e V(\rho')]J(\Omega_{t+1},\rho,\rho')}_{\text{not poached or poached by a low vacancy}} \Bigg]\Bigg\}. \end{aligned}$$

The match surplus relative to unemployment is given by

$$Surplus(\Omega_t,\rho,\rho') = W(\Omega_t,\rho,\rho') - U(\Omega_t) + J(\Omega_t,\rho,\rho'). \tag{A.5.9}$$

A.5.3 Illustration of Value Functions

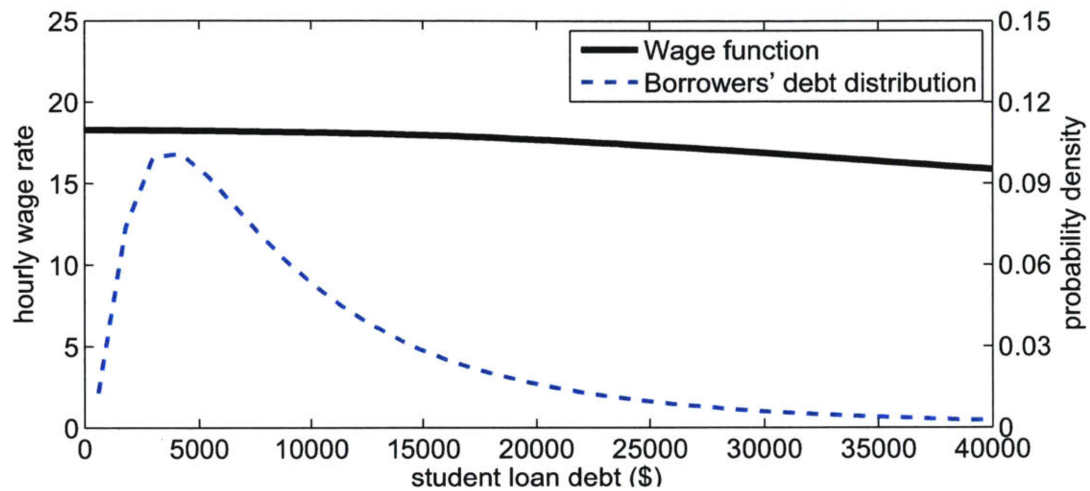
In this subsection, I illustrate the underlying mechanism of IBR by plotting the value functions. In panel A of Figure A.4, I plot the value function under the fixed repayment plan for an unemployed agent and the value function that could be achieved if the agent accepts a job with productivity $\rho = 0.35$ and negotiation benchmark $\rho' = 0.35$. Panel A illustrates the key mechanism of student loan debt by showing that the unemployed value function decreases faster with debt compared to the employed value function. As a result, there is an intersection between the two curves. In this example, when the level of student loan debt is below \$18,000, the agent rejects the job offer and stays unemployed. When the level of debt is above \$18,000, the agent takes the job. My theoretical analysis has revealed the mechanism for this observation: when the agent is burdened with more debt, she not only becomes more risk averse but also has larger liquidity needs, both increasing the marginal utility of consumption and pushing her toward accepting the job offer.

Panel B plots the value functions under IBR. It shows that under IBR, a higher level of debt reduces the value only slightly for both unemployed agents and employed agents. This is because there is much better risk sharing provided by IBR. First, IBR allows the agent to repay less when she has lower income, especially during unemployment. Second, there is debt forgiveness after 25 years, which convexifies the value functions. In this particular example, there is a sharp comparison as the unemployed agent always rejects the job offer with productivity ($\rho = 0.35$), and continues job search.

A.5.4 Wage Function

The wage rate is renegotiated in every period, reflecting the change in Ω. The assumption of Nash bargaining links workers' wage rates to their characteristics, implying that wealth, student loan debt, and labor productivity can influence income. As argued by Krusell, Mukoyama and Sahin [155], it is logical to assume that workers have the incentive to bargain for higher wages if outside options are strong. Moreover, the results under Nash bargaining are useful for comparison with the existing literature, because it is the most commonly used assumption under risk neutrality.

One concern of applying Nash bargaining to model wage determination is that the change in student loan debt could change the wage rate that maximizes the bargaining problem (1.4.13). This confounds the mechanism I hope to quantify, which is how student loan debt affects wage income by affecting job search decisions. As shown in Figure A.5, the wage rate derived from Nash bargaining is not very responsive to the level of debt. This is due to the existence of two countervailing forces in problem (1.4.13). On the one hand, a larger debt repayment reduces the value of the outside option $U(\Omega)$ more than the reduction in $W(\Omega, \rho, w)$ because the marginal value of liquidity is higher during unemployment when income is relatively lower. This increases worker's surplus from the match, $W(\Omega, \rho, w) - U(\Omega)$, reducing the wage rate for the worker. On the other hand, a larger debt repayment increases the marginal value of liquidity for the worker at the current job due to the reduction in consumption. This increases the sensitivity of the worker's employment value with respect to the wage rate, $\partial W(\Omega, \rho, w)/\partial w$, increasing the wage rate for the worker. The impact of the bargaining channel could be large when the level of student loan debt is very high, which is not the case in my estimation sample. This result is also consistent with Krusell, Mukoyama and Sahin [155]'s finding that wage differentials created by the heterogeneity of asset and Nash bargaining are small.



This figure plots the wage function for a typical agent in the first year after college graduation for different amounts of student loan debt. I consider the agent having average wealth ($4,500) and being employed at a job with average productivity (0.725) and with the negotiation benchmark's productivity being set at the reservation productivity (0.325). It shows that increasing the amount of student loan debt from $0 to the average amount ($19,000) reduces the wage rate by about 1.0% (from $18.26 to $18). For agents with other job productivity and negotiation benchmark, the sensitivity of wage rates with respect to student loan debt is similar.

Figure A.5: An illustration of the wage function under the fixed repayment plan.

In the figure, I consider the agent having average wealth ($4,500) and being employed at a job with average productivity (0.725) and with the negotiation benchmark's productivity being set at the reservation productivity (0.325). It shows that increasing the amount of student loan debt from $0 to the average amount ($19,000) reduces the wage rate by about 1.0% (from $18.26 to $18). For agents with other job productivity and negotiation benchmark, the sensitivity of wage rates with respect to student loan debt is similar. This suggests that the bargaining channel confounds the mechanism but quantitatively it is much less important. Specifically, as shown in Table 1.6, normalized borrowers' wage income is 4.2% lower compared to that of non-borrowers, suggesting that roughly three quarters of the reduction in wage income is caused by the mechanism that reduces the reservation wage, and one quarter is caused by the nash bargaining channel which reflects the change in outside options.

Figure A.5 also indicates that the wage rate is more sensitive to student loan debt when the amount of debt is very high. When student loan debt increases from $0 to $40,000, the reduction in wage rates caused by the nash bargaining channel alone is as large as 13.1%. However, these rare cases are not driving the quantitative results of my model, because most students have loan amounts below $20,000 according to the estimated distribution.

Finally, I would like to point out that the strength of the bargaining channel in determining the wage rate also depends on the worker's bargaining parameter ξ. Loosely speaking, the wage rate becomes less sensitive with respect to student loan debt when ξ increases. In the extreme case with $\xi = 1$, the worker's wage rate is always equal to the marginal product of labor $z\rho$, and is therefore not varying with student loan debt at all. When $\xi = 0$, the worker's wage rate is set such that the employment value is equal to the unemployment value. In this case, the sensitivity of wage rate with respect to student loan debt closely depends on the sensitivity of the worker's unemployment value with respect to student loan debt. As a result, the strength of the bargaining channel is comparable to the strength of the liquidity channel of debt burden.

A.6 Background Information for Federal Student Loan Programs

The U.S. federal student loan programs include the William D. Ford Federal Direct Loan Program, the Federal Family Education Loan (FFEL) Program, and the Federal Perkins Loan Program.

The Direct Loan Program is the largest program whose lender is the U.S. Department of Education. This program includes Direct Subsidized/Unsubsidized Loans (also called Stafford Loans), Direct PLUS Loans, and Direct Consolidation Loans. The FFEL Program was the second largest program, funded through a public/private partnership administered at the state and local level. This program includes Subsidized/Unsubsidized Federal Stafford Loans, Federal PLUS Loans, and Federal Consolidation Loans. Following the passage of the Health Care and Education Reconciliation Act of 2010 on March 26, 2010, the FFEL Program was eliminated, and no subsequent loans were permitted to be made under the program after June 30, 2010.[18] The Perkins Program is a school-based loan program for undergraduate and graduate students with exceptional financial need. Under this program, the school is the lender. The Perkins Program only accounts for about 1% of the outstanding federal student loan debt.

Below, I introduce the main features of federal student loans. More detailed information can be found at Federal Student Aid, Fin Aid, and The Institute For College Access & Success.

A.6.1 Grace Period

Under both the Direct Loan Program and the FFEL Program, the borrowers do not have to begin repaying most federal student loans until after they leave college or drop below half-time enrollment. This so-called grace period gives borrowers time to get financially settled and to select their repayment plans. For most loans, interest will accrue during the grace period. However, not all federal student loans have a grace period: Direct Subsidized/Unsubsidized Loans and Subsidized/Unsubsidized Federal Stafford Loans have a 6-month grace period before payments are due. PLUS loans have no grace period. They enter repayment once they are fully disbursed, but may be eligible for a deferment.

[18]On April 24, 2009, President Barack Obama called for an end to the FFEL program, calling it a wasteful and inefficient system of "taxpayers...paying banks a premium to act as middlemen–a premium that costs the American people billions of dollars each year....a premium we cannot afford."

A.6.2 Consolidation Loan

Federal student loan borrowers can consolidate their federal education loans into a consolidation loan. Loan consolidation can greatly simplify loan repayment by centralizing loans to one bill and can lower monthly payments by giving borrowers up to 30 years to repay their loans. After consolidation, borrowers will be able to switch their variable interest rate loans to a fixed interest rate. The downside of loan consolidation is that borrowers might lose benefits offered with the original loans, which may include interest rate discounts, principal rebates, or some loan cancellation benefits.

A.6.3 Repayment Plans

Both the Direct Loan Program and the FFEL Program allow borrowers to choose from among different repayment plans: standard repayment plan, graduated repayment plan, extended repayment plan, and income-driven repayment plan. Borrowers can switch among these plans once a year. All federal education loans allow prepayment without penalty. For loans that are not in default, any excess payment is applied first to interest and then to principal.

Standard Repayment Plan

The default option for federal student loan borrowers is the standard repayment plan. Under this plan, monthly payments are a fixed amount of at least $50 each month and made for up to 10 years for all loan types except Direct Consolidation Loans and FFEL Consolidation Loans. If borrowers have a Direct Consolidation Loan or FFEL Consolidation Loan, the length of the repayment period will depend on the amount of total education loan indebtedness. The maximum repayment period is 30 years.

Graduated Repayment Plan

Under the graduated repayment plan, monthly payments start out low and increase every two years. The repayment period is 10 years for all loan types except for Direct Consolidation Loans and FFEL Consolidation Loans, which allow an extension of the repayment period to 30 years depending on the amount of total education loan indebtedness. Moreover, monthly payments can at least cover the amount of interest that accrues between payments, and they are never more than three times greater than any other payment.

Extended Repayment Plan

Under the extended repayment plan, monthly payments are either fixed or graduated. The repayment period can be extended up to 25 years. As a result, monthly payments are generally lower than those made under the standard and graduated repayment plans. To qualify for the extended repayment plan, borrowers must have had no outstanding balance on a Direct

Loan/FFEL Loan as of October 7, 1998, or on the date they obtained a Direct Loan/FFEL Loan after October 7, 1998. Moreover, borrowers must have more than $30,000 in outstanding Direct Loans or in FFEL Loans.

Income-Driven Repayment Plans

The goal of income-driven repayment plans is to help make borrowers' monthly payments more affordable by basing them on their income and family size.

The income-based application now includes four different income-driven repayment plans: income-contingent repayment plan (ICR Plan), income-based repayment plan (IBR Plan), pay as you earn repayment plan (PAYE Plan), and revised pay as you earn repayment plan (REPAYE Plan).

ICR ICR is available since 1994 and applies only to the Direct Loan Program. It does not have an eligibility requirement; any borrower with an eligible Direct Loan may choose to repay under the ICR Plan. However, ICR is generally less favorable due to the high repayment ratio. The required monthly payment is the lessor of 20% of the borrower's discretionary income[19] or the amount the borrower would pay under a 12-year fixed repayment plan, multiplied by an "income percentage factor". As a result, the monthly payment could be higher than that under the standard repayment plan. Moreover, ICR only applies to the Direct Loan Program, accounting for less than 30% of outstanding loans during late 90s and early 00s [157] and 20% during late 00s [215], which limits its popularity.[20] In fact, between 1996 to 2000, fewer than 1% of borrowers who borrow from Direct Loans choose ICR, although it is 17% of all Direct Loans projected by the Secretary of Education [209].

IBR IBR has been available since July 1, 2009, to borrowers with either Direct or FFEL Loans. To initially enter IBR, borrowers need to have a "partial financial hardship", which means that the payment under IBR is less than the required payment under a standard 10-year repayment plan. Any loan balance that remains after 25 years will be forgiven.[21] The monthly payment under IBR is either 15% of discretionary income or the payment under the standard repayment plan, whichever is smaller. The amendment made by the Health Care and Education Reconciliation Act of 2010 allows borrowers who take out their first loans on or after July 1, 2014, to pay 10%

[19]For ICR, discretionary income is the difference between adjusted gross income and 100% of the poverty guideline amount, adjusted by family size and state-dependent. This differs from the standard used by other income-driven repayment plans, in which discretionary income is based on 150% of the poverty guideline amount.

[20]Schrag [209] provides an detailed discussion about the low participation rate in ICR. One important reason is due to the competition between FFEL lenders and the Direct Loan Program. The FFEL lenders eager to preserve billions of dollars of virtually risk-free federally guaranteed profit, bitterly fought against the creation of the federal direct lending program. Because ICR offers a competitive edge to Direct Loans, the FFEL industry targeted that plan for ridicule and attack. In 1996, three industry groups issued a report attacking the plan as too costly for students. It purported to show that income-contingent repayment is "an expensive option", compared to other repayment plans.

[21]The discharged loan amounts are considered as taxable income, a sort of anti-moral-hazard for excessive borrowing.

of discretionary income, and the forgiveness period is shortened to 20 years (known as the new IBR plan). The participation rate in loan amounts has been steadily increasing due to better publicity.[22]

PAYE PAYE was introduced on December 21, 2012, and it has similar terms to the new IBR plan. However, it is not available to students with older loans–those who borrowed before September 30, 2007, or those who have not borrowed since September 30, 2011.

REPAYE In June 2014, President Obama issued a Presidential Memorandum directing the Department of Education to propose regulations to further ease the burden of student loan debt. On October 27, 2015, the Department of Education issued a final regulation establishing REPAYE. REPAYE improves upon PAYE and extends its protections to all student borrowers with Direct Loans.

Under REPAYE, monthly payments are 10% of discretionary income for all Direct Loan borrowers regardless of their loan origination dates. REPAYE forgives remaining debt after 20 years for those who borrowed only for undergraduate study and 25 years for those who borrowed for graduate study.

Which Income-Driven Repayment Plan to Choose? The older ICR is less borrower-friendly, and usually does not deserve a concern. FFEL Loan borrowers who borrowed before September 30, 2007, are only eligible for IBR. Direct Loan borrowers are usually eligible for multiple income-driven repayment plans and face a choice to make.

New Direct Loan borrowers since July 2014 are eligible for IBR, PAYE, and REPAYE. All three offer the same 10% repayment ratio, but PAYE is slightly more favorable than IBR because all outstanding interest is capitalized under IBR but the amount is capped under PAYE. Moreover, borrowers under IBR are required to spend at least one month in the standard repayment plan before switching to a new plan, but borrowers under PAYE face no such hurdle.

Whether PAYE is better than REPAYE depends on individual circumstances and concerns. REPAYE has three main drawbacks: first, borrowers with any Direct Loans obtained for graduate or professional school face a 25-year repayment period under REPAYE, five years longer than the 20-year PAYE period. Second, spousal income is generally included in calculating "discretionary

[22]In fact, the participation rate was low for a few years after the initialization of IBR in 2009. This is because enrolling in the income-based repayment plan requires working with a loan servicer to complete a 12-page application. As shown by the Consumer Financial Protection Bureau, this is often a bumpy process that can take months. In the meantime, the bills keep coming and millions of borrowers end up in default. Government officials had criticized loan servicers for not being aggressive in enrolling borrowers into IBR and for not informing enough borrowers about the initiative. It is possibly due to an incentive misalignment. The reduced monthly payments under IBR could bite into the revenues of companies under the FFEL Program or hurt holders of FFEL Loans that these companies have securitized and sold to investors. For example, an article published on Huffpost Politics criticized Sallie Mae, the nation's largest servicer of federal student loans, for the failure to enroll many of its distressed borrowers into debt relief programs. Another article on U.S. News reported that student debt relief companies profit from borrowers' confusion.

income". Third, REPAYE does not cap monthly payments at the "standard" repayment amount, as under PAYE and IBR. REPAYE has its own advantages: first, all Direct Loan borrowers are eligible. Second, borrowers whose payments are insufficient to cover the interest accrued during the month will only be charged for up to 50% of the unpaid interest, and interest is not capitalized if the borrower no longer has a partial financial hardship. Borrowers who would face negative amortization can thus save on interest under REPAYE.

Borrowers with older Direct Loans may face a choice between REPAYE and the pre-July 2014 IBR formulation. Most will do better under REPAYE as it offers a lower repayment ratio and a shorter repayment period for undergraduate loans. However, REPAYE's inclusion of spousal income and the lack of a payment cap, may nonetheless make IBR a better option.

Expanding Enrollment in Income-Driven Repayment Plans The income-driven repayment plans alleviate the burden from inflexible repayment. But due to the reasons above, the enrollment is still quite low. As suggested by The Executive Office of the President of the United States [214], continuing to expand enrollment in income-driven repayment plans remains a key priority for the administration. In fact, the administration has used several tools to increase enrollment, such as behavioral "nudges", improved loan servicer contract requirements, efforts associated with the President's Student Aid Bill of Rights, a student debt challenge to gather commitments from external stakeholders, and increased and improved targeted outreach to key borrower segments who would benefit from income contingency. The participation rate in income-driven repayment plans has quadrupled over the last four years from 5% in 2012 to 20% in 2016. In April 2016, the administration announced a series of new actions to further expand the enrollment in income-driven repayment plans.

A.6.4 Deferment and Forbearance

Under certain circumstances, borrowers can receive a deferment or forbearance that allows them to temporarily postpone or reduce their federal student loan payments. But many borrowers do not use these options because applying for deferment and forbearance involves bureaucratic hurdles and detailed paper work.

A deferment is a period during which repayment of the principal and interest of the loan is temporarily delayed. During a deferment, borrowers do not need to make payments. If loans are Federal Perkins Loan, Direct Subsidized Loan, or Subsidized Federal Stafford Loan, the federal government may pay the interest on the loan during a period of deferment. For other types of loans, borrowers are responsible for paying the interest that accrues during the deferment period; otherwise, it may be capitalized.

Borrowers are eligible for a deferment if they are in college, career schools, graduate fellowship programs, approved rehabilitation training program for the disabled, or active duty military service during a war, military operation, or national emergency. Borrowers are eligible for a deferment for up to 3 years if they are experiencing a period of economic hardship, including a period of unemployment or inability to find full-time employment.

With forbearance, borrowers may be able to stop making payments or reduce their monthly payment for up to 12 months. Interest will continue to accrue on all loan types, including subsidized loans. To be eligible for a mandatory forbearance, borrowers must meet certain criteria including financial hardship, illness, national or teaching services, etc. If none of these criteria are satisfied, borrowers can request a discretionary forbearance, and the lender decides whether to grant forbearance or not.

A.6.5 Default and Delinquency

Unlike virtually all other forms of credit, student loans are generally not underwritten: they are frequently offered to young borrowers who have little or no credit history and little to no current income; second, they are accumulated during college studies before individuals enter the job market; and third, lenders (now primarily the taxpayers) are given additional security in that student loans, unlike other forms of debt, are not dischargeable in bankruptcy.[23] This makes the default cost of student loans very high.

Student loans become delinquent the first day after borrowers miss a payment. The delinquency will continue until all payments are made to bring loans current. After a payment reaches 90 days past due, the delinquent status will be reported to the three major credit bureaus and a negative mark will be added to the borrower's credit report. Default occurs when borrowers are delinquent for 270 days. At this point, the debt will be put into collections and payment will be required from collection agencies.

The consequences of default can be severe, and can include: (1). Loss of eligibility for forgiveness plans. (2). The entire unpaid balance and any interest is immediately due and payable. (3). Borrowers' federal and state taxes may be withheld through a tax offset. (4). Student loan debt will increase because of the late fees, additional interest, court costs, collection fees, attorney's fees, and any other costs associated with the collection process. The total increase in cost could be up to 25% of the unpaid balance. (5). Could result in a wage garnishment of 15% of disposable pay.

Once in default, borrowers can get out of default either by repaying the loan in full or through loan rehabilitation. Loan rehabilitation is a one-time opportunity to clear the default on a defaulted federal education loan and regain eligibility for federal student aid. If borrowers redefault on the loan, they will not be able to rehabilitate the loan a second time. If a judgment has been obtained on the defaulted loan, it is not eligible for rehabilitation. To rehabilitate the Direct Loans or FFEL Loans, borrowers and U.S. Department of Education must agree on a reasonable and affordable payment plan. Generally, a monthly payment is considered to be

[23]Student loans can only be discharged if borrowers prove "undue hardship" through a court determination. The undue hardship standard is generally difficult to meet, making student loans practically non-dischargeable through bankruptcy.

reasonable and affordable if it is at least 1.0% of the current loan balance.

A.7 Robustness Check

I conduct three robustness checks for the quantitative results reported in Tables 1.6-1.7. In each robustness check, I reestimate all internally-estimated parameters following the procedure in subsection 1.5.2.

A.7.1 Risk Aversion

One important parameter that determines the effect of debt burden on job search is risk aversion γ. In my baseline specification, γ is set to be 3 according to the precautionary savings literature. I now reduce its value to 1.5, according to the macro development literature on financial frictions. The simulation results in Table A.5 indicate that with lower risk aversion, the reduction in wage income is 72.3% of the baseline under the fixed repayment plan. IBR alleviates the debt burden by 44.7% and increases wage income by 1.1%, compared to 52.1% and 1.9% in the baseline. 22.7% of the reduction in debt burden is attributed to higher reservation wages as opposed to 37.0% in the baseline.

In general equilibrium (Table A.6), IBR increases college entry rate and overall welfare by 4.2% and 2.0% compared with 6.1% and 2.4% in the baseline. The decomposition shows that increased job postings, college entry rate, and insurance in the labor market contribute to 0.4%, 0.9%, and 0.7% of the welfare increase, compared with 0.5%, 1.1%, and 0.8% in the baseline.

Table A.5: Robustness check (lower risk aversion $\gamma = 1.5$): Evaluation of IBR.

	Non-borrowers	Normalized borrowers			Difference	
		FIX	IBR	IBR(w^*_{FIX})	IBR-FIX	IBR(w^*_{FIX})-IBR
Compensation ($)	N/A	5,524	2,890	3,489	-2,634	599
Unemp. dur.	24.2	22.8	23.9	22.9	1.1	-1.0
(week)		(-5.8%)	(-1.2%)	(-5.4%)	(4.6%)	(-4.2%)
Match quality	0.844	0.826	0.836	0.827	0.010	-0.009
		(-2.1%)	(-0.9%)	(-2.0%)	(1.3%)	(-1.1%)
Wage income	48,234	46,782	47,340	45,912	558	-1,428
($)		(-3.0%)	(-1.9%)	(-4.8%)	(1.1%)	(-2.9%)
Output	61,182	59,289	60,098	58,212	809	-1,886
($)		(-3.1%)	(-1.8%)	(-4.9%)	(1.3%)	(-3.1%)
Labor supply	1,737	1,726	1,720	1,699	-6	-21
(hour)		(-0.6%)	(-1.0%)	(-2.2%)	(-0.4%)	(-1.2%)

Table A.6: Robustness check (lower risk aversion $\gamma = 1.5$): General Equilibrium Implications of Student Debt.

	FIX	IBR		
		(1)	(2)	(3)
Fraction of college graduates	41.2%	45.4%	45.5%	41.2%
Fraction of borrowers	63.2%	65.8%	65.9%	63.2%
Average debt among borrowers ($)	10,432	16,135	16,231	10,432
Job contact rate	0.82	0.86	0.82	0.82
Wage income ($)	37,835	38,890	38,590	38,019
		(2.8%)	(2.0%)	(0.5%)
Output ($)	46,593	47,396	47,107	46,811
		(1.7%)	(1.1%)	(0.5%)
Welfare (%)		2.0%	1.6%	0.7%

A.7.2 Elasticity of Labor Supply

The elasticity of labor supply determines the distortion of IBR on the number of hours. In my baseline specification, σ is set to be 2.59 so that the tax-modified Frisch elasticity is 0.33. I now check the model's implication by setting $\sigma = 0.78$ and $\sigma = 88.89$, corresponding to 1 and 0.01 tax-modified labor supply elasticities.

When elasticity is 1, the simulation results in Table A.7 indicate that IBR barely alleviates the debt burden or increases wage income due to the large distortion on labor supply. Borrowers' labor supply is on average reduced by 3.5% relative to non-borrowers, compared to 1.5% in the baseline. The reservation wage effect is still positive, as the wealth compensation would increase by $1,100 for borrowers if reservation wages are fixed. As shown in Table A.8, there is not much response in general equilibrium as the debt burden is almost unchanged.

When elasticity is 0.01, there is almost no response in labor supply when borrowers switch to IBR. As a result, IBR becomes very effective in alleviating the debt burden. The wealth compensation is reduced by 56.4% on average when all borrowers switch to IBR, compared to 52.1% in the baseline. In general equilibrium (A.10), IBR increases college entry rate and overall welfare by 6.6% and 2.6% compared with 6.1% and 2.4% in the baseline. The decomposition shows that increased job postings, college entry rate, and insurance in the labor market contribute to 0.6%, 1.1%, and 0.9% of the welfare increase, compared with 0.5%, 1.1%, and 0.8% in the baseline.

Table A.7: Robustness check (higher labor supply elasticity $\sigma = 0.78$): Evaluation of IBR.

	Non-borrowers	**Normalized borrowers**			**Difference**	
		FIX	IBR	IBR(w^*_{FIX})	IBR-FIX	IBR(w^*_{FIX})-IBR
Compensation ($)	N/A	6,530	6,412	7,512	-118	1,100
Unemp. dur.	24.0	22.1	23.5	22.4	1.4	-1.1
(week)		(-7.9%)	(-2.1%)	(-6.7%)	(5.8%)	(-4.6%)
Match quality	0.877	0.858	0.869	0.857	0.011	-0.012
		(-2.2%)	(-0.9%)	(-2.3%)	(1.3%)	(-1.4%)
Wage income	48,121	45,976	46,012	44,545	36	-1,467
($)		(-4.5%)	(-4.4%)	(-7.4%)	(0.1%)	(-3.0%)
Output	56,387	53,887	53,940	52,173	53	-1,677
($)		(-4.4%)	(-4.3%)	(-7.5%)	(0.1%)	(-3.0%)
Labor supply	1,657	1,621	1,599	1,575	-22	-24
(hour)		(-2.2%)	(-3.5%)	(-4.9%)	(-1.3%)	(-1.4%)

Table A.8: Robustness check (higher labor supply elasticity $\sigma = 0.78$): General Equilibrium Implications of Student Debt.

	FIX	**IBR**		
		(1)	(2)	(3)
Fraction of college graduates	41.3%	41.6%	41.6%	41.7%
Fraction of borrowers	62.0%	62.2%	62.2%	62.3%
Average debt among borrowers ($)	10,315	10,539	10,578	10,315
Job contact rate	0.82	0.82	0.82	0.82
Wage income ($)	37,489	37,635	37,603	37,521
		(0.4%)	(0.3%)	(0.1%)
Output ($)	42,529	42,647	42,613	42,589
		(0.3%)	(0.2%)	(0.1%)
Welfare (%)		0.3%	0.2%	0.1%

Table A.9: Robustness check (lower labor supply elasticity $\sigma = 88.89$): Evaluation of IBR.

	Non-borrowers	Normalized borrowers			Difference	
		FIX	IBR	IBR(w^*_{FIX})	IBR-FIX	IBR(w^*_{FIX})-IBR
Compensation ($)	N/A	6,512	2,842	4,237	-3,670	1,395
Unemp. dur.	23.7	22.0	23.5	22.2	1.5	-1.3
(week)		(-7.2%)	(-0.8%)	(-6.3%)	(6.4%)	(-5.5%)
Match quality	0.811	0.787	0.807	0.788	0.020	-0.019
		(-3.0%)	(-0.5%)	(-2.8%)	(2.5%)	(-2.3%)
Wage income	46,547	45,198	46,210	45,179	1,021	-1,031
($)		(-2.9%)	(-0.7%)	(-2.9%)	(2.2%)	(-2.2%)
Output	64,628	62,685	64,256	62,730	1,571	-1,526
($)		(-3.0%)	(-0.6%)	(-2.9%)	(2.4%)	(-2.3%)
Labor supply	1,732	1,732	1,731	1,731	-1	0
(hour)		(-0.0%)	(-0.1%)	(-0.1%)	(-0.1%)	(-0.0%)

Table A.10: Robustness check (lower labor supply elasticity $\sigma = 88.89$): General Equilibrium Implications of Student Debt.

	FIX	IBR		
		(1)	(2)	(3)
Fraction of college graduates	41.4%	48.0%	48.2%	41.4%
Fraction of borrowers	62.2%	67.8%	67.9%	62.2%
Average debt among borrowers ($)	10,451	17,865	18,010	10,451
Job contact rate	0.82	0.89	0.82	0.82
Wage income ($)	36,423	37,765	37,342	36,679
		(3.7%)	(2.5%)	(0.7%)
Output ($)	48,138	49.264	48,996	48,425
		(2.3%)	(1.8%)	(0.6%)
Welfare (%)		2.6%	2.0%	0.9%

A.7.3 Access to Other Credit

Credit access alleviates the liquidity problem, which would attenuate the effect of debt burden on job search. In the baseline specification, agents cannot borrow. I now relax this assumption. Using data from the Survey of Consumer Finances (SCF), Kaplan and Violante [145] estimate that the median ratio of credit limit to annual labor income is 18.5% for households aged 22 to

59. Based on this estimate, I allow employed agents to borrow 18.5% of their wage income, and unemployed agents to borrow 18.5% of UI benefits (i.e., \$1,500). The simulation results in Table A.11 indicate that credit access slightly alleviates the debt burden. The reduction in wage income is 90% of the baseline under the fixed repayment plan. The small difference comes from the fact agents cannot borrow much due to the low income during unemployment. IBR alleviates the debt burden by 45.8% and increases wage income by 1.8%, compared to 52.1% and 1.9% in the baseline. 27.6% of the reduction in debt burden is attributed to higher reservation wages as opposed to 37.0% in the baseline.

In general equilibrium (Table A.12), IBR increases college entry rate and overall welfare by 5.4% and 2.2% compared with 6.1% and 2.4% in the baseline. The decomposition shows that increased job postings, college entry rate, and insurance in the labor market contribute to 0.5%, 1.0%, and 0.7% of the welfare increase, compared with 0.5%, 1.1%, and 0.8% in the baseline.

Table A.11: Robustness check (credit access): Evaluation of IBR.

	Non-borrowers	**Normalized borrowers**			**Difference**	
		FIX	IBR	IBR(w^*_{FIX})	IBR-FIX	IBR(w^*_{FIX})-IBR
Compensation (\$)	N/A	7,623	4,135	5,098	-3,488	963
Unemp. dur.	24.0	22.5	23.6	22.5	1.1	-1.1
(week)		(-6.3%)	(-1.7%)	(-6.3%)	(4.6%)	(-4.6%)
Match quality	0.840	0.817	0.831	0.818	0.014	-0.013
		(-2.7%)	(-1.1%)	(-2.6%)	(1.6%)	(-1.5%)
Wage income	47,356	45,550	46,399	44,976	849	-1,423
(\$)		(-3.8%)	(-2.0%)	(-5.0%)	(1.8%)	(-3.0%)
Output	59,489	57,343	58,212	56,562	869	-1,650
(\$)		(-3.6%)	(-2.1%)	(-4.9%)	(1.5%)	(-2.8%)
Labor supply	1,722	1,712	1,705	1,695	-7	-10
(hour)		(-0.6%)	(-1.0%)	(-1.6%)	(-0.4%)	(-0.6%)

Table A.12: Robustness check (credit access): General Equilibrium Implications of Student Debt.

	FIX	IBR		
		(1)	(2)	(3)
Fraction of college graduates	41.3%	46.7%	46.7%	41.3%
Fraction of borrowers	62.1%	66.9%	67.0%	62.1%
Average debt among borrowers ($)	10,333	16,894	16,978	10,333
Job contact rate	0.82	0.87	0.82	0.82
Wage income ($)	37,015	38,124	37,764	37,203
		(3.0%)	(2.0%)	(0.5%)
Output ($)	45,198	46,026	45,845	45,367
		(1.8%)	(1.4%)	(0.4%)
Welfare (%)		2.2%	1.7%	0.7%

Appendix B

Appendix for External Financing and Customer Capital: A Financial Theory of Markups

B.1 Benchmark Cases with Constant Markups

We conduct theoretical experiments by setting two benchmark cases in order to illustrate the role of financial frictions in shaping the dynamics of markups desired by the firm.

B.1.1 The Optimal Static Monopolistic Price

Under the common setup adopted by the New Keynesian literature [e.g. 80, 97, for a review], the firm has no external financing costs and there are no customer flows among sellers. By mapping these assumptions onto our model, we have $\phi = \gamma = 0$ (i.e. no external financing costs), $\lambda = 0$ (i.e. no cash holding costs), and $\alpha = 0$ (i.e. no customer flows among sellers). Thus, on the financial side, the Modigliani and Miller [183] Theorem holds, and on the product side, the price elasticity of demand only shows up in the short run. The intra-temporal profit optimization, as in traditional New Keynesian (DSGE) models, leads to the equilibrium where the firm chooses $p = p^* \equiv \frac{\eta}{\eta - 1}\bar{c}$ once it gets the chance to reset its price and keeps this price forever. The desired markup is constant over time and purely determined by the intra-temporal elasticity of demand.[1] In other words, we have $w_0^p = \infty$. Since there are no external financing costs, the marginal value of cash held by the firm is one. Thus, it is reasonable to guess that the value function of the firm has the following form

$$u(w, p) \equiv u(p) + w. \tag{B.1.1}$$

[1] Note that in traditional New Keynesian DSGE models [e.g. 97], although the desired markup is constant over time, the realized product price can fluctuate over time which is purely driven by the time-variation in marginal costs. In order to highlight the forces driving time-varying desired markups, we postulate a fixed marginal cost. This helps us to clearly illustrate the key mechanisms of the model.

By plugging (B.1.1) into the coupled ODEs, we can get

$$u(p_H) = \theta(\delta + r) - \theta\sqrt{(\delta + r)^2 - 2\left[\mu_H - (r + \delta)\right]/\theta + 1}, \quad \text{and} \tag{B.1.2}$$

$$u(p_L) = \theta(\delta + r + \xi) - \theta\sqrt{(\delta + r + \xi)^2 - 2\left[\mu_L + \xi u(p_H) - (r + \delta + \xi)\right]/\theta + 1}, \tag{B.1.3}$$

where $\mu_H \equiv (p_H - \bar{c})\mu(p_H)$ and $\mu_L \equiv (p_L - \bar{c})\mu(p_L)$ are expected current profits for the firm with p_H and p_L, respectively. For illustrative purposes, we assume that p_H is the optimal static monopolistic product price, $p_H = p^*$. It implies that $\mu_H > \mu_L$. Therefore, the optimal investment is

$$i(w, p_H) = (r + \delta) - \sqrt{(r + \delta)^2 - 2\left[\mu_H - (r + \delta)\right]/\theta}, \quad \text{and} \tag{B.1.4}$$

$$i(w, p_L) = (r + \delta + \xi) - \sqrt{(r + \delta + \xi)^2 - 2\left[\mu_L + \xi u(p_H) - (r + \delta + \xi)\right]/\theta}. \tag{B.1.5}$$

In this case, the steady-state price is deterministic with $p \equiv p_H$. We highlight the following implications arising from this simple benchmark case. First, it is apparent that the firm's investment decisions only focus on keeping investment on the optimal growth path, while its price setting decisions focus on maximizing expected current profits. This is called the "growth-profit separation effect". Second, the desired markup is constant at $\frac{\eta}{1-\eta}$. Third, price stickiness has no impact on the firm's value or its decisions since the optimal price p_H is constant overtime. Fourth, the efficiency cost of price stickiness is just a pass-through from whatever cost (e.g. menu costs) resulting a sticky product price, if the initial price is not p_H. That is, the value function is deteriorated exactly by the same amount of menu costs.

At last, in order to make the above equilibrium solution rigorous, we show, in Proposition 11, that it is indeed the case that $u(p_H) > u(p_L)$.

Proposition 11. *Suppose the parameters satisfy $(\delta + r)^2 - 2\left[\mu_H - (r + \delta)\right]/\theta > 0$, then $u(p_L)$ in (B.1.3) is well defined and $u(p_H) > u(p_L)$.*

Proof. First, we show that

$$(r + \delta + \xi)^2 - 2\left[\mu_L + \xi u(p_H) - (r + \delta + \xi)\right]/\theta > \left\{\sqrt{(r + \delta)^2 - 2\left[\mu_H - (r + \delta)\right]/\theta} + \xi\right\}^2 > 0. \tag{B.1.6}$$

Rearranging the terms, we get

$$\begin{aligned}&(r + \delta + \xi)^2 - 2\left[\mu_L + \xi u(p_H) - (r + \delta + \xi)\right]/\theta - \left\{\sqrt{(r + \delta)^2 - 2\left[\mu_H - (r + \delta)\right]/\theta} + \xi\right\}^2 \\ &= 2(\mu_H - \mu_L) > 0.\end{aligned} \tag{B.1.7}$$

It is straightforward to see that $u(p_L)$ converges to $u(p_H)$ as ξ goes to infinity. The partial derivative of $u(p_L)$ with respect to ξ is

$$\frac{\partial u(p_L)}{\partial \xi} = 1 - \frac{\sqrt{(r+\delta)^2 - 2\left[\mu_H - (r+\delta)\right]/\theta + \xi}}{\sqrt{(r+\delta+\xi)^2 - 2\left[\mu_L + \xi u(p_H) - (r+\delta+\xi)\right]/\theta}} > 0. \tag{B.1.8}$$

Therefore, the value function $u(p_L)$ is monotonically increasing in ξ, and $u(p_H) > u(p_L)$. □

B.1.2 The Optimal Inter-temporal Monopolistic Price in Customer Market Models

We incorporate the customer market [e.g. 194, 205] into traditional New Keynesian models. The only difference from the previous irrelevance benchmark is that now the firm's customer base affects its profits and customer flows are allowed among sellers, i.e. $\alpha > 0$ and $h'(p) < 0$. The Modigliani and Miller [183] Theorem still holds, but in this case the price elasticity of demand shows up both in the short run and in the long run. As a result, price setting decisions not only affect the firm's current profits but also influence the long-run growth rate in customer base, thus leading to an equilibrium price lower than the optimal static monopolistic price $p^* \equiv \frac{\eta}{\eta - 1}\bar{c}$. In equilibrium, the firm sets p_L once it gets the chance to reset its price, and the firm keeps the product price at p_L forever.[2] In other words, we have $W_0^P = 0$. As in traditional New Keynesian models, the desired markup is constant over time, however, it is lower than p^* as emphasized by the traditional customer market model [e.g. 194, 205]. Since there are no external financing costs, the marginal value of cash is one. Thus, it is reasonable to guess that the value function of the firm has the following form:

$$u(w,p) \equiv u(p) + w. \tag{B.1.9}$$

By plugging (B.1.1) into the coupled ODEs, we can get

$$u(p_L) = \frac{\theta(\delta(1-\alpha)+r)+(1-\alpha)}{(1-\alpha)^2} - \frac{\theta\sqrt{\left[\delta(1-\alpha)+r\right]^2 - 2(1-\alpha)\left[\mu_L - r - \delta(1-\alpha)\right]/\theta}}{(1-\alpha)^2}, \tag{B.1.10}$$

and

$$u(p_H) = \frac{\theta\left[\delta(1-\alpha)+r+\xi-\alpha h(p_H)\right]+(1-\alpha)}{(1-\alpha)^2} - \frac{\theta\sqrt{\left[\delta(1-\alpha)+r+\xi-\alpha h(p_H)+(1-\alpha)/\theta\right]^2 - (1+2\theta\mu_H+2\theta\xi u(p_L))(1-\alpha)^2/\theta^2}}{(1-\alpha)^2}. \tag{B.1.11}$$

Therefore, the optimal investment is

$$i(w,p_L) = \frac{\delta(1-\alpha)+r}{1-\alpha} - \frac{\sqrt{\left[\delta(1-\alpha)+r\right]^2 - 2(1-\alpha)\left[\mu_L - r - \delta(1-\alpha)\right]/\theta}}{1-\alpha}, \text{ and} \tag{B.1.12}$$

[2] We require $p_L < p^*$ but p_L should not be very small so that, for example, the condition (B.1.14) holds.

$$i(w, p_H) = \frac{\delta(1-\alpha) + r + \xi - \alpha h(p_H)}{(1-\alpha)^2} - \frac{\sqrt{[\delta(1-\alpha) + r + \xi - \alpha h(p_H) + (1-\alpha)/\theta]^2 - (1 + 2\theta\mu_H + 2\theta\xi u(p_L))(1-\alpha)^2/\theta^2}}{1-\alpha}. \tag{B.1.13}$$

If customer flows are fast enough so that the deterioration of customer base due to a high product price is relatively more significant than the gain from current profits, i.e.

$$(1-\alpha)^2(\mu_H - \mu_L) + \alpha h(p_H)\left[\theta\delta(1-\alpha) + \theta r + (1-\alpha)\right] \leq 0. \tag{B.1.14}$$

In this case, the steady-state price is deterministic with $p \equiv p_L$. We highlight four important aspects of this simple benchmark case with the customer market. First, the firm's investment decisions only focus on the optimal growth path, but price setting decisions are affected by both short-run profits and long-run growth rates. When the firm weights more on long-run growth, it is optimal to choose the low price p_L in order to build up customer base at the cost of reducing current profits. Second, the desired markup is constant at $\frac{p_L}{c} < \frac{p_H}{c} \equiv \frac{\eta}{1-\eta}$, which is lower than the optimal static monopolistic markup $\frac{\eta}{1-\eta}$ due to the existence of customer flows. Third, price stickiness has no impact on the firm's value or its decisions. Lastly, the efficiency cost of price stickiness is just a pass-through to the firm's value from whatever cost (e.g. menu costs) resulting a sticky product price, if the initial price is not p_L.

Similarly, in order to make the above equilibrium solution rigorous, we show, in Proposition 12, that it is indeed the case that $u(p_H) > u(p_L)$.

Proposition 12. *Suppose the parameters satisfy the restriction (B.1.14) and* $[\delta(1-\alpha) + r]^2 - 2(1-\alpha)\left[\mu_L - r - \delta(1-\alpha)\right]/\theta > 0$, *then* $u(p_H)$ *in (B.1.11) is well defined and* $u(p_L) > u(p_H)$.

Proof. First, it is straightforward to see that

$$\begin{aligned}
&[\delta(1-\alpha) + r + \xi - \alpha h(p_H) + (1-\alpha)/\theta]^2 - [1 + 2\theta\mu_H + 2\theta\xi u(p_L)]\,(1-\alpha)^2/\theta^2 \\
&\quad - \left\{\xi - \alpha h(p_H) - \sqrt{[\delta(1-\alpha) + r]^2 - (1-\alpha)\left[\mu_L(1-\alpha) - \delta(1-\alpha) - r\right]/\theta}\right\}^2 \\
&\geq -2\alpha h(p_H)\left[\delta(1-\alpha) + r + (1-\alpha)/\theta\right] + 2\frac{(1-\alpha)^2}{\theta}(\mu_L - \mu_H) > 0.
\end{aligned} \tag{B.1.15}$$

Obviously, $u(p_H)$ converges to $u(p_L)$ as ξ goes to infinity. The partial derivative of $u(p_H)$ with respect to ξ is $\frac{\partial u(p_H)}{\partial \xi}$ which has the following expression

$$\frac{\theta}{(1-\alpha)^2}\left\{1 - \frac{\xi - \alpha h(p_H) - \sqrt{[\delta(1-\alpha) + r]^2 - (1-\alpha)\left[\mu_L(1-\alpha) - \delta(1-\alpha) - r\right]/\theta}}{\sqrt{[\delta(1-\alpha) + r + \xi - \alpha h(p_H) + (1-\alpha)/\theta]^2 - [1 + 2\theta\mu_H + 2\theta\xi u(p_L)]\,(1-\alpha)^2/\theta^2}}\right\}.$$

Based on the inequality (B.1.15) above, we know that $\frac{\partial u(p_H)}{\partial \xi} > 0$. Therefore, the value function $u(p_H)$ is monotonically increasing in ξ, and $u(p_L) > u(p_H)$. $\square$

B.2 Data

Our empirical analysis is conducted using quarterly series at the industry level over the period 1998 – 2012. We use three major datasets: Labor Productivity and Costs (LPC) database and the industry-level Producer Price Index (PPI) from the Bureau of Labor Statistics (BLS), the U.S. national and industry economic accounts from the Bureau of Economic Analysis (BEA), and the Compustat/CRSP firm-level quarterly series. In order to measure price stickiness, we also incorporate the median frequency of price change for 14 industries from Nakamura and Steinsson [188, Table VI], which is collected from the PPI micro dataset.

B.2.1 Industry Categories

We focus on the manufacturing sector, and consider broadly defined 18 industries according to the first three digits of the NAICS code. Our industry categories are consistent with BEA industry accounts except for the transportation industry. The transportation industry in BEA accounts is divided into two finer industries, motor vehicles and other transportation. Since the first three digits of the NAICS code for the two sub-industries are the same, we merge the two and use a more general category—transportation industry. Table B.1 lists all the industries and their associated three digits NAICS codes. Among the 18 industries, 10 industries produce durable goods, and the rest produce non-durable goods.

B.2.2 Construct Industry-Level Corporate Firms' Variables

Following Korajczyk and Levy [153], we use firm-level quarterly series from Compustat/CRSP. All series are converted to real values in 1998 dollars using the consumer price index (CPI).

All firms are sorted into their associated industries according to the NAICS code.[3] We focus on relatively large and mature firms that had been existing for at least 16 quarters between 1998 and 2012 and with assets not below the 30% quantile in each industry. As in Korajczyk and Levy [153, footnote 3], we exclude firms' data if any of the following conditions is satisfied.[4] The capital letters wrote in parentheses refer to the variable name in Compustat dataset.

[3]The number of firms in our sample for each industry is provided in parentheses—Wood products (32), Nonmetallic mineral products (44), Primary metals (110), Fabricated metal products (97), Machinery (330), Computer and electronic products (1123), Electrical equipment, appliances, and components (114), Transportation (184), Furniture and related products (27), Miscellaneous manufacturing (257), Food and beverage and tobacco products (145), Textile mills and textile product mills (31), Apparel and leather and allied products (112), Paper products (74), Printing and related support activities (34), Petroleum and coal products (69), Chemical products (821), Plastics and rubber products (83).

[4]Note, if firm i satisfies any condition in period t, we only remove the data point for firm i in period t. Firm i's data points in other periods remain in the sample.

Table B.1: Industry Categories

Industry name	**NAICS code**
Durable goods	
Wood product	321
Nonmetallic mineral products	327
Primary metals	331
Fabricated metal products	332
Machinery	333
Computer and electronic products	334
Electrical equipment, appliances, and components	335
Transportation	336
Furniture and related products	337
Miscellaneous manufacturing	339
Non-durable goods	
Food and beverage and tobacco products	311, 312
Textile mills and textile product mills	313, 314
Apparel and leather and allied products	315, 316
Paper products	322
Printing and related support activities	323
Petroleum and coal products	324
Chemical products	325
Plastics and rubber products	326

(i) Firms with negative selling, general and administrative expenses (XSGAQ), or book assets (ATQ) or property plant and equipment (PPEGTQ).

(ii) Firms with market-to-book ratios (Tobin's average q, constructed following Davis, Fama and French [71], see below) greater than 19.51 or less than 0.31.

(iii) Firms with average operating income (OIBDPQ) to book assets ratios (ATQ) over the previous four quarters greater than 0.35 and less than -0.12.

(iv) Firms with average selling expense (XSGAQ) to sales (SALEQ) ratios over the previous four quarters greater than 1.18 and less than 0.019.

(v) Firms with accounts receivables (RECTQ) to book assets ratios (ATQ) greater than 0.74.

(vi) Firms with average depreciation (DPQ) to book assets ratios (ATQ) over the previous four quarters greater than 0.042.

(vii) Firms with average income taxes (TXTQ) to book assets ratios (ATQ) over the previous four quarters greater than 0.046 and less than -0.016.

(viii) Firms with one year excess stock returns (see below for the variable construction) greater than 338% and less than -112%.

Financing variables

Equity Repurchase

We measure the amount of equity repurchases in period t as the increase in the amount of treasure stock (TSTKQ) from period $t-1$ to period t.[5] Therefore, for firm f in industry i, the repurchase ratio variable in period t, $REPR^f_{i,t}$, is constructed as $REPR^f_{i,t} = \frac{TSTKQ^f_{i,t} - TSTKQ^f_{i,t-1}}{ATQ^f_{i,t}}$. If either $TSTKQ^f_{i,t}$ or $TSTKQ^f_{i,t-1}$ is missing, or if $TSTKQ^f_{i,t} - TSTKQ^f_{i,t-1} < 0$, we set $REPR^f_{i,t} = 0$. Each industry's repurchase ratio is constructed as the average of all firms' repurchase ratios within that industry weighted by sales. For industry i,

$$REPR_{i,t} = \frac{\sum_{f \in industry\ i} REPR^f_{i,t} \times SALEQ^f_{i,t}}{\sum_{f \in industry\ i} SALEQ^f_{i,t}}.$$

Equity Issuance

We measure the amount of issued equity as the number of common shares issued (CSHIQ) times the stock price at the end of the quarter (PRCCQ). Since the value of preferred stock is rarely reported in Compustat at the quarterly frequency, we do not distinguish common equity from preferred equity. For firm f in industry i, the amount of equity issued is constructed as $EI^f_{i,t} = CSHIQ^f_{i,t} \times PRCCQ^f_{i,t}$. The amount of equity issued in period t is $\Delta EI^f_{i,t} = EI^f_{i,t} - EI^f_{i,t-1}$ and the equity issuance ratio is $\Delta EIR^f_{i,t} = \frac{\Delta EI^f_{i,t}}{ATQ^f_{i,t}}$. Each industry's equity issuance ratio is constructed as the average of all firms' equity issuance ratios within that industry weighted by sales. For industry i,

$$\Delta EIR_{i,t} = \frac{\sum_{f \in industry\ i} \Delta EIR^f_{i,t} \times SALEQ^f_{i,t}}{\sum_{f \in industry\ i} SALEQ^f_{i,t}}.$$

Debt Change

We measure the total amount of debt as the sum of long term debt (DLTTQ) and the amount of debt in current liabilities (DLCQ). For firm f in industry i, the total amount of debt in period t, $DEBT^f_{i,t}$, is constructed as $DEBT^f_{i,t} = DLTTQ^f_{i,t} + DLCQ^f_{i,t}$. The change in debt level in period t, is constructed as the change in the total amount of debt from period $t-1$ to period t, $\Delta DEBT^f_{i,t} = DEBT^f_{i,t} - DEBT^f_{i,t-1}$. The change in debt

[5]We construct equity repurchases based on the change in the amount of treasure stock instead of using total shares repurchased (CSHOPQ) for more accurate measurement.

ratios is, $\Delta DEBTR^f_{i,t} = \frac{\Delta DEBT^f_{i,t}}{ATQ^f_{i,t}}$. Each industry's change in debt ratios is constructed as the average of all firms' change in debt ratios within that industry weighted by sales. For industry i,

$$\Delta DEBTR_{i,t} = \frac{\sum_{f \in industry\ i} \Delta DEBTR^f_{i,t} \times SALEQ^t_{i,f}}{\sum_{f \in industry\ i} SALEQ^f_{i,t}}.$$

Other variables

Tobin's average q

Following Davis, Fama and French [71], we construct Tobin's average q as the ratio of market equity over book equity. The book equity is constructed as the sum of stockholder's equity (SEQQ) and deferred taxes and investment tax credit (TXDITCQ) minus preferred/preference stock (PSTKQ). The market equity is constructed as the stock price at the end of the quarter (PRCCQ) times the number of common shares outstanding (CSHOQ). For firm f in industry i in period t, the book equity is $BK^f_{i,t} = SEQQ^f_{i,t} + TXDITCQ^f_{i,t} - PSTKQ^f_{i,t}$, the market equity is $MK^f_{i,t} = PRCCQ^f_{i,t} \times CSHOQ^f_{i,t}$, and the Tobin's q is $q^f_{i,t} = \frac{MK^f_{i,t}}{BK^f_{i,t}}$.

Investment rate

We measure the amount of investment using capital expenditure. Compustat database provides the amount of year-to-date capital expenditure (CAPXY). For firm f in industry i in period t, we construct investment rate $IR^f_{i,t}$ as

$$IR^f_{i,t} = \begin{cases} \frac{CAPXY^f_{i,t}}{ATQ^f_{i,t}} & \text{if } t \text{ is the first quarter of the year.} \\ \frac{CAPXY^f_{i,t} - CAPXY^f_{i,t=1}}{ATQ^f_{i,t}} & \text{otherwise.} \end{cases}$$

Cash holdings

Cash holdings are measured as cash and short-term investments (CHEQ), which consist of cash (CHQ) and short-term investments (IVSTQ). It includes, among others, the following items: cash in escrow; government and other marketable securities; letters of credits; time, demand, and certificates of deposit; restricted cash.

For firm f in industry i in period t, the change in cash is constructed as the increase in the firm's cash holdings from period t to period $t+1$, $\Delta CASH^f_{i,t} = CHEQ^f_{i,t+1} - CHEQ^f_{i,t}$. The cash to book equity ratio is $CASHR^f_{i,t} = \frac{CHEQ_{i,t+1}}{SEQQ^f_{i,t} + TXDITCQ^f_{i,t} - PSTKQ^f_{i,t}}$.

Stock return

The stock return is measured to include both dividend returns and capital gains. We measure dividend returns as the dividend per share at the ex-dividend date (DVPSXQ), and capital gains as the difference in the stock price from period $t-1$ to period t divided by the stock price in period $t-1$. For firm f in industry i in period t, the stock return is constructed as, $R_{i,t}^{f} = DVPSXQ_{i,t}^{f} + \frac{PRCCQ_{i,t}^{f} - PRCCQ_{i,t-1}^{f}}{PRCCQ_{i,t-1}^{f}}$.

Each industry's stock return is constructed as the average of all firms' stock returns within that industry weighted by sales. For industry i,

$$R_{i,t} = \frac{\sum_{f \in industry\ i} R_{i,t}^{f} \times SALEQ_{i,t}^{f}}{\sum_{f \in industry\ i} SALEQ_{i,t}^{f}}.$$

Abnormal return

We calculate the average equity abnormal return (AR_i) to correct for the observed equity price run-up prior to equity issue announcements. The average abnormal return is approximated by the constant term of the Fama and French 3-factor model. Specifically, for each industry i, we run the following regression

$$R_{i,t} = \alpha_i + \beta_{i,1} R_t^m + \beta_{i,2} SMB_t + \beta_{i,3} HML_t + \epsilon_t,$$

where R^m is the excess return of the market portfolio, SMB stands for "Small (market capitalization) Minus Big", and HML for "High (book-to-market ratio) Minus Low". The constant term α_i is used to proxy industry i's average equity abnormal return.

B.2.3 Measure of Price Stickiness

We follow the empirical literature [e.g. 23, 37, 57, 188] and proxy the degree of price stickiness by measuring how frequently firms change their prices. Firms facing larger price stickiness adjust their prices less frequently. In this sense, the average frequency of price change can be regarded as an inverse measure of price stickiness. There are three factors that largely determine the price change frequency. The first factor is the industry's Herfindahl index. Gaskins [99] and Eliashberg and Jeuland [87] argue that firms' pricing decisions are affected by the threat of competitive entries. Therefore, firms with more monopoly power are facing smaller price stickiness in their products. The second factor is the Tobin's q, which reflects firms' future perspectives, and has predicative power for future investment opportunities. The third factor is the sensitivity of prices to firms' business conditions. We proxy this factor at the industry level using the sensitivity of industry price indexes with respect to industry specific productivity shocks.

We estimate the industry-level price change frequency in two steps.

In the first step, we collect the value of the three relevant factors for all the 18 industries over the period 1998 – 2012. The Herfindahl index is quite stable over time, and thus we use its value from the 2007 survey conducted by the U.S. census. The industry-level Tobin's q is estimated using firm-level data from Compustat, as described in Appendix B.2.2. Appendix B.2.3 elaborates the estimation procedures for the sensitivity of industry price indexes with respect to industry specific productivity shocks.

In the second step, we characterize and estimate the relationship between the price change frequency and the three factors using a linear model. In Nakamura and Steinsson [188, Table VI], the median price change frequencies measured from the Producer Price Index (PPI) micro-data for 14 industries are provided. These industries are broadly consistent with some of the industries we use, and thus form the basis for our estimation (see Appendix B.2.3 below).

Estimate the Price Sensitivity of Productivity Shocks

To estimate the sensitivity of price indexes with respect to industry-level productivity shocks, we run the following regression for each industry i and for the whole economy

$$\pi_{i,t} = \alpha_{i,0} + \alpha_{i,1}\mu_{i,t} + \alpha_{i,2}log(q_{i,t}/q_{i,t-1}) + \epsilon_{i,t}, \tag{B.2.1}$$

where $\mu_{i,t}$ is the estimated innovation in productivity in industry i in period t, $q_{i,t}$ is the industry i's average Tobin's q in period t. Thus $log(q_{i,t}/q_{i,t-1})$ measures the innovation in industry i's Tobin's q, which reflects the change in investment opportunities. The coefficient $\alpha_{i,1}$ is to our interest, which measures the price sensitivity of productivity shocks (the change in business conditions) in industry i.

Industry i's productivity innovation, $\mu_{i,t}$ is estimated by filtering out the economy's Solow residual from industry i's Solow residual.[6] To be specific, we assume that an industry's aggregate production function is in its general form, $Y_i(t) = F[K_i(t), L_i(t), A_i(t)]$. The output growth can be decomposed into growth in technology, capital and labor:

$$\frac{\dot{Y}_i(t)}{Y_i(t)} = \frac{F_{A,i}(t)A_i(t)}{Y_i(t)}\frac{\dot{A}_i(t)}{A_i(t)} + \frac{F_{K,i}(t)K_i(t)}{Y_i(t)}\frac{\dot{K}_i(t)}{K_i(t)} + \frac{F_{L,i}(t)L_i(t)}{Y_i(t)}\frac{\dot{L}_i(t)}{L_i(t)}. \tag{B.2.2}$$

With competitive factor markets, $F_{K,i}(t) = R(t)$ and $F_{L,i}(t) = w(t)$. Denote factor shares as $\alpha_{i,K}(t) \equiv \frac{R(t)K_i(t)}{Y_i(t)}$ and $\alpha_{i,L}(t) \equiv \frac{w(t)L_i(t)}{Y_i(t)}$. Denote growth rates of output, capital, and labor as $g_i(t) \equiv \frac{\dot{Y}_i(t)}{Y_i(t)}$, $g_{i,K}(t) \equiv \frac{\dot{K}_i(t)}{K_i(t)}$, $g_{i,L}(t) \equiv \frac{\dot{L}_i(t)}{L_i(t)}$. The contribution of technology to growth, or the Solow residual is $x_i(t) = \frac{F_{A,i}(t)A_i(t)}{Y_i(t)}$. We obtain

$$x_i(t) = g_i(t) - \alpha_{i,K}(t)g_{i,K}(t) - \alpha_{i,L}(t)g_{i,L}(t). \tag{B.2.3}$$

[6] Gabaix [96] applies a similar approach to estimate firm-level granular residuals by taking out the mean growth rate of the sample from each firm's growth rate.

In discrete time, the analog of equation (B.2.3) is

$$x_i^{t\to t+1} = g_i^{t\to t+1} - \bar{\alpha}_{i,K}^{t\to t+1} g_{i,K}^{t\to t+1} - \bar{\alpha}_{i,L}^{t\to t+1} g_{i,L}^{t\to t+1}, \tag{B.2.4}$$

where $\bar{\alpha}_{i,K}^{t\to t+1} \equiv (\alpha_{i,K}(t) + \alpha_{i,K}(t+1))/2$ and $\bar{\alpha}_{i,L}^{t\to t+1} \equiv (\alpha_{i,L}(t) + \alpha_{i,L}(t+1))/2$ are average capital and labor shares from t to $t+1$. The productivity innovation μ_t^i is estimated as the residual of the following regression

$$x_i(t) = \beta_i x(t) + \mu_{i,t}, \tag{B.2.5}$$

where $x(t)$ refers to the economy's Solow residual, which is estimated using a similar approach. Our estimation of Solow residuals is conducted using annual series from BLS and BEA. The labor growth rate, $g_{i,L}(t)$, is constructed using the number of hours-worked from the BLS Labor Productivity and Costs (LPC) database. The capital growth rate, $g_{i,K}(t)$, is constructed by adding up the net stock of private equipment (Table 3.2E) and the net stock of private structure (Table 3.2S) from BEA's fixed assets accounts. The output growth rate, $g_i(t)$, is constructed using value-added by industry from BEA's industry accounts. The labor share, $\alpha_{i,L}(t)$, is constructed by dividing total compensation of employees by value-added. Both variables are available from BEA's Gross-Domestic-Product-(GDP)-by-Industry Data.

We construct industry-level price indexes using PPI (Discontinued Industry Data – (NAICS basis)). There are two issues that complicate our exercise. First, PPI provides price indexes at a finer level than the industry categories we use. Second, for each industry, the price index for some sub-groups are not recorded. To deal with these issues, we simply take the average of price indexes for all available sub-groups, relying on the fact that inflation rates across sub-groups within the same industry are highly correlated.

Estimate the Price Change Frequency

Nakamura and Steinsson [188, Table VI] present the median frequencies of price change for 15 major industries over the period 1998 – 2005, which is constructed from the PPI micro dataset. Although industries defined in this paper are based on the SIC code, 14 of them are broadly consistent with our industry categories. For example, the wood product industry in our paper is named lumber and wood product industry in Nakamura and Steinsson [188], and machinery industry is named machinery and equipment there, etc. Denote $FREQ_i$ as the price change frequency in industry i. To obtain the price change frequency for the 4 unknown industries, we estimate a linear specification using the data provided by Nakamura and Steinsson [188]. The estimation result is (with standard errors in parentheses)

$$log(FREQ_i) = \underset{[0.7206]}{-0.8033} + \underset{[0.1333]}{-0.2133} log(\mid \alpha_{i,1} \mid) + \underset{[0.3879]}{1.1988} log(\bar{q}_i) + \underset{[0.0008]}{0.0045} H_i, \qquad Adj.R^2 = 0.7119, \tag{B.2.6}$$

where $\bar{q}_i$ is the average industry i's Tobin's q over the period 1998 – 2012. H_i is industry i's

Table B.2: Estimated Price Change Frequency at the Industry-level

Our sample		Sample of Nakamura and Steinsson [188]	
Industry name	**Freq.**	**Industry name**	**Freq.**
	Durable goods		
Wood product	1.3	Lumber and wood products	1.3
Nonmetallic mineral products	3.6	Nonmetallic mineral products	4.1
Primary metals	3.7	Metals and metal products	3.8
Fabricated metal products	2.0	N/A	
Machinery	3.9	Machinery and equipment	3.7
Computer and electronic products	16.7	N/A	
Electrical equipment, appliances, and components	8.7	N/A	
Transportation	16.6	Transportation equipment	27.3
Furniture and related products	2.3	Furniture and Household Durables	5.1
Miscellaneous manufacturing	9.5	Miscellaneous manufacturing	16.5
	Non-durable goods		
Food and beverage and tobacco products	24.7	Processed foods and feeds	26.3
Textile mills and textile product mills	5.3	Textile products and apparel	2.3
Apparel and leather and allied products	6.3	Hides, skins, leather, and related products	3.8
Paper products	9.9	Pulp, paper and allied products	4.4
Printing and related support activities	2.9	N/A	
Petroleum and coal products	33.2	Fuels and related products and power	48.7
Chemical products	9.4	Chemicals and allied products	6.1
Plastics and rubber products	2.6	Rubber and plastic products	3.2

Herfindahl index. Table B.2 exhibits the estimated price change frequencies for the 18 industries. The estimation result is pretty reliable due to a large adjusted R^2. Moreover, the coefficient on the log price sensitivity of productivity shocks is significant at the 10% level, while the coefficients on the log Tobin's average q and the Herfindahl index are all significant at the 1% level.

B.3 Additional Numerical Results

B.3.1 The Impact of Financing Costs

In this section, we analyze the impact of the fixed and variable financing costs on the firm's enterprise value, financing, payout, investment, and price setting decisions.

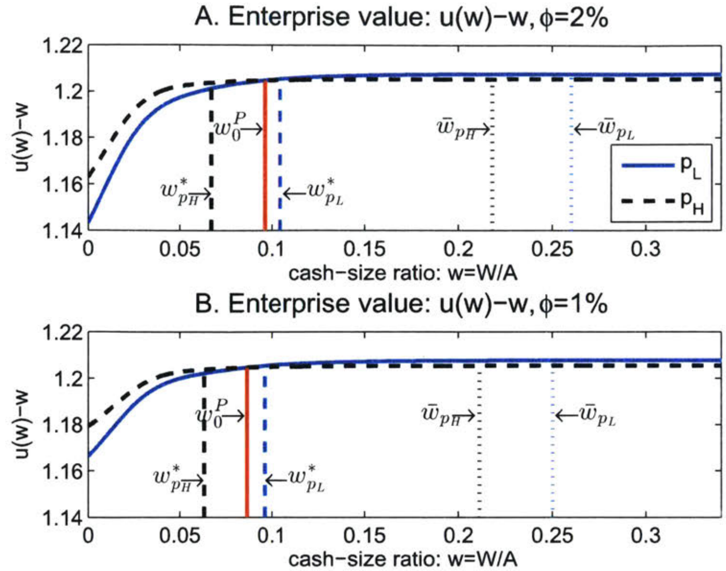


Figure B.1: The firm's enterprise value for different values of the fixed financing cost. Panel A is plotted for $\phi = 2\%$, and Panel B is plotted for $\phi = 1\%$.

The Fixed Financing Cost Figure B.1 presents the firm's normalized enterprise value when the fixed financing cost ϕ varies. When ϕ is increased from 1% to 2%, the enterprise value slightly decreases. Moreover, the payout boundaries and the issuance amounts shift to the right for both p_L and p_H, indicating that the firm will hold more cash on its balance sheet. This is because a higher fixed financing cost increases the marginal value of cash by dampening the external financing channel. The price resetting boundary shifts to the right when the fixed financing cost increases. This implies that the firm is more likely to set its price to p_H when external financing costs are high. The response of investment is similar to Bolton, Chen and Wang [41], i.e. investment is lower when the fixed financing cost increases (see Figure B.2).

Panel A of Figure B.3 shows that the normalized firm value increases with the fixed financing cost. The underlying force is similar to the one driving the impact of price stickiness. That is, the firm increases its cash holdings (as shown in panel B) due to a higher marginal value of cash when the fixed financing cost increases.[7] Panel C shows that the normalized enterprise value (or the average q) decreases as external financing becomes more costly.

The Variable Financing Cost Figure B.4 shows the response of the normalized enterprise value when the variable financing cost varies. Similar to the effect of the fixed financing cost, the normalized enterprise value is lower when the variable financing cost increases.

The firm is less likely to payout when γ increases as the marginal value of cash is higher.

[7]Note that the results presented in Panel A of Figure B.3 are about the average firm's value when the firm is in steady state. A larger fixed financing cost indeed reduces the firm's value at the impact of the change in financing costs.

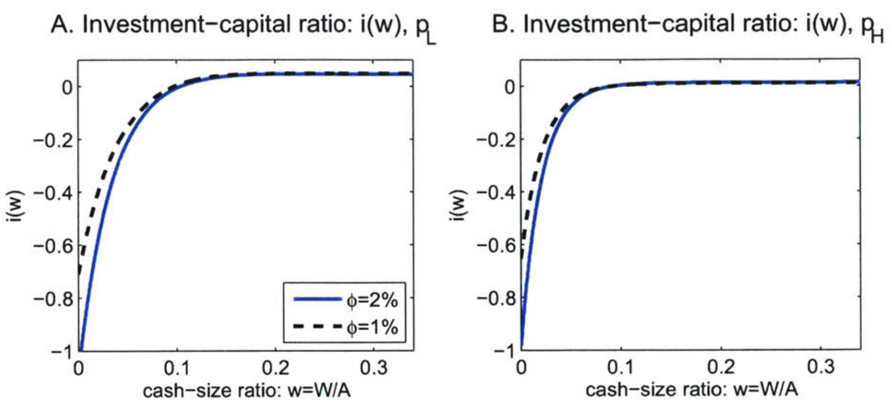


Figure B.2: The firm's investment for different values of the fixed financing cost. Panel A/B plots investment for the p_L/p_H case.

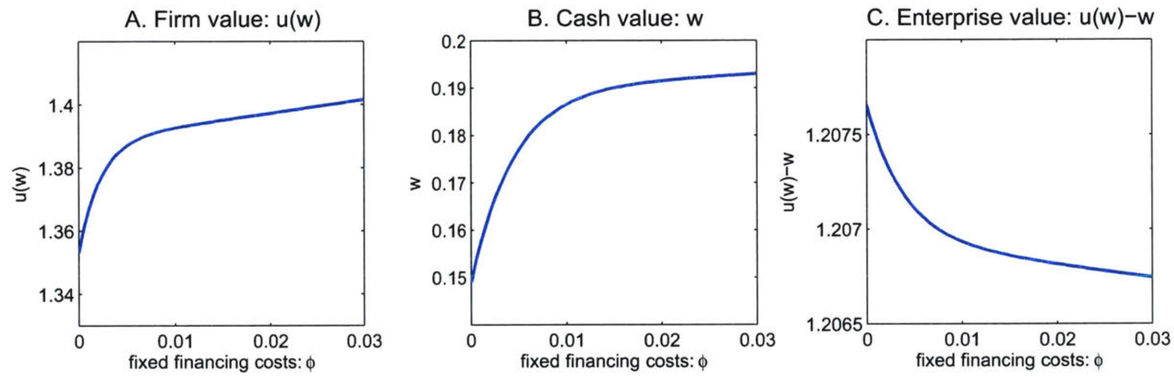


Figure B.3: The average steady-state normalized firm value, cash holdings and normalized enterprise value for different values of the fixed financing cost.

This can be seen in Figure B.4, which shows that the payout boundaries (the vertical dotted lines) shift to the right from panel A ($\gamma = 1\%$) to panel C ($\gamma = 16\%$). The optimal issuance amounts are located at the points where the marginal value of cash is equal to one plus the variable financing cost γ. Therefore, when γ increases, the firm issues less equity when hitting the financing boundary (the vertical dashed lines shift to the left).

Notice that the relative locations of the issuance amounts and the price resetting boundary (the vertical solid line) is indeterminate. Panel A plots the case for $\gamma = 1\%$. The issuance amount for both the firm with p_L and the firm with p_H are to the right of the price resetting boundary ($w^*_{p_L} > w^*_{p_H} > w^P_0$). This suggests that when the firm runs out of cash and issues equity, it takes advantage of the low variable financing cost and raises sufficient external funds in order to not distort its pricing behavior (i.e. set the price to p_L after external financing). In panel B, the variable financing cost is set at $\gamma = 6\%$. In this case, only the firm with p_L raises sufficient cash through external financing to go beyond the price resetting boundary ($w^*_{p_H} < w^P_0 < w^*_{p_L}$). In panel C, the variable financing cost is high, $\gamma = 16\%$. Both the firm with p_L and p_H remain in the high price

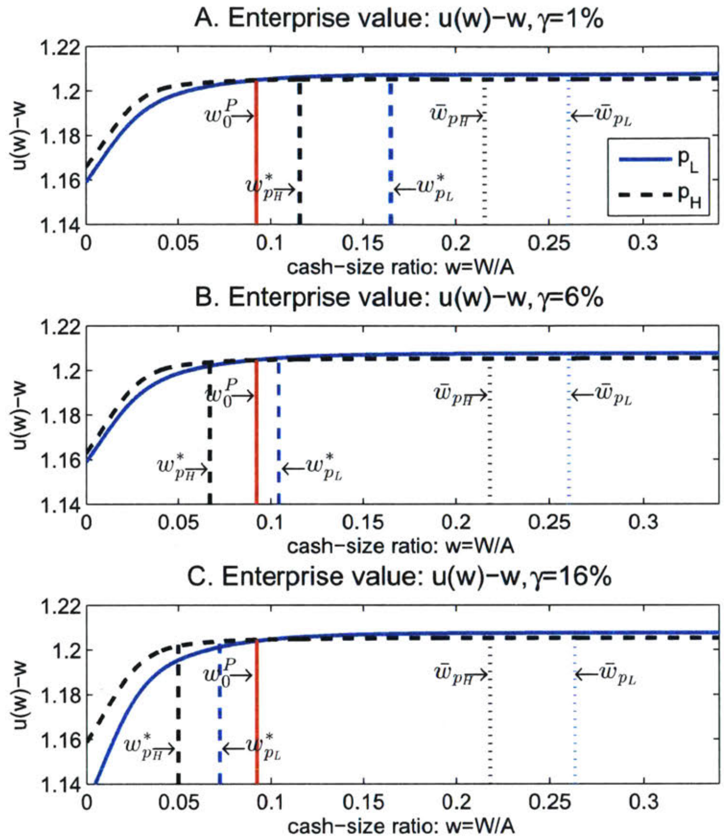


Figure B.4: The firm's enterprise value for different values of the variable financing cost. Panel A is plotted for $\gamma = 1\%$, panel B is plotted for $\gamma = 6\%$, and panel C is plotted for $\gamma = 16\%$.

region after external financing ($w_0^P > w_{p_L}^* > w_{p_H}^*$).

B.3.2 Countercyclical Markups

In this section, we present the economy's transitional dynamics to illustrate that our model provides new forces which generate countercyclical markups. Since our model is a partial equilibrium model, without taking into account the feedback effect from firms' optimal decisions on industry average prices or on the equilibrium interest rate, our analysis about countercyclical markups is only intended to be illustrative. A more systematic analysis of the cyclicality of markups based on a general equilibrium model with heterogeneous firms is left for future research.

Demand Shocks As elaborated in Section 2.3, the firm's desired product price is p_H when it is financially constrained, since setting a higher price increases the incremental operating revenue. Consider an economy-wide negative demand shock which reduces all firms' incremental operating revenue. Consequently, firms' have to run down their cash holdings, and some of them become financially constrained. To prevent costly external financing, these financially constrained firms raise their product prices, resulting in a higher aggregate price level. This generates countercyclical markups/desired markups.

To illustrate this idea, we consider a simple scenario where the steady-state economy, under the benchmark calibration, is hit by an unexpected aggregate negative demand shock which lasts for one quarter.[8] We assume that for firm i, the nominal shock σdZ_t^i in equation (2.2.2) consists of an aggregate component $\sigma_A dZ_t^A$ and an idiosyncratic component $\sigma_I dZ_t^i$, i.e., $\sigma dZ_t^i = \sigma_A dZ_t^A + \sigma_I dZ_t^i$ and $\sigma^2 = \sigma_A^2 + \sigma_I^2$. dZ_t^A and dZ_t^i are standard Brownian motion and are independent from each other. We generate the aggregate demand shock by assuming $dZ_t^A = -4$ in the first quarter, and $dZ_t^A = 0$ thereafter. Hence, the first quarter of the economy is hit by a negative aggregate demand shock. Figure B.5 shows the transitional dynamics for the average price and cash holdings of a large number of firms when $\sigma_I = \sigma_A = \sigma/\sqrt{2}$. During the first quarter, the negative aggregate demand shock reduces operating revenue and most of the firms have to run down cash to maintain investment. Moreover, firms who find themselves financially constrained increase their prices (i.e. set their prices to p_H), which drives up the economy's price level and generates a higher average markup since the marginal cost of production is unchanged. When the demand shock subsides at the end of the first quarter, financially constrained firms start to accumulate cash and reset their prices to p_L, allowing the economy's price level to decrease and gradually converge to its steady-state value.

Notice that the underlying mechanism for the countercyclical markups is different from standard New Keynesian models. In New Keynesian models, a negative demand shock reduces the marginal cost, but as firms cannot adjust prices immediately, the markups would be temporarily high. This relies on the assumptions that the marginal cost of production is increasing in demand and that firms face nominal rigidity. In our model, these two assumptions are not necessary, and we emphasize the roles played by the customer market and financial frictions in generating countercyclical markups. Moreover, customer base can be considered as a force that generates real rigidity. Therefore, the desired markup is also countercyclical in our model, while it is constant in standard New Keynesian models. This addresses the concern raised by Blanchard [38]. Moreover, in contrast to the classical customer market models without financial frictions [e.g. 194, 195], the countercyclical markups are more robust in our model. As pointed out by Klemperer [152], the customer market model can generate both countercyclical and procyclical markups depending on the nature of demand shocks. If positive demand shocks come from existing customers increasing their demand, then firms would be motivated to increase their prices to profit from locked-in customers. But, if demand shocks come from the arrival of new customers, then firms would decrease their prices to invest in customer base. In a related paper, Chevalier and Scharfstein [60]

[8]The steady state of the economy is defined as the state where the joint distribution of firms' cash holdings and product prices is invariant over time. To obtain this, we simulate 500,000 firms for 10 years.

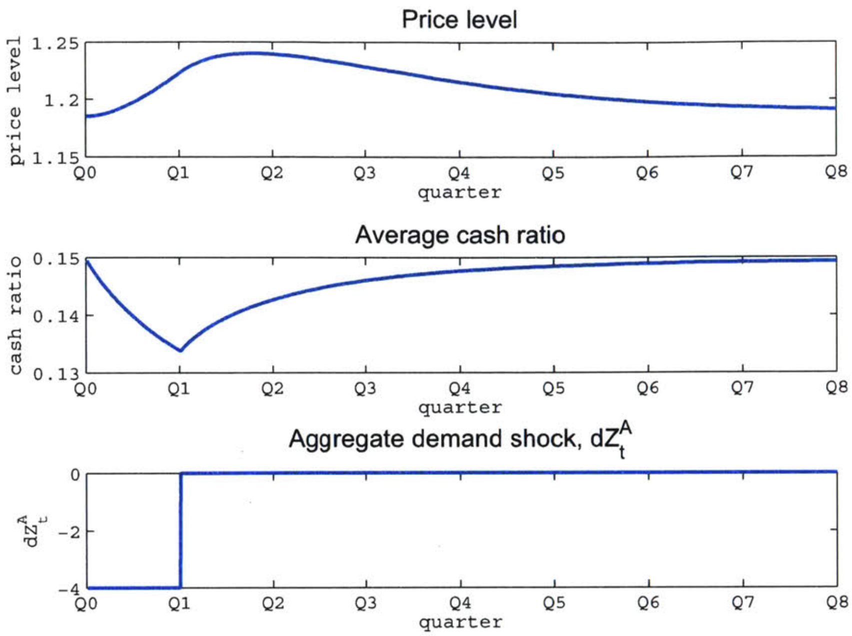


Figure B.5: The dynamics of the average cash ratio and the price level after a demand shock.

show that the customer market model without financial frictions can only generate procyclical markups.[9] In our model, by introducing financial frictions, the endogenous marginal value of liquidity can interact with the customer market, which generates an additional force pushing towards countercyclical markups.

Financial Shocks As shown in Figure B.1, the price resetting boundary shifts to the right when the fixed financing cost increases. This implies that the firm is more likely to set its price to p_H when external financing costs are high. Since most recessions are associated with a credit crunch in the financial market [64], the increase in external financing costs provide an additional force pushing towards countercyclical markups.

To show this, we start from the steady-state economy with the benchmark calibration and consider an unexpected one-quarter financial shock which increases the fixed financing cost from 2% to 3%. The results are shown in Figure B.6.

The left three panels of Figure B.6 are based on the benchmark value of parameter $\xi = 2.8$. It shows that firms tend to raise their prices to accumulate more cash when the financial shock arrives in order to avoid costly external financing, resulting in an increase in the average price level. When the financial shock vanishes, both the price level and the average cash ratio converge to their steady-state values. Note that during the first quarter, the price level increases only gradually because firms are facing nominal rigidity. By contrast, when the price is more flexible,

[9] This is considered as the benchmark model which motivates them to introduce financial frictions in order to generate countercyclical markups. However, in fact, enriching their benchmark model with multiple periods enables countercyclical markups when the increase in demand is from new customers or when demand shocks are persistent.

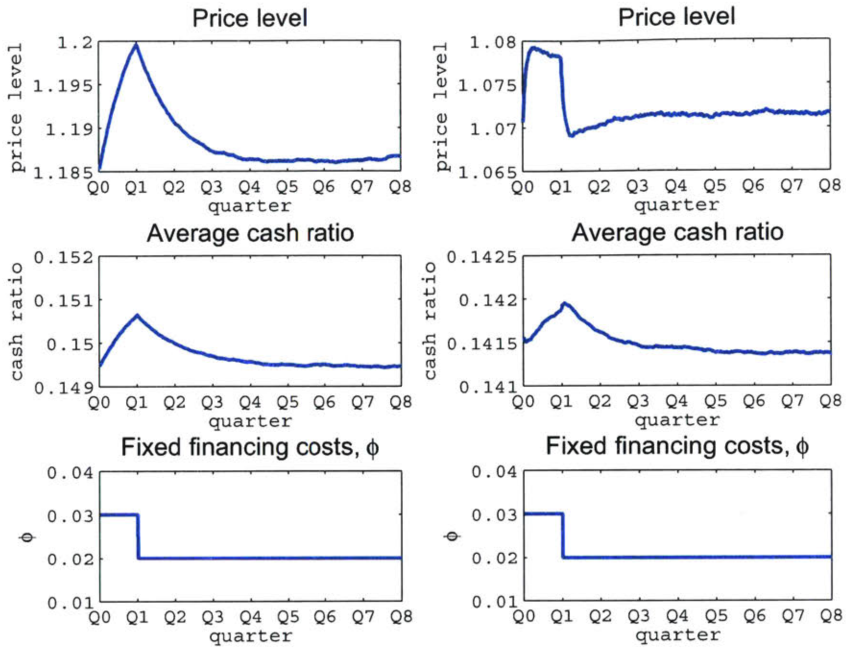


Figure B.6: The dynamics of the average cash ratio and the price level after a financial shock. The left three panels are for the case with $\xi = 2.8$, the right three panels are for the case with $\xi = 40$.

the increase in the price level is the highest at the impact of the financial shock, and the price level gradually decreases as firms accumulate more cash and become less financially constrained. This is shown in the right three panels of Figure B.6, where the Calvo parameter is set to $\xi = 40$. Therefore, in our model, nominal rigidity essentially dampens the response of the price level to a financial shock, and the immediate increase in the price level following a negative financial shock is larger when the price is more flexible. Hence, the mechanism delivered by our model has the potential to explain the lack of deflationary pressure during the "Great Recession" in the United States not by appealing to nominal rigidity or large unobservable shocks to the markup.

B.3.3 Price Setting and Product Market Characteristics

Following the discussion of equation (2.2.10), when κ is fixed, parameter ν captures how sensitive customer flows are to the change in product prices. By varying the value of ν, our model is able to capture markets with different degrees of competition, and further shed light on the firm's behavior when the characteristics of the underlying product market change. Intuitively, a larger ν implies that customer base is more sensitive to the price, reflecting a competitive product market. On the contrary, a smaller ν captures the feature of a customer-based product market, which is associated with a higher degree of consumption inertia (or higher switching/information costs). In a market with a relatively small ν, the firm loses customer base slowly even if a high price is temporarily charged.

Our model implies that firms in a customer-based product market tend to set their prices to

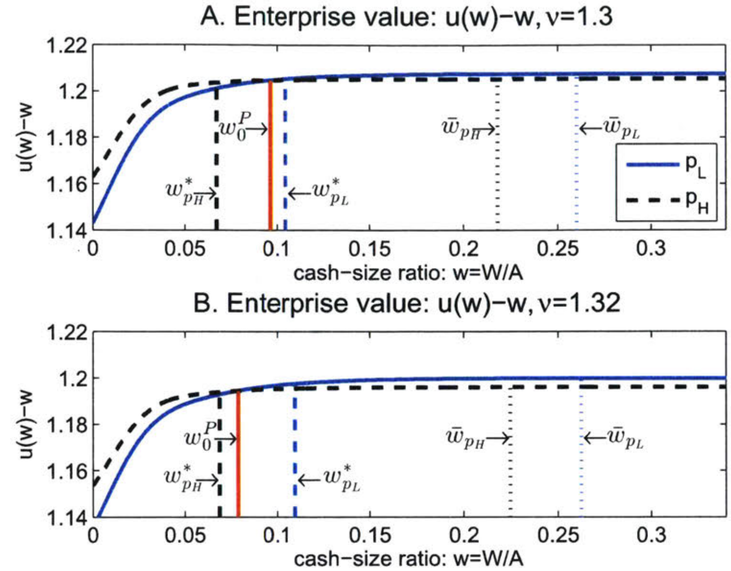


Figure B.7: The firm's enterprise value in different product markets. Panel A is plotted for $\nu = 1.3$, and Panel B is plotted for $\nu = 1.32$.

p_H during a liquidity-constrained period. This is because setting a high price imposes less cost on the firm if the underlying product market is more customer based—the demand is less elastic in the short run and customer base only decreases slowly as consumers are reluctant to change their consumption habits or switch to other brands. However, setting a high price benefits the firm through the current profit channel by increasing short-term operating revenue. Thus, the firm is more inclined to set its price to p_H when the cash-size ratio is low.

Figure B.7 illustrates this result. As parameter ν decreases from 1.32 to 1.3, the price resetting boundary (vertical solid line) shifts to the right, indicating that the firm is more likely to set its price to p_H when experiencing a liquidity problem. This is consistent with the empirical findings in Gilchrist et al. [102], which show that firms operating in a more customer-based market (as captured by high SG&A expenses or advertising expenses[10]) raise their product prices relative to the industrial average prices during the recent U.S. financial crisis.

B.3.4 The Volatility of Productivity Shocks

In this section, we discuss the impact of the volatility of productivity shocks on the firm's enterprise value, investment, payout boundaries, optimal issuance amounts, and the price resetting boundary. As shown in Figure B.8 and B.9, when productivity shocks become more volatile (parameter σ increases from 12% to 14%), the firm has a lower enterprise value and lower investment. Moreover, the payout boundaries $\overline{w}_{p_L}$ and $\overline{w}_{p_H}$, and the optimal equity issuance amounts $w^*_{p_L}$ and $w^*_{p_H}$ shift to the right, reflecting a higher marginal value of cash. The price

[10]Gourio and Rudanko [109] use SG&A ratios as an indicator for a customer market environment.

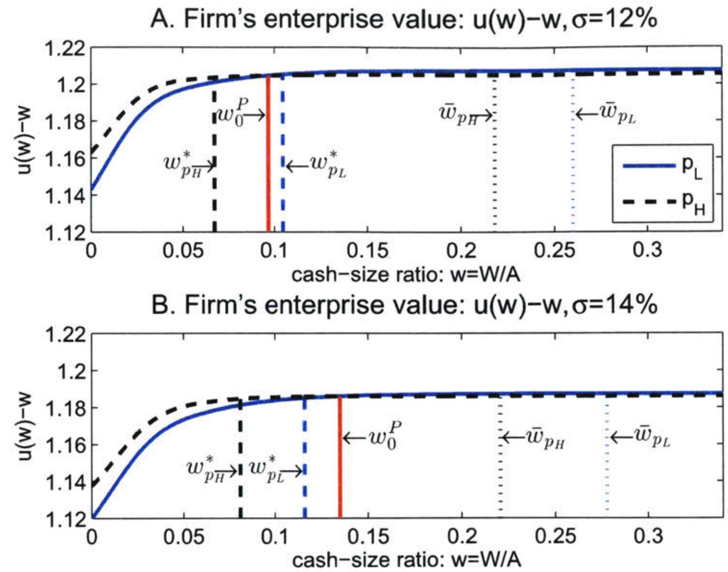


Figure B.8: The firm's enterprise value for different volatilities of productivity shocks. Panel A is plotted for $\sigma = 12\%$, and Panel B is plotted for $\sigma = 14\%$.

resetting boundary (w_0^P) shifts to the right, indicating that the firm is more likely to raise its price because a higher volatility of productivity shocks increases the chance of hitting the financing boundary.

B.4 A Model with Menu Costs

In our baseline model, price stickiness is modeled as in Calvo [52]. In this section, we modify this aspect by assuming that price adjustment is totally under the control of the firm's manager. However, whenever the price is adjusted, a fixed "menu cost", ζ is incurred. This modification enables us to quantitatively measure the direct cost of price stickiness, and to address the concern that the qualitative predictions of the baseline model on the firm's pricing strategy is driven by the mechanical Calvo pricing rule.

Introducing a fixed cost of price adjustment changes the HJB equation (2.2.14) to the following

$$\begin{aligned} rU(A,W,p) = \max_{I,p^+\in\{p_L,p_H\}} & \left[\alpha h(p) + (1-\alpha)(I/K-\delta)\right] AU_A \\ & + \left[(r-\lambda)W + A(p-\bar{c})\mu(p) - \Gamma(I,K,A)\right] U_W + \frac{1}{2}\sigma^2 A^2 U_{WW} \\ & + \left[U(A,W,p^+) - U(A,W,p)\right] - \zeta \mathbb{1}_{p^+\neq p}, \end{aligned} \tag{B.4.1}$$

where $\mathbb{1}_{p^+\neq p}$ is an indicator function, which equals one if $p^+ \neq p$ and zero otherwise. The financing and payout boundary conditions are the same as the baseline model.

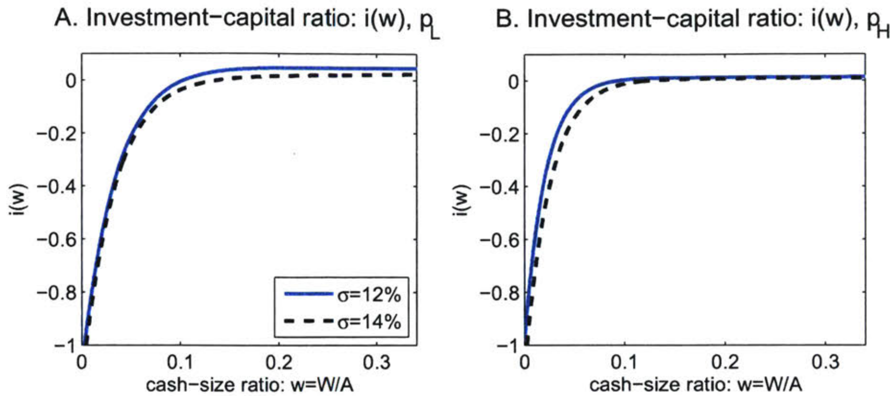


Figure B.9: The firm's investment for different volatilities of productivity shocks. Panel A/B plots investment for the p_L/p_H case. Solid/Dashed lines refer to the $\sigma = 12\%/\sigma = 14\%$ case.

B.4.1 Quantitative Results

We set the menu cost parameter $\zeta = 0.002$ to match the firm's average normalized enterprise value in the menu cost model with the one in the baseline model. The other parameters are set to be the same as the baseline model. Quantitatively, this implies that the menu cost amounts to about 0.15% of the average normalized enterprise value, or 1% of the firm's average cash holdings.

Figure B.10 plots the normalized enterprise value for the menu cost model. Similar to the baseline model (Figure 2-2), the normalized enterprise value of the firm with p_L is higher than the firm with p_H when the cash-size ratio is high, and the reverse is true when the cash-size ratio is low. However, the difference in the normalized enterprise value between the firm with p_L and the firm with p_H is bounded by the amount of the menu cost. This is intuitive since whenever the difference is larger than the menu cost, the firm will choose to reset its price immediately, which ensures that the resulting difference in the enterprise value cannot exceed the menu cost.

There are two price resetting boundaries in the menu cost model, as marked by the vertical solid lines in the figure. The right boundary ($w_0^{p_H \to p_L} = 0.13$) captures the threshold of the cash-size ratio at which the firm switches from p_H to p_L, while the left boundary ($w_0^{p_L \to p_H} = 0.05$) captures the threshold where the firm switches from p_L to p_H. When the cash-size ratio is between the two boundaries ($w_0^{p_L \to p_H} \leq w \leq w_0^{p_H \to p_L}$), the firm is in the "inaction" region and does not change its price at all because the benefit obtained from resetting the price is smaller than the menu cost. When the cash-size ratio is below the left price resetting boundary ($w < w_0^{p_L \to p_H}$), the firm always sets p_H, while when the cash-size ratio is above the right price resetting boundary ($w > w_0^{p_H \to p_L}$), the firm always sets p_L.

Compared to the baseline model, the menu cost model implies a one-to-one mapping from the cash-size ratio to the product price out of the inaction region. While in the baseline model, the mapping only exists between the cash-size ratio and the "desired" product price. Whether the firm can achieve the desired price or not depends on the arrival of price resetting opportunities,

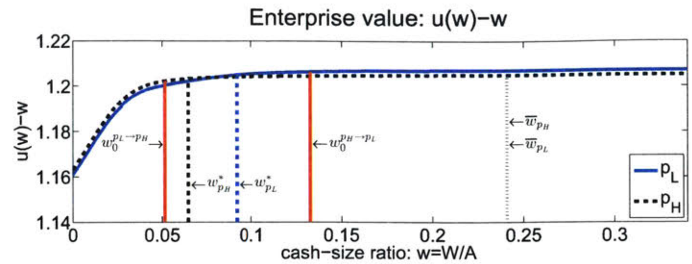


Figure B.10: The enterprise value in the model with menu costs.

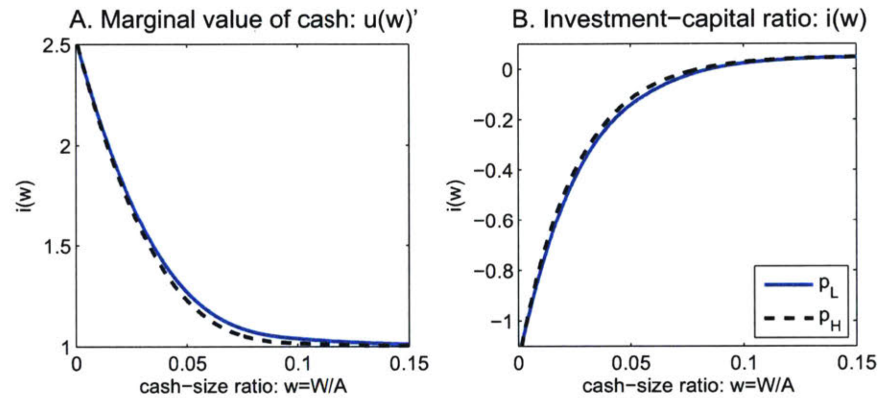


Figure B.11: Marginal value of cash and investment in the model with menu costs.

which is out of the manager's control. Within the inaction region, the firm's price is indeterminate, depending on the inherited price when the firm first enters the region. The firm facing a larger menu cost (or a stickier price) is associated with a wider inaction region and the shift in issuance amounts and payout boundaries are exactly consistent with the implications of the benchmark model (see Section 2.3.3) following the same intuitions.

Figure B.11 plots the marginal value of cash (panel A) and the optimal investment-capital ratio (panel B). Consistent with the baseline model, the marginal value of cash is high for the firm with p_L. However, the difference only shows up within the inaction region. Outside this region, the difference in the enterprise value is locked by the menu cost, resulting in the same marginal value of cash irrespective of the firm's price. Similarly, the firm with p_H has a higher investment ratio (or a smaller disinvestment ratio) compared to the firm with p_L only within the inaction region, reflecting the impact of the current profit channel.

B.5 Numerical Methods

Due to the large non-linearity introduced by price stickiness and financing costs, the solution of the coupled ODE problem is not robust and subject to large numerical errors. To mitigate this problem, we reformulate the continuous time coupled ODEs into a discrete recursive problem, which is solved using a standard dynamic programming method.

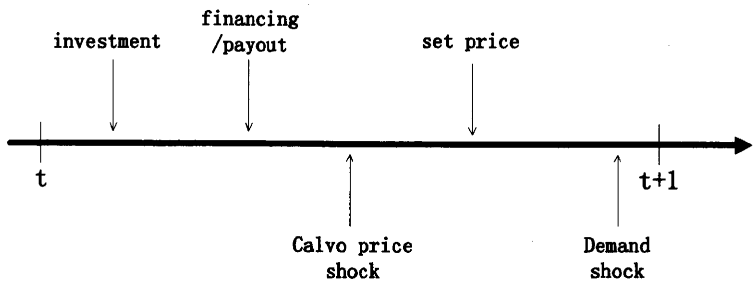


Figure B.12: Timing assumption for the recursive problem.

Time is discrete with interval Δ, The aggregate demand shock Q_t is i.i.d. and follows a normal distribution, $N(\mu\Delta, \sigma^2\Delta)$. In period t, the state variables for the recursive problem are cash W_t and effective size A_t.

Let $U^L(W, A)$ and $U^H(W, A)$ be the value functions for the low price and high price, respectively. Price resetting opportunities arrive with probability $\xi\Delta$. We assume that price resetting opportunities arrive before the realization of demand shocks, Q. The firm makes investment and financing decisions before they know whether they can adjust the price in the current period. After the realization of price resetting opportunities, if the firm obtains the chance to reset its price, either p_L or p_H will be chosen to maximize the objective function. Otherwise, the firm has to stick to the price inherited from the previous period. Demand shocks are realized after price setting decisions are made. Figure B.12 illustrates the timing.

Note that the timing assumption used here is to simplify calculations. Alternatively, we could assume that the firm makes optimal decisions after the realization of demand shocks. However, this would complicate computations because now the firm's optimal decisions also depend on the realized value of demand shocks [see 82]. When Δ approaches zero, the timing assumption does not matter, and the solution will converge to the solution of the coupled ODEs. In our calculation, we set $\Delta = 0.02$. The resulted numerical errors are negligible.

The recursive formulation for $U^L(W, A)$ is

$$U^L(W,A) = \max_{i,U} U + \frac{\xi\Delta}{1+r\Delta}\max\{E[U^L(W',A')|Q], E[U^H(W'',A'')|Q]\} + \frac{1-\xi\Delta}{1+r\Delta}E[U^L(W',A')|Q],$$

subject to

$$\begin{aligned}
W' &= W + p_L AQ - g(i)\Delta A + (r-\lambda)\Delta W - U - ((\phi A - \gamma U)1_{U<0}), \\
A' &= (1+h(p_L)\Delta)^\alpha[(1-\delta\Delta) + i\Delta]^{1-\alpha}A, \\
W'' &= W + p_H AQ - g(i)\Delta A + (r-\lambda)\Delta W - U - ((\phi A - \gamma U)1_{U<0}), \\
A'' &= (1+h(p_H)\Delta)^\alpha[(1-\delta\Delta) + i\Delta]^{1-\alpha}A.
\end{aligned}$$

Similarly, for $V^h(W, A)$, the recursive formulation is

$$
\begin{aligned}
U^H(W,A) &= \max_{i,U} U + \frac{\xi\Delta}{1+r\Delta}\max\{E[U^L(W',A')|Q], E[U^H(W'',A'')|Q]\} \\
&+ \frac{1-\xi\Delta}{1+r\Delta}E[U^H(W'',A'')|Q],
\end{aligned}
$$

subject to

$$
\begin{aligned}
W' &= W + p_L AQ - g(i)\Delta A + (r-\lambda)\Delta W - U - ((\phi A - \gamma U)1_{U<0}), \\
A' &= (1+h(p_L)\Delta)^\alpha[(1-\delta\Delta)+i\Delta]^{1-\alpha}A, \\
W'' &= W + p_H AQ - g(i)\Delta A + (r-\lambda)\Delta W - U - ((\phi A - \gamma U)1_{U<0}), \\
A'' &= (1+h(p_H)\Delta)^\alpha[(1-\delta\Delta)+i\Delta]^{1-\alpha}A.
\end{aligned}
$$

Since $U(W,A)$ is homogeneous in A, we write $U^L(W,A) = u^L(w)A$, $U^H(W,A) = u^H(w)A$, and define $w = \frac{W}{A}$, and $u = \frac{U}{A}$.

The recursive problems for the normalized value functions, u^L and u^H are

$$
\begin{aligned}
u^L(w) &= \max_{i,u} u + \tfrac{[(1-\delta\Delta)+i\Delta]^{1-\alpha}}{1+r\Delta}\{(1-\xi\Delta)(1+h(p_L)\Delta)^\alpha E[u^L(w')|Q] + \\
&\xi\Delta\max\{(1+h(p_L)\Delta)^\alpha E[u^L(w')|Q], (1+h(p_H)\Delta)^\alpha E[u^H(w'')|Q]\}\},
\end{aligned}
$$

subject to

$$
\begin{aligned}
w' &= \frac{w + p_L Q - g(i)\Delta + (r-\lambda)\Delta w - u - ((\phi - \gamma u)1_{u<0})}{[(1-\delta\Delta)+i\Delta]^{1-\alpha}(1+h(p_L)\Delta)^\alpha}, \\
w'' &= \frac{w + p_H Q - g(i)\Delta + (r-\lambda)\Delta w - u - ((\phi - \gamma u)1_{u<0})}{[(1-\delta\Delta)+i\Delta]^{1-\alpha}(1+h(p_H)\Delta)^\alpha},
\end{aligned}
$$

and

$$
\begin{aligned}
u^H(w) &= \max_{i,u} u + \tfrac{[(1-\delta\Delta)+i\Delta]^{1-\alpha}}{1+r\Delta}\{(1-\xi\Delta)(1+h(p_H)\Delta)^\alpha E[u^H(w'')|Q] + \\
&\xi\Delta\max\{(1+h(p_L)\Delta)^\alpha E[u^L(w')|Q], (1+h(p_H)\Delta)^\alpha E[u^H(w'')|Q]\}\},
\end{aligned}
$$

subject to

$$
\begin{aligned}
w' &= \frac{w + p_L Q - g(i)\Delta + (r-\lambda)\Delta w - u - ((\phi - \gamma u)1_{u<0})}{[(1-\delta\Delta)+i\Delta]^{1-\alpha}(1+h(p_L)\Delta)^\alpha}, \\
w'' &= \frac{w + p_H Q - g(i)\Delta + (r-\lambda)\Delta w - u - ((\phi - \gamma u)1_{u<0})}{[(1-\delta\Delta)+i\Delta]^{1-\alpha}(1+h(p_H)\Delta)^\alpha}.
\end{aligned}
$$

Table B.3 lists the parameters for the discretization of state space.

Table B.3: Discretization of state space

Parameter	Value	Description
Δ	0.017	Time interval
n_w	2000	Number of cash grid points (equally spaced)
n_i	1000	Number of investment grid points (equally spaced)
n_u	2000	Number of financing grid points (equally spaced)
$[\underline{w}\ \bar{w}]$	[0 0.50]	Cash range
$[\underline{i}\ \bar{i}]$	$[-1.00\ 0.50]$	Investment range
$[\underline{u}\ \bar{u}]$	$[-0.20\ 0.20]$	Financing range

Appendix C

Appendix for Distinguishing Constraints on Financial Inclusion and Their Impact on GDP, TFP, and Inequality

C.1 Proofs

C.1.1 Proof of Proposition 4

For any level of capital, the optimal labor employed by entrepreneurs is given by the first order condition of problem (3.3.4),

$$l = [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}. \tag{C.1.1}$$

Plugging l into (3.3.4), entrepreneurs solve

$$\begin{aligned} \pi^S(b,z) = \max_k \quad & (1-p)[(\tfrac{(1-\nu)(1-\alpha)}{w} z)^{\frac{1}{\alpha(1-\nu)+\nu}} w \tfrac{v+\alpha(1-\nu)}{(1-\nu)(1-\alpha)} k^{\frac{\alpha(1-\nu)}{\nu+\alpha(1-\nu)}} - \delta k + k] \\ & + p\eta(1-\delta)k + (1+r^d)(b-k), \\ \text{subject to} \quad & k \ \leq \ b. \end{aligned}$$

Solving this problem without imposing the wealth constraint, the unconstrained capital demand is

$$\tilde{k}^S(z) = [\frac{1-p}{r^d + (1-p)\delta - p\eta(1-\delta) + p} \frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)}{w} z)^{\frac{1}{\nu}}. \tag{C.1.2}$$

Since end-of-period wealth $\pi^S(b,z)$ is increasing in k for $k \leq \tilde{k}^S(z)$, the optimal capital for the constrained problem is

$$k^*(b,z) = \min\{b, \tilde{k}^S(z)\}. \tag{C.1.3}$$

C.1.2 Proof of Lemma 2

First we compute the default boundary $\bar{\lambda}$. For entrepreneurs with no default risk, the recovered capital when production fails plus the amount of collateral (including interest earnings) should be higher than the face value of the loan. Therefore

$$\eta(1-\delta)k + (1+r^d)(b-\psi) \geq \Omega. \tag{C.1.4}$$

When condition (C.1.4) is satisfied, the zero profit condition (3.3.7) implies that $\Omega = (1+r^d)k$. Substituting this into (C.1.4), we obtain

$$\eta(1-\delta)k + (1+r^d)(b-\psi) \geq (1+r^d)k. \tag{C.1.5}$$

Using the definition of leverage ratio (3.3.6), (C.1.5) can be written as

$$\tilde{\lambda} = \frac{k}{b-\psi} \leq \frac{1+r^d}{1+r^d-\eta(1-\delta)}. \tag{C.1.6}$$

Hence, the default boundary is $\overline{\lambda} = \dfrac{1+r^d}{1+r^d-\eta(1-\delta)}$. Note that limited commitment imposes the constraint, $\tilde{\lambda} \leq \lambda$. To obtain a positive default rate for the model economy, we require $\dfrac{1+r^d}{1+r^d-\eta(1-\delta)} < \lambda$. This determines the range of endogenous interest rates $r^d > \dfrac{\eta(1-\delta)\lambda}{\lambda-1} - 1$.

Next, we compute the lending rate for entrepreneurs with leverage ratio $\tilde{\lambda}$.

If $\tilde{\lambda} \leq \overline{\lambda}$, entrepreneurs do not default. As stated above, the lending rate is equal to the deposit rate, $r^l = r^d$.

If $\tilde{\lambda} > \overline{\lambda}$, entrepreneurs default when production fails and condition (C.1.4) is violated. The zero profit condition (3.3.7) implies that the face value of the loan is

$$\Omega = \frac{(1+r^d)k + p\chi k - p\eta(1-\delta)k - p(1+r^d)(b-\psi)}{1-p}. \tag{C.1.7}$$

The lending rate defined by (3-3) is

$$r^l = \frac{1+r^d+p\chi-p\eta(1-\delta)-p(1+r^d)/\tilde{\lambda}}{1-p} - 1. \tag{C.1.8}$$

Note that the lending rate is discontinuous at $\overline{\lambda}$, $\lim_{\tilde{\lambda}\to\overline{\lambda}_+} r^l = r^d + \frac{p\chi}{1-p} \neq \lim_{\tilde{\lambda}\to\overline{\lambda}_-} = r^d$. This is due to the discontinuity of the optimal verification strategy as described in Proposition 5.

C.1.3 Proof of Lemma 3

For entrepreneurs in the low-leverage region ($\tilde{\lambda} \leq \overline{\lambda}$), default never happens, and entrepreneurs pay the interest rate $r^l = r^d$ regardless of whether production fails or not. Hence, the cost of capital is $R = r^d$.

For entrepreneurs in the high-leverage region ($\tilde{\lambda} > \overline{\lambda}$), when production succeeds, entrepreneurs pay the face value of the loan, Ω; when production fails, entrepreneurs default and pay $\eta(1-\delta)k + (1+r^d)(b-\psi)$. The cost of capital is equal to the expected amount of payment divided by the total amount of borrowing,

$$R = \frac{(1-p)\Omega + p[\eta(1-\delta)k + (1+r^d)(b-\psi)]}{k} - 1. \tag{C.1.9}$$

Substituting the zero profit condition (3.3.7) into (C.1.9), we obtain $R = r^d + p\chi$.

C.1.4 Proof of Proposition 6

In the credit regime, entrepreneurs solve problem (3.3.8) subject to the zero profit condition (3.3.7). This problem is non-convex because in the high-leverage region, banks monitor when production fails, increasing the cost of capital.

We solve the problem facing entrepreneurs of type (b, z) by converting problem (3.3.8) into two convex sub-problems: in one problem, entrepreneurs do not default, and leverage ratios are restricted by $\tilde{\lambda} \leq \overline{\lambda}$. In the other problem, entrepreneurs default when production fails, and leverage ratios are restricted by $\tilde{\lambda} \leq \lambda$.

The wealth function for each sub-problem is convex. The highest end-of-period wealth that can be obtained by entrepreneurs is the upper envelope of the two wealth functions, which is non-convex.

In the following, we first characterize the solution to each sub-problem and then provide the solution to the original problem.

The first sub-problem

In this sub-problem, entrepreneurs do not default. As shown in the proof of Lemma 2, $\Omega = (1+r^d)k$. Problem (3.3.8) can be written as

$$\begin{aligned} \pi^l(b,z) = \max_{k,l} \quad & 1-p)(z(k^{\alpha}l^{1-\alpha})^{1-\nu} - wl - \delta k - r^d k) + p\eta(1-\delta)k + (1+r^d)(b-\psi) - p(1+r^d)k \\ \text{subject to} \quad & k \;\; \leq \;\; \overline{\lambda}(b-\psi). \end{aligned}$$

By applying a similar analysis used in the proof of Proposition 4, we obtain the unconstrained level of capital

$$\tilde{k}^l(z) = [\frac{1-p}{r^d + (1-p)\delta - p\eta(1-\delta) + p}\frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)}{w}z)^{\frac{1}{\nu}}.$$

The optimal amount of capital is

$$k^*(b,z) = \min(\bar{\lambda}(b-\psi), \tilde{k}^l(z)),$$

and the optimal amount of labor is

$$l^*(b,z) = [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^*(b,z)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}.$$

The wealth function is

$$\pi^l(b,z) = (1-p)(z(k^{*\alpha}l^{*1-\alpha})^{1-\nu} - wl^* - \delta k^* - r^d k^*) + p\eta(1-\delta)k^* + (1+r^d)(b-\psi) - p(1+r^d)k^*. \tag{C.1.10}$$

The second sub-problem

In this sub-problem, entrepreneurs default when production fails. As shown in the proof of Lemma 2,

$$\Omega = \frac{(1+r^d)k + p\chi k - p\eta(1-\delta)k - p(1+r^d)(b-\psi)}{1-p}.$$

Substituting this into (3.3.8), entrepreneurs solve

$$\begin{aligned} \pi^h(b,z) = \max_{k,l} \quad & (1-p)[z(k^\alpha l^{1-\alpha})^{1-\nu} - wl + (1-\delta)k + (1+r^d)(b-\psi)] \\ & -[(1+r^d)k + p\chi k - p\eta(1-\delta)k - p(1+r^d)(b-\psi)] \\ \text{subject to} \quad & k \ \leq \ \lambda(b-\psi). \end{aligned}$$

Similarly, we obtain the unconstrained level of capital

$$\tilde{k}^h(z) = [\frac{1-p}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p}\frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)}{w}z)^{\frac{1}{\nu}}.$$

The optimal amount of capital is

$$k^*(b,z) = \min(\lambda(b-\psi), \tilde{k}^h(z)),$$

and the optimal amount of labor is

$$l^*(b,z) = [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^*(b,z)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}.$$

The wealth function is

$$\begin{aligned} \pi^h(b,z) = \quad & (1-p)[z(k^{*\alpha}l^{*1-\alpha})^{1-\nu} - wl^* + (1-\delta)k^* + (1+r^d)(b-\psi)] \\ & -[(1+r^d)k^* + p\chi k^* - p\eta(1-\delta)k^* - p(1+r^d)(b-\psi)]. \end{aligned} \tag{C.1.11}$$

The solution to the original problem is the upper envelop of the two sub-problems:

$$\pi^C(b,z) = \max\{\pi^l(b,z), \pi^h(b,z)\}.$$

To provide some intuitions on the optimal choice of the leverage ratio, consider two extreme cases:

(i) As $b \to \psi$, $\pi^h(b,z) \geq \pi^l(b,z)$. This is because the production function satisfies the inada condition. The marginal return is high when b is small and equality holds only when $b = \psi$.

(ii) As $b \to \infty$, $\pi^h(b,z) < \pi^l(b,z)$. This is because when entrepreneurs have sufficient wealth to operate businesses at the unconstrained scale, there is no reason to borrow from banks, which potentially increases the cost of capital due to monitoring.

Since both $\pi^l(b,z)$ and $\pi^h(b,z)$ are concave and increasing in b, there exists a unique intersection of the two curves, which defines the threshold level of wealth $\overline{b}(z)$. When b is below $\overline{b}(z)$, $\pi^l(b,z) < \pi^h(b,z)$, and the entrepreneur chooses high-leverage. Otherwise, the entrepreneur chooses low-leverage.

C.1.5 Proof of Proposition 7

Consider entrepreneurs with talent z, whose leverage ratios are denoted by $\tilde{\lambda}$. Note that $\overline{b}(z)$ is the threshold of wealth at which entrepreneurs are indifferent between taking low leverage and high leverage, i.e. $\pi^l(\overline{b}(z),z) = \pi^h(\overline{b}(z),z)$. Therefore, entrepreneurs with $b \in [\overline{b}(z), \overline{b}(z)+\epsilon)$ (with ϵ very small), are always hitting the borrowing constraint defined in the low-leverage region (i.e. $\tilde{\lambda} = \overline{\lambda}$).[1] Hence, the optimal amount of capital is $k_l^* = \overline{\lambda}(\overline{b}(z) - \psi)$ when $b = \overline{b}(z)$. Let l_l^* be the corresponding optimal amount of labor. The wealth function is

$$\begin{aligned}\pi^l(\overline{b}(z),z) = {} & (1-p)(z((k_l^*)^\alpha(l_l^*)^{1-\alpha})^{1-\nu} - wl_l^* - \delta k_l^* - r^d k_l^*) + p\eta(1-\delta)k_l^* \\ & + (1+r^d)(\overline{b}(z) - \psi) - p(1+r^d)k_l^*.\end{aligned}$$

In the high-leverage region, entrepreneurs with $b \in (\overline{b}(z)-\epsilon, \overline{b}(z))$ may or may not hit the borrowing constraint λ. Therefore, the optimal amount of capital is $k_h^* = \min(\lambda(b-\psi), \tilde{k}^h(z))$.

Let l_h^* be the corresponding optimal amount of labor. The wealth function is

$$\begin{aligned}\pi^h(\overline{b}(z),z) = \lim_{b \to \overline{b}(z)_-} \pi^h(b,z) \quad = \quad & (1-p)[z((k_h^*)^\alpha(l_h^*)^{1-\alpha})^{1-\nu} - wl_h^* + (1-\delta)k_h^* + (1+r^d)(\overline{b}(z)-\psi)] \\ & -[(1+r^d)k_h^* + p\chi k_h^* - p\eta(1-\delta)k_h^* - p(1+r^d)(\overline{b}(z)-\psi)].\end{aligned}$$

[1] The logic is as follows: a non-binding borrowing constraint implies that entrepreneurs can achieve the unconstrained level of capital $\tilde{k}^l(z)$ when $b \in [\overline{b}(z), \overline{b}(z)+\epsilon)$. Therefore, when $b \in (\overline{b}(z)-\epsilon, \overline{b}(z))$, entrepreneurs should also be able to achieve the unconstrained level of capital in the high-leverage region, $\tilde{k}^h(z)$, by borrowing only up to the leverage ratio $\overline{\lambda}$, since $\tilde{k}^h(z) < \tilde{k}^l(z)$. This is in contradiction to Proposition 6, which requires $\tilde{\lambda} > \overline{\lambda}$ when $b \in (\overline{b}(z)-\epsilon, \overline{b}(z))$.

Since $\pi^l(\bar{b}(z), z) = \pi^h(\bar{b}(z), z)$, $\bar{b}(z)$ is characterized implicitly by the following equation:

$$\begin{aligned} p\chi k_h^* =& (1-p)[z[((k_h^*)^\alpha (l_h^*)^{1-\alpha})^{1-\nu} - ((k_l^*)^\alpha (l_l^*)^{1-\alpha})^{1-\nu}] - w(l_h^* - l_l^*)] \\ &+ [p\eta(1-\delta) - p - (1-p)\delta - r^d](k_h^* - k_l^*). \end{aligned}$$

Substituting l_l^* and l_h^*, we get

$$p\chi k_h^* = Ez^{\frac{1}{\alpha(1-\nu)+\nu}}((k_h^*)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} - (k_l^*)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}) + F(k_h^* - k_l^*), \tag{C.1.12}$$

where $E = (1-p)w\dfrac{\alpha(1-\nu)+\nu}{(1-\alpha)(1-\nu)}[\dfrac{(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} > 0$, and $F = p\eta(1-\delta) - p - (1-p)\delta - r^d < 0$.

To show that $\bar{b}(z)$ is increasing in z, we consider the following two cases.

Case 1

The borrowing constraint is binding for $b \in (\bar{b}(z) - \epsilon, \bar{b}(z))$, $k_h^* = \lambda(b - \psi)$.

Substituting k_l^* and k_h^* into equation (C.1.12), we obtain

$$[p\chi\lambda - F(\lambda - \bar{\lambda})](\bar{b}(z) - \psi) = Ez^{\frac{1}{\alpha(1-\nu)+\nu}}(\lambda^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} - \bar{\lambda}^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}})(\bar{b}(z) - \psi)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}.$$

Take the first derivative with respect to z,

$$\bar{b}(z)' = \frac{\bar{b}(z) - \psi}{\nu z} > 0.$$

Case 2

The borrowing constraint is not binding for $b \in (\bar{b}(z) - \epsilon, \bar{b}(z))$.

In this case, $k_h^* = \tilde{k}^h(z) = [\dfrac{1-p}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p}\dfrac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\dfrac{(1-\nu)(1-\alpha)}{w}z)^{\frac{1}{\nu}} = Gz^{\frac{1}{\nu}}$, where $G = [\dfrac{1-p}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p}\dfrac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\dfrac{(1-\nu)(1-\alpha)}{w})^{\frac{1}{\nu}} > 0$.

Substituting k_l^* and k_h^* into equation (C.1.12), we obtain

$$Ez^{\frac{1}{\alpha(1-\nu)+\nu}}\bar{\lambda}^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}(\bar{b}(z) - \psi)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} = [EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z) - \psi).$$

Take the first derivative with respect to z,

$$\bar{b}(z)' = \frac{\dfrac{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F-p\chi)G]\frac{1}{\nu}z^{\frac{1}{\nu}}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F-p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z)-\psi)}\dfrac{1}{z} - \dfrac{1}{\alpha(1-\nu)+\nu}\dfrac{1}{z}}{\dfrac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}\dfrac{1}{\bar{b}(z)-\psi} + \dfrac{F\bar{\lambda}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F-p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z)-\psi)}}. \tag{C.1.13}$$

Next we show that both the numerator and denominator are smaller than 0, so that $\bar{b}(z)' > 0$.

Numerator:

$$F < 0 \Rightarrow \frac{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]\frac{1}{\nu}z^{\frac{1}{\nu}}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z) - \psi)}\frac{1}{z} - \frac{1}{\alpha(1-\nu)+\nu}\frac{1}{z} < \frac{1}{z} - \frac{1}{\alpha(1-\nu)+\nu}\frac{1}{z} < 0.$$

Denominator:

$$\begin{aligned}
& [EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]z^{\frac{1}{\nu}} \\
= & -(\frac{1}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p})^{\frac{\alpha(1-\nu)}{\nu}}[\frac{(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\nu}}[\frac{(1-p)\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)}{\nu}}(1-p)\frac{\alpha w}{1-\alpha}\frac{\nu}{(1-\nu)} \\
< & \ 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}\frac{1}{\bar{b}(z) - \psi} + \frac{F\bar{\lambda}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z) - \psi)} \\
< & \ \frac{1}{\bar{b}(z) - \psi} + \frac{F\bar{\lambda}}{-F\bar{\lambda}(\bar{b}(z) - \psi)} \\
= & \ 0.
\end{aligned}$$

In conclusion, $\bar{b}(z)' < 0$.

C.1.6 Proof of Lemma 4

When $\nu = 0$, according to equations (C.1.10) and (C.1.11), the wealth functions in the low-leverage and high-leverage regions are

$$\begin{aligned}
\pi^l(b,z) &= (1-p)[\frac{\alpha}{1-\alpha}w[\frac{z(1-\alpha)}{w}]^{\frac{1}{\alpha}} + (1-\delta)]k \\
& \quad -(1+r^d)k + p\eta(1-\delta)k + (1+r^d)(b-\psi), & \text{(C.1.14)} \\
\pi^h(b,z) &= (1-p)[\frac{\alpha}{1-\alpha}w[\frac{z(1-\alpha)}{w}]^{\frac{1}{\alpha}} + (1-\delta)]k \\
& \quad -(1+r^d)k + p\eta(1-\delta)k - p\chi k + (1+r^d)(b-\psi). & \text{(C.1.15)}
\end{aligned}$$

Substituting $k = \bar{\lambda}(b-\psi)$ and $k = \lambda(b-\psi)$ into equations (C.1.14) and (C.1.15), respectively. We obtain

$$\begin{aligned}
\pi^h(b,z) - \pi^l(b,z) &= [(1-p)(\frac{\alpha}{1-\alpha}w(\frac{z(1-\alpha)}{w})^{\frac{1-\alpha}{\alpha}} + (1-\delta)) - (1+r^d) + p\eta(1-\delta)](\lambda - \bar{\lambda})(b-\psi) \\
& \quad -p\chi\lambda(b-\psi).
\end{aligned}$$

$\pi^h(b,z) - \pi^l(b,z) > 0$ implies that

$$z > \bar{z} = \frac{w}{1-\alpha}[(\frac{p\chi\lambda}{(\lambda-\bar{\lambda})} + (1+r^d) + p\eta(1-\delta) - (1-p)(1-\delta))\frac{1-\alpha}{\alpha w(1-p)}]^{\alpha}. \tag{C.1.16}$$

Therefore, talented agents choose the maximum leverage ratio λ.

C.1.7 Proof of Theorem 1

We focus on the case with $z > \bar{z}$. All proofs apply naturally to the case with $z \leq \bar{z}$.

When $\nu = 0$, using the results of Proposition 4, we obtain the wealth function in the savings regime as

$$\pi^S(b,z) = (1-p)[(\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)]b + p\eta(1-\delta)b. \tag{C.1.17}$$

Following Lemma 4, the wealth function in the credit regime is

$$\begin{aligned}\pi^C(b,z) &= (1-p)[(\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)]\lambda(b-\psi) \\ &\quad -[1+r^d+p\chi-p\eta(1-\delta)]\lambda(b-\psi) + (1+r^d)(b-\psi).\end{aligned}$$

Let $\pi^C(b,z) > \pi^S(b,z)$, we obtain

$$b > \underline{b} = \frac{(1-p)\lambda[(\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)] + 1 + r^d - \lambda(1+r^d+p\chi-p\eta(1-\delta))}{(\lambda-1)[(1-p)((\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)) - (1+r^d-p\eta(1-\delta))] - p\chi\lambda}\psi.$$

Let

$$\begin{aligned}m_1(z) &= (\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta), \\ m_2 &= 1+r^d-p\eta(1-\delta).\end{aligned}$$

Then we can write $\bar{b}$ as

$$\underline{b} = \frac{\lambda((1-p)m_1(z)-m_2-p\chi)+1+r^d}{(\lambda-1)[(1-p)m_1(z)-m_2]-p\chi\lambda}\psi, \tag{C.1.18}$$

$$\pi^S(b,z) = [(1-p)m_1(z)-m_2]b + (1+r^d)b, \tag{C.1.19}$$

$$\pi^C(b,z) = [(1-p)m_1(z)-m_2-p\chi]\lambda(b-\psi) + (1+r^d)(b-\psi). \tag{C.1.20}$$

$z > \bar{z}$ implies that $\pi^h(b,z) - \pi^l(b,z) > 0$. Substituting $\pi^l(b,z)$ and $\pi^h(b,z)$ with equation (C.1.14) and (C.1.15), we obtain

$$[(1-p)m_1(z)-m_2]\lambda - p\chi\lambda > (1-p)m_1(z)\bar{\lambda} - m_2.$$

$\overline{\lambda} = \frac{1+r^d}{1+r^d-\eta(1-\delta)} > 1$ implies that

$$[(1-p)m_1(z)-m_2]\lambda - p\chi\lambda > (1-p)m_1(z)-m_2.$$

Thus,

$$(\lambda-1)[(1-p)m_1(z)-m_2] > p\chi\lambda.$$

Therefore, $\underline{b}$ is well defined and $\underline{b} > 0$.

Take the first derivative,

$$\frac{\partial \underline{b}}{\partial \psi} = \frac{(1-p)m_1(z)+1+r^d-\lambda(m_2+p\chi)}{(\lambda-1)[(1-p)m_1(z)-m_2]-p\chi\lambda} > 0, \tag{C.1.21}$$

$$\frac{\partial \underline{b}}{\partial \lambda} = -\frac{[(1-p)m_1(z)-m_2-p\chi][1+r^d+(1-p)m_1(z)-m_2]}{[(\lambda-1)((1-p)m_1(z)-m_2)-p\chi\lambda]^2}\psi < 0, \tag{C.1.22}$$

$$\frac{\partial \underline{b}}{\partial \chi} = \frac{(1-p)m_1(z)-m_2+1+r^d}{[(\lambda-1)((1-p)m_1(z)-m_2)-p\chi\lambda]^2}\lambda p\psi > 0. \tag{C.1.23}$$

The second derivatives are:

$$\frac{\partial^2 \underline{b}}{\partial \chi \partial \lambda} = \frac{\partial^2 \underline{b}}{\partial \lambda \partial \chi} = -\frac{[1+r^d+(1-p)m_1(z)-m_2][(\lambda+1)((1-p)m_1(z)-m_2)-\lambda p\chi]}{[(\lambda-1)((1-p)m_1(z)-m_2)-p\chi\lambda]^3}p\psi < 0, \tag{C.1.24}$$

$$\frac{\partial^2 \underline{b}}{\partial \lambda \partial \psi} = \frac{\partial^2 \underline{b}}{\partial \psi \partial \lambda} = -\frac{[(1-p)m_1(z)-m_2-p\chi][1+r^d+(1-p)m_1(z)-m_2]}{[(\lambda-1)((1-p)m_1(z)-m_2)-p\chi\lambda]^2} < 0, \tag{C.1.25}$$

$$\frac{\partial^2 \underline{b}}{\partial \chi \partial \psi} = \frac{\partial^2 \underline{b}}{\partial \psi \partial \chi} = \frac{(1-p)m_1(z)-m_2+1+r^d}{[(\lambda-1)((1-p)m_1(z)-m_2)-p\chi\lambda]^2}\lambda p > 0. \tag{C.1.26}$$

C.1.8 Proof of Theorem 2

We focus on the case with $z > \bar{z}$. All proofs apply naturally to the case with $z \leq \bar{z}$.

When $\nu = 0$ and $z > \bar{z}$, the optimal labor demand is

$$l^*(b,z) = [\frac{z(1-\alpha)}{w}]^{1/\alpha}\lambda(b-\psi),$$

and

$$y^C = \pi^C + r^d\lambda(b-\psi) + (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha}\lambda(b-\psi) - (1+r^d)b. \tag{C.1.27}$$

Hence,

$$\frac{\partial y^C}{\partial \psi} = -\lambda[(1-p)m_1(z) - m_2 - p\chi] - (1+r^d) - r^d\lambda - (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha}\lambda < 0, \quad \text{(C.1.28)}$$

$$\frac{\partial y^C}{\partial \lambda} = [(1-p)m_1(z) - m_2 - p\chi + r^d](b-\psi) + (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha}(b-\psi) > 0, \quad \text{(C.1.29)}$$

$$\frac{\partial y^C}{\partial \chi} = -p\lambda(b-\psi) < 0. \quad \text{(C.1.30)}$$

$$\frac{\partial^2 \pi^C}{\partial \psi \partial \lambda} = \frac{\partial^2 \pi^C}{\partial \lambda \partial \psi} = -[(1-p)m_1(z) - m_2 - p\chi + r^d] - (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha} < 0, \quad \text{(C.1.31)}$$

$$\frac{\partial^2 \pi^C}{\partial \psi \partial \chi} = \frac{\partial^2 \pi^C}{\partial \chi \partial \psi} = p\lambda > 0, \quad \text{(C.1.32)}$$

$$\frac{\partial^2 \pi^C}{\partial \lambda \partial \chi} = \frac{\partial^2 \pi^C}{\partial \chi \partial \lambda} = -p(b-\psi) < 0. \quad \text{(C.1.33)}$$

C.2 A Model with Forward-Looking Agents

Our model is based on an overlapping generation framework with a constant savings rate (or bequest rate). Therefore, the model is unable to capture one important way of coping with financial frictions: self financing. Buera, Kaboski and Shin [47] and Moll [184] all emphasize that the effects of financial frictions are amplified if a self-financing channel is precluded. As a result, it is expected that in the presence of forward-looking agents and endogenous savings rates, the impact of financial inclusion on GDP could be smaller.

In this appendix, we extend the model with forward-looking agents to address this concern.

C.2.1 Model Setup

We modify the baseline model with endogenous savings rates. The main modeling ingredients are similar to the baseline model, thus we state them briefly and only highlight the difference.

There is a continuum of agents living indefinitely. Population is constant and there is no aggregate uncertainty. Agents are heterogeneous in terms of wealth b and talent z. Wealth evolves endogenously, which is determined by agents' forward-looking decisions. Productivity z follows an exogenous Markov process. With probability γ, agents retain their productivities in the previous period; with probability $1-\gamma$, agents draw new entrepreneurial productivities. The new draw is from a time-invariant Pareto distribution governed by parameter θ and is independent of agents' previous productivities.

Agents have preference at time t,

$$E_t \sum_{s=t}^{\infty} \beta^{s-t} \frac{c_s^{1-\sigma} - 1}{1-\sigma}, \quad \text{(C.2.1)}$$

where β is the time discount factor.

Agents can choose occupation, either to become workers or entrepreneurs. Each worker supplies one unit of labor inelasticly and earns the equilibrium wage. Entrepreneurs use capital and hire labor to produce goods. Capital depreciates at rate δ after use. Entrepreneurs have access to the following production technology,

$$f(z,k,l) = z(k^{\alpha}l^{1-\alpha})^{1-\nu}. \tag{C.2.2}$$

Production fails with probability p, in which case output is zero and agents are able to recover only a fraction $\eta < 1$ of installed capital net of depreciation at the end of the period. Again, we assume workers get paid only when production is successful. Therefore, each worker receive the wage with probability $1-p$.

All agents can make a deposit at banks so as to transfer wealth across periods. However, agents need to pay a fixed credit participation cost ψ to borrow from banks. Borrowing is subject to limited commitment and asymmetric information problems as stated in the baseline model. Similarly, we use parameter λ and χ to capture the tightness of the borrowing constraint and the bank monitoring cost. Consistent with the myopic-agent model, we assume that agents' credit participation status is only maintained for one period. Therefore, agents who obtained credit in period t, still have to pay ψ in period $t+1$ if they want to borrow.

To simplify the problem (and also to be consistent with our myopic-agent model), we assume that agents choose occupation, credit participation, capital and labor to maximize expected end-of-period wealth (or expected income). Then, agents choose consumption and savings to maximizes utility. Notice that this allows us to solve the problem in two separate steps. In the first step, we solve a static problem to obtain the optimal occupation choice, credit participation, capital and labor inputs conditional on the beginning-of-period wealth and talent. In the second step, we solve a dynamic problem to obtain optimal consumption and savings. Thus the only difference from the baseline myopic-agent model is that we are endogenizing the consumption and savings decisions instead of using a constant savings rate.

Since the first part of the problem is solved exactly in the same way as the baseline myopic-agent model, below we only formulate the endogenous consumption/savings decisions, while taking the occupation choice, credit participation, capital and labor inputs as given.

Let $V(b,z,t)$ be the value function for agents of type (b,z) at the beginning of period t. Let I_t^s be the income if production succeeds, and I_t^f be the income if production fails in period t.

Therefore, given the occupation choice, credit participation, capital and labor inputs solved in the first part of the problem, I_t^s and I_t^f can be expressed as

$$I_t^s = \begin{cases} w_t & \text{Workers,} \\ z_t(k_t^{\alpha}l_t^{1-\alpha})^{1-\nu} - w_t l_t + (1-\delta-r_t^d)k_t + r_t^d b_t & \text{Entrepreneurs, savings regime,} \\ z_t(k_t^{\alpha}l_t^{1-\alpha})^{1-\nu} - w_t l_t + (1-\delta)k_t - \Omega + r_t^d b_t - \psi(1+r_t^d) & \text{Entrepreneurs, credit regime,} \end{cases}$$

$$I_t^f = \begin{cases} 0 & \text{Workers.} \\ -k_t + \eta(1-\delta)k_t + r_t^d(b_t - k_t) & \text{Entrepreneurs, savings regime.} \\ \max(0, \eta(1-\delta)k_t + (1+r_t^d)(b_t - \psi) - \Omega) - b_t & \text{Entrepreneurs, credit regime.} \end{cases}$$

Then agents choose consumption and savings to maximize life-time utilities. Denote c_t^s/c_t^f be the consumption when production succeeds/fails. Taking I_t^s and I_t^f as given, the recursive formuation for the second part of the problem is

$$\begin{aligned} V(b_t, z_t, t) = \max_{c_t^s, c_t^f, b_{t+1}^s, b_{t+1}^f} \quad & (1-p)[\frac{(c_t^s)^{1-\sigma}-1}{1-\sigma} + \beta E[V(b_{t+1}^s, z_{t+1}, t+1)|z_t]] \\ & +p[\frac{(c_t^f)^{1-\sigma}-1}{1-\sigma} + \beta E[V(b_{t+1}^f, z_{t+1}, t+1)|z_t]], \\ \text{subject to} \quad & c_t^s + b_{t+1}^s = b_t b_t + I_t^s, \\ & c_t^f + b_{t+1}^f = b_t + I_t^f. \end{aligned}$$

C.2.2 Calibration and Simulation Results

In the model with forward-looking agents, two extra parameters are introduced, the time discount rate β and the risk-aversion parameter σ. Following the standard practice, we set $\beta = 0.96$ and $\sigma = 1.5$. Computation complexity increases tremendously for the model with forward-looking agents. For tractability, we do not match the employment distribution, instead we set $\theta = 4.15$ for all the six countries following Buera and Shin [46], which is selected to match the U.S. employment distribution. We choose parameters λ, ψ, p, χ, and η to match the collateral to loan ratio, the percent of firms with credit, the NPL ratio, the interest rate spread, and the bank overhead costs to assets ratio (see Table C.1). To obtain consistent comparisons with the baseline model, we re-calibrate all the parameters in the baseline model for each country under the parameter restriction, $\theta = 4.15$ (see Table C.2).[2]

The simulation results shown in Table C.3 indicate that, in general, the impact of relaxing the borrowing constraint on GDP and TFP is smaller (except for Mozambique), which is consistent with the results of Buera, Kaboski and Shin [47]. But the difference is not as large partly because in our model production fails with probability p, in which case entrepreneurs' wealth is wiped out. This constrains the self-financing channel. However, there are other differences. Notably, a reduction in the intermediation cost has a much larger GDP and TFP boosting effect as compared to the baseline model.[3] One reason for this could be that entrepreneurs respond to a lower intermediation cost by saving more, which is not captured if the savings rate is constant.[4] The

[2]By comparing the baseline model's simulation results with θ being calibrated to match country-specific firm-size distribution and θ being fixed at 4.15, we can analyze the impact of parameter θ on macroeconomic variables. However, according to the simulation results presented in Table C.3, the impact of this parameter on the model-predicted changes in GDP, TFP and the Gini coefficient is not as clear cut, depending on other country-specific characteristics.

[3]This seems to be consistent with the results in Greenwood, Sanchez and Wang [112], which capture a wide range of GDP per worker across different countries by purely varying the intermediation cost.

[4]This is confirmed by the simulated response of the credit access ratio. In the model with forward-looking agents,

Table C.1: Calibration of the model with forward-looking agents

Target Moments	Uganda Data	Uganda Model	Uganda Parameter	Kenya Data	Kenya Model	Kenya Parameter	Mozambique Data	Mozambique Model	Mozambique Parameter
Collateral (% of loan)	173	173	$\lambda = 1.58$	120.8	120.8	$\lambda = 1.83$	92	92	$\lambda = 2.09$
Firms with credit (%)	17.2	17.9	$\psi = 0.07$	25.4	22.93	$\psi = 0.08$	14.2	14.3	$\psi = 0.21$
Non-perfor. loan (%)	2.3	3.7	$p = 0.15$	10.6	10.9	$p = 0.18$	3.1	3.7	$p = 0.14$
Interest rate spread	10.9	9.5	$\chi = 0.80$	8.5	9.2	$\chi = 0.20$	8.2	7.4	$\chi = 0.7$
Overhead costs/assets	6.9	6.6	$\eta = 0.37$	6.6	6.5	$\eta = 0.37$	7.4	7.3	$\eta = 0.54$
	Malaysia			The Philippines			Egypt		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Collateral (% of loan)	64.6	64.6	$\lambda = 2.56$	238.4	238.4	$\lambda = 1.42$	85.5	85.5	$\lambda = 2.17$
Firms with credit (%)	60.4	62.2	$\psi = 0.08$	33.2	31.6	$\psi = 0.07$	17.4	14.6	$\psi = 0.08$
Non-perfor. loan (%)	8.5	7.6	$p = 0.15$	4.5	3.4	$p = 0.11$	19.3	16.6	$p = 0.25$
Interest rate spread	3.3	5.0	$\chi = 0.05$	4.3	8.3	$\chi = 0.10$	6.1	8.8	$\chi = 0.01$
Overhead costs/assets	1.5	1.5	$\eta = 0.37$	3.2	3.1	$\eta = 0.29$	1.5	1.4	$\eta = 0.44$

same most-binding-constraints that constrain GDP are identified for all six countries by the two models. In Uganda, Kenya, Mozambique, and the Philippines, GDP increases more if the borrowing constraint is relaxed, while in Malaysia and Egypt, GDP is most responsive to a lower participation cost. The prediction on the change in income inequality is broadly consistent with the baseline model for reducing the participation cost and relaxing the borrowing constraint. However, the change in Gini coefficient has a flipped sign when the intermediation cost is lowered.

Table C.2: Calibration of the baseline model with $\theta = 4.15$

Target Moments	Uganda Data	Uganda Model	Uganda Parameter	Kenya Data	Kenya Model	Kenya Parameter	Mozambique Data	Mozambique Model	Mozambique Parameter
Savings (% of GDP)	8	8	$\omega = 0.08$	15.4	15.4	$\omega = 0.15$	7.1	7.1	$\omega = 0.08$
Collateral (% of loan)	173	173	$\lambda = 1.58$	120.8	120.8	$\lambda = 1.83$	92	92	$\lambda = 2.09$
Firms with credit (%)	17.2	17.3	$\psi = 0.04$	25.4	25.4	$\psi = 0.10$	14.2	14.2	$\psi = 0.08$
Non-perfor. loan (%)	2.3	2.6	$p = 0.15$	10.6	5.7	$p = 0.17$	3.1	5.7	$p = 0.14$
Interest rate spread	10.9	9.7	$\chi = 0.85$	8.5	11.5	$\chi = 0.61$	8.2	10.8	$\chi = 0.73$
Overhead costs/assets	6.9	6.8	$\eta = 0.37$	6.6	6.4	$\eta = 0.45$	7.4	7.5	$\eta = 0.54$
	Malaysia			The Philippines			Egypt		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	39	39	$\omega = 0.39$	25.7	25.7	$\omega = 0.26$	24.5	24.5	$\omega = 0.25$
Collateral (% of loan)	64.6	64.6	$\lambda = 2.56$	238.4	238.4	$\lambda = 1.42$	85.5	85.5	$\lambda = 2.17$
Firms with credit (%)	60.4	60.5	$\psi = 0.24$	33.2	33.2	$\psi = 0.07$	17.4	17.3	$\psi = 0.25$
Non-perfor. loan (%)	8.5	7.6	$p = 0.12$	4.5	3.2	$p = 0.11$	19.3	15.7	$p = 0.28$
Interest rate spread	3.3	5.7	$\chi = 0.15$	4.3	5.6	$\chi = 0.62$	6.1	8.0	$\chi = 0.08$
Overhead costs/assets	1.5	1.5	$\eta = 0.37$	3.2	3.2	$\eta = 0.27$	1.5	1.5	$\eta = 0.44$

Note: In this calibration, we set $\theta = 4.15$ for all six countries to be consistent with the calibration of the model with forward-looking agents. We choose parameters λ, ψ, p, χ and η to match the collateral to loan ratio, the percent of firms with credit, the NPLs ratio, the interest rate spread, and the bank overhead costs to assets ratio.

We prefer to retain our baseline overlapping generations framework as our primary specification. First, the theorems and propositions are clear in the baseline model and would not have closed-form expressions in a more general forward-looking-agents model, due to the concavity of the value function. Second, computational complexity increases tremendously in the model with forward-looking agents.[5] This precludes the possibility of matching the employment distribution

agents save to pay the credit participation cost as the monitoring cost decreases, which increases the credit access ratio significantly.

[5]For a computer with i7-4700MQ CPU (2.40GHz), it takes 20 minutes to compute the steady state of the baseline

in every country.[6]

Table C.3: Comparing the impact of financial inclusion generated by the baseline model and the model with forward-looking agent.

		Participation cost ψ			Borrowing constraint λ			Intermediation cost χ		
		GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini
Uganda	Baseline	5.77	5.67	-0.0210	17.94	10.41	-0.0034	0.74	0.42	0.0018
	($\theta = 4.15$)	4.66	5.50	-0.0231	21.68	11.93	0.0233	0.74	0.35	0.0011
	FL model	7.46	1.84	-0.0087	9.65	3.80	-0.0119	6.89	2.11	-0.0165
Kenya	Baseline	5.16	6.50	-0.0314	12.28	9.30	-0.0203	1.93	0.74	0.0082
	($\theta = 4.15$)	4.79	5.39	-0.0300	13.90	10.28	-0.0014	1.70	0.49	0.0040
	FL model	6.98	4.24	-0.0206	10.68	6.16	-0.0263	6.18	2.56	-0.015
Mozambique	Baseline	12.72	10.16	-0.0267	10.30	4.83	0.0217	0.88	0.32	0.0033
	($\theta = 4.15$)	8.11	9.90	-0.0376	10.91	6.73	0.0081	0.87	0.24	0.0032
	FL model	12.08	6.70	-0.0174	13.32	6.70	-0.0111	10.78	6.36	-0.0241
Malaysia	Baseline	8.44	10.94	-0.0696	4.52	2.85	0.0059	1.26	0.00	0.0013
	($\theta = 4.15$)	6.62	6.92	0.0091	4.00	3.73	-0.0013	1.45	0.20	0.0048
	FL model	3.10	2.15	-0.0128	2.88	1.68	-0.0082	1.25	0.11	-0.0045
The Philippines	Baseline	2.56	3.40	-0.0165	20.21	16.45	-0.0336	1.48	0.58	0.0033
	($\theta = 4.15$)	2.16	3.34	-0.0167	22.43	18.02	-0.0124	1.48	0.73	0.0018
	FL model	10.98	5.02	-0.0226	18.58	8.56	-0.0550	7.75	2.17	-0.0092
Egypt	Baseline	7.04	11.31	-0.0590	7.78	6.61	0.0026	0.69	0.02	0.0033
	($\theta = 4.15$)	7.63	7.34	-0.0216	7.48	5.83	0.0025	0.74	0.08	0.0026
	FL model	5.81	2.00	-0.0502	5.68	2.00	0.0053	1.50	0.16	0.0015

Note: In all cases, we consider financial inclusion that moves the country to world financial sector frontier for one of the three parameters.

model using matlab (2014a). However, for the model with forward-looking agents, it takes more than 24 hours when a certain accuracy level is required. To promote the computation speed, we code the value function iteration and wealth distribution iteration parts of the program in C++ (VS studio 2014) and use matlab to call these scripts. At the same time, we use a 20-core server to parallel the computation of heterogeneity. This reduces the computation time of the steady state to 40 minutes. Hence, more complicated dynamics are within reach, but require more hardware and coding, which limits their wide applicability.

[6] As noted before, the calibration for the forward-looking-agents model is done by selecting parameter θ from the literature, which is calibrated using the U.S. employment distribution, not the distribution of each country. Similar approaches are also used in several other quantitative papers with forward-looking heterogeneous agents [e.g. 46, 47, 112].

Bibliography

[1] **Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L. Violante.** 2016. "Education Policy and Intergenerational Transfers in Equilibrium." National Bureau of Economic Research, Inc NBER Working Papers 18782.

[2] **Abraham, Arpad, and Nicola Pavoni.** 2005. "The Efficient Allocation of Consumption under Moral Hazard and Hidden Access to the Credit Market." *Journal of the European Economic Association,* 3(2-3): 370–381.

[3] **Acemoglu, Daron, and David Autor.** 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." , ed. Orley Ashenfelter and David Card Vol. 4 of *Handbook of Labor Economics,* 1043–1171. Elsevier.

[4] **Acemoglu, Daron, and Robert Shimer.** 1999. "Efficient Unemployment Insurance." *Journal of Political Economy,* 107(5): 893–928.

[5] **Acemoglu, Daron, and Robert Shimer.** 2000. "Productivity Gains From Unemployment Insurance." *European Economic Review,* 44(7): 1195–1224.

[6] **Addo, Fenaba.** 2014. "Debt, Cohabitation, and Marriage in Young Adulthood." *Demography,* 51(5): 1677–1701.

[7] **Aghion, Philippe, and Patrick Bolton.** 1997. "A Theory of Trickle-Down Growth and Development." *Review of Economic Studies,* 64(2): 151–72.

[8] **Ahlin, Christian, and RobertM. Townsend.** 2007. "Using Repayment Data to Test Across Models of Joint Liability Lending." *Economic Journal,* 117(517): F11–F51.

[9] **Ai, Hengjie, and Dana Kiku.** 2013. "Growth to Value: Option Exercise and the Cross Section of Equity Returns." *Journal of Financial Economics,* 107(2): 325–349.

[10] **Ai, Hengjie, Mariano Massimiliano Croce, and Kai Li.** 2013. "Toward a Quantitative General Equilibrium Asset Pricing Model with Intangible Capital." *Review of Financial Studies,* 26(2): 491–530.

[11] **Albuquerque, Rui, and Hugo A. Hopenhayn.** 2004. "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies,* 71(2): 285–315.

[12] **Alem, Mauro, and Robert M. Townsend.** 2013. "An Evaluation of Financial Institutions: Impact on Consumption and Investment Using Panel Data and the Theory of Risk-Bearing." *Annals Issue of the Journal of Econometrics in Honor of Bill Barnett,* , (forthcoming).

[13] **Algan, Yann, Arnaud Cheron, Jean-Olivier Hairault, and Francois Langot.** 2003. "Wealth Effect on Labor Market Transitions." *Review of Economic Dynamics*, 6(1): 156–178.

[14] **Altinkilic, Oya, and Robert S Hansen.** 2000. "Are There Economies of Scale in Underwriting Fees? Evidence of Rising External Financing Costs." *Review of Financial Studies*, 13(1): 191–218.

[15] **Amaral, Pedro S., and Erwan Quintin.** 2010. "Limited Enforcement, Financial Intermediation, And Economic Development: A Quantitative Assessment." *International Economic Review*, 51(3): 785–811.

[16] **Andreoni, James.** 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447–58.

[17] **Assuncao, Juliano, Sergey Mityakov, and Robert M. Townsend.** 2012. "Ownership Matters: the Geographical Dynamics of BAAC and CommercialBanks in Thailand." Working Paper.

[18] **Athey, Susan, Kyle Bagwell, and Chris Sanchirico.** 2004. "Collusion and Price Rigidity." *Review of Economic Studies*, 71(2): 317–349.

[19] **Atkeson, Andrew, and Patrick J. Kehoe.** 2005. "Modeling and Measuring Organization Capital." *Journal of Political Economy*, 113(5): 1026–1053.

[20] **Ayyagari, Meghana, Asli DemirgÃijc-Kunt, and Vojislav Maksimovic.** 2008. "How Well Do Institutional Theories Explain Firms' Perceptions of Property Rights?" *Review of Financial Studies*, 21(4): 1833–1871.

[21] **Backus, David K, Patrick J Kehoe, and Finn E Kydland.** 1994. "Dynamics of the Trade Balance and the Terms of Trade: The J-Curve?" *American Economic Review*, 84(1): 84–103.

[22] **Bagger, Jesper, Francois Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin.** 2014. "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics." *American Economic Review*, 104(6): 1551–96.

[23] **Baharad, Eyal, and Benjamin Eden.** 2004. "Price Rigidity and Price Dispersion: Evidence from Micro Data." *Review of Economic Dynamics*, 7(3): 613–641.

[24] **Ball, Laurence, and David Romer.** 1990. "Real Rigidities and the Non-neutrality of Money." *Review of Economic Studies*, 57(2): 183–203.

[25] **Banerjee, Abhijit V, and Andrew F Newman.** 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy*, 101(2): 274–98.

[26] **Banerjee, Abhijit V, and Andrew F. Newman.** 2003. "Inequality, Growth and Trade Policy." MIT Working Paper.

[27] **Beck, Thorsten, Asli DemirgÃijÃğ-Kunt, and Ross Levine.** 2007. "Finance, Inequality and the Poor." *Journal of Economic Growth*, 12(1): 27–49.

[28] **Beck, Thorsten, Asli DemirgÃijÃğ-Kunt, and Vojislav Maksimovic.** 2005. "Financial and Legal Constraints to Growth: Does Firm Size Matter?" *Journal of Finance*, 60(1): 137–177.

[29] **Beggs, Alan W, and Paul Klemperer.** 1992. "Multi-period Competition with Switching Costs." *Econometrica*, 60(3): 651–66.

[30] **Belo, Frederico, Xiaoji Lin, and Maria Ana Vitorino.** 2014. "Brand Capital and Firm Value." *Review of Economic Dynamics*, 17(1): 150–169.

[31] **Benabou, Roland.** 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70(2): 481–517.

[32] **Bernstein, Asaf.** 2016. "Household Debt Overhang and Labor Supply." MIT.

[33] **Bharadwaj, Anandhi S., Sundar G. Bharadwaj, and Benn R. Konsynski.** 1999. "Information Technology Effects on Firm Performance as Measured by Tobin's q." *Management Science*, 45(7): 1008–1024.

[34] **Bhattacharya, Jayanta, and William B Vogt.** 2003. "A Simple Model of Pharmaceutical Price Dynamics." *Journal of Law and Economics*, 46(2): 599–626.

[35] **Bils, Mark.** 1987. "The Cyclical Behavior of Marginal Cost and Price." *American Economic Review*, 77(5): 838–55.

[36] **Bils, Mark.** 1989. "Pricing in a Customer Market." *The Quarterly Journal of Economics*, 104(4): 699–718.

[37] **Bils, Mark, and Peter J. Klenow.** 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy*, 112(5): 947–985.

[38] **Blanchard, Olivier.** 2009. "The State of Macro." *Annual Review of Economics*, 1(1): 209–228.

[39] **Bloemen, Hans G, and Elena G F Stancanelli.** 2001. "Individual Wealth, Reservation Wages, and Transitions into Employment." *Journal of Labor Economics*, 19(2): 400–439.

[40] **Bolton, Patrick, and David S Scharfstein.** 1990. "A Theory of Predation Based on Agency Problems in Financial Contracting." *American Economic Review*, 80(1): 93–106.

[41] **Bolton, Patrick, Hui Chen, and Neng Wang.** 2011. "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management." *Journal of Finance*, 66(5): 1545–1578.

[42] **Bolton, Patrick, Hui Chen, and Neng Wang.** 2013. "Market Timing, Investment, and Risk Management." *Journal of Financial Economics*, 109(1): 40–62.

[43] **Braun, Matias, and Claudio Raddatz.** 2012. "Financial Constraints, Competition, and Markup Cyclicality."

[44] **Browning, Martin, Thomas F. Crossley, and Eric F. Smith.** 2007. "Asset Accumulation and Short Term Employment." *Review of Economic Dynamics*, 10(3): 400–423.

[45] **Brown, Jennifer, and David A. Matsa.** 2016. "Locked in by Leverage: Job Search during the Housing Crisis." National Bureau of Economic Research, Inc NBER Working Papers 22929.

[46] **Buera, Francisco J., and Yongseok Shin.** 2013. "Financial Frictions and the Persistence of History: A Quantitative Exploration." *Journal of Political Economy*, 121(2): 221 – 272.

[47] **Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin.** 2011. "Finance and Development: A Tale of Two Sectors." *American Economic Review*, 101(5): 1964–2002.

[48] **Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin.** 2012. "The Macroeconomics of Microfinance." National Bureau of Economic Research, Inc NBER Working Papers 17905.

[49] **Burgess, Robin, and Rohini Pande.** 2005. "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment." *American Economic Review,* 95(3): 780–795.

[50] **Caballero, Ricardo J.** 1991. "On the Sign of the Investment-Uncertainty Relationship." *American Economic Review,* 81(1): 279–88.

[51] **Cagetti, Marco, and Mariacristina De Nardi.** 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy,* 114(5): 835–870.

[52] **Calvo, Guillermo A.** 1983. "Staggered prices in a utility-maximizing framework." *Journal of Monetary Economics,* 12(3): 383–398.

[53] **Card, David, Andrew Johnston, Pauline Leung, Alexandre Mas, and Zhuan Pei.** 2015. "The Effect of Unemployment Benefits on the Duration of Unemployment Insurance Receipt: New Evidence from a Regression Kink Design in Missouri, 2003-2013." National Bureau of Economic Research, Inc NBER Working Papers 20869.

[54] **Carlson, John A, and R Preston McAfee.** 1983. "Discrete Equilibrium Price Dispersion." *Journal of Political Economy,* 91(3): 480–93.

[55] **Caselli, Francesco, and Nicola Gennaioli.** 2013. "Dynastic Management." *Economic Inquiry,* 51(1): 971–996.

[56] **Castaneda, Ana, Javier Diaz-Gimenez, and Jose-Victor Rios-Rull.** 2003. "Accounting for the U.S. Earnings and Wealth Inequality." *Journal of Political Economy,* 111(4): 818–857.

[57] **Cecchetti, Stephen G.** 1986. "The Frequency of Price Adjustment : A Study of the Newsstand Prices of Magazines." *Journal of Econometrics,* 31(3): 255–274.

[58] **Chen, K. C., and J. Lee.** 1995. "Accounting Measures of Business Performance and Tobin's q Theory." *Journal of Accounting, Auditing and Finance,* 10(3): 587–609.

[59] **Chetty, Raj.** 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy,* 116(2): 173–234.

[60] **Chevalier, Judith A, and David S Scharfstein.** 1996. "Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence." *American Economic Review,* 86(4): 703–25.

[61] **Chintagunta, Pradeep, Ekaterini Kyriazidou, and Josef Perktold.** 2001. "Panel Data Analysis of Household Brand Choices." *Journal of Econometrics,* 103(1-2): 111–153.

[62] **Chirinko, R S, and S Fazzari.** 1994. "Economic Fluctuations, Market Power, and Returns to Scale: Evidence from Firm-Level Data." *Journal of Applied Econometrics,* 9(1): 47–69.

[63] **Choudhary, Ali, and J. Michael Orszag.** 2007. "Costly Customer Relations and Pricing." *Oxford Economic Papers,* 59(4): 641–661.

[64] **Claessens, Stijn, M Ayhan Kose, and Marco E Terrones.** 2009. "What Happens during Recessions, Crunches and Busts?" *Economic Policy,* 24(60): 653–700.

[65] **Clarke, Georg R. G., Lixin Colin Xu, and Heng fu Zou.** 2006. "Finance and Income Inequality: What Do the Data Tell Us?" *Southern Economic Journal,* 72(3): 578–596.

[66] **Clementi, Gina Luca, and Hugo A Hopenhayn.** 2006. "A Theory of Financing Constraints and Firm Dynamics." *The Quarterly Journal of Economics,* 121(1): 229–265.

[67] **Collado, M. Dolores, and Martin Browning.** 2007. "Habits and Heterogeneity in Demands: A Panel Data Analysis." *Journal of Applied Econometrics,* 22(3): 625–640.

[68] **Danforth, John P.** 1974. "Wealth and The Value of Generalized Lotteries." University of Minnesota.

[69] **Danforth, John P.** 1979. "On the Role of Consumption and Decreasing Absolute Risk Aversion in the Theory of Job Search." In *Studies in the Economics of Search.* Vol. 123 of *Contributions to Economic Analysis,* , ed. S. A. Lippman and J. J. McCall, Chapter 6, 109–131. North-Holland.

[70] **Davies, James B.** 1982. "The Relative Impact of Inheritance and Other Factors on Economic Inequality." *The Quarterly Journal of Economics,* 97(3): 471–98.

[71] **Davis, James L., Eugene F. Fama, and Kenneth R. French.** 2000. "Characteristics, Covariances, and Average Returns: 1929 to 1997." *Journal of Finance,* 55(1): 389–406.

[72] **Dearden, Lorraine, Emla Fitzsimons, Alissa Goodman, and Greg Kaplan.** 2008. "Higher Education Funding Reforms in England: The Distributional Effects and the Shifting Balance of Costs." *Economic Journal,* 118(526): F100–F125.

[73] **DeMarzo, Peter M., Michael J. Fishman, Zhiguo He, and Neng Wang.** 2012. "Dynamic Agency and the q Theory of Investment." *Journal of Finance,* 67(6): 2295–2340.

[74] **Diamond, Douglas W., and Zhiguo He.** 2014. "A Theory of Debt Maturity: The Long and Short of Debt Overhang." *Journal of Finance,* 69(2): 719–762.

[75] **Dobbie, Will, and Jae Song.** 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." *American Economic Review,* 105(3): 1272–1311.

[76] **Dobbie, Will, and Jae Song.** 2016. "Debt Relief or Debt Restructuring? Evidence from an Experiment with Distressed Credit Card Borrowers." Princeton University.

[77] **Doepke, Matthias, and Robert M. Townsend.** 2006. "Dynamic Mechanism Design with Hidden Income and Hidden Actions." *Journal of Economic Theory,* 126(1): 235–285.

[78] **Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Petersen.** 1986. "Business Cycles and the Relationship Between Concentration and Price-Cost Margins." *RAND Journal of Economics,* 17(1): 1–17.

[79] **Donaldson, Jason Roderick, Giorgia Piacentino, and Anjan Thakor.** 2016. "Household Debt and Unemployment." Washington University in St Louis.

[80] **Dou, Winston Wei, Lars Peter Hansen, Andrew W. Lo, Harald Uhlig, and Ameya Muley.** 2014. "Macroeconomic Models for Monetary Policies: A Critical Review from a Finance Perspective." MIT Sloan and Chicago University Working Paper.

[81] **Dumas, Bernard.** 1991. "Super Contact and Related Optimality Conditions." *Journal of Economic Dynamics and Control*, 15(4): 675–685.

[82] **Dumas, Bernard, and Andrew Lyasoff.** 2012. "Incomplete-Market Equilibria Solved Recursively on an Event Tree." *Journal of Finance*, 67(5): 1897–194.

[83] **Dynarski, Susan, and Daniel Kreisman.** 2013. "Loans for Educational Opportunity: Making Borrowing Work for Today's Students." The Hamilton Project Discussion Paper.

[84] **Eberly, Janice, Sergio Rebelo, and Nicolas Vincent.** 2009. "Investment and Value: a Neoclassical Benchmark."

[85] **Eggertsson, Gauti B., and Paul Krugman.** 2012. "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *The Quarterly Journal of Economics*, 127(3): 1469–1513.

[86] **Eisfeldt, Andrea L., and Dimitris Papanikolaou.** 2013. "Organization Capital and the Cross-Section of Expected Returns." *Journal of Finance*, 68(4): 1365–1406.

[87] **Eliashberg, Jehoshua, and Abel P. Jeuland.** 1986. "The Impact of Competitive Entry in a Developing Market upon Dynamic Pricing Strategies." *Journal of Economic Theory*, 5(1): 20–36.

[88] **Elliott, William, Michal Grinstein-Weiss, and Ilsung Nam.** 2013*a*. "Does Outstanding Student Debt Reduce Asset Accumulation?" Center for Social Development Working Paper 13-32.

[89] **Elliott, William, Michal Grinstein-Weiss, and Ilsung Nam.** 2013*b*. "Student Debt and Declining Retirement Savings." Center for Social Development Working Paper 13-34.

[90] **Farrell, Joseph, and Carl Shapiro.** 1988. "Dynamic Competition with Switching Costs." *RAND Journal of Economics*, 19(1): 123–137.

[91] **Farrell, Joseph, and Paul Klemperer.** 2007. "Coordination and Lock-In: Competition with Switching Costs and Network Effects." *Handbook of Industrial Organization*, , ed. Mark Armstrong and Robert Porter Vol. 3, Chapter 31, 1967–2072. Elsevier.

[92] **Feldstein, Martin, and James M. Poterba.** 1984. "Unemployment Insurance and Reservation Wages." *Journal of Public Economics*, 1-2(23): 141–167.

[93] **Field, Erica.** 2009. "Educational Debt Burden and Career Choice: Evidence from a Financial Aid Experiment at NYU Law School." *American Economic Journal: Applied Economics*, 1(1): 1–21.

[94] **Fischer, Edwin O., Robert Heinkel, and Josef Zechner.** 1989. "Dynamic Capital Structure Choice: Theory and Tests." *Journal of Finance*, 44(1): 19–40.

[95] **Friend, Irwin, and Marshall E Blume.** 1975. "The Demand for Risky Assets." *American Economic Review*, 65(5): 900–922.

[96] **Gabaix, Xavier.** 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica*, 79(3): 733–772.

[97] **Galí, Jordi.** 2008. *Introduction to Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework.* Princeton University Press.

[98] **Galí, Jordi, Mark Gertler, and J. David López-Salido.** 2007. "Markups, Gaps, and the Welfare Costs of Business Fluctuations." *The Review of Economics and Statistics*, 89(1): 44–59.

[99] **Gaskins, Darius Jr.** 1971. "Dynamic Limit Pricing: Optimal Pricing under Threat of Entry." *Journal of Economic Theory, Elsevier*, 3(3): 306–322.

[100] **Gervais, Martin, and Nicolas L. Ziebarth.** 2016. "Life after Debt: Post-Graduation Consequences of Federal Student Loans." University of Iowa.

[101] **Ghosal, Vivek.** 2000. "Product Market Competition and the Industry Price-Cost Markup Fluctuations: Role of Energy Price and Monetary Changes." *International Journal of Industrial Organization*, 18(3): 415–444.

[102] **Gilchrist, Simon, Jae Sim, Raphael Schoenle, and Egon Zakrajsek.** 2016. "Inflation Dynamics During the Financial Crisis." *American Economic Review*, forthcoming.

[103] **Gine, Xavier, and Robert M. Townsend.** 2004. "Evaluation of financial liberalization: a general equilibrium model with constrained occupation choice." *Journal of Development Economics*, 74(2): 269–307.

[104] **Gomes, Joao, Leonid Kogan, and Lu Zhang.** 2003. "Equilibrium Cross Section of Returns." *Journal of Political Economy*, 111(4): 693–732.

[105] **Gottfries, Nils.** 1986. " Price Dynamics of Exporting and Import-Competing Firms." *Scandinavian Journal of Economics, Wiley Blackwell*, 88(2): 417–36.

[106] **Gottfries, Nils.** 1991. "Customer Markets, Credit Market Imperfections and Real Price Rigidity." *Economica*, 58(231): 317–23.

[107] **Gourieroux, Christian, and Alain Monfort.** 1997. *Simulation-Based Econometric Methods.* Oxford University Press.

[108] **Gourinchas, Pierre-Olivier, and Jonathan A. Parker.** 2002. "Consumption Over the Life Cycle." *Econometrica*, 70(1): 47–89.

[109] **Gourio, Francois, and Leena Rudanko.** 2014. "Customer Capital." *Review of Economic Studies*, 81(3): 1102–1136.

[110] **Green, Edward J, and Robert H Porter.** 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica*, 52(1): 87–100.

[111] **Greenwood, Jeremy, and Boyan Jovanovic.** 1990. "Financial Development, Growth, and the Distribution of Income." *Journal of Political Economy*, 98(5): 1076–1107.

[112] **Greenwood, Jeremy, Juan Sanchez, and Cheng Wang.** 2013. "Quantifying the Impact of Financial Development on Economic Development." *Review of Economic Dynamics*, 16(1): 194–215.

[113] **Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman.** 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review*, 78(3): 402–17.

[114] **Grenadier, Steven R., and Neng Wang.** 2005. "Investment timing, agency, and information." *Journal of Financial Economics*, 75(3): 493 – 533.

[115] **Griffin, John M., and Michael L. Lemmon.** 2002. "Book-to-Market Equity, Distress Risk, and Stock Returns." *Journal of Finance*, 57(5): 2317–2336.

[116] **Guerrieri, Veronica, and Guido Lorenzoni.** 2015. "Credit Crises, Precautionary Savings, and the Liquidity Trap." University of Chicago.

[117] **Guler, Bulent, Fatih Guvenen, and Giovanni L. Violante.** 2012. "Joint-Search Theory: New Opportunities and New Frictions." *Journal of Monetary Economics*, 59(4): 352–369.

[118] **Hall, Robert E.** 2001. "Struggling to Understand the Stock Market." *American Economic Review*, 91(2): 1–11.

[119] **Hall, Robert E.** 2005. "Employment Fluctuations with Equilibrium Wage Stickiness." *American Economic Review*, 95(1): 50–65.

[120] **Hall, Robert E.** 2012. "The Cyclical Response of Advertising Refutes Counter-Cyclical Profit Margins in Favor of Product-Market Frictions." National Bureau of Economic Research, Inc NBER Working Papers 18370.

[121] **Haltiwanger, John, and Joseph E. Harrington.** 1991. "The Impact of Cyclical Demand Movements on Collusive Behavior." *RAND Journal of Economics*, 22(1): 89–106.

[122] **Hansen, Gary D, and Ayse Imrohoroglu.** 1992. "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard." *Journal of Political Economy*, 100(1): 118–42.

[123] **Hayashi, Fumio.** 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, 50(1): 213–24.

[124] **Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante.** 2014. "Consumption and Labor Supply with Partial Insurance: An Analytical Framework." *American Economic Review*, 104(7): 2075–2126.

[125] **Heckman, James, Lance Lochner, and Christopher Taber.** 1998. "Explaining Rising Wage Inequality: Explanations With A Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents." *Review of Economic Dynamics*, 1(1): 1–58.

[126] **Hennessy, Christopher A., and Toni M. Whited.** 2007. "How Costly Is External Financing? Evidence from a Structural Estimation." *Journal of Finance*, 62(4): 1705–1745.

[127] **Herkenhoff, Kyle.** 2015. "The Impact of Consumer Credit Access on Unemployment." University of Minnesota.

[128] **Herkenhoff, Kyle F., and Lee E. Ohanian.** 2015. "The Impact of Foreclosure Delay on U.S. Employment." NBER Working Papers 21532.

[129] **Herkenhoff, Kyle, Gordon Phillips, and Ethan Cohen-Cole.** 2016. "How Credit Constraints Impact Job Finding Rates, Sorting and Aggregate Output." University of Minnesota.

[130] **He, Zhiguo, and Konstantin Milbradt.** 2014. "Endogenous Liquidity and Defaultable Bonds." *Econometrica*, 82(4): 1443–1508.

[131] **He, Zhiguo, and Wei Xiong.** 2012. "Rollover Risk and Credit Risk." *The Journal of Finance*, 67(2): 391–430.

[132] **Hsieh, Chang-Tai, and Peter J. Klenow.** 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448.

[133] **Hubbard, R Glenn, Jonathan Skinner, and Stephen P Zeldes.** 1995. "Precautionary Saving and Social Insurance." *Journal of Political Economy*, 103(2): 360–99.

[134] **Huggett, Mark.** 1996. "Wealth Distribution in Life-cycle Economies." *Journal of Monetary Economics*, 38(3): 469–494.

[135] **Ionescu, Felicia.** 2009. "The Federal Student Loan Program: Quantitative Implications for College Enrollment and Default Rates." *Review of Economic Dynamics*, 12(1): 205–231.

[136] **Ionescu, Felicia, and Marius Ionescu.** 2014. "The Interplay Between Student Loans and Credit Card Debt: Implications for Default in the Great Recession." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2014-66.

[137] **Jarosch, Gregor.** 2015. "Searching for Job Security and the Consequences of Job Loss." Stanford University.

[138] **Jeong, Hyeok, and Robert M. Townsend.** 2008. "Growth And Inequality: Model Evaluation Based On An Estimation-Calibration Strategy." *Macroeconomic Dynamics*, 12(S2): 231–284.

[139] **Jeong, Hyeok, and Robert Townsend.** 2007. "Sources of TFP Growth: Occupational Choice and Financial Deepening." *Economic Theory*, 32(1): 179–221.

[140] **Joensen, Juanna, and Elena Mattana.** 2016. "Student Aid, Academic Achievement, and Labor Market Behavior."

[141] **Johnson, Matthew T.** 2013. "Borrowing Constraints, College Enrollment, and Delayed Entry." *Journal of Labor Economics*, 31(4): 669 – 725.

[142] **Kaboski, Joseph P., and Robert M. Townsend.** 2011. "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative." *Econometrica*, 79(5): 1357–1406.

[143] **Kane, Thomas J.** 2006. "Public Intervention in Post-Secondary Education." , ed. Erik Hanushek and F. Welch Vol. 2 of *Handbook of the Economics of Education*, Chapter 23, 1369–1401. Elsevier.

[144] **Kaplan, Greg.** 2012. "Moving Back Home: Insurance against Labor Market Risk." *Journal of Political Economy*, 120(3): 446 – 512.

[145] **Kaplan, Greg, and Giovanni L. Violante.** 2014. "A Model of Consumption Response to Fiscal Stimulus Payments." *Econometrica*, 82(4): 1199–1239.

[146] **Karaivanov, Alexander, and Robert M. Townsend.** 2014. "Dynamic Financial Constraints: Distinguishing Mechanism Design From Exogenously Incomplete Regimes." *Econometrica*, 82(3): 887–959.

[147] **Keane, Michael P.** 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature*, 49(4): 961–1075.

[148] **Keane, Michael P, and Kenneth I Wolpin.** 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review*, 42(4): 1051–1103.

[149] **King, Robert G, and Ross Levine.** 1993. "Finance and Growth: Schumpeter Might Be Right." *The Quarterly Journal of Economics*, 108(3): 717–37.

[150] **Kinnan, Cynthia.** 2014. "Distinguishing Barriers to Insurance in Thai Villages." Northwestern University Working Papers 1276.

[151] **Klemperer, Paul.** 1987. "Markets with Consumer Switching Costs." *The Quarterly Journal of Economics*, 102(2): 375–94.

[152] **Klemperer, Paul.** 1995. "Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade." *Review of Economic Studies*, 62(4): 515–39.

[153] **Korajczyk, Robert A., and Amnon Levy.** 2003. "Capital Structure Choice: Macroeconomic Conditions and Financial Constraints." *Journal of Financial Economics*, 68(1): 75–109.

[154] **Koszegi, Botond, and Paul Heidhues.** 2008. "Competition and Price Variation When Consumers Are Loss Averse." *American Economic Review*, 98(4): 1245–68.

[155] **Krusell, Per, Toshihiko Mukoyama, and Aysegull Sahin.** 2010. "Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations." *Review of Economic Studies*, 77(4): 1477–1507.

[156] **Lang, Larry H.P., and Robert H. Litzenberger.** 1989. "Dividend Announcements: Cash Flow Signalling vs. Free Cash Flow Hypothesis?" *Journal of Financial Economics*, 24(1): 181âĂŞ191.

[157] **Lee, Stephanie, and Max Egan.** 2009. "United States: Student Loans And Student Loan Asset-Backed Securities: A Primer." NERA Economic Consulting.

[158] **Leland, Hayne E.** 1994. " Corporate Debt Value, Bond Covenants, and Optimal Capital Structure." *Journal of Finance*, 49(4): 1213–52.

[159] **Leland, Hayne E.** 1998. "Agency Costs, Risk Management, and Capital Structure." *The Journal of Finance*, 53(4): 1213–1243.

[160] **Leland, Hayne E., and Klaus Bjerre Toft.** 1996. "Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads." *The Journal of Finance*, 51(3): 987–1019.

[161] **Leslie, Larry L., and Paul T. Brinkman.** 1987. "Student Price Response in Higher Education: The Student Demand Studies." *The Journal of Higher Education*, 58(2): 181–204.

[162] **Levine, Ross.** 2005. "Finance and Growth: Theory and Evidence." In *Handbook of Economic Growth*. Vol. 1 of *Handbook of Economic Growth*, , ed. Philippe Aghion and Steven Durlauf, Chapter 12, 865–934. Elsevier.

[163] **Liew, Jimmy, and Maria Vassalou.** 2000. "Can Book-to-Market, Size and Momentum be Risk Factors That Predict Economic Growth?" *Journal of Financial Economics*, 57(2): 221–245.

[164] **Lise, Jeremy.** 2013. "On-the-Job Search and Precautionary Savings." *Review of Economic Studies*, 80: 1086–1113.

[165] **Lise, Jeremy, and Jean-Marc Robin.** 2017. "The Macro-Dynamics of Sorting between Workers and Firms." *American Economic Review*, 107(4): 1104–1135.

[166] **Lise, Jeremy, Costas Meghir, and Jean-Marc Robin.** 2016. "Matching, Sorting and Wages." *Review of Economic Dynamics*, 19(Jan): 63–87.

[167] **Livshits, Igor, James MacGee, and Michele Tertilt.** 2007. "Consumer Bankruptcy: A Fresh Start." *American Economic Review*, 97(1): 402–418.

[168] **Ljungqvist, Lars, and Thomas J. Sargent.** 1998. "The European Unemployment Dilemma." *Journal of Political Economy*, 106(3): 514–550.

[169] **Lloyd-Ellis, Huw, and Dan Bernhardt.** 2000. "Enterprise, Inequality and Economic Development." *Review of Economic Studies*, 67(1): 147–68.

[170] **Lochner, Lance, and Alexander Monge-Naranjo.** 2016. "Student Loans and Repayment: Theory, Evidence and Policy." , ed. S. Machin E. Hanushek and L. Woessmann Vol. 5 of *Handbook of the Economics of Education*, Chapter 8, 397–478. Elsevier.

[171] **Luo, Mi, and Simon Mongey.** 2016. "Student Debt and Job Choice: Wages vs. Job Satisfaction." New York University.

[172] **Machin, Stephen, and John Van Reenen.** 1993. "Profit Margins and the Business Cycle: Evidence from UK Manufacturing Firms." *Journal of Industrial Economics*, 41(1): 29–50.

[173] **Manso, Gustavo.** 2008. "Investment Reversibility and Agency Cost of Debt." *Econometrica*, 76(2): 437–442.

[174] **Martin, Alberto, and Filippo Taddei.** 2013. "International Capital Flows and Credit Market Imperfections: A Tale of Two Frictions." *Journal of International Economics*, 89(2): 441–452.

[175] **Mazumder, Sandeep.** 2014. "The Price-marginal Cost Markup and its Determinants in U.S. Manufacturing." *Macroeconomics Dynamics*, 18: 783–811.

[176] **McCall, J. J.** 1970. "Economics of Information and Job Search." *The Quarterly Journal of Economics*, 84(1): 113 – 126.

[177] **McGrattan, Ellen R., and Edward C. Prescott.** 2010. "Technology Capital and the US Current Account." *American Economic Review*, 100(4): 1493–1522.

[178] **Mian, Atif, and Amir Sufi.** 2014. "What Explains the 2007-2009 Drop in Employment?" *Econometrica*, 82: 2197–2223.

[179] **Midrigan, Virgiliu, and Daniel Yi Xu.** 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review*, 104(2): 422–58.

[180] **Midrigan, Virgiliu, and Thomas Philippon.** 2016. "Household Leverage and the Recession." New York University.

[181] **Minicozzi, Alexandra.** 2005. "The short term effect of educational debt on job decisions." *Economics of Education Review*, 24(4): 417–430.

[182] **Mirrlees, J. A.** 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38(2): 175–208.

[183] **Modigliani, Franco, and Merton Miller.** 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *The American Economic Review*, 48(3): 261–297.

[184] **Moll, Benjamin.** 2014. "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *American Economic Review*, , (forthcoming).

[185] **Moll, Benjamin, Robert M. Townsend, and Victor Zhorin.** 2014. "Economic Development and the Equilibrium Interaction of Financial Frictions." Working Papers.

[186] **Mulligan, Casey B.** 2009. "Means-Tested Mortgage Modification: Homes Saved or Income Destroyed?" National Bureau of Economic Research, Inc NBER Working Papers 15281.

[187] **Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny.** 1989. "Building Blocks of Market Clearing Business Cycle Models." In *NBER Macroeconomics Annual 1989, Volume 4. NBER Chapters*, 247–302. National Bureau of Economic Research, Inc.

[188] **Nakamura, Emi, and Jón Steinsson.** 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *The Quarterly Journal of Economics*, 123(4): 1415–1464.

[189] **Nardi, Mariacristina De.** 2004. "Wealth Inequality and Intergenerational Links." *Review of Economic Studies*, 71(3): 743–768.

[190] **Nekarda, Christopher J., and Valerie A. Ramey.** 2013. "The Cyclical Behavior of the Price-Cost Markup." National Bureau of Economic Research, Inc NBER Working Papers 19099.

[191] **Okuda, Hidenobu, and Fumiharu Mieno.** 1999. "What Happened to Thai Commercial Banks in the Pre-Asian Crisis Period: Microeconomic Analysis of Thai Banking Industry." *Hitotsubashi Journal of Economics*, 40(2): 97–121.

[192] **Papanikolaou, Dimitris.** 2011. "Investment Shocks and Asset Prices." *Journal of Political Economy*, 119(4): 639 – 685.

[193] **Paulson, Anna L., Robert M. Townsend, and Alexander Karaivanov.** 2006. "Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship." *Journal of Political Economy*, 114(1): 100–144.

[194] **Phelps, Edmund, and Sidney G. Winter.** 1970. "Optimal Price Policy under Atomistic Competition." In *In E.S. Phelps et al., Microeconomic Foundations of Employmnet and Infaltion Theory*. Norton, New York.

[195] **Phelps, Edmund S.** 1992. "Consumer Demand and Equilibrium Unemployment in a Working Model of the Customer-Market Incentive-Wage Economy." *The Quarterly Journal of Economics*, 107(3): 1003–32.

[196] **Pissarides, Christopher A.** 1994. "Search Unemployment with On-the-Job Search." *Review of Economic Studies*, 61(3): 457–75.

[197] **Pissarides, Christopher A., and Barbara Petrongolo.** 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature*, 39(2): 390–431.

[198] **Postel-Vinay, Fabien, and Jean-Marc Robin.** 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica*, 70(6): 2295–2350.

[199] **Pratt, John W.** 1964. "Risk Aversion in the Small and in the Large." *Econometrica*, 32(1): 122–136.

[200] **Reinhart, Carmen M., and Kenneth S. Rogoff.** 2011. "From Financial Crash to Debt Crisis." *American Economic Review*, 101(5): 1676–1706.

[201] **Reinhart, Carmen M., and Kenneth S. Rogoff.** 2014. "This Time is Different: A Panoramic View of Eight Centuries of Financial Crises." *Annals of Economics and Finance*, 15(2): 1065–1188.

[202] **Rotemberg, Julio J, and Garth Saloner.** 1986. "A Supergame-Theoretic Model of Price Wars during Booms." *American Economic Review*, 76(3): 390–407.

[203] **Rotemberg, Julio J., and Michael Woodford.** 1991. "Markups and the Business Cycle." In *NBER Macroeconomics Annual 1991, Volume 6. NBER Chapters*, 63–140. National Bureau of Economic Research, Inc.

[204] **Rotemberg, Julio J, and Michael Woodford.** 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy*, 100(6): 1153–1207.

[205] **Rotemberg, Julio J., and Michael Woodford.** 1993. "Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets." National Bureau of Economic Research, Inc NBER Working Papers.

[206] **Rothstein, Jesse, and Cecilia Elena Rouse.** 2011. "Constrained After College: Student Loans and Early-Career Occupational Choices." *Journal of Public Economics*, 95(1-2): 149–163.

[207] **Rubinstein, Yona, and Yoram Weiss.** 2006. "Post Schooling Wage Growth: Investment, Search and Learning." , ed. Erik Hanushek and F. Welch Vol. 1 of *Handbook of the Economics of Education*, Chapter 1, 1–67. Elsevier.

[208] **Saez, Emmanuel.** 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68: 205–229.

[209] **Schrag, Philip G.** 2001. "The Federal Income-Contingent Repayment Option for Law Student Loans." *Hofstra Law Review*, 29(1): 733–862.

[210] **Shimer, Robert.** 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95(1): 25–49.

[211] **Silvio, Rendon.** 2006. "Job Search And Asset Accumulation Under Borrowing Constraints ." *International Economic Review*, 47(1): 233–263.

[212] **Stiglitz, Joseph E.** 2015. "Remarks on Income Contingent Loans: How Effective Can They Be at Mitigating Risk?" , ed. Joseph E. Stiglitz, Bruce Chapman and Timothy Higgins *Income Contingent Loans: Theory, Practice and Prospects*, 31–38. Houndmills, UK and New York: Palgrave Macmillan.

[213] **Stiglitz, Joseph E., Timothy Higgins, and Bruce Chapman.** 2014. *Income Contingent Loans: Theory, Practice and Prospects.* Palgrave Macmillan.

[214] **The Executive Office of the President of the United States.** 2016. "Investing in Higher Education: Benefits, Challenges, and the State of Student Debt." Executive Office of the President of the United States.

[215] **The Institute for Higher Education Policy.** 2014. "2014 Release of Federal Student Aid Data." Institute for Higher Education Policy.

[216] **Townsend, Robert M.** 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory*, 21(2): 265–293.

[217] **Townsend, Robert M., and Kenichi Ueda.** 2006. "Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation." *Review of Economic Studies*, 73(1): 251–293.

[218] **Townsend, Robert M., and Kenichi Ueda.** 2010. "Welfare Gains From Financial Liberalization." *International Economic Review*, 51(3): 553–597.

[219] **Volkwein, J. Fredericks, Alberto F. Cabrera, Bruce P. Szelest, and Michelle R. Napierski.** 1998. "Factors Associated with Student Loan Default among Different Racial and Ethinic Groups." *Journal of Higher Education*, 69(2): 206–2372.

[220] **Weber, Michael.** 2014. "Nominal Rigidities and Asset Pricing."

[221] **Weidner, Justin.** 2016. "Does Student Debt Reduce Earnings?" Princeton University.

[222] **Wernerfelt, Birger, and Cynthia A Montgomery.** 1988. "Tobin's q and the Importance of Focus in Firm Performance." *American Economic Review*, 78(1): 246–50.

[223] **Whited, Toni M.** 1992. " Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data." *Journal of Finance*, 47(4): 1425–60.

[224] **Yannelis, Constantine.** 2015. "Asymmetric Information in Student Loans." New York University.

[225] **Zimmermann, Christian.** 1997. "International Real Business Cycles Among Heterogeneous Countries." *European Economic Review*, 41(2): 319–356.