From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2024

Essays in Macroeconomics with Micro Frictions and Heterogeneity

Li, Lintong

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01ks65hg59n

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays in Macroeconomics with Micro Frictions and Heterogeneity

Lintong Li

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Economics
Adviser: Nobuhiro Kiyotaki

September 2024

© Copyright by Lintong Li, 2024.

All Rights Reserved

Abstract

The thesis consists of three essays on heterogeneous firms and households in general equilibrium. The broad theme of the thesis is to understand how idiosyncratic and aggregate shocks to firms and households affect the aggregate economy and income distribution when there are frictions in markets and technology.

The first chapter studies the propagation of aggregate shocks through the interaction between financial frictions and the partial irreversibility of fixed capital and relevant policies. At the individual level, the interaction between financial frictions and irreversibility reinforces each other, leading to misallocation, inefficient liquidation, and default. At the aggregate level, lower aggregate demand enters revenue productivity and creates a larger distortion in production in the presence of monopolistic competition. Quantitative results show that 2 percent TFP shock generates a maximum of 3.9 percent drop in output. A counter-cyclical fiscal policy can mitigate the recession by boosting aggregate demand, resulting from a rich interaction between liquidity, asset price, and financial frictions.

The second chapter, a joint work with Qing Huang, documents the trend in time allocation in China from 2008 to 2020 and explores the explanations behind the observed empirical pattern. We find weekly market hours per person increase by 3 to 6 in urban areas while time spent on core housework decreases by a comparable magnitude. Quantitatively, we find that rising income uncertainty, changing demographic structures, and capital augmenting productivity growth in home production explain the observed empirical trend in market hours and non-market hours and the correlation between market hours and wage rate reasonably well.

The third chapter studies the heterogeneity in residual inequality by education groups and its implications in macroeconomics. It documents that high-skilled labor has experienced a larger increase in wage rate residual inequality than low-skilled labor from 1980 to 2018, while the variances of transitory shocks, permanent shocks, and initial dispersion of

permanent components have been widening between the two groups. Quantitative results suggest the model can match the changes in empirical moments for two groups over time, and this heterogeneity across skill groups has significant quantitative implications for aggregate variables.

Acknowledgements

I am deeply indebted to my advisor, Nobuhiro Kiyotaki, for his continuous guidance and support during my doctoral studies. His invaluable advice has greatly improved the quality of my research and inspires me to work on what I love. I thank him for his patience and trust when I had an unclear path toward the completion of my dissertation. He has been dedicated to so much time advising me that I felt I possibly wasted his time sometimes. His generosity and pure love for research empower me to pursue excellence in academics and to become a better person. Nobu once told me that his advisor said that the only thing a professor can teach students is the passion for research, while I think Nobu is much more successful than that.

I am much obliged to Richard Rogerson for his advice and backing throughout the program. Richard's comments always bring me closer to the fundamentals of economics. He uses the simplest words, examples, and knowledge structure to guide me in observing, analyzing, and describing the nature of the world. Also, I gratefully acknowledge invaluable teaching and feedback from the faculty at Princeton. Special thanks to Mark Aguiar, John Grigsby, Moritz Lenel, Ezra Oberfield, Jonathan Payne, Mikkel Plagborg-Moller, Stephen Redding, Karthik Sastry, Gianluca Violante, Wei Xiong, and Motohiro Yogo. I would like to thank Laura Hedden and Jennifer Bello for their help in navigating life at Princeton. I would like to extend my gratitude to Richard Freeman and Xiaobo Zhang, who guided me in pursuing graduate studies in economics, and Justin Yifu Lin, who motivated me to study economics at Peking University.

During my graduate years, I have been fortunate to meet with many extraordinary students and become friends with them. I thank my friends Narek Alexanian, Shushen Hou, Dake Li, Ziang Li, Lucas Mann, Georgios Nikolakoudis, Don Noh, Shumiao Ouyang, Jeffrey Sun, Yucheng Yang, Fangyuan Yi, Mu Zhang for all the support. I am deeply thankful to my co-author, Qing Huang, for our joint paper. I have been lucky enough to have fantastic

roommates, Zhang Chen, Lunyang Huang, and Haiyue Yu, over the years for helping each other get through life and debating economics. Special thanks to Jing Qian, Xueqi Li, Sifan Xue, Tianhan Zhang, and Yantao Wu for inviting me to lots of great dinners and sincere conversations.

I thank my parents for their unconditional love over the past 30 years. They teach me to pursue what I love. I am fortunate to have grown up with my brother, Qitong Li, who is always my best friend. We started to explore the world together from the very beginning. In the end, I cannot express enough gratitude to my girlfriend, Mengqi Yang, for always staying by my side. It is my great fortune to embark on a new chapter of my life together with her. This dissertation is a dedication to them.

To my family

Contents

Abstract 3

Acknowledgements 5

1 Aggregate Demand and Irreversible Investment **16**

1.1 Introduction 16

1.2 Model 25

1.2.1 Environment 25

1.2.2 Technology 26

1.2.3 Intermediary Firms' Problem 28

1.2.4 Bank and Debt Price 32

1.2.5 New entry 33

1.2.6 Capital Goods Producer 33

1.2.7 Households 34

1.2.8 Government 36

1.2.9 Definition of Recursive Competitive Equilibrium 36

1.3 Model Characterization and Empirical Evidence 37

1.3.1 Default, Constrained and Unconstrained Zones 38

1.3.2 Decision Rules 40

1.3.3 Empirical Evidence 48

1.4 Calibration 56

1.4.1 Externally Calibrated Parameters . . . 57
1.4.2 Internally Calibrated Parameters . . . 57
1.5 Quantitative Results . . . 59
1.5.1 Aggregate Responses to Productivity Shock . . . 60
1.5.2 Counter-Cyclical Fiscal Policy . . . 64
1.5.3 Counter-Cyclical Credit Policy . . . 68
1.6 Conclusion . . . 69

2 Richer and Busier? The Facts, Causes and Consequences of Labor Supply in China 71

2.1 Introduction . . . 72
2.2 Data . . . 79
2.3 Empirical Facts . . . 82
2.3.1 Aggregate Hours Worked . . . 82
2.3.2 Extensive Margin vs. Intensive Margin . . . 85
2.3.3 Hours Worked for Different Demographic Groups . . . 86
2.3.4 Hours and Wages . . . 90
2.4 Why Hours Still Increase While Wages Grow? . . . 92
2.5 Model . . . 96
2.5.1 Demographics . . . 97
2.5.2 Production . . . 97
2.5.3 Preferences . . . 98
2.5.4 Households' Problem . . . 99
2.5.5 Firm's Problem . . . 101
2.5.6 Government . . . 101
2.5.7 Competitive Equilibrium . . . 102

2.6 Calibration to the Chinese Economy . . . 103
2.6.1 First-Stage Calibration . . . 104
2.6.2 Second and Third Stage Calibration . . . 109
2.7 Quantitative Results . . . 110
2.8 Conclusion . . . 115

3 Macroeconomic Implications of Heterogeneity in Residual Inequality 116
3.1 Introduction . . . 116
3.2 Literature Review . . . 120
3.3 Data . . . 122
3.3.1 CPS . . . 122
3.3.2 PSID . . . 122
3.4 Heterogeneity in Income Process by Education Groups . . . 123
3.4.1 Heterogeneity in Residual Variance . . . 123
3.4.2 Income Process Estimation: stochastic component . . . 127
3.4.3 Deterministic Component . . . 133
3.5 Model . . . 136
3.5.1 Demographics . . . 136
3.5.2 Income process . . . 137
3.5.3 Production . . . 137
3.5.4 Household Problem . . . 138
3.5.5 Definition of Stationary RCE . . . 138
3.6 Quantitative Results . . . 139
3.6.1 Calibration . . . 140
3.6.2 Correlation between Log wage and Log hour . . . 142
3.6.3 Variance of Log (hour) . . . 146

3.6.4 Variance of Log (earning) . . . 148
3.6.5 Implications on Aggregate Moments . . . 150
3.7 Conclusion . . . 151

A Appendix to Chapter 1 **154**
A.1 Data Construction . . . 154
A.1.1 Summary Statistics . . . 154
A.2 More on Empirical Results . . . 155
A.3 More on Theoretical Results . . . 156
A.3.1 Toy Example . . . 156
A.3.2 Proof . . . 157
A.4 More on Quantitative Results . . . 162

B Appendix to Chapter 2 **164**
B.1 More on Empirical Results . . . 164

List of Figures

1.1 Timeline of Intermediary Firms' Problem 28
1.2 Default and Unconstrained Threshold 40
1.3 Decision Rules Among Unconstrained Firms 42
1.4 Decision Rules Among Constrained Firms 46
1.5 Counter Cyclical Dispersion in MRPK 56
1.6 IRFs to TFP Shock 63
1.7 Decision Rules under Productivity Shock 64
1.8 IRFs to TFP Shock: Two Sector Model 65
1.9 IRFs to TFP Shock: Fiscal Policy 66
1.10 Consumption Equivalent Welfare Gains under Fiscal Policy 68
1.11 IRFs to TFP Shock: Credit Policy 70

2.1 Market Hours Per Worker Across Countries 74
2.2 Hours Per Worker and Wage Rate, 2010-2020 91
2.3 Correlation of Hour and Wages: 2010 vs. 2020 92
2.4 The Decline of Non-Market Hours: Food Production and Washing Machine . 94
2.5 Wage Inequality Between 2010 and 2020 95
2.6 Age-specific Death Rate and Birth Rate 96

3.1 Varinace of Log Male Wage, CPS 124

3.2 Heterogeneity in Residual Inequality by Skill Groups 126
3.3 Variance of Transitory Shocks 130
3.4 Variance of Permanent Shocks 131
3.5 Variance of Log Male Wage Residuals at Age 25 132
3.6 College/High Graduate Wage Premium 134
3.7 Growth Rate of Relative College/High Labor Supply 134
3.8 Secular Trend of Experience Premium 135
3.9 Correlation between Log (wage) and Log (hour) 142
3.10 Variance of Log (hour) 146
3.11 Variance of Log (earning) 148

A.1 Sector Level Capital Recovery Rate 155
A.2 Toy Example of Investment Decisions 156
A.3 The Role of Monopolistic Competition 162

B.1 Tax Rate, 2010 vs. 2020 164

List of Tables

1.1 Exit Rate and Firm Characteristics 51

1.2 Marginal Productivity and Financial Frictions under Heterogeneous Investment Decisions 52

1.3 Sector Level Fixed Capital Recovery Rate and Outcomes in Recessions . . . 54

1.4 Summary of Externally Calibrated Parameters 58

1.5 Summary of Internally Calibrated Parameters 59

1.6 Calibration Target and Model Fit 59

2.1 Time Allocation by Area and Gender: Hours Per Person 84

2.2 Employment Rate and Hours Per Worker by Area and Gender 86

2.3 Hours Per Worker: Heterogeneity 89

2.4 Summary of Externally Calibrated Parameters 106

2.5 Income Process Estimation 108

2.6 Summary of Internally Calibrated Parameters 110

2.7 Model Outcomes 112

3.1 Within-Group Residual Variance of Wages by Experience-Education Cell for Men 125

3.2 Income Process Estimation 131

3.3 Calibration Parameters 142

3.4 Correlation between Log (wage) and Log (hour) 143
3.5 Variance of Log (hour) 146
3.6 Variance of Log (earning) 149
3.7 Implications on Aggregate Moments 151

A.1 Average of Key Variables by Firm Size 154
A.2 The Correlation between Tightness of Financing Constraint and Marginal Revenue of Inputs 156

B.1 Elasticity of Hours to Individual Income 165

Chapter 1

Aggregate Demand and Irreversible Investment

[1]

1.1 Introduction

One of the key research questions in macroeconomics is to what extent micro-level frictions and heterogeneity are important to understanding the dynamics of aggregate variables. This article studies the propagation of first-moment aggregate shocks through the interaction between financial frictions and the partial irreversibility of fixed capital and policies to mitigate recessions. Partial irreversibility and financial frictions have been separately studied exten-

[1]I am highly indebted to Nobuhiro Kiyotaki for all his guidance and support and to Richard Rogerson, Mark Aguiar, and Wei Xiong for their insightful comments and discussions. I also thank David Argente, Zhang Chen, John Grigsby, Ricardo Lagos, Moritz Lenel, Ezra Oberfield, Pablo Ottonello, Jonathan Payne, Stephen Redding, Karthik Sastry, Andreas Schaab, Florian Trouvain, Gianluca Violante, Motohiro Yogo, Sifan Xue, Haonan Zhou for their helpful comments, as well as all the participants in the Princeton Macro Student Workshops. Any errors that remain are solely the responsibility of the authors.

sively in the literature [2,3]. Although the financial frictions and irreversibilities have been shown to be matching micro data on investment and financing decisions [4], each has been found to generate relatively little amplification of aggregate shocks to the level of revenue productivity. Specifically, aggregate dynamics are quantitatively similar to those in a standard RBC model with first-moment shocks [5,6]. One requires exogenous financial shocks and uncertainty (second-moment) shocks to make the micro-level frictions matter for aggregate dynamics[7].

In this article, I revisit these irrelevance perspectives and argue that the interaction between partial irreversibility and financial frictions is essential to determine individual firms' investment decisions and generate substantial amplification effects in response to first-moment shocks. To make my argument, I build a quantitative general equilibrium model populated by heterogeneous firms that face idiosyncratic productivity shocks and endogenous default risks under monopolistic competition. Firms use internal funds and external borrowing to finance both fixed capital and working capital investment. External borrowing is limited since equity issuance is not allowed, and external debt needs to be secured with

[2]Technically speaking, a wedge between the purchasing and selling price of capital summarizes frictions, including asset specificity, indivisibility, and search frictions. Kermani and Ma (2023) shows the wedge is significant. They document that the liquidation value of fixed assets is 35% of the net book value in the average industry in the U.S. It is also motivated by the observations that gross investment data contains many zeros and is highly asymmetric, see Cooper and Haltiwanger (2006).

[3]It is often modeled as collateral constraints following Kiyotaki and Moore (1997) as debt is not secured with limited commitment. There is a large literature in corporate finance. For example, see Whited (1992) and Bolton et al. (2011).

[4]For example, see Cooper et al. (1999),Cooper and Haltiwanger (2006); Hennessy et al. (2007),Riddick and Whited (2009), Erickson and Whited (2000) and Cao et al. (2019).

[5]The irrelevance result on non-convex adjustment cost is established in a sequential work by Thomas (2002), Khan and Thomas (2008) and House (2014). In short, investment is forward-looking in the presence of non-convex adjustment costs and very sensitive to the change in the price of capital.

[6]Although Kiyotaki and Moore (1997) articulates theoretically how financial frictions affect investment and create capital misallocation, Kocherlakota et al. (2000) shows limited amplification effect of financial frictions in a calibrated model with labor. Khan and Thomas (2013) shows limited misallocation is created in response to aggregate productivity shocks in a quantitative model with collateral constraints.

[7]For example, Bloom (2009) shows ignoring capital adjustment costs would lead to substantial bias in accounting the impacts of uncertainty shocks and Bloom et al. (2007) shows higher uncertainty reduces the responsiveness of investment to demand shocks in the presence of irreversibility.

collateral. Fixed capital is partially irreversible as a constant fraction of capital will be lost when the ownership is transferred.

When calibrated to the Chinese economy, with parameters disciplined by firm-level data in the Chinese manufacturing sector, I find a two percent negative TFP shock leads to a maximum of 3.9 percent drop in output, amplified by a 20 percent increase in default density and around 2 percent increase in the standard deviation of log MRPK. This amplification effect is much smaller in models where I shut down either financial frictions or irreversibility.

To understand the interaction of irreversibility and financial friction and how the two frictions reinforce each other, consider a firm that receives a temporary negative shock to the level of revenue productivity. Suppose there is no friction; it is optimal for the firm to downsize fixed capital and working capital instantly and simultaneously. When we only add partial irreversibility, firms find immediate downsizing costly and might want to hold the fixed capital and enjoy the option value, delivering an investment inaction zone. Meanwhile, firms will also choose a higher working capital as fixed capital and working capital are complements. One key assumption I made is both fixed capital and working capital investment are subject to financing constraints in my model. This is in contrast to most previous work that assumes labor and intermediate inputs are free from financial frictions[8].

With financial frictions on top of irreversibility, suppose firms do not choose to sell fixed capital immediately. Firms would not generate cash flow by liquidating fixed capital, but the funding demand to finance working capital would be higher. As a result, the existence of irreversibility makes financing constraints tighter. Conversely, when the financing constraint on working capital binds, MRPK will be lower since working capital is not financed to the optimal level, and the optimal level of fixed capital holding will be lower, which tightens the "irreversibility constraint" and enlarges the investment inaction zone. Consequently, firms

[8]Though researchers often get abstraction from financial frictions on working capital, Fazzari and Petersen (1993) do find that working capital competes with fixed capital investment for a limited pool of finance. This argument is also used in the international macro literature for amplification effect see Mendoza (2010).

would have too much fixed capital but too little working capital in the presence of the interaction of the two frictions, and misallocation rises. The production would be very unbalanced and unprofitable, while firms are expected to lose net worth quickly. This inefficient input choice is difficult to solve since the fixed capital choice is constrained by irreversibility, and the working capital choice is constrained by financial frictions.

By recognizing that firms are losing net worth, default risks and credit spread pick up. When a firm is close to the default threshold with a high marginal value of liquid wealth, a fire sale is triggered and one has to sell the fixed capital even though it is very costly. Firms are gambling between deleveraging, selling fixed capital at a shabby price with lower expected default risks but lower option value, and keeping fixed capital stock with a higher expected default risk but compensated by a higher option value conditional on surviving. Formally, endogenous default rises due to limited commitment and incomplete markets as debt is non-state contingent and defaultable. Firms with too weak fundamentals are not able to roll over the debt. The default threshold on the book value of net worth and risk premium can be solved simultaneously in a fixed point problem. I show that this default threshold is weakly increasing in fixed capital conditional on productivity. These results highlight the additional risks of debt overhang when capital is not fully reversible. In response to aggregate adverse shocks to the level of revenue productivity, the default threshold will be higher and rise more for firms with high fixed capital. As a result, a more significant density of firms have to default. Because firms do not produce and are replaced by small and unproductive firms when they default, the effect of default on aggregate output is persistent.

My model features monopolistic competition, while most previous related studies utilize a decreasing return to scale production functions. With monopolistic competition, aggregate demand externality rises à la Blanchard and Kiyotaki (1987). In other words, aggregate demand (aggregate output) enters individual firms' revenue productivity. As a result, output loss caused by misallocation and default will have a spillover effect since aggregate demand

is shifted inward. This shift can be viewed as a quasi-negative productivity shock and creates an endogenous feedback loop. Furthermore, the imperfect market setting allows the possibilities of analyzing demand shocks[9] and policies that affect aggregate demand. Additionally, my model assumes a fixed measure of agents are hand-to-mouth workers who only supply labor and have no access to the financial market. When aggregate adverse shocks hit, the labor demand decreases due to exogenous shocks and endogenous feedback through tighter financing constraints, extra default, and lower aggregate demand. A lower labor demand pushes down the wage rate and wage bill for the hand-to-mouth workers. As a result, the consumption demand also contracts, amplified by the supply-side distortion and the existence of hand-to-mouth agents.

Next, I utilize Chinese manufacturing data and provide supportive empirical evidence for my model. First, I test the model-implied decision rules on default and exit. I find that a firm's exit probability is only positively correlated with fixed capital when the firm adjusts its capital stock non-positively, conditional on net worth and productivity. This is in line with the model that suggests net worth and productivity are not sufficient statistics in deciding the default threshold only if upward fixed capital adjustment is not optimal. Second, I test the model-implied decision rules on fixed and working capital investment. I find that the correlation between the tightness of financial frictions and MRPK is positive for the firms with active investment adjustment, while the correlation is reversed for the firms in the investment inaction zone, a unique feature generated by the interaction between financial frictions and irreversibility implied by the model. Firms equalize the marginal value of liquid wealth and the marginal benefit of investment when making fixed capital investments. When firms are outside the investment inaction zone, a higher marginal value of liquid wealth (tighter financing constraints) is associated with a lower fixed capital choice and higher MRPK. In contrast, when firms are inside the investment inaction zone, a higher marginal

[9]See Kiyotaki (2011) for discussion on this.

value of liquid wealth is associated with a lower working capital choice and lower MRPK. At the sector level, I find that sectors with higher capital recovery rates experience smaller increases in the dispersion of MRPK and a smaller drop in output in recessions, highlighting the role of irreversibility in propagating macroeconomic shocks through misallocation. At the aggregate level, I show that dispersion in MRPK is countercyclical in the manufacturing sector in both China and the US.

Then, I calibrate the model to the Chinese economy. The capital recovery rate, defined as the ratio of liquidation value to book value of fixed capital, governs i) the size of the investment inaction zone and how costly it is to shrink in response to negative shocks, ii) the collateral value of fixed capital, the tightness of financial frictions and how costly of default to the lenders. In the exercises, The capital recovery rate is chosen to match the fraction of inactive firms in the steady state. The default density in the steady state is jointly disciplined by choosing the fixed cost in operation and capital recovery rate to match the empirical non-performing loan rate. With the calibrated model, I fit in ex-ante unexpected and ex-post fully anticipated negative productivity shocks and study the transition dynamics. I find that a 2 percent negative productivity shock generates a maximum of 3.9 percent drop in output, amplified by a 20 percent increase in default density and a 2 percent increase in the standard deviation of MRPK from the steady state. I also find that endogenous feedback through aggregate demand externality accounts for around one-third of the total amplification effect. In contrast, when the irreversibility or financial frictions are shut down, the maximum decline in output is close to 2 percent. In an economy without financial frictions, dispersion in MRPK only increases modestly, and the default channel is shut down. The investment inaction zone does not move by much because investment is a long-lived object and very elastic to the changes in capital price in the presence of real frictions. In an economy without real friction. Default density only rises modestly, and dispersion in MRPK even goes down. Financial frictions tend to be slightly looser on average in response to aggregate adverse productivity

shocks. This is because the decline in asset demand outweighs the decline in collateral value. Capital demand is very elastic to productivity shocks since it becomes a static choice without real frictions, while capital price is forward-looking.

My model also contains novel policy implications. The inefficiency comes from financial frictions and is amplified by the irreversibility. I show that it is beneficial to boost aggregate demand even though my model does not carry any nominal rigidity. In contrast to sticky price, my model features a combination of sticky quantity and financial friction. A higher aggregate demand releases firms from investment inaction zones and relaxes financing constraints endogenously. Better credit conditions improve supply-side efficiency and higher output feedback as additional positive demand shock with self-amplification. Quantitative results show that a counter-cyclical fiscal policy can mitigate recessions by lowering misallocation and default density. Modest welfare gains are achievable even though government expenditure does not enter household utility functions. Along the transition path, the optimal fiscal policy suggests that government expenditure should increase by 2.1 percent in response to a 1 percent drop in GDP. I also show that counter-cyclical credit policy can make the economy more resilient to adverse shocks. However, the tightening financial frictions is associated with significant welfare loss in the steady state.

Related Literature This article relates to a broad research agenda that studies the implications of micro frictions and heterogeneity in macroeconomics. This article mainly contributes to four strands of literature. First, this paper contributes to our understanding of the propagation and amplification of aggregate shocks through micro-level frictions. Abel and Eberly (1994),Abel and Eberly (1996) and Bertola and Caballero (1994) characterize the optimal decision rules under irreversibility[10]. In response to the first moment shocks, the

[10]Baley et al. (2022) and Chen et al. (2023) investigate the impact of tax policies in the presence of irreversibility.

sole presence of non-convex adjustment cost does not affect aggregate investment by much as shown by Thomas (2002), Khan and Thomas (2008) and House (2014) while Winberry (2021) articulates the elasticity is pro-cyclical. Similarly, in response to first-moment shocks, the sole presence of financial frictions does not deliver a large amplification effect. In contrast, second-moment shocks or financial shocks are required to make micro-level frictions relevant.(Gilchrist et al. (2014), Arellano et al. (2019), Khan and Thomas (2013), Bloom et al. (2018)) Existing work that jointly investigates aggregate implications of the two frictions together is limited, with several exceptions. Cui (2022) uses this interaction to rationalize the "capital reallocation puzzle"[11] and Caggese (2007) is the closest to this article, which illustrates this interaction is key to explaining why input inventories and material deliveries are so volatile, procyclical, and asymmetric in industry dynamics. This article shows this interaction is essential in understanding the endogenous amplification in response to first-moment shocks.

Second, this paper contributes to models with financial heterogeneity and endogenous default risk. My model builds on Khan et al. (2014). By applying this framework, Ottonello and Winberry (2020) studies the transmission of monetary policy, and Guntin (2022) studies rollover risks. Compared to their work, I add irreversibility on fixed capital to show that portfolio choice matters for default risks. Additionally, I add monopolistic competition which delivers endogenous amplification through aggregate demand and allows possibilities of analyzing demand shocks.

Third, this paper contributes to our understanding of cyclical properties of misallocation both empirically and theoretically. The effect of misallocation on aggregate output with heterogeneous production units is highlighted by seminal work Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Moll (2014) and Midrigan and Xu (2014) argue that the

[11]Eisfeldt and Rampini (2006) find capital reallocation is procyclical and the benefits to capital reallocation appear countercyclical. Lanteri (2018) rationalize the same thing without financial frictions but the endogenous price for used capital.

efficiency loss caused by financial frictions in steady states is small if productivity shocks are persistent. Secular trends of misallocation have been studied in the context of development Buera et al. (2011), Buera and Shin (2013), transition in response to interest rate shocks Gopinath et al. (2017) and heterogeneous markups Baqaee and Farhi (2020). The evidence on the cyclicality of misallocation is relatively thin except for Sandleris and Wright (2014) and Oberfield (2013). I contribute to this literature by characterizing how irreversibility reinforces financial frictions and providing novel empirical evidence on misallocation over business cycles.

Finally, this paper contributes to the growing literature on the distributional effect of macroeconomic policies. Kaplan et al. (2018) studies the transmission of monetary policy with heterogeneous households in the presence of illiquid assets and borrowing constraints. My model shares a similar spirit in the supply side. Financially constrained firms would not liquidate fixed capital to finance working capital and behave like "rich hand-to-mouth" firms. To boost aggregate demand in the absence of nominal rigidity, I choose fiscal policy which basically redistributes wealth from agents with low MPC to agents with high MPC (the government). Farhi and Werning (2016) shows the macro-prudential policy can do the same thing by redistributing wealth implicitly when monetary policy is constrained. This distributive externality is also important on the supply side as shown by Lanteri and Rampini (2023). Baqaee et al. (2023) is the closest to this article which studies the transmission of demand shock into supply-side allocative efficiency. The difference is the friction in their work is heterogeneous pass-throughs, while the friction in my paper is the combination of real and financial frictions.

Road Map Section 2 lays out the model. Section 3 characterizes the model and presents related empirical evidence. Section 4 calibrates and validates the model. Section 5 presents the quantitative results and Section 6 concludes.

1.2 Model

In this section, I develop a heterogeneous firm model that features financial frictions and partially irreversible fixed capital to study the propagation of first-moment shocks. Following Khan et al. (2014) and Ottonello and Winberry (2020), the core of the economy is a continuum of risk-neutral firms facing uninsurable productivity shocks. They use internal and external resources to finance fixed and working capital investments. External financing is constrained as no equity issuance is allowed, and borrowing is not possible unless the debt is secured with collateral. Fixed capital investment is irreversible, as a constant fraction of capital will be lost when the ownership is transferred. I focus on the steady state equilibrium without aggregate uncertainty and study the transition dynamics in response to an ex-ante unexpected and ex-post fully anticipated

1.2.1 Environment

Time is discrete and infinite. There are three types of goods in the economy. There exists a continuum of intermediate firms that produce differentiated goods using fixed capital and working capital (labor) under monopolistic competition with a measure of one. There exists a final good producer who aggregates differentiated goods into final goods under perfect competition. Final goods can be used for consumption, investment, government expenditure, and production of capital goods. At the aggregate level, capital goods cannot be transferred back to final goods. At the individual level, once invested, capital goods become specific to intermediate firms. A constant fraction of fixed capital will be lost for technological reasons when it is sold to second-hand users. After losing a fraction, used capital and new capital are perfect substitutes..

There are two representative households in the economy. There is one representative entrepreneur family with a measure of M that consists of a fixed measure of entrepreneurs.

Each family member owns an intermediate firm and operates on their own. However, the entrepreneurs would not eat on their own but pay dividends to the family and make the consumption choice together with full insurance within the family. There is also one representative worker family with a measure of $1 - M$ who only supply labor and consume. The worker's family does not have access to the financial market and they are pure hand-to-mouth ("HTM"). A representative capital goods producer uses aggregate capital and new investment to produce new capital with aggregate adjustment cost. A representative bank issues one-period risk-free bonds and makes one-period non-state contingent defaultable loans at rates that take default risks into account. The representative entrepreneur family owns the capital goods producer and the bank.

1.2.2 Technology

Intermediate firm i utilizes fixed capital k_{it} and working capital l_{it} to produce differentiated intermediate goods with a constant return to scale technology at time t. However, fixed capital k_{it} and working capital l_{it} are fully financed and chosen one period before at $t-1$. In our model setting, labor input is the working capital. A_t is the aggregate total factor productivity. The persistent component of idiosyncratic productivity z_{it} follows log AR(1) process and the transitory component of idiosyncratic productivity ε_{it} is drawn from a log-normal distribution i.i.d..

$$y_{it} = A_t z_{it} \varepsilon_{it} k_{it}^{\alpha} l_{it}^{1-\alpha}$$

$$\log z_{i,t} = \rho \log z_{i,t-1} + \nu_{it}, \quad \nu_{it} \sim N(0, \sigma_{\nu}^2) \tag{1.1}$$

$$\log \varepsilon_{it} \sim N(0, \sigma_{\varepsilon}^2) \tag{1.2}$$

The quantity of aggregate output, which can be used as either a consumption good or investment good, is a function of many differentiated goods as:

$$Y_t = (\int_0^1 (y_{it})^{\frac{\theta-1}{\theta}} di)^{\frac{\theta}{\theta-1}}$$

where θ is the elasticity of substitution between differentiated goods and $\theta > 1$. Let p_{it} be the price of intermediate good, and we normalize the price index that corresponds to the above aggregate output to be one:

$$P_t = (\int_0^1 p_{it}^{1-\theta} di)^{\frac{1}{1-\theta}} = 1$$

Under monopolistic competition, intermediary firms face a downward-sloping demand curve and make decisions taking aggregate output as given:

$$\frac{y_{it}}{Y_t} = (\frac{p_{it}}{P_t})^{-\theta}$$

We assume that the model does not carry any nominal rigidity, and there is no inventory. As a result, as inputs are predetermined, the price for differentiated goods is implied by the demand curve directly. Then the revenue for firm i in terms of final good can be written as:

$$\frac{p_{it} y_{it}}{P_t} = (A_t z_{it} \varepsilon_{it} k_{it}^{\alpha} l_{it}^{1-\alpha})^{1-\frac{1}{\theta}} \times Y_t^{\frac{1}{\theta}} \tag{1.3}$$

Aggregate demand enters into an individual's revenue production function. As a result, supply-side distortions and demand shocks can shift the aggregate demand Y_t and will affect the real revenue of firms, even though there is no price stickiness.

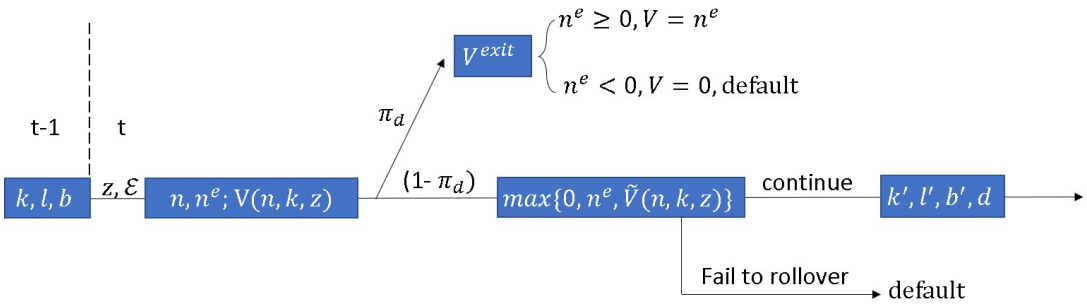
1.2.3 Intermediary Firms' Problem

Firms are risk-neutral, and the firm i's objective is to maximize the expected discounted value of dividends:

$$V_{i0} = \mathbb{E}_0 \sum_{t \geq 0} \Lambda_t d_{it} \tag{1.4}$$

where Λ_t is the stochastic discounting factor of the representative entrepreneur family and d_{it} is the dividends paid to the entrepreneur family.

Figure 1.1: Timeline of Intermediary Firms' Problem



Notes: This figure illustrates the timeline of a typical intermediary firm.

Timeline At the beginning of the period t, intermediary firm i inherits predetermined fixed capital k_{it}, working capital l_{it}, defaultable non-state contingent loan with debt obligation b_{it} from the end of last period. Aggregate and individual level of productivity is realized. From now on, I drop the subscript and use state variables to describe the problem in the steady state. Notice that we have three state variables. persistent component of productivity z enters into the value function because it is persistent and evolves exogenously. Because of the financial frictions, net worth positions matter for the cost of external funding. We define the book value of the net worth as:

$$n = (Az\varepsilon k^{\alpha} l^{1-\alpha})^{1-\frac{1}{\theta}} Y^{\frac{1}{\theta}} + (1-\delta)qk - b \tag{1.5}$$

where δ is the depreciation rate of fixed capital, and q is the price of fixed capital. Since fixed capital is irreversible, portfolio choices also matter for the firms' choices, and k enters the value function as an endogenous variable.

Then firms receive exogenous death shock with the probability of π_d. The death shock ensures that not all firms save out of financing constraints since they can always delay dividend payments. For firms who receive the exogenous death shock, they choose between exiting after production, paying back debt and dividends, or defaulting on debt straightly. We assume only the one who starts the project can finish the project. A final touch is required to finish production and that will cause some disutility to the firm owners. We also assume there is no time for renegotiation and debt restructuring. As a result, there will be no production under default, leading to deadweight loss. Notice that fixed capital and hired labor will rest for one period.[12] Let $\eta \leq 1$ be the recovery rate when the used capital is sold in the secondary market by firms or the bank. Suppose the firm finishes production, repays debt, and sells the depreciated used capital in the capital market. The liquidation value is:

$$n^e(n, k, z) = n - (1 - \eta)q(1 - \delta)k \tag{1.6}$$

The wedge between the book value of the net worth and the liquidation value comes from the partial irreversibility. And there could be a significant fraction of firms that have a positive book value of the net worth but a negative liquidation value.

For firms with non-negative liquidation value $n^e \geq 0$, firms will finish the production and pay n^e as dividends to the family. If the random draws are not so good and firms' fundamentals are also weak such that $n^e < 0$, then firms have no incentive to finish the production and will default on debt. When default happens, fixed capital is seized by creditors as collateral. Firms will exit the economy permanently and pay zero dividends to

[12]Ottonello (2017) studies capital unemployment with search frictions. Here I assume that fixed capital will not be used in the production in the period associated with default event. The fixed capital will be sold to a new firm at the end of the period.

equity holders. The default value is 0. The exit value can be expressed as:

$$V^{exit}(n,k,z) = \max\{n^e, 0\} = \max\{n - (1-\delta)(1-\eta)qk, 0\} \tag{1.7}$$

With prob $1-\pi_d$, firms can decide on continuing, exiting endogenously, or defaulting on debt on their own will. Let $V(n,k,z)$ be the value of the firm at the beginning of each period, before receiving the exogenous death shock but after the productivity shocks are realized.

Let $V^{cont}(n,k,z)$ be the continuation value in the middle of the period, and we can connect $V(n,k,z)$ and $V^{cont}(n,k,z)$ as:

$$V(n,k,z) = \pi_d V^{exit}(n,k,z) + (1-\pi_d)\max\{V^{exit}(n,k,z), V^{cont}(n,k,z)\} \tag{1.8}$$

Continuation Firms' s Problem Let us describe the firm problem conditional on continuing. Firms will choose current period dividends d, next period capital qk', and working capital wl' by raising fund $F(b',k',l',z)$ from the bank where b' is the non-state contingent debt obligation in the next period. Equation 1.8 and 1.9 characterize the dynamic programming problem.

Financial frictions take two forms. First, equity issuance is not possible [13], and firms are subject to a non-negative dividend payment constraint 1.11. This assumption is standard in the literature, and indirect financing through banks dominates in the Chinese economy, which I calibrate to. Secondly, external borrowing $F(b',k',l',z)$ is bounded because firms don't have the full commitment to repay the debt. When default happens, the bank can only seize all the installed capital as collateral with liquidation value $(1-\delta)\eta k$ but not the potential output even if the materials are ready. $F(b',k',l',z)$ specifies how much one

[13]This assumption can be relaxed to a costly equity issuance. The Lagrangian multiplier on non-negative dividends payment constraint is equivalent to the extra marginal cost of equity issuance. For example, see Lanteri and Rampini (2023) and Guntin (2022).

can borrow in this period when one is scheduled to repay b' in the next period conditional on making input choices k', l' and current productivity z. The implied average borrowing rate is $Q(b', k', l', z) = b'/F(b', k', l', z)$. This object is endogenous, and I will discuss the determination of the borrowing rates in detail in the next sections. These lead to a financial constraint with respect to external resources.

Firms are subject to a flow of cash constraint 1.10. Firms use internal funds n and external borrowing $F(b', k', \ell', z)$ to finance operation fixed cost ξ, investments in fixed capital, working capital $w\ell'$ and dividend payments d. Since fixed capital is irreversible when firms make new investments, the purchasing price of capital is q, and the selling price is ηq when firms make disinvestments.

$$V^{cont}(n, k, z) = \max_{d,k',b',l'} d + \mathbb{E}[\Lambda V(n', k', z')|z] \tag{1.9}$$

Subject to:

$$d = n + F(b', k', l', z) - \xi - wl' - \mathbb{I}_{k'>(1-\delta)k} \cdot qk' - \mathbb{I}_{k'<(1-\delta)k} \cdot q((1-\eta)(1-\delta)k + \eta k') \tag{1.10}$$

$$d \geq 0 \tag{1.11}$$

Default Threshold For Continuing firms Not all continuation firms' problems yield a solution since some of them can't satisfy non-negative dividend payment constraints no matter what choices are made.[14] We can define default threshold $\underline{n}(z, k)$ as the minimum starting net worth such that the maximum of dividends payment one can make is zero.

$$\max_{b',k',l'} \{\underline{n}(z,k) + F(b', k', l', z) - wl' - \xi - \mathbb{I}_{k'>(1-\delta)k} \cdot qk' - \mathbb{I}_{k'<(1-\delta)k} \cdot q((1-\eta)(1-\delta)k + \eta k')\} = 0$$

[14] We let $\tilde{V}(n, k, z) = -\infty$ for firms with $n < \underline{n}(z, k)$.

Rearrange:

$$\underline{n}(z,k) \equiv -\max_{b',k',l'}\left\{F(b',k',l',z) - wl' - \xi - \mathbb{I}_{k'>(1-\delta)k}\cdot qk' - \mathbb{I}_{k'<(1-\delta)k}\cdot q((1-\eta)(1-\delta)k + \eta k')\right\} \tag{1.12}$$

Notice that due to the existence of irreversibility constraint, the default threshold depends not only on net worth but also on capital. When $\eta = 1$, $\underline{n}(z,k)$ will be only dependent on productivity and collapse to $\underline{n}(z)$. We will characterize the properties of this default threshold in the next section.

1.2.4 Bank and Debt Price

The bank is owned by the representative entrepreneur family. It raises funds by issuing one-period risk-free bonds and making one-period non-state contingent deductible loans. The bank takes default risk into consideration when providing financial contracts. When the firms default, the creditors can seize the firms' fixed capital fully. In the case of default, the payment to the creditor will be $\min(b', (1-\delta)\eta q k')$. We assume the banking sector is competitive and not subject to financial frictions. Therefore, it is expected to earn zero profit and risk-neutral. The price of debt is pinned down by the break-even condition:

$$F(b',k',l',b') = \left(\Lambda\left\{\begin{array}{l} b'\cdot[\pi_d\mathbb{I}_{n^{e\prime}>0} + (1-\pi_d)(1-\mathbb{I}_{n'<\underline{n}'}\mathbb{I}_{n^{e\prime}<0})] + \\ \min(b', \eta q(1-\delta)k)\cdot[\pi_d\mathbb{I}_{n^{e\prime}<0} + (1-\pi_d)\mathbb{I}_{n'<\underline{n}'}\mathbb{I}_{n^{e\prime}<0}] \end{array}\right\}\right) \tag{1.13}$$

LHS specifies how much the bank pays at date t, and LHS gives how much the bank is expected to get paid at date t+1, taking the representative entrepreneur's stochastic discounting factor into account. In the first row of equation 1.12, banks get paid back entirely when the liquidation value is larger zero $\mathbb{I}_{n^{e\prime}>0}$ conditional on receiving the death shock, or when the following two does not hold at the same time conditional on not receiving

the death shock: (i) liquidation value is negative (ii) net worth is below the default threshold.

In the second row of equation 1.12, firms would default if the liquidation value is negative conditional on receiving the death shock; or when both the following two holds: (i) liquidation value is negative (ii) the net worth is lower than the default threshold conditional on not receiving the death shock.

Equation A.20 and equation 1.12 constitute a fixed point problem. Debt price and default threshold can be solved simultaneously.

1.2.5 New entry

In each period, a measure of the same mass firms will replace exit firms and keep the mass of continuing firms equal to one. New entrants will be injected with n_0 equity by the representative household and draw initial productivity z_0 from the ergodic distribution $\Gamma(z)$. Then the firm value for the new entrant is $V(n_0, 0, z_0)$. Note that the initial draw of productivity is only useful to form the expectation. Let μ be the measure of new entrants:

$$\mu_t = \pi_d + (1 - \pi_d) \int_0^1 \mathbb{I}[V_{it}^{exit} > V_{it}^{cont}] di \tag{1.14}$$

1.2.6 Capital Goods Producer

There is a representative capital goods producer who produces new capital goods using technology

$$\Phi(\frac{I}{K})K = (\frac{\tilde{\delta}^{1/\phi}}{1 - 1/\phi}(\frac{I_t}{K_t})^{1-1/\phi} - \frac{\tilde{\delta}}{\phi - 1})K \tag{1.15}$$

where I is the amount of final goods used to produce capital, K is the aggregate capital stock in the beginning of the period, $\tilde{\delta}$ is the aggregate depreciation rate in the steady state that takes capital loss due to irreversibility into account, and $\Phi()$ is the aggregate capital adjustment cost function. The capital goods producer chooses I to maximize the profit,

where q is the price of capital goods.

$$\max_{I} q\Phi(\frac{I}{K})K - I$$

FOC:

$$q = \frac{1}{\Phi'(\frac{I}{K})} = (\frac{I/K}{\tilde{\delta}})^{1/\phi} \tag{1.16}$$

In the steady state, $q = 1$ and the price of capital moves with investment in the same direction to capital ratio follows the financial accelerator effect in Bernanke et al. (1999). Furthermore, this function form implies that the capital goods producer makes zero profit in the steady state. The capital goods producer might make net gains or losses D^B which will be transferred back to the representative entrepreneur family lump sum in the transition dynamics.

1.2.7 Households

There are two representative agents on the demand side with the same GHH preferences over consumption C and Labor supply:

$$\mathbb{E}_0 \sum_{t\geq 0} \beta^t Log(C_t^j - \Psi \frac{{L_t^j}^{1+\nu}}{1+\nu}), \quad j \in R, K \tag{1.17}$$

where ψ governs the disutility of labor and ν measures inverse of elasticity of labor supply.

Representative Entrepreneur Family There exists a representative entrepreneur family with a measure of $1 - M$, and we denote variables she chooses with superscript R. The representative entrepreneur family owns all the intermediate firms, banks, and the representative capital goods producer. The family has access to financial markets and trade one-period risk-free bonds. The flow of cash constraint is described in equation 1.18. The family

uses prepaid labor income $w_t L^R_{t+1}$, dividend payment $\int_0^1 d_{it}di$, saving in bond $B^R_t(1+r_t)$, lump sum transfer from the bank and the capital goods producer D^B_t and D^K_t to finance consumption C^R_t, net tax payment to the government T^R_t and bond purchase B^R_{t+1}.

$$C^R_t + B^R_{t+1} = w_t L^R_{t+1} - T^R_t + \int_0^1 d_{it}di + B^R_t(1+r_t) - \mu_t n_0 + D^K_t + D^B_t \tag{1.18}$$

The stochastic discounting factor is defined as:

$$\Lambda_t = \beta \frac{U_c(C^R_{t+1}, L^R_{t+1})}{U_c(C^R_t, L^R_t)} \tag{1.19}$$

And F.O.C. implies:

$$1 = \mathbb{E}_t[\beta \frac{U_c(C^R_{t+1}, L^R_{t+1})}{U_c(C^R_t, L^R_t)} R_{t+1}] \tag{1.20}$$

$$w_t = \mathbb{E}[\frac{\beta U_L(C^R_{t+1}, L^R_{t+1})}{U_C(C^R_t, L^R_t)}] \tag{1.21}$$

Representative Worker There exists a representative worker with a measure of M, and we denote variables she chooses with superscript K. we assume that they do not have access to financial markets. Henceforth referred to as "Hand to Mouth," who just consume their labor income net of taxes (or transfers) each period.

$$C^K_t = w_t L^K_{t+1} - T^K_t \tag{1.22}$$

1.2.8 Government

The government uses lump sum tax collecting from households, tax revenue collecting from banks, and new bond issuance to finance government expenditure and old bond payments.

$$G_t + B^G_{t+1} = T^R_t + T^K_t + (1 + R_t)B^G_t$$

Assumption 1 *If the government only collects lump sum tax from the representative entrepreneur family such that* $\{T^K_t\}^\infty_{t=0} = 0$*, Ricardian equivalence holds. Without a loss of generosity, we assume* $\{B^G_t\}^\infty_{t=0} = 0$

In the benchmark case, we assume $G_t = \overline{G}$.

1.2.9 Definition of Recursive Competitive Equilibrium

Now we describe the steady state recursive competitive equilibrium. The state space is denoted by $\mathcal{S} \equiv \mathcal{N} \times \mathcal{K} \times \mathcal{Z}$. Let $\Sigma_\mathcal{S}$ be the sigma algebra on $\mathcal{S}$ and $(\mathcal{S}, \Sigma_\mathcal{S})$ the corresponding measurable space. Denote the stationary distribution as λ.

A competitive equilibrium is a value function $\{V(s), V^{cont}(s), V^{exit}(s)\}$; intermediate firm decision rules $\{d(s), k'(s), b'(s), l'(s)\}$; prices $\{r, w, q\}$, debt price schedule $F(b', k', \ell', z)$, default threshold $\underline{n}(k, z)$, Aggregates $\{C^R, C^K, L^R, L^K, I, G\}$ and and measures of agents λ, such that:

1. Household choices are determined by (1.19), (1.20), (1.21)
2. Capital good producer optimize (1.16)
3. Given prices, intermediate firm's decision rules solve the continuing firm's problem 1.9 while $V(s), \tilde{V}(s), V^{exit}(s)$ are associated value functions following (1.8), (1.7).

4. The bank price default risk competitively (A.20), consistent with default threshold (1.12)

5. Labor market clears

$$(1-M)L^{R'} + ML^{K'} = (1-\pi_d)\int \mathbb{I}_{V^{cont}>V^{exit}} l' d\lambda + \mu \int l'(n_0, 0, z) d\Gamma(z) \tag{1.23}$$

6. Capital market clears

$$\Phi(\frac{I}{K})K = \underbrace{(1-\pi_d)\int \mathbb{I}_{V^{cont}>V^{exit}}[\mathbb{I}_{k'>(1-\delta)k}(k'-(1-\delta)k) - \mathbb{I}_{k'<(1-\delta)k})\eta((1-\delta)k-k')]\ d\lambda}_{\text{investment by incumbent firms}}$$
$$+\underbrace{[\mu_t \int k'(n_0,0,z)d\Gamma(z)]}_{\text{investment by entry firms}} - \underbrace{[\pi_d \int (1-\delta)\eta k\ d\lambda_t + (1-\pi_d)\int \mathbb{I}_{V^{cont}<V^{exit}}(1-\delta)\eta k\ d\lambda]}_{\text{selling of used capital by exiting firms}} \tag{1.24}$$

7. Goods market clears

$$Y = C + I + G \tag{1.25}$$

8. The distribution of intermediate firms is consistent with the law of motion Γ implied by decision rules and constitutes a fixed point

1.3 Model Characterization and Empirical Evidence

In this section, we focus on steady-state and characterize the decision rules in different scenarios. We first split the state space into three zones following 1 and then characterize decision rules in each case. At the same time, we utilize Chinese manufacturing data to examine the model-implied empirical predictions.

1.3.1 Default, Constrained and Unconstrained Zones

Considering the continuation problem for the intermediate firms in the state space $\mathcal{S}$, not all of the firms are able to rollover the debt and satisfy non-negative dividend payment constraints.

Proposition 1 *In the steady state, consider a firm that does not receive the death shock, has idiosyncratic productivity z, capital k, and has net worth n. There exists a threshold $\underline{n}(k,z)$, such that the firm defaults when $n < \underline{n}(k,z)$. $\underline{n}(z,k)$ is weakly increasing in k. And there exists $\underline{k}(z)$ such that $\underline{n}(z,k) = \underline{n}(z,0)$ all for $k \leq \underline{k}(z)$. $\underline{k}(z)$ is increasing in z. $V(n,k,z) = V^{exit}$ for $n < \underline{n}(k,z)$.*

Proof: See Appendix A.

To understand this proposition, it is convenient to see (1.12). If there is no irreversibility where $\eta = 1$, the default threshold collapses to be just a function of productivity z[15]. With irreversibility where $\eta < 1$, the default threshold if and only if be a function of both productivity z and fixed capital k when capital downward adjustment solves the maximization problem in (1.12). $k(z)$ specifies the threshold when downward adjustment maximizes the payment of the dividends. The key message is that the default threshold will be independent of k when upward adjustment is the optimal choice. But that will be dependent on k when downward adjustment is the optimal choice.

Proposition 2 *Let ϕ be the Lagrangian multiplier on the non-negative dividend constraint 1.11. There exists a threshold $\overline{n}(k,z)$, such that the non-negative dividend constraint is not binding, which implies that the borrowing constraint does not bind onward when $n > \overline{n}(k,z)$. $\overline{n}(k,z)$ is weakly increasing in k.*

$$V^{cont}(n,k,z) = n - \overline{n}(k,z) + V^{cont}(\overline{n}(k,z),k,z)$$

[15]This is the case in Ottonello and Winberry (2020)

Proof: See Appendix A.

Firms who are able to roll over the debt and continue may face financing frictions in making investments. Firms can borrow at a risk-free rate when the current borrowing is associated with no default risk in the next period. However, that does not imply the firm will never face any default risks in the future. Therefore, If a firm places non-zero probability weight on encountering a state in which default risks are non-zero, we identify it as constrained; otherwise, it is unconstrained. To be clear, a constrained firm need not face a debt price with a risk premium in the current period; the definition includes firms perceiving the risks of facing a default probability in the future.

Since intermediate firms are risk-neutral, it is always optimal for firms to retain profit and delay dividend payments, except if one has saved out of financing constraints. Let ϕ_t be the Lagrange multipliers on non-negative dividend payment constraint in a sequential problem. It can be shown that ϕ_t is supermartingale.[16] $\phi = 0$ corresponds to the case where firms save out of financing constraints, and $1 + \phi$ summarizes the marginal value of liquid wealth.

Proposition 2 illustrates the subset of state space where firms save out of financing constraints. For firms that save out of financing constraints, they are indifferent between retaining the profit within the firm or paying back to the firm owners as long as they maintain a minimum net worth position $\overline{n}(k, z)$. As a result, the decision rules of unconstrained firms will be collapsed into a two-dimensional object.

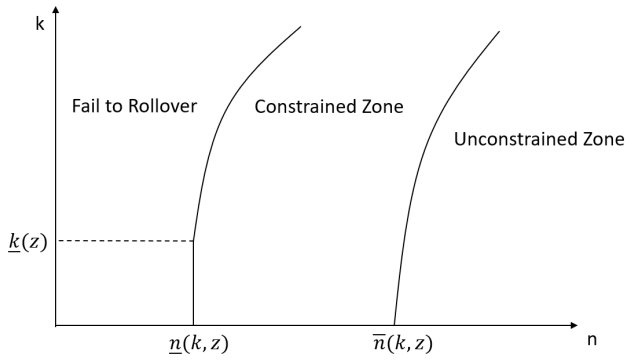
Proposition 3 *When* $\overline{n}(k, z) > n > \underline{n}(z, k)$, $\phi > 0$ *and constrained firms pay zero dividends. The marginal value of liquid wealth is* $1 + \phi$.

Proof: See Appendix A.

[16] See Appendix A for the proof.

For firms with a net worth between the two thresholds, financial constraints bind, and they would choose not to invest to the optimal level. The tightness of financial frictions can be proxied by the marginal value of liquid wealth. Figure 1.2 illustrates a plane in the state space $\mathcal{N} \times \mathcal{K}$ can be divided into three regions, following the previous propositions.

Figure 1.2: Default and Unconstrained Threshold



Notes: This figure illustrates the state space $\mathcal{N} \times \mathcal{K}$ conditional on z and the thresholds for default, constrained, and unconstrained zones.

1.3.2 Decision Rules

Unconstrained Firms For unconstrained firms in the subset of state space $\mathcal{S}$, the problem reduces to a classical investment problem with non-convex adjustment costs. The continuing firm's problem reduces to:

$$V^{cont}(n,k,z) = n - \overline{n}(k,z) + V^{cont}(\overline{n}(k,z),k,z)$$

Dividend payment is undetermined since the marginal value of liquid wealth is identical between the unconstrained intermediate firm and the representative entrepreneur family. As a result, $d \in [0, n - \overline{n}(k,z)]$. To consider the optimal investment policy, let us define the marginal $Q^m(k';z)$ as the present value by having one additional unit of fixed capital [17].

[17]Notice that we define marginal Q slightly differently from the classical way in two aspects: (i) Our marginal Q does not include current period benefit because net worth is defined after receiving the revenue

$$
\begin{aligned}
Q^m(k';z) &\equiv \mathbb{E}[\Lambda' \frac{dV(n',k',z')}{dk'}|z] \\
&= \mathbb{E}[\Lambda'(\underbrace{\frac{\partial V(n',k',z')}{\partial k'}}_{<0:\text{irreversibility punishment}} + \underbrace{\frac{\partial V(n',k',z')}{\partial n'} \times \text{MRPK}'}_{>0:\text{liquid wealth}} + \underbrace{\frac{\partial V(n',k',z')}{\partial n'} \times (1-\delta)q)}_{>0:\text{illiquid wealth}} |z] \\
&= \mathbb{E}[\Lambda'((\frac{\partial V(n',k',z')}{\partial k'} + \text{MRPK}') + (1-\delta)q)|z]
\end{aligned}
$$

With one additional unit of capital, firms will generate more liquid wealth by the marginal revenue product of capital at date t+1 by $\text{MRPK}' = \frac{\partial}{\partial k'}[(A'z'\varepsilon' k'^{\alpha} l'^{1-\alpha})^{1-\frac{1}{\theta}} Y'^{\frac{1}{\theta}}]$. Firms will also retain fixed capital by $(1-\delta)q$. $\frac{\partial V(n',k',z')}{\partial k'}$ is the marginal effect of extra capital, controlling the book value of net worth at date $t+1$. Because having a larger capital for a given net worth makes it more difficult for the firm to repay debt and adjust production, this compensated term that summarizes the cost of irreversibility is negative. $\frac{\partial V(n',k',z')}{\partial k'} = 0$ when the fixed capital is fully reversible. The last equality comes from the property that the marginal value of liquid wealth is equal to one when the firms are financially unconstrained.

Following the first-order conditions and the envelop conditions. We have:

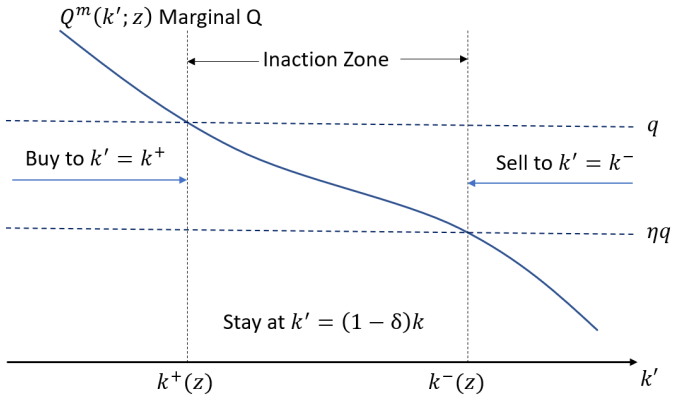
$$
Q^m(k^+;z) = \mathbb{E}[\Lambda' \frac{dV'}{dk'}|k' = k^+] = q, \quad Q^m(k^-;z) = \mathbb{E}[\Lambda' \frac{dV'}{dk'}|k' = k^-] = \eta q
$$

Figure 1.3 displays the investment policy. Firms will

$$
\begin{cases}
\text{purchase capital to } k' = k^+(z), \text{ if } (1-\delta)k < k^+(z) \\
\text{sell capital to } k' = k^-(z), \text{ if } (1-\delta)k > k^-(z) \\
\text{maintain capital at } k' = (1-\delta)k, \text{ if } k^+(z) < (1-\delta)k < k^-(z)
\end{cases}
$$

payment. (ii) We do not divide the replacement cost of fixed capital because we want to display in diagrams how the price of capital affects investment decisions.

Figure 1.3: Decision Rules Among Unconstrained Firms



Constrained Firms Among constrained firms in the subset of state space $\mathcal{S}$, investment policy will depend on both (n, k, z) and $d(n, k, z) = 0$.

F.O.C. with respect to b' delivers the optimal borrowing:

$$\frac{\partial F(b', k', l', z)}{\partial b'} \cdot \frac{\partial V^{cont}(n, k, z)}{\partial n} = \mathbb{E}[\Lambda' \frac{\partial V(n', k', z')}{\partial n'} | z] \tag{1.26}$$

The left hand side (LHS) is the marginal benefit of raising funds at date t by issuing extra debt to pay at date t+1: The first term is the extra fund raised, and the second term is the marginal value of date-t liquidity. If the non-negativity constraint of dividend (1.11) is binding at date t so that Lagrangian multiplier ϕ is positive, the marginal value of date-t liquidity exceeds the consumption value:

$$\frac{\partial V^{cont}(n, k, z)}{\partial n} = 1 + \phi > 1$$

The right hand side (RHS) of (1.26) is the marginal cost of issuing extra debt, as date t+1 book value of the net worth decreases by unity by extra debt issue. We can rewrite the first

order condition as:

$$\underbrace{\frac{\partial F(b',k',l',z)}{\partial b'} - \mathbb{E}[\Lambda']}_{<0\text{: price wedge in risky bond over safe bond}} = \mathbb{E}\Bigg[\Lambda' \underbrace{\left(\frac{\frac{\partial V(n',k',z')}{\partial n'}}{\frac{\partial V^{cont}(n,k,z)}{\partial n}} - 1\right)}_{<0\text{: tightness of the financing constraint at date t}} |z\Bigg]$$

The LHS is the difference between the marginal prices of risky bonds and the safe bond, which is negative if there is a positive probability of default. In RHS, the term inside the parenthesis $\frac{\frac{\partial V(n',k',z')}{\partial n'}}{\frac{\partial V^{cont}(n,k,z)}{\partial n}} - 1$ is the gap between the marginal rate of substitution between date-t+1 book value of net worth and date-t liquidity and unity—a relative measurement of the tightness of the financing constraint at date t. This term tends to be negative when the firm faces a tight liquidity constraint at date t so that the marginal value of date-t liquidity tends to exceed the marginal date t+1 book value of net worth. Thus, this equation says the firm is willing to issue risky debt if it faces a significant liquidity constraint at date t.

Since working capital is also subject to financial frictions, the labor wedge can be obtained by taking F.O.C. with respect to date-t+1 labor l', which is chosen and financed at date t:

$$\left(w - \frac{\partial F(b',k',l',z)}{\partial l'}\right) \cdot \frac{\partial V^{cont}(n,k,z)}{\partial n} = \mathbb{E}\left[\Lambda' \text{MRPL}' \frac{\partial V(n',k',z')}{\partial n'} |z\right]$$

The LHS is the marginal cost of hiring extra workers. The first term is the real wage rate minus the extra fund raised by increasing employment (with a constant face value of debt), and the second term is the marginal value of date-t liquidity. The RHS is the marginal benefit of having one extra worker: the product of the entrepreneur family's stochastic discounting factor Λ', the marginal revenue product of labor $\text{MRPL}' = \frac{\partial}{\partial l'}[(A'z'\varepsilon' k'^{\alpha} l'^{1-\alpha})^{1-\frac{1}{\theta}} Y'^{\frac{1}{\theta}}]$, and the marginal value of liquidity at date t+1. We can rewrite the first order condition as:

$$\underbrace{w - \frac{\partial F(b', k', l', z)}{\partial l'}}_{\text{marginal cost of hiring worker}} = \mathbb{E}[\Lambda' \text{MRPL}'] + \underbrace{\mathbb{E}[\Lambda' \overbrace{(\frac{\frac{\partial V}{\partial n'}}{\frac{\partial V^{cont}}{\partial n}} - 1)}^{<0\text{: tightness of financing constraint}} \text{MRPL}'|z]}_{\text{Labor Wegde}} \quad (1.27)$$

The first term of RHS is the standard expected discounted marginal revenue product of labor. The second term tends to be negative if the firm issues risky debt with non-zero default probability, following equation 1.26. When the firm faces a large idiosyncratic uncertainty, the extra funds raised by increasing employment with a constant face value of debt are limited. Therefore, the firm tends to cut employment when it faces a tight financing constraint, partly because it cannot easily adjust fixed capital downward, as we see below.

Finally, let us solve the decision rules for fixed capital investment. Similar to the unconstrained case, let us define the marginal Q as the marginal benefit of investment:

$$\begin{aligned} Q^m(k'; n, k, z) &\equiv \mathbb{E}[\Lambda' \frac{dV(n', k', z')}{dk'}|z] + \frac{\partial F}{\partial k'} \\ &= \mathbb{E}\left[\Lambda'\left(\underbrace{\frac{\partial V(n', k', z')}{\partial k'}}_{<0\text{:irreversibility punishment}} + \underbrace{\frac{\partial V}{\partial n'} \times \text{MRPK}'}_{>0\text{:liquid wealth}} + \underbrace{\frac{\partial V}{\partial n'} \times (1-\delta)q}_{>0\text{:illiquid wealth}}\right)|z\right] + \frac{\partial F}{\partial k'} \end{aligned} \quad (1.28)$$

Compared with the case without financial frictions, having one additional unit of fixed capital increases the current period borrowing by $\frac{\partial F}{\partial k'}$. This term would be zero if the current period borrowing does not involve any default risks. Following the first-order conditions w.r.t k' and the envelop conditions. The optimal choice of fixed capital is:

- buy capital to $k' = k^+$, if $(1-\delta)k < k^+$

$$Q^m(k^+; n, k, z) = \frac{\partial V^{cont}(n, k, z)}{\partial n} = q(1+\phi)$$

- sell capital to $k' = k^-$, if $(1-\delta)k > k^-$

$$Q^m(k^-; n, k, z) = \frac{\partial V^{cont}(n,k,z)}{\partial n} = \eta q(1+\phi)$$

- maintain capital at $k' = (1-\delta)k$ if

$$\eta q(1+\phi) < [Q^m(k^-; n, k, z) < q(1+\phi)$$

Figure 1.4 displays the investment policy in a diagram. Similarly, we can define the fixed capital investment wedge for the upward adjustment and downward adjustment separately.

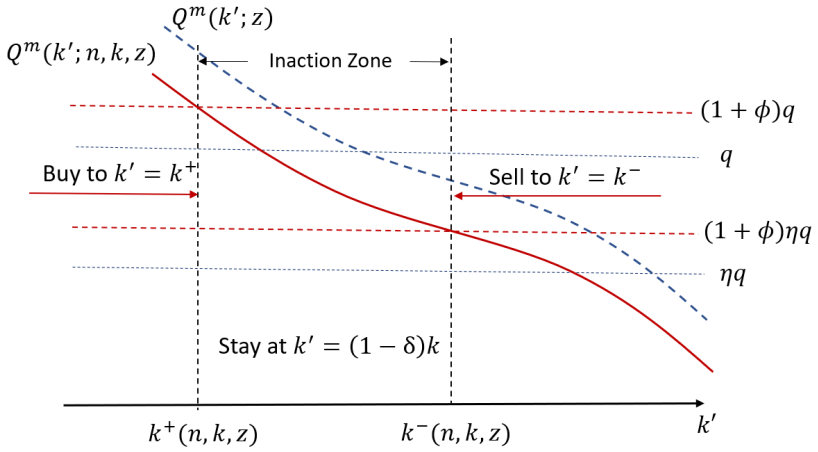
$$\underbrace{q(1-\mathbb{E}[\Lambda'(1-\delta)])}_{\text{use cost: upward adjustment}} = \mathbb{E}[\Lambda' \text{MRPK}'] + \underbrace{\mathbb{E}[\Lambda' (\frac{\overbrace{\frac{\partial V}{\partial n'}}^{<0\text{: fin. cons.}}}{\frac{\partial \tilde{V}}{\partial n}} - 1)(\text{MRPK}' + (1-\delta)q)] + \mathbb{E}[\Lambda' \frac{\frac{\partial V}{\partial k'}}{\frac{\partial V^{cont}}{\partial n}}] + \frac{\frac{\partial F}{\partial k'}}{\frac{\partial \tilde{V}}{\partial n}}}_{\text{Investment Wedge}} \tag{1.29}$$

$$\underbrace{\eta q(1-\mathbb{E}[\Lambda'(1-\delta)])}_{\text{use cost: downward adjustment}} = \mathbb{E}[\Lambda' \text{MRPK}'] + \underbrace{\mathbb{E}[\Lambda' (\frac{\overbrace{\frac{\partial V}{\partial n'}}^{<0\text{: fin. cons.}}}{\frac{\partial \tilde{V}}{\partial n}} - 1)(\text{MRPK}' + (1-\delta)q)] + \mathbb{E}[\Lambda' \frac{\frac{\partial V}{\partial k'}}{\frac{\partial V^{cont}}{\partial n}}] + \frac{\frac{\partial F}{\partial k'}}{\frac{\partial \tilde{V}}{\partial n}}}_{\text{Investment Wedge}} \tag{1.30}$$

The LHS is the standard definition of the user cost of fixed capital. The first term in the RHS is the present value of the marginal revenue product of capital. With financial frictions and irreversibility, an investment wedge rises. The first term in the investment wedge captures the default premium. The second term in the investment wedge is the punishment of having irreversibility, which pushes down the investment level.

Let us compare the decision rules with unconstrained firms or what happens when the

Figure 1.4: Decision Rules Among Constrained Firms



financing constraint becomes tighter. First, let us consider the movement of the marginal benefit of investment, which is governed by three forces: (i) Because working capital is subject to financing constraints, insufficient working capital lowers expected MRPK. The second term in (1.28) decreases, which shifts down $Q^m(n,k,z)$ (ii) Constrained firms are more likely to sell capital under fire sale. As a result, the punishment of irreversibility tends to be higher, and the first term $\frac{\partial V(n',k',z')}{\partial k'}$ in (1.28) decreases. This shifts down $Q^m(n,k,z)$ (iii) Since firms are financially constrained, $\frac{\partial V(n',k',z')}{\partial n'}$ is larger and this shifts up $Q^m(n,k,z)$. However, a higher expected marginal value of liquid wealth implies a higher current marginal value of liquid wealth. This upward shift will be canceled by the upward shift in $(1+\phi)q$ when the current borrowing constraint does not bind. When the borrowing constraints bind, the upward shift in $(1+\phi)q$ will dominate.

As a result, the purchasing threshold k^+ decreases because of the downward shift in Q^m and the upward shift in the shadow price of fixed capital. The two powers discourage firms from investing. Additionally, the selling threshold decreases for the same reason, which implies selling is encouraged compared to the case without financial frictions. For firms with tight financial constraints with a high marginal value of liquid wealth, the selling threshold

would be very low and create inefficient liquidation. This extra selling is inefficient since it is triggered by financial frictions. Finally, for the firms in the investment inaction zone, tighter financial frictions imply a lower MRPK, holding all else equal. For firms outside the inaction zone, tighter financial frictions imply a higher MRPK. We will test these unique predictions from our model in the next subsection.

Next, let us discuss the dynamic properties by considering a firm receives an adverse shock to its level of revenue productivity. First, for the firms in the purchasing zone where Equation (1.29) is satisfied, the adverse shock shifts Q^m down. The default risks shoot up since the expected net worth is lower, and default thresholds become higher. Additionally, the existence of financial friction makes the punishment of irreversibility $\frac{\partial V}{\partial k}$ even stronger because of the involved default risks. The three forces push the purchasing threshold to the left hand side.

Second, let us consider the firms in the inaction zone, where firms evaluate the RHS in Equation (1.29) at $k' = (1-\delta)k$ and find the present value of marginal investment is lower than the upward user cost of capital but higher than the downward user cost of capital. As a result, they would not sell fixed capital immediately. The adverse shock will lower the MRPK and tighten the financing constraint. The tightening financing constraint will also affect the choice of labor through the labor wedge. As a result, firms will find they have too much fixed capital and too little working capital input in the short run. This will additionally decrease MRPK because working capital can't be financed to the optimal level. Firms are expected to lose net worth under this unbalanced and unprofitable production structure. As a result, the default premium further rises. An endogenous feedback loop at the individual level is created. A higher default premium lowers the MRPK, and a lower MRPK pushes up the default premium. Notice that the feedback loop does not exist if we do not have irreversibility because firms can increase MRPK and decrease default risks by selling fixed capital at the same time. As the firm goes into this spiral, the present value of marginal

investment will decrease quickly, and it will finally equalize with the downward user cost of capital. At that point, fire sale and inefficient liquidation will be triggered. When firms are close to the default threshold with a high marginal value of liquid wealth. Firms have to sell fixed capital even though it is very costly. In extreme cases, firms have to default due to a low net worth value. Eventually, the firm will face the trade off between a lower default risk with a lower option value by selling fixed capital early and a higher default risk with a higher option value by holding fixed capital stock. This is the critical implication when we combine financial friction and irreversibility together.

Next, let us consider the role of the capital recovery rate in this model. It affects the financing constraint, irreversibility, and the interaction between the two frictions. Firstly, a lower capital recovery rate implies a lower collateral value, and the credit spread will be higher, conditional on the same default risks and pledgeability of fixed capital. Consequently, a lower capital recovery rate delivers a higher default threshold. Secondly, a Lower capital recovery rate implies a wider investment inaction zone. Thirdly, a lower capital recovery rate implies the firm tends to bear higher default risks and lower MRPK in the investment inaction zone. The above mechanisms imply an economy/sector with a lower capital recovery rate tends to have large distortions in the steady state and be more vulnerable to aggregate shocks.

1.3.3 Empirical Evidence

Data We utilize the Annual Survey of Industrial (ASI) Firms in China from 1998 to 2013 to examine empirical predictions from our model. [18] The data contains several advantages. First, it covers all the firms with revenue over 5 million RMB (700K USD) from 1998 to 2010 and all the firms with revenue over 20 million RMB (3M USD) from 2011-2013, leaving a

[18] 2010 data is not available, and not all variables are available in all years. Please see Appendix A for a detailed description

large sample size from 100 thousand to 300 thousand a year. Second, more than 99 percent of firms are not publicly listed firms. As a result, financial frictions are a main obstacle among small and medium-sized firms. Meanwhile, mergers and acquisitions are not the main source of exiting choices for those firms which we do not consider in our model. Therefore, this sample is a good fit for our model. Third, this data set contains variables in both production and balance sheets. This allows us to study investment dynamics and financial frictions together. See more introduction about the data set in Appendix A.

Firm Level Evidence The first object we examine is the decisions on default. Our model offers three properties of the default threshold. (i) Higher productivity implies lower default risks as stochastic monotonicity holds, (ii) higher net worth implies a lower default risk, and (iii) higher fixed capital is associated with a higher default risk conditional on net worth if and only if downward adjustment solves the maximum dividend payment problem. Our firm-level data set does not carry any information on default, so we use exit decisions as a proxy to measure default risks using panel data where E_{it} is a dummy variable that equals one if the firm does not appear in our sample onward. Fixed capital k_{it} includes assets in equipment, buildings, structures, and other productive capital measured in terms of the book value, deflated by the national price index of fixed-asset investment. The investment i_t is the fixed capital investment defined as the capital stock in the end of this year minus the capital stock in the end of last year plus depreciation in this year, $i_t = k_{i,t+1} - k_{it} + \delta_{it}$. Net worth n_{it} is total assets minus total liabilities. TFPR z_{it} is the residuals from a production function estimation.

We estimate variants of the baseline empirical specification with a linear probability model.

$$E_{it} = \alpha_i + \alpha_{st} + \beta_1 \log n_{it} + \beta_2 \log z_{it} + \beta_3 \log k_{it} + \beta_4 \mathbb{I}\{i_t \leq 0\} \times \log k_{it} + \Gamma' Z_{i,t-1} + e_{it} \quad (1.31)$$

where α_i is a firm i fixed effect, α_{st} is a year-two digit sector fixed effect, $\mathbb{I}\{i_t \leq 0\}$ is an indicator specifies whether one makes a positive adjustment in fixed capital or not. Table 1.1 reports the results. Property (i) and (ii) are verified with data since higher productivity or higher net worth is associated with a higher lower exit probability, holding all else equal. Columns 1 and 3 also suggest that higher fixed capital is associated with a lower default probability conditional on productivity and net worth. [19]. However, this effect is heterogeneous depending on the investment actions by looking at the coefficients on interaction term. By looking at β_4 in columns 2 and 4, higher fixed capital implies a higher probability of exit in the next period if and only if firms make a non-positive investment. This is exactly consistent with Proposition 1, which shows higher fixed capital is associated with a higher default threshold conditional on net worth if and only if downward adjustment solves the maximum dividend payment problem in the presence of irreversibility.

The second object we examine is how the interaction between irreversibility and financial friction affects capital misallocation. Normally, tighter financial constraint is associated with a higher MRPK, and MRPK is particularly high for firms with higher productivity, but not enough net worth can be pledged as collateral. As a result, financial frictions dampen the speed of accumulating wealth in the transition in response to a persistent positive shock. However, asset demand goes down in response to negative shocks, and financial constraint is likely to be loosened and would not have much effect on investment policies. Interstingly, adding irreversibility makes things more sophisticated. According to Figure 1.4, for firms outside the investment inaction zone, a tighter financial constraint shifts the shadow price of capital $(1+\phi)q$ or $(1+\phi)\eta q$ up. Resulting in a lower level of fixed capital and higher marginal benefit of investment associated with higher expected MRPK. In contrast, for

[19]This is not the prediction from our model. To rationalize this fact, one reason is the definition of net worth in our model is not fully consistent with what in the data. In the model, the definition of net worth is assets plus revenue minus liability, while in the data, net worth is assets minus liability. Therefore, a higher fixed capital implies a higher potential revenue and higher net worth in the model conditional on the same net worth in the data.

Table 1.1: Exit Rate and Firm Characteristics

	(1) Exit	(2) Exit	(3) Exit	(4) Exit
Log(Net Worth)× D(Net Worth>0)	-0.010***	-0.010***	-0.010***	-0.010***
	(0.000)	(0.000)	(0.000)	(0.000)
Log(-Net Worth)× D(Net Worth<0)	0.008***	0.008***	0.020***	0.020***
	(0.000)	(0.000)	(0.000)	(0.000)
Log(TFPR)	-0.011***	-0.011***	-0.016***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.001)
Log(Fixed Capital)	-0.014***	-0.011***	-0.008***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)
D(Investment≤0)× Log(Fixed Capital)		**0.016*** **		**0.018*** **
		(0.000)		(0.000)
Constant	0.309***	0.299***	0.230***	0.213***
	(0.001)	(0.001)	(0.004)	(0.004)
Year × Sector FE	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes
Observations	1932468	1932468	1779519	1779519
R-Square	0.062	0.064	0.370	0.373

Notes: Results from estimating (1.31). Fixed capital k_{it} includes assets in equipment, buildings, structures, and other productive capital measured in terms of the book value. The investment is fixed capital investment $i_t = k_{i,t+1} - k_{it} + \delta_{it}$. Net worth is total assets minus total liabilities. TPFR is residuals from a production function estimation. Data source: Chinese Annual Survey of Industrial Firms, 1998-2007. Standard errors are clustered at the firm level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

firms in the investment inaction zone, tighter financial constraints would not affect the choice of the level of fixed capital. However, a tighter financing constraint would shift the $Q^m(n, k, z)$ down as working capital can't be invested to the optimal level. The lower marginal benefit of investment is associated with a lower expected MPRK. As a result, tighter financing constraint is associated with higher MRPK outside the investment inaction one and associated with lower MRPK inside the investment inaction zone.

To test this unique model's implications, we estimate variants of the baseline empirical specification:

$$\log y_{it} = \alpha_i + \alpha_{st} + \beta_1 \log x_{it} + \beta_2 \mathbb{I}\{i_{it} = 0\} \times \log x_{it} + \Gamma' Z_{i,t-1} + e_{it} \tag{1.32}$$

where α_i is a firm i fixed effect, α_{st} is a year-two digit sector fixed effect, $\mathbb{I}\{i_t = 0\}$ is an indicator specifies whether one makes a positive adjustment in fixed capital or not. we use

Table 1.2: Marginal Productivity and Financial Frictions under Heterogeneous Investment Decisions

	(1) Log(ARPK)	(2) Log(ARPK)
Log(Debt/Asset)	0.054***	
	(0.002)	
D(Investment=0)× Log(Debt/Asset)	**-0.074****	
	(0.002)	
Log (Borrowing Rate)		0.033***
		(0.001)
D(Investment=0) × Log(Borrowing Rate)		**-0.017****
		(0.001)
Constant	0.160***	0.072***
	(0.001)	(0.004)
Year × Sector FE	Yes	Yes
Firm FE	Yes	Yes
Observations	2875653	2140368
R-Square	0.839	0.855

Notes: Results from estimating (1.32). α_i is a firm i fixed effect, α_{st} is a year-two digit sector fixed effect, $\mathbb{I}\{i_t = 0\}$ is an indicator specifies whether one makes a positive adjustment in fixed capital or not. ARPK y_{it} is defined as value-added over fixed capital to measure MRPK. The tightness of financing constraint is proxied by two variables $x_{it} \in \{\text{leverage ratio}, \text{borrowing rate}\}$. The leverage ratio is defined as total liabilities over total assets, and the borrowing rate is defined as total interest payment over total liabilities. Data source: Chinese Annual Survey of Industrial Firms, 1998-2007. Standard errors are clustered at the firm level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ARPK y_{it} defined as value-added over fixed capital to measure MRPK. The tightness of financing constraint is proxied by two variables $x_{it} \in \{\text{leverage ratio}, \text{borrowing rate}\}$. The leverage ratio is defined as total liabilities over total assets, and the borrowing rate is defined as total interest payment over total liabilities.

Table 1.2 reports the results. As predicted in the model, ARPK is positively correlated with the leverage ratio and borrowing rate for the firms that make active capital adjustments. In contrast, APRK is negatively correlated with the leverage ratio for firms that are in the investment inaction zone. Though the results on borrowing rate are not strong enough to reverse the positive correlation, it dampens the positive correlation, and the empirical results are generally in line with the properties of steady-state decision rules implied by the interaction between financial frictions and irreversibility.

Sector Level Evidence Up to now, we have shown the firm-level empirical evidence, which is in line with the model properties coming from the interaction between financial frictions and partial irreversibility in steady-state. These results support the validity of the model we built. Back to our original question, we want to understand the role of irreversibility and financial frictions in propagating aggregate shocks. Ideally, if there are various economies with different capital recovery rates in the lab, it is straightforward to examine the dynamics with respect to various economies in response to aggregate shocks. Although we do not have such a research lab in economics, we might be able to utilize the sectoral differences in capital recovery rates and examine the differences in sectoral dynamics in recessions. Fortunately, Kermani and Ma (2023) computes sector-specific fixed capital recovery rate in the US using bankruptcy documentation. Though I do not have the same measurement using Chinese data, I argue this capital recovery rate is a sector-specific feature, and the variance should be informative in another economy. To examine the relationship between capital recovery rate and dynamics in response to aggregate shocks, we estimate variants of the following empirical specification:

$$y_{it} = \alpha_i + Q(t, t^2) + \beta_1 D_t + \beta_2 D_t \times \eta_i + \beta_3 D_t \times s_i + e_{it} \tag{1.33}$$

where α_i is a two-digit sector-level fixed effect. $Q(t, t^2)$ is a quadratic form that controls the non-linear aggregate trends. $y_{it} \in \{\text{St.d.}(\log \text{ARPK}), \log \text{Output}\}$ is two sector-year level outcomes. We use standard deviation in ARPK within a sector to measure the misallocation and sector-level output to measure aggregate impacts. D_t is a dummy variable that equals one in the years of 2008 and 2009, the only recession in the span of our data coverage. η_i is the sector-specific capital recovery rate. s_i is the sector-level export share defined as export value over output before the recession in 2007. This measures the degree of demand shock in the recession.

Table 1.3: Sector Level Fixed Capital Recovery Rate and Outcomes in Recessions

	(1)	(2)	(3)	(4)
Panel A - Dep. Var. St.d. (log ARPK)				
D(Recession=1)	0.032***	0.076***	0.020***	0.063***
	(0.005)	(0.019)	(0.008)	(0.021)
D(Recession=1)×Recovery Rate		-0.128**		-0.117**
		(0.053)		(0.053)
D(Recession=1)×Export Share			0.050*	0.043
			(0.026)	(0.027)
Constant	1,748.759***	1,748.759***	1,748.759***	1,748.759***
	(341.223)	(339.110)	(340.109)	(338.419)
Observations	420	420	420	420
R-Square	0.759	0.763	0.761	0.764
Panel B - Dep. Var. (log Output)				
D(Recession=1)	-0.070***	-0.377***	-0.035***	-0.345***
	(0.029)	(0.108)	(0.045)	(0.117)
D(Recession=1)×Recovery Rate		0.884***		0.859**
		(0.299)		(0.302)
D(Recession=1)×Export Share			-0.156	-0.103
			(0.151)	(0.151)
Constant	-1,035.363	-1,035.363	-1,035.363	-1,035.363
	(1,943.420)	(1,924.397)	(1,943.276)	(1,925.731)
Observations	420	420	420	420
R-Square	0.979	0.979	0.979	0.979

Notes: The table reports results from estimating 1.33. where α_i is a two-digit sector-level fixed effect. $Q(t, t^2)$ is a quadratic form that controls the non-linear common trend. $y_{it} \in \{\text{St.d.}(\log \text{ARPK}), \log \text{Output}\}$ is two sector-year level outcomes. D_t is a dummy variable that equals one in the years of 2008 and 2009. η_i is the sector-specific capital recovery rate taken from Kermani and Ma (2023). s_i is the sector-level export share defined as export value over output in 2007. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3 reports the results. From Panel A, the standard deviation in log $ARPK$ rises by 3 percent on average in the recession relative to the trend, suggesting misallocation rises in the recession. However, this relationship is heterogeneous and heavily depends on the capital recovery rate. An 0.1 increase in capital recovery rate is associated with a 1.3 percent decline in dispersion in ARPK in the recession. In column 3, sectors with higher export share pre-recession are also associated with a higher dispersion. This result suggests that a higher demand shock is associated with a higher misallocation. Panel B conveys a similar set of information. Relative to the trend, the average output loss is around 7 percent in the two

recession years. This loss is heterogeneous across sectors. An 0.1 increase in capital recovery rate is associated with a 9 percent gain in output in the recession. Sectors with higher export share pre-recession are also associated with a higher output loss in the recession, though not statistically significant.

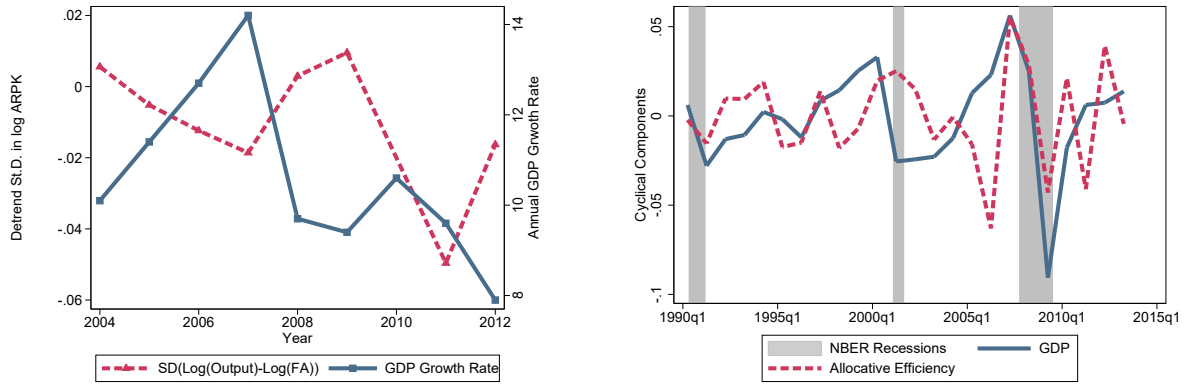
To summarize, the results in Table 1.3 suggest a significant role of irreversibility in propagating aggregate shocks, partially through the misallocation channel. The aggregate shock we consider here in China is likely to be a first-moment demand shock rather than financial shocks or uncertainty shocks. In the quantitative exercises, we will simulate an economy with two sectors to connect the empirical evidence we found here.

Aggregate Level Evidence In this paragraph, we present some empirical evidence about the cyclicality of misallocation. We find that within-sector dispersion in MRPK is counter-cyclical in both China and the U.S. in the manufacturing sector. To measure the cyclicality of misallocation in China, I compute the standard deviation of ARPK within 4-digit sector and take a weighted average by sector share in output. To measure the cyclicality of misallocation in the U.S., I utilize a data set constructed in Bils et al. (2021) where they provide a measurement error-adjusted annual sequence of allocative efficiency in the US manufacturing sector. The allocative efficiency is defined as the current output divided by the potential output once the marginal revenue of inputs is equalized.

In Figure 1.5 (a), I plot the detrend standard deviation of ARPK (removing linear trend) versus GDP growth rate in China from 2004 to 2012. These two variables are negatively correlated. In Figure 1.5 (b), I plot the HP-filtered allocative efficiency versus GDP in the U.S. manufacturing sector. In the past three recessions in the U.S., the allocative efficiency went down. The correlation between the cyclical component of allocative efficiency and GDP is 0.32. The allocative efficiency also seems a bit lagged. Though it is hard to identify what kind of shocks drive the business cycles, misallocation is suggested to have a role in

propagating the aggregate shocks.

Figure 1.5: Counter Cyclical Dispersion in MRPK



(a) China (b) U.S.

Notes: Panel A: Data Source: St.D. of ARPK is computed from the Annual Survey of Industrial firms 2004-2012 (no 2010) in China within the 4-digit sector at annual frequency.
Panel B: Log deviations from HP trend (smoothing parameter $\lambda = 6.25$) of (i) Allocation efficiency data is from (Bils et al. (2021)). (ii) US real GDP, manufacturing sector. Annual frequency. Corr(allocative efficiency, GDP)=0.32

1.4 Calibration

In this section, we calibrate the model we describe in section 2 to the Chinese economy. The calibration process is conducted in two steps. We first select a set of externally calibrated parameters from the literature or our estimation from the micro-level data. Secondly, we discipline our model by choosing a set of internally calibrated parameters to match a set of simulated moments with empirical moments in the data. Then, we examine some untargeted moments to verify the calibrated model is a validated lab that represents the distribution and behaviors of the economy in the steady-state.

1.4.1 Externally Calibrated Parameters

Table 1.4 presents the choices of externally calibrated parameters. The quarterly discounting factor is chosen as 0.99, implying an annual risk-free rate of 4 percent in the steady state. As we choose GHH (Greenwood et al. (1988)) preferences, the elasticity of labor supply is independent of the marginal utility of consumption. We choose the Frisch elasticity of labor supply to be 1. Though the number is greater than some of the early micro estimations surveyed by Blundell and MaCurdy (1999), it is consistent with compensated elasticities at the macro level Keane and Rogerson (2015). We choose the measure of the representative entrepreneur family to be 0.5, implying half of the population in China is hand-to-mouth. Though it is an arbitrary assumption, it fits the fact that a large share of households are still relatively poor with little financial assets and net worth in China.

Capital share and elasticity of substitution are chosen from the production estimation in (1.3); please see Appendix A for the results. We choose the quarterly depreciation rate to be 0.025, which is standard in the literature. Notice that due to the irreversibility, the macro aggregate depreciation rate $\tilde{\delta}$ is larger than the micro depreciation rate δ, which implies an annual I/K ratio well above 0.1 in the steady state. The aggregate adjustment cost of capital goods is governed by Ψ. We choose it to be 2, implying a one percent change in I/K is associated with a 0.5 percent change in the price of capital. The idiosyncratic process of revenue productivity is estimated from the residual dynamics of production estimation using the Annual Survey of Industrial Firms in China. Compared to the literature. the micro-level uncertainty is higher in China than in advanced economies like the U.S.

1.4.2 Internally Calibrated Parameters

In the second step, we choose a set of parameters to match a set of empirical moments jointly. Though some of the parameters are more informative to a particular moment than others, the

Table 1.4: Summary of Externally Calibrated Parameters

Parameters	Description/Sources	Value
Households		
β	Discounting factor	0.99
ν	Inverse of elasticity of labor supply	1
M	Measure of entrepreneur family	0.5
Firms		
α	Capital share	0.30
δ	Depreciation rate	0.025
θ	Elasticity of substitution	5.0
Φ	Elasticity of q to I/K	2
Idiosyncratic Processes		
ρ	From micro estimates	0.75
σ_ν	From micro estimates	0.08
σ_ε	From micro estimates	0.18
Policies		
λ^G	Elasticity of fiscal policy	1
λ^{τ^D}	Elasticity of credit policy	4

model is nonlinear, and one-on-one mapping is not possible. Table 1.5 reports the values of internally calibrated parameters, and table 1.6 reports the comparison between the targeted model simulated moments and the empirical moments. First, the fixed operation cost shifts the default threshold directly following (1.12). A higher fixed operation cost implies a higher default threshold and higher default density in the steady state. Our empirical default rate is the fraction of non-performing loans in China reported by the China Banking Regulatory Commission. We choose the quarterly exogenous exit rate to be 0.02, combined with the annual default rate of 0.04, contributing to a total annual exit rate of 0.12, which is observed from our calculation from the Annual Survey of Industrial Firms in China. Next, we assume exiting firms are replaced by newborn firms with the same initial equity n_0. Newborn firms are smaller on average and take time to grow. As a result, the default cost is persistent and is governed by the size of newborn firms. We choose initial equity n_0 to match the ratio of average net worth for young firms with ages less than 3 years to the average net worth of all the firms. The higher n_0 is, the higher this ratio would be in the simulated data. Finally, we identify the capital recovery rate to match the ratio of firms that make zero fixed capital investments. In the firm-level investment data, we allow for small capital adjustment

when we define investment inaction zone. We classify firms with absolute annual empirical investment rates less than 0.01 as inactive firms, and we find the share of inactive firms is 0.21. In the model, a lower capital recovery rate implies a wider investment inaction zone and a higher density of firms that make zero investments in fixed capital. The capital recovery rate also determines the collateral value. It is informative to default thresholds, default density in the steady state, and credit spread in the model. Therefore, the default rate is pinned down jointly by the fixed operation cost and capital recovery rate. The calibrated capital recovery rate is 0.43, close to what is suggested in Kermani and Ma (2023). Finally, $\overline{G}$ is chosen to match the government expenditure ratio of 0.25.

Table 1.5: Summary of Internally Calibrated Parameters

Parameters	Description	Value
ξ	Operating cost	0.15
π_d	Exogenous exit rate	0.02
n_0	Initial equity	0.18
η	Recovery rate	0.43
$\overline{G}$	Government expenditure	0.42

Table 1.6: Calibration Target and Model Fit

Moments	Description	Data	Model
$\mathbb{E}$[default rate]	Annual default rate	0.04	0.04
$\mathbb{E}$[Exit rate]	Annual exit rate	0.12	0.12
$\mathbb{E}[\frac{n(age \leq 3\text{years})}{\overline{n}}]$	Ratio of net worth at age 1-3 to mean	0.49	0.48
$\mathbb{E}$[Investment=0]	Ratio of annual inactive firms	0.21	0.21
$\overline{G}/Y = 0.25$	Ratio of government expenditure	0.25	0.25

1.5 Quantitative Results

In this section, with the calibrated model, we examine the macroeconomic implications of the interaction between financial frictions and the partial irreversibility of fixed capital by

simulating the transition dynamics in response to aggregate shocks under variants of the model.

1.5.1 Aggregate Responses to Productivity Shock

We first study the transition dynamics in response to an ex-ante unexpected but ex-post fully anticipated aggregate productivity shock. To assess the amplification effect from the interaction between financial frictions and irreversibility, I simulate the same shock in three economies. The baseline case features both financial friction and irreversibility. The economy only features irreversibility where financial frictions are shut down by assuming external debt payment is fully committed and equity issuance is not associated with any additional cost. And the economy only features financial frictions where I set $\eta = 1$. The TFP in the steady state is $\overline{A} = 1$, the shock I consider is:

$$\log A_{t+1} = \rho_A \log A_t + (1 - \rho_A) \log \overline{A} + \varepsilon_t^A$$

where the economy starts from the steady state and TFP shock hits at t_0 unexpectedly while $\varepsilon_{t_0}^A = -0.02$ and $\varepsilon_{t \neq t_0}^A = 0$. The persistence of the shock is $\rho_A = 0.7$. The definition of the equilibrium in the transition path is defined in Appendix A.

Proposition 4 *Along the transition dynamics with $\varepsilon_{t_0}^A < 0$, the default threshold is higher than in the steady state*

$$\{\underline{n}_t(k,z)\}_{t=t_0}^{\infty} > \underline{n}(k,z)$$

and the threshold specifies the default threshold dependency on k is lower

$$\{\underline{k}_t(z)\}_{t=t_0}^{\infty} < \underline{k}(z)$$

Proof: See Appendix A.

Figure 1.6 displays the results. The baseline simulation shows a sizeable amplification effect is generated from the interaction between financial frictions and irreversibility compared to a model with only financial frictions or real frictions. A 2 percent TFP shock leads to a maximum of 3.9 percent and a more persistent decline in output. Since inputs are predetermined, the drop in output when the aggregate shock hits comes from the decline in TFP and extra default. Surviving firms choose investment decisions.

As illustrated in Figure 1.7, negative TFP shock lowers the marginal benefit of investment and pushes a more significant density of firms into the investment inaction zone unexpectedly from the left-hand side. As the shock is temporary, selling the capital immediately is very costly. The financing constraint on working capital becomes tighter for three reasons. First, a higher fixed capital level increases the demand for working capital. Second, default risks rise as the expected cash flow is lower. Third, a lower capital price decreases the collateral value, which drives up the credit spread conditional on default risks. As a result, working capital can't be financed to an optimal level. Consequently, misallocation, measured by the standard deviation in ARPK, increases by 2 percent. Then, if the financing constraint becomes even tighter with a high marginal value of liquid wealth, a fire sale is triggered even though the capital price is lower. In extreme cases, firms have to default. As shown in Proposition 4, the default zone expands along the transition path, and a larger density of firms into the default zone. Since inputs are pre-determined, the output should drop by 2 percent if there is no extra default in the initial period that the shock hits. Extra output loss in that initial period is only associated with extra default. Indeed, default density increases by over 20 percent in the baseline simulation but decays quickly as there are no additional unexpected adverse shocks. Inefficient liquidation comes from default and fire sale, which pushes down the capital price, and the lower capital price leads to more inefficient liquidation. Since exiting firms will be replaced by small firms which take time to grow, the effect of default on output and labor demand will accumulate and be persistent. Liquidating firms also take time

to restock fixed capital. The process is lengthy because irreversibility makes them cautious, and financial frictions make investments costly when the net worth is low.

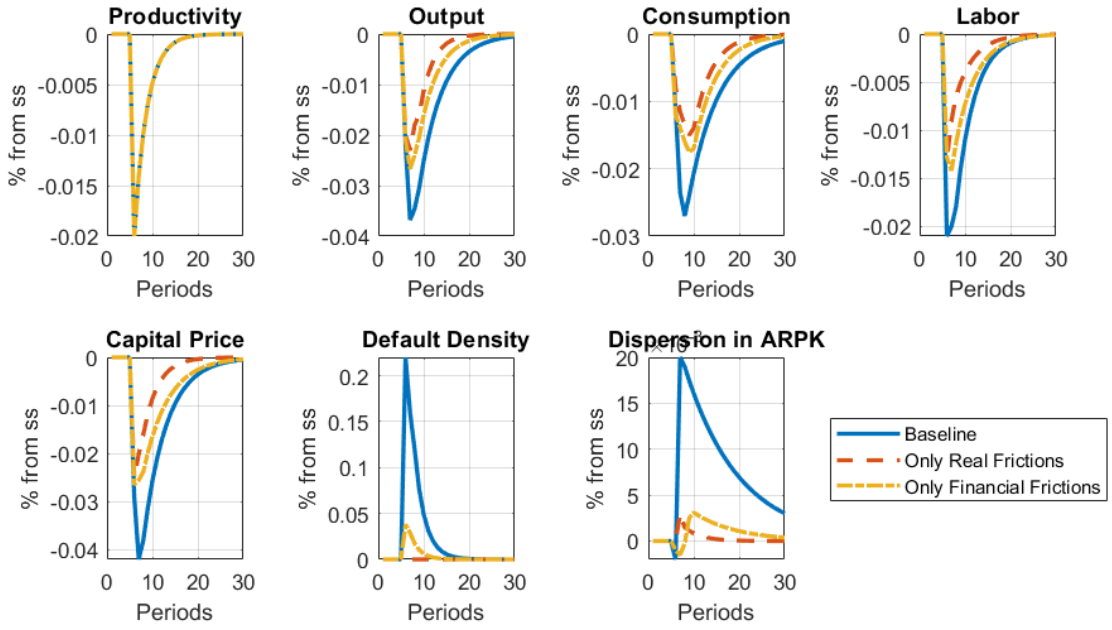
Labor demand drops for two reasons: the negative TFP shock and the tightening financing constraint on working capital. As a result, the wage rate goes down, and the labor supply shrinks proportionally since there is no wealth effect with GHH preference. Hand-to-month workers can't smooth their consumption, and therefore, the demand for consumption drops by more. Aggregate output is dragged down by extra default and misallocation in addition to TFP shocks. It is also affected by the demand side since hand-to-month agents consume less. This lower aggregate demand will enter the aggregate revenue productivity according to equation (1.3). As a result, it works as an additional negative quasi-TFP shock and constitutes an endogenous feedback loop. Notice that amplification will not exist if Y_t moves with A_t at the same pace. But a more profound drop in Y_t relative to A_t will be amplified endogenously.

Then let us examine the economy with no financial frictions. The default channel is shut down entirely without financial friction. There would also be no misallocation. The existence of real frictions will lead to a slightly larger dispersion in ARPk because a larger share of firms are pushed into the investment inaction zone. However, the decline in capital price and interest rate pushes down the purchasing threshold and stimulates investment in the general equilibrium. As a result, the investment inaction zone would not move by much. Furthermore, firms in the inaction zone can always finance enough working capital, and therefore, dispersion in ARPK would not increase by much. There would be no extra power to shift down the labor demand relative to a standard RBC model. As a result, output only decreases by precisely 2 percent when the shock hits, and there is no significant amplification effect along the transition path.

In a model with only financial frictions, an inaction zone does not exist. Firms would shrink immediately in response to negative TFP shock. The default threshold, only de-

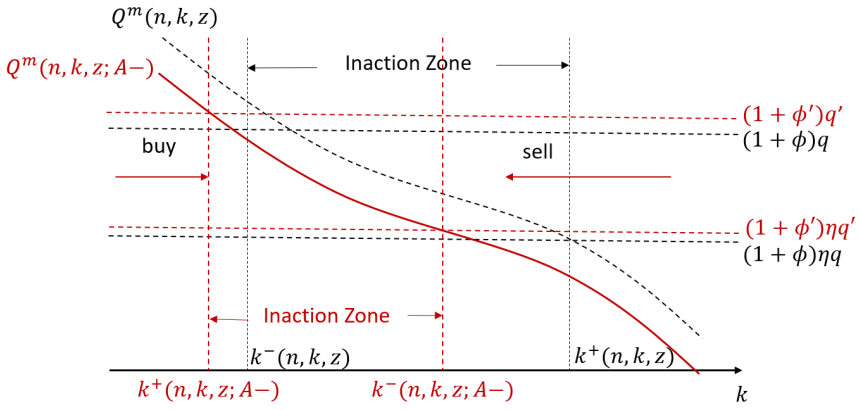
pending on net worth now, will increase, but default density only rises modestly without irreversibility. The dispersion in ARPK even goes down slightly when the aggregate shock just hits because the decline in fixed capital demand outweighs the decline in collateral value. This is because the demand for inputs is a static choice without adjustment cost, while the capital price is forward looking. Along the transition path, firms are losing net worth, and the demand for fixed capital rebounds. Consequently, the decline in collateral value gradually overweighs the decline in capital demand. Misallocation picks up in the middle and makes the recession more persistent than the economy with only real frictions.

Figure 1.6: IRFs to TFP Shock



Notes: Aggregate impulse responses to 2 percent negative TFP shock which decays at a rate of 0.7. Computed as the perfect foresight transition in response to a series of unexpected shocks starting from a steady state at t=6.

Figure 1.7: Decision Rules under Productivity Shock

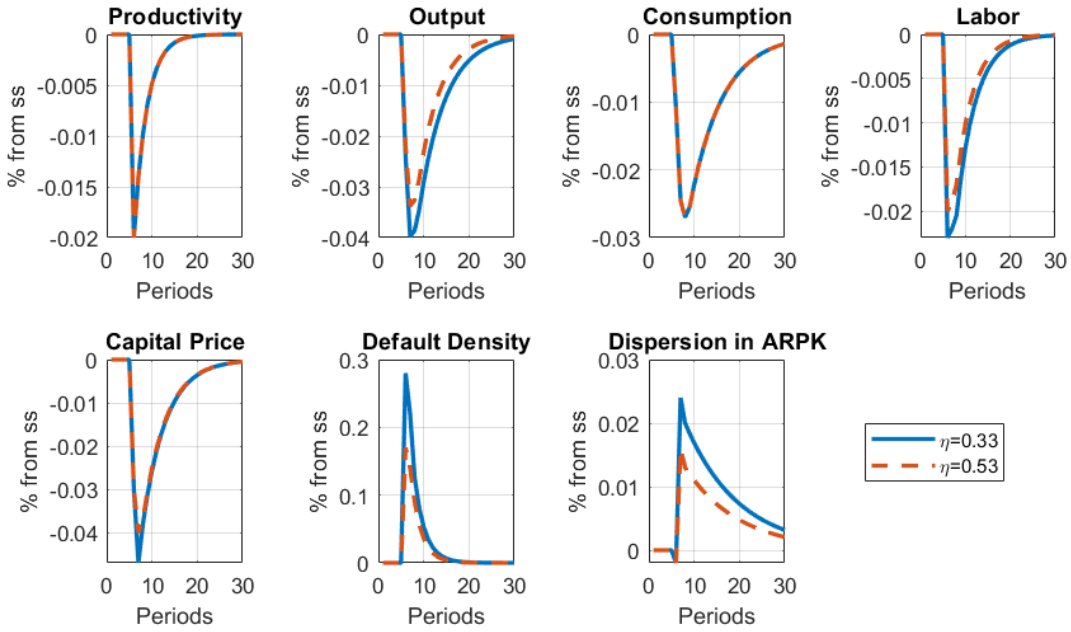


Multisector Model Motivated by the regression results in Table 1.3, we modify the model with two sectors. We simply assume that intermediate firms are split into two sectors with different capital recovery rates, 0.33 and 0.53, respectively. All else is untouched compared with the baseline model. The simulation results show the sector with a lower capital recovery rate experiences a deeper and more persistent decline in output with a larger amplification effect in response to the same TFP shock. The sector with a lower capital recovery rate also experiences a smaller spike in default probability and a lower increase in dispersion in ARPK. As firms in the sector with lower capital recovery rates face tighter financing constraints and liquidate more aggressively, there is also an associated deeper drop in labor demand. Our model simulation results are generally in line with the regression results.

1.5.2 Counter-Cyclical Fiscal Policy

To mitigate the recession, we consider a countercyclical fiscal policy. In contrast to New Keynesian models where firms face sticky prices, firms face a combination of "sticky quantity" and financial frictions in our model. As the adverse shock shifts the demand curve inward, sticky quantity limits the firms' ability to adjust the capital stock and exacerbates financial frictions. A higher aggregate demand can push the demand curve outward, release firms

Figure 1.8: IRFs to TFP Shock: Two Sector Model



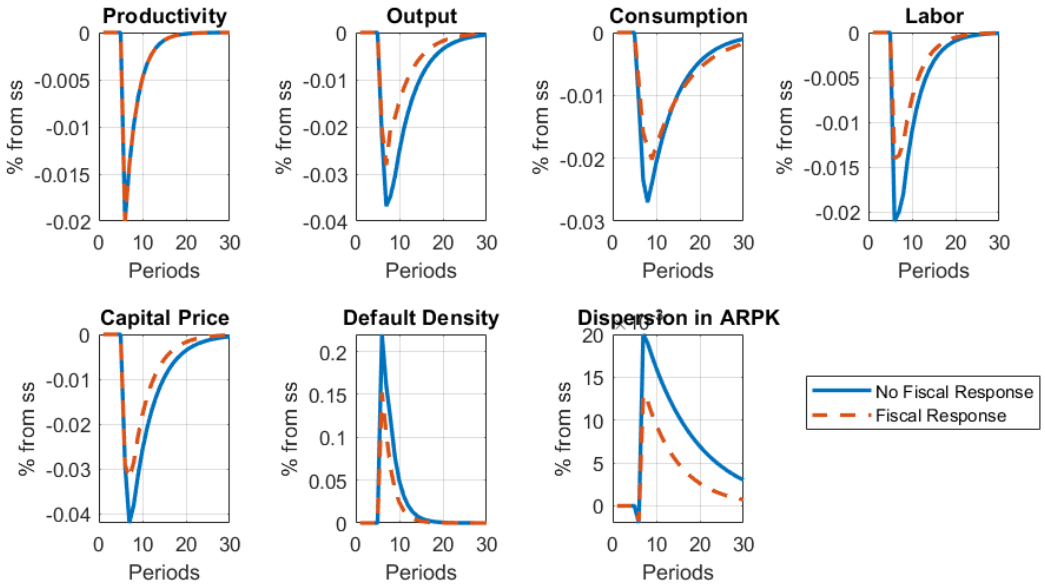
Notes: Aggregate impulse responses to 2 percent negative TFP shock, which decays at a rate 0.7. Computed as the perfect foresight transition in response to a series of unexpected shocks starting from a steady state at t=6.

from the investment inaction zone, and restore the efficient positions of fixed capital holding. Financing constraints on working capital will be relaxed endogenously due to higher expected cash flow. To boost the aggregate demand during the recession, we consider the fiscal policy rule specified by the following equation :

$$\log G_t = \log \overline{G} - \lambda^G (\log Y_{t-1} - \log \overline{Y})$$

where a 1 percent downward deviation in GDP relative to steady state at $t-1$ leads to a (λ^G) percent upward deviation in government expenditure relative to steady state at $t+1$. Figure 1.9 displays the comparison of IRFs between the baseline economy where government expenditure is constant and the economy where countercyclical fiscal policy is active with $\lambda^G = 1$. By boosting aggregate demand, the economy experiences a lower decline in output, capital price, and a smaller rise in default density, as well as dispersion in ARPK. As the

Figure 1.9: IRFs to TFP Shock: Fiscal Policy



Notes: Aggregate impulse responses to 2 percent negative TFP shock, which decays at a rate 0.7. Computed as the perfect foresight transition in response to a series of unexpected shocks starting from a steady state at t=6. The solid lines depict the IRFs in the baseline model with no active fiscal policy, and the dashed lines depict the IRFs in the model with counter-cyclical fiscal policy as described in the text. $\lambda^G = 1$

financing constraints on working capital become looser, labor demand and the wage rate are boosted together. Higher wage bills translate into higher consumption demand with a large proportion of hand-to-mouth agents, which delivers another upward push to aggregate demand. Notice that even though the government does not increase expenditure in the period when the negative shock hits because the fiscal policy is lagged for one period. It effectively lowers the default density at $t = t_0$ because the bank and firms are forward-looking if the policy is fully committed.

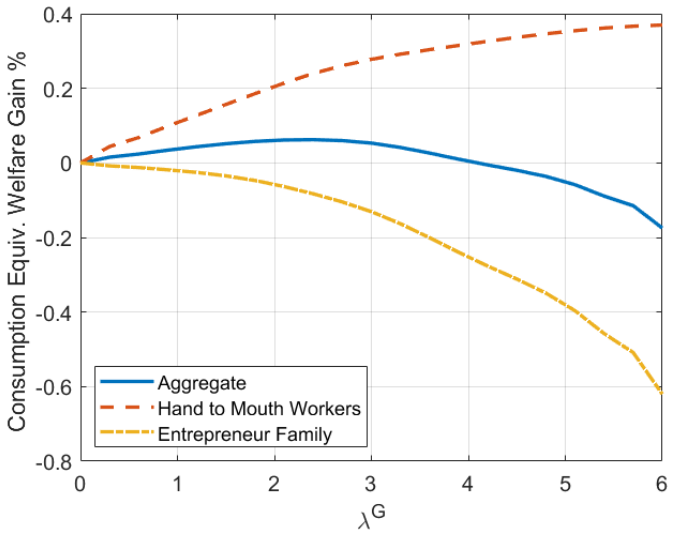
Optimal Fiscal Policy As shown in Figure 1.9, counter-cyclical fiscal policy is effective in mitigating the recession, but it is associated with higher government expenditure, which is financed by lump sum tax on the entrepreneur family. As we assume that government expenditure does not affect private agents' utility level, the impact of the credit policy on

welfare is ambiguous. Although our model does not contain recurrent aggregate shocks, we can study the optimal fiscal policy in the transition path with full policy commitment in response to a particular type of exogenous shock. We study the optimal fiscal policy within the class of the policy rules we specify with various λ^G. We assume the economy goes back to the original steady state in 10 years (40 periods) and focus on the consumption equivalent welfare gains along the transition path. Though it is not the most ideal environment to study optimal fiscal policy, our positive exercises can recover the fundamental mechanisms in this economy.

As Hand-to-Month workers do not have access to the financial market to smooth their consumption, a drop in labor demand leads to a relatively significant welfare loss. Higher government expenditure pushes up aggregate demand and improves supply-side efficiency endogenously. Labor demand is higher, resulting from less misallocation, inefficient liquidation, and default. Since Hand-to-Month workers do not pay taxes, the higher λ^G is, the higher welfare gain they get. As the improvement in allocative efficiency is diminishing, The welfare gain of the representative worker family diminishes with λ^G.

In contrast, the entrepreneurs' family bears the cost of financing government expenditures. The direct effect is it would crowd out their consumption C^R. However, they also gain because the intermediate firms' value and labor demand would improve as the government purchases more goods, which crowds in their consumption. The relative strength between the two powers depends on how many firms can be saved from the default zone and inefficient liquidation. The quantitative results in Figure 1.10 show that the crowding out effect is increasing with λ^G. When the fiscal policy is not so aggressive, a slight increase in government expenditure can generate a more potent impact on the decline in default probability and allocative efficiency. As a result, the crowding-out effect is not that strong, and the aggregate welfare gains are positive. When the fiscal policy is already very aggressive, an increase in government expenditure can't improve the supply-side efficiency, and the crowding-out effect

Figure 1.10: Consumption Equivalent Welfare Gains under Fiscal Policy



Notes: Consumption equivalent welfare gains under various λ^G along the transition path in responses to 2 percent negative TFP shock, which decays at a rate of 0.7. Computed as the perfect foresight transition in response to a series of unexpected shocks starting from a steady state. The solid line presents welfare gains for the aggregate economy with equal weight, the dashed line presents welfare gains of the Hand-to-Mouth workers, and the dash-dot line presents welfare gains of the entrepreneur family.

dominates. We find the optimal fiscal policy can deliver a 0.06% increase in consumption equivalent welfare along the transition path with $\lambda^G = 2.1$. It implies that government expenditure should increase by 2.1% in response to a 1 % decline in GDP. The welfare gain is modest in our calculation, but it is still positive even if government expenditure does not enter the utility function.

1.5.3 Counter-Cyclical Credit Policy

To model credit and macroprudential policies that control credit supply, we assume the government will impose a tax τ^D on the bank when it sells seized capital on the secondary market; the tax revenue will be transferred back to the government in a lump sum way. [20] Government can also choose $\tau^D \in [0, 1]$ to control credit supply where the policy rule can

[20] This setup is identical to the one where the tax revenue is transferred back to the household if Ricardian Equivalence holds.

be specified as:

$$\tau_t^D = \max\{\overline{\tau^D} + \lambda^{\tau^D}(\log Y_{t-1} - \log \overline{Y}), 0\}$$

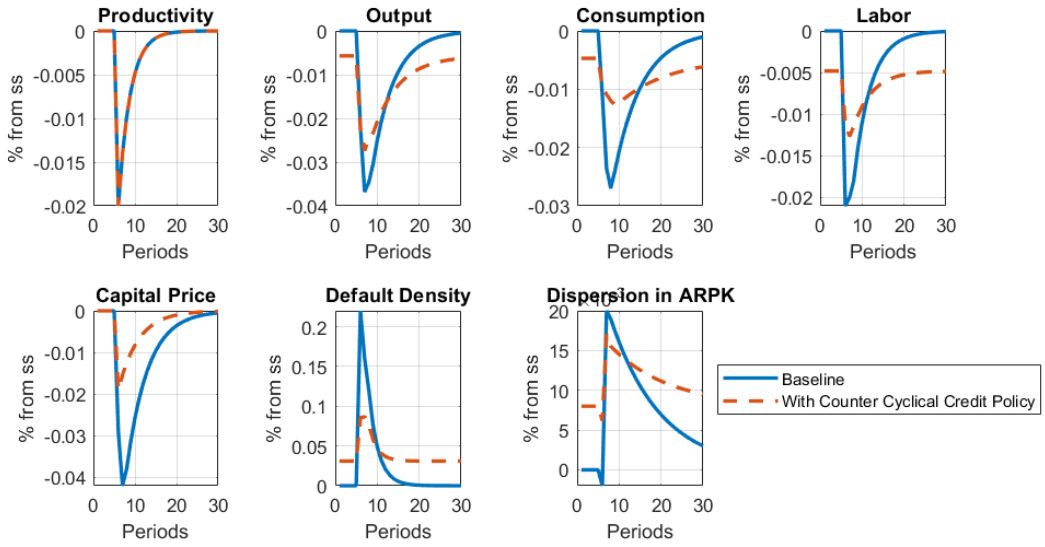
where the government relaxes credit supply when the output is lower than the steady-state level but faces an upper bound because the subsidy is not allowed. In Figure 1.11, we compare two scenarios: the baseline case where the credit policy is relaxed at all time $\tau^D = 0$, and a tighter credit policy in the steady state $\tau^D = 0.2$ with a countercyclical credit policy that loosens the credit constraint in the recession following the rule we specify.

We find that the economy with tighter credit constraints is associated with steady-state efficiency loss as the credit spreads are higher conditional on the same default risks. However, along the transition path, the economy with a tighter credit constraint in the steady state suffers less severe debt overhang problems and has a buffer where credit policy can be relaxed. As a result, the recession was less severe with a counter-cyclical credit policy. The policymakers face a trade-off between steady-state efficiency loss and a more resilient economy in response to adverse shocks.

1.6 Conclusion

In this article, we argue that the interaction between financial frictions and the partial irreversibility of fixed capital can propagate the first-moment aggregate shocks with a sizeable amplification effect as the two frictions reinforce each other. In recessions, firms tend to reduce fixed capital as well as working capital if there are no frictions. However, irreversibility makes fixed capital costly to adjust downward. This tightens financing constraints on working capital. The marginal benefit of investment decreases as expected MRPK goes down, and the marginal value of liquidity picks up. Higher default risks and lower expected MRPK create an endogenous feedback loop at the individual level. Constrained firms experience misallocation, inefficient liquidation, and default as the financial constraint becomes tighter

Figure 1.11: IRFs to TFP Shock: Credit Policy



Notes: Aggregate impulse responses to 2 percent negative TFP shock, which decays at a rate 0.7. Computed as the perfect foresight transition in response to a series of unexpected shocks starting from a steady state at t=6. The solid lines depict the IRFs in the baseline model with no active credit policy, and the dashed lines depict the IRFs in the model with counter-cyclical credit policy as described in the text. $\lambda^{\tau^D} = 4$.

gradually. With monopolistic competition, lower aggregate demand further decreases revenue productivity and creates an endogenous feedback loop at the aggregate level with an amplification effect. This result implies that micro-level frictions, the interaction between financial frictions and partial irreversibility, heterogeneity, financial heterogeneity, and portfolio choice are essential to understanding the dynamics of aggregate variables.

Our policy recipe is simple and intuitive. Higher aggregate demand releases firms from the investment inaction zone and relaxes financing constraints endogenously. We find a counter-cyclical fiscal policy is beneficial, and modest welfare gains are achievable as hand-to-mouth workers gain on higher labor demand and entrepreneurs gain on more efficient production and less inefficient liquidation. In contrast to sticky price models, where the policy recipe's direction is opposite in response to a negative productivity shock or demand shock, our model suggests the policy in the same direction, possibly more intuitive to policymakers' choices.

Chapter 2

Richer and Busier? The Facts, Causes and Consequences of Labor Supply in China

[1]

[1]I thank my coauthor Qing Huang for her permission to reproduce our work here. We thank Mark Aguiar, Nobuhiro Kiyotaki, Richard Rogerson, Kjetil Storesletten, and Xiaobo Zhang for their helpful comments and discussions. Any errors that remain are solely the responsibility of the authors.

2.1 Introduction

> Three-hour shifts or a fifteen-hour week may put off the problem for a great while. For three hours a day is quite enough to satisfy the old Adam in most of us!
>
> *John Maynard Keynes, 1930*
> *Economic Possibilities for our Grandchildren*

> I personally think that 996 is a huge blessing. How do you achieve the success you want without paying extra effort and time?
>
> *Jack Ma, 2019*
> *In an interview as CEO of Alibaba*

Working hours and leisure are crucial in determining welfare level in economic growth beyond GDP or consumption level (Jones and Klenow, 2016). Recent studies have documented that high-income countries work less than low-income countries (Bick et al., 2018), and aggregate hours worked decrease in income levels within advanced economies (Boppart and Krusell, 2020). China has experienced forty-year rapid economic growth, but less is known about the secular trend in hours and leisure. Most of the growth accounting assumes labor supply is inelastic. Do Chinese work for fewer hours and enjoy a higher welfare growth rate than just looking at consumption? Or do Chinese work for longer hours, and hence-

forth, the estimated TFP and welfare growth is upward-biased if not considering the varying labor supply margin? And what explains the changing pattern in time allocation? We try to answer those questions.

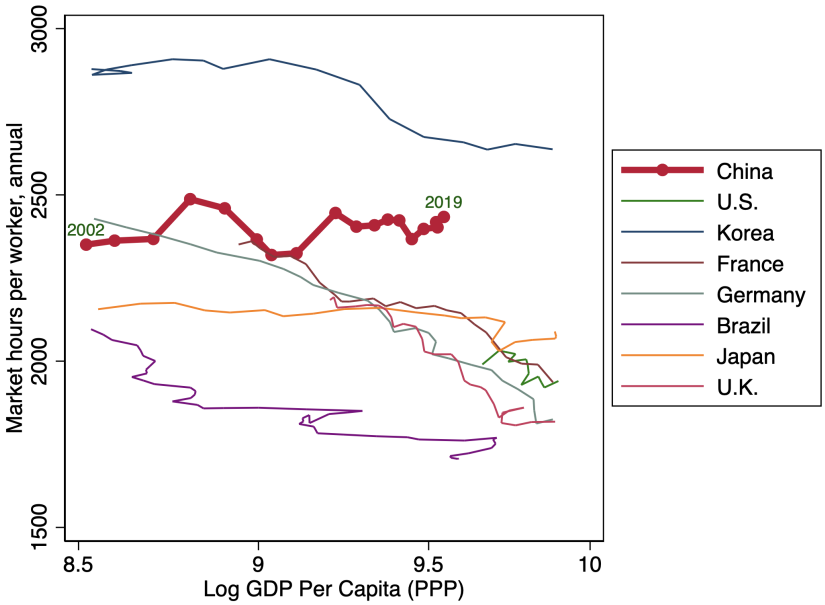
In this paper, we study the secular trend in time allocation within China in the last decade. We document a novel and robust empirical finding: Market hours have become much longer in the urban area as China has become richer in the past decade. This fact is opposite to what is observed in many other economies. In Figure 2.1, we plot market hours per worker versus GDP per capita for a set of selected major economies.[2] Market hours per worker gradually decline in all the major economies when their national income levels increase at the same development stage, except for China. For China, market hours increase with GDP growth if anything happens. We consolidate this finding by utilizing the Chinese Time Use Survey (CTUS), the best possible data to study time allocation and just becoming available in China, and the China Family Panel Studies (CFPS), a nationally representative longitudinal survey.

In the first part of the paper, we lay out the main empirical finding that working hours have increased significantly in China in the 2010s for urban, non-agriculture workers. Using the Chinese Time Use Survey in 2008 and 2018, we segment time allocation exclusively into five broad categories: time spent in market hours, time spent in housework, time spent in child care, time spent in education, and leisure. Urban and rural samples have significantly different secular patterns. For the urban sample, We find that market hours per person increased by 6.6 hours per week for males and 2.9 hours per week for females. The increase in market hours is jointly compensated by a decline in core housework and leisure[3] while time spent on childcare rises rapidly. For the rural sample, market hours per person declined by 4

[2]The data for countries except China is taken from Penn world table, and the data for China is *average market hours for urban employees*, the only related statistics released by National Bureau of Statistics. The CDP per capita in the horizontal axis is PPP adjusted.

[3]Leisure is strictly decreasing if we take child care as a separate category. The trend in leisure will be ambiguous if we take some portion of child care as leisure.

Figure 2.1: Market Hours Per Worker Across Countries



Notes: This figure plots the sequences of the average annual market hours per worker corresponding to the logarithm of GDP per capita in different countries.
Data source: Penn World Table 10.0, OECD database, and National Bureau of Statistics (China).

and 4.2 hours per week for males and females, respectively. Leisure increases by around 3.2 hours per week for males and 3.6 hours per week for females in rural areas. The inequality in terms of leisure is shrinking between rural and urban areas.

We then study the sources of observed dynamics in market hours along extensive-margin and intensive-margin. We calculate the employment rate as the extensive-margin measure and the average market hours per worker among employed as the intensive-margin measure. Overall, the increase in market hours per person in urban area is mainly driven by the changes in the intensive margin. There is a modest increase in the employment rate among urban males (2.1 percent) and a modest decline among urban females (0.9 percent). Weekly market hours per worker increase by 5.6 and 3.7 hours. For the rural sample, the decline in market hours per person is driven by both intensive and extensive margins. One interesting finding is that agricultural workers worked much less in 2018 than in 2008, which drives the

difference between rural and urban samples. The time allocation for agriculture workers is understudied in the literature due to a lack of data.

We investigate further heterogeneity related to this fact using the China Family Panel Studies (CFPS). As a comprehensive and longitudinal survey, CFPS allows us to measure working hours, wage rates, and various demographic characteristics. Using CFPS, we first find market hours per worker increased by around 3 hours from 2010 to 2020, a similar magnitude as we find using the CTUS. Then, we find that the substantial increase in market hours is broad-based in all groups. It holds across people by age, gender, education, sector, and childbearing. We also use CFPS to document the relationship between individual-level hours and wages. We find a negative relationship between hours and wages. Low-wage workers work more than their high-wage counterparts. However, the hours-wage elasticity is exceptionally stable from 2010 to 2020. These facts suggest that the increasing working hours are unlikely to be explained by different compositions of the population or the change in individual-level hours-wage elasticity.

To reconcile the tension between rises in both hourly wage rate and market hours among urban wage workers, we build a quantitative life cycle heterogeneous agent incomplete market model featuring endogenous labor supply, home production, and pay-as-you-go pension plan to account for underlying factors behind the puzzle. The model is a natural extension of Aiyagari (1994) and Huggett (1996) with the Beckerian model of household production Becker (1965).

First, we consider rising wage uncertainty to explain why Chinese workers work longer even though they are richer. We first document that wage inequality has risen sharply in China in the past decade, using the individual wage data from CFPS. By investigating wage residual dynamics from CFPS data, we find rising wage inequality is mainly attributed to an increase in initial wage dispersion and variance in persistent shocks. The increase in initial dispersion potentially captures deterministic profiles like growing college premiums,

occupation premiums, and regional inequality. The increase in the variance of persistent shocks implies individuals are facing higher uninsurable income uncertainty. Ex-ante, higher inequality, and uncertainty imply a higher density of individuals hitting borrowing limits in the steady state, delivering a stronger precautionary saving motive. Ex-post, when borrowing constraint binds, endogenous labor supply becomes the only way to provide insurance. Both channels encourage higher hours.

The second set of factors we consider is the changing age structures as a combination of birth and death rates decline. The expected worsening working-age population ratio can change consumption, saving, and labor supply decisions in a simultaneous way. As the pension system is Pay-as-you-go, a decline in the replacement ratio of pension benefits is expected. Therefore, individuals have to work longer as the extensive margin and retirement age are relatively inelastic in China. While relatively poor agents rely more on pension benefits, they have to adjust their policy functions more aggressively, leading to a decline in the correlation between wages and hours.

Motivated by the fact that non-market hours have declined sharply, the third factor we explore is the substitution between market and non-market hours. As the market goods used in home production and non-market hours are substitutes for producing home goods, growth in wage rate and capital augmenting productivity growth increases the comparative advantage of market hours over non-market hours in home production.

With the model calibrated to the Chinese economy, we find that an increase in TFP would generate lower total hours per worker and a higher correlation between total hours and wages, as the income effect dominates when we choose relative risk aversion bigger than 1. Rising initial dispersion, the variance of persistent shocks, and changing birth and death rates all contribute to an increase in the average total hours, as the model predicts. A TFP growth, as well as an increase in capital augmenting productivity in producing home goods, shifts up the ratio of market hours to non-market hours. When we combine the three

moving factors, the model can replicate the dynamics in both market and non-market hours reasonably well. The model also successfully reproduces a modest decline in the correlation between market hours and wages. This is because agents with low productivity will be more sensitive to the three factors we analyze above.

Related Literature

This paper is related to several strands of literature. First, our paper relates to empirical documentation of secular trends in time allocation. Boppart and Krusell (2020) documents that across a set of advanced countries and historically, hours fall steadily. The evidence on the US counts from the classical book-length reference Ghez and Becker (1975) to a systematic retrospect accounting by Aguiar and Hurst (2007) and McGrattan et al. (2004). The US evidence is different from most advanced economies since the market hours are relatively stable as the economy grows, partially driven by an increase in the labor force participation rate for females. One exception is that leisure is also growing in these economies, which is the other side of the coin due to a drop in non-market hours. By looking at the secular trend in time allocation in China, this paper shows a unique trend in that market hours increase as leisure decreases.

Second, our paper relates to the literature explaining differentials in hours across countries and over time, mainly motivated by the difference between the US and Western Europe. For example, Prescott (2004), Ohanian et al. (2008), and Rogerson (2010) argue that differentials in labor income taxes explain variations in market hours. We find that in our sample period, labor income tax is low for most workers and does not change by much in China.[4] Additionally, Rogerson (2006) suggests the role of government receipts to GDP ratio, and Alesina et al. (2005) suggests that labor-market regulations motivated by culture matter for

[4]For a reference, we plot the average tax rate and the marginal tax rate of personal income between 2010 and 2020 in Figure **??** in the Appendix B. Using the CFPS sample, we find the mean of real average labor income tax rate workers face is consistently below 2 percent from 2010 to 2020.

labor supply.

Thirdly, our research relates to the literature regarding income uncertainty and its implications on labor supply. It is well documented in previous studies that income uncertainty has been rising in the US since 1980 and the rising part mostly comes from transitory shocks (Moffitt and Gottschalk, 1994) (Heathcote et al., 2010a). Since transitory shocks are highly insurable, a rise in variance of transitory shocks does not transfer to consumption inequality (Krueger and Perri, 2006) or aggregate hours. For example, Pijoan-Mas (2006) finds households make ample use of endogenous labor supply as a consumption smoothing mechanism. Individuals work for much longer hours compared with a complete market scenario. Santaeulalia-Llopis and Zheng (2018) and Chamon et al. (2013) use the China Health and Nutrition Survey to estimate the income process and find a rising earning uncertainty from 1989 to 2009. We contribute to the literature by showing rising income uncertainty is also important in explaining the secular trend in market hours.

Finally, our research contributes to our understanding of the role of home production in macroeconomics. Though most of the macroeconomic models get abstract from home production. Benhabib et al. (1991) and Greenwood and Hercowitz (1991) show that real business cycle models with explicit household production sectors perform better than the standard real business cycle model. Our paper highlights that a model with explicit home production also helps to explain the secular trend of total and market hours better.

The rest of this paper is organized as follows. Section 2.2 introduces the data sources and measurement. Section 2.3 establishes stylized facts in time allocation and rising working hours in China, while section 2.4 discusses potential driving factors empirically. Section 2.5 presents the theoretical model and section 2.6 shows the calibration results. Section 2.7 conducts quantitative estimation. Section 2.8 concludes.

2.2 Data

We are interested in whether and how Chinese people change their time allocation between market work and other activities, and what explains the change in working hours. To empirically investigate these questions, we mainly use two nationally representative surveys in China: the Chinese Time Use Survey and the China Family Panel Studies. We describe the data source as well as measurement in detail below.

Chinese Time Use Survey We utilize the Chinese Time Use Survey (CTUS) at the individual level to document the time allocation across different activities for Chinese people. The CTUS is conducted by the National Bureau of Statistics and has been implemented twice: the first time in May 2008 and the second time in April 2018. The two rounds of CTUS are repeated cross-sectional and intended to be nationally representative.[5] One key advantage of using CTUS is that it is based on 24-hour diaries where respondents report the activities from the previous day in detailed time intervals.[6] Each activity is assigned to a specific category in the CTUS's set classification scheme.[7]

We segment the time allocation of each individual into five broad time-use categories: market hours, home production, child care, education, and leisure. We construct the categories to be mutually exclusive and to sum to the individuals' entire time endowment (168 hours per week).[8] These categories are based on the response for the primary time-use activ-

[5]The 2008 wave surveyed around 45,000 individuals aged 15-74 from 18,000 households covering 10 provinces. These provinces include Beijing, Hebei, Heilongjiang, Zhejiang, Anhui, Henan, Guangdong, Sichuan, Yunnan, and Gansu. The 2018 wave included around 60,000 individuals at ages 15+ from 20,000 households living in the same 10 provinces.

[6]For each household, the CTUS collects their 24-hour diaries on two days, a workday and a weekend day. For how to determine the two days more specifically, the CTUS first assigns any continuous week in the survey month, or continuous 7 days, for each household, and then chooses a day from Monday to Friday in this week as the workday and chooses Saturday or Sunday in this week as the weekend day.

[7]For example, there are 114 detailed time-use subcategories falling into 10 categories in the TUS 2008.

[8]We calculate the "daily average" time allocation for each individual by adding up the working day time multiplied by 5/7 and the weekend day time multiplied by 2/7. The "weekly average" time allocation for

ity. Market hours measure all time spent on full-time jobs, part-time jobs, apprenticeships or internships, job searches, work-related training, and family production and business activities. Home production includes any time spent on meal preparation and cleanup, doing laundry, ironing, dusting, vacuuming, indoor household cleaning, shopping/obtaining goods and services, and taking care of adult families. Child care measures all time spent by the individual caring for, educating, or playing with their children. Travel time associated with each activity is embedded in the total time spent on the activity.

We measure hours per person as the average hours spent among all persons aged 15-74.[9] In addition, the CTUS allows us to construct employment rates and hours worked per worker according to respondents' employment status and market hours. The employment rate is the fraction of adults who report being employed. The hours worked per worker are the average market hours among all those who are employed. One caveat worth noting is that CTUS is only cross-sectional, containing two periods without any information on wages. So we resort to another longitudinal household survey collected in a similar period as CTUS.

China Family Panel Studies We retrieve individual-level working hours and wages from the 2010-2020 waves of the China Family Panel Studies (CFPS). The CFPS is carried out by Peking University every two years and is a nearly nationwide, comprehensive, longitudinal household survey conducted in mainland China.[10] In each round of the survey, the CFPS identifies each respondent to be employed (as a salaried worker, agricultural worker, or self-employed), unemployed, or exited the labor market according to their self-reported employment status. For those employed, CFPS contains a *job module* to survey their working

each individual is the "daily average" time multiplied by 7.

[9]We limit the analysis sample to respondents aged 15-74 to ensure comparability between CTUS 2008 and 2018 since the CTUS 2008 only surveyed individuals in this age group.

[10]In 2010, the CFPS implemented the baseline survey covering around 57,000 individuals in 20,000 households from 25 provinces (excluding Tibet, Xinjiang, Qinghai, Ningxia, Inner Mongolia, and Hainan). In subsequent waves, the CFPS followed up with all the family members from the baseline sample. The successful tracking rates at the individual level between any two rounds of the survey were above 80%.

hours and income for each job they worked in the past 12 months.

We only focus on salaried employees using CFPS due to two major reasons. First, there are large measurement errors regarding the self-reported working hours of farmers and self-employed. Not like the Time Use Survey which is based on 24-hour diaries to record each activity, the CFPS does not require individuals to report the time allocated to other kinds of activities such as entertainment or housework, except working hours in a reference period.[11] In this case, farmers and self-employed intend to assign the time when they are working and also have other activities to working time, and thus overestimate the working time due to the flexibility of their jobs. Second, we observe individual earnings only for those who are currently employed as salaried employees. The income from agriculture or self-employment is regarded as family income in the CFPS. It is challenging to assign these incomes to each family member according to the share of time they have worked.

For salaried employees in CFPS, we calculate their actual hours worked in all jobs in the last week or in a reference week.[12] We also construct individual wages for non-agricultural salaried workers with no self-employed income.[13] In particular, we calculate the annual income for each salaried worker, including wages, bonuses, cash benefits, and in-kind subsidies, and excluding tax and insurance fees from all jobs. We obtain the hourly wage rate via the annual income divided by the annual hours worked.[14] Via providing the six-round panel

[11]For example, the CFPS asked respondents how many hours per week on average they worked in the last 12 months.

[12]Instead of asking the hours worked in a reference week, CFPS 2010 and 2012 surveyed respondents on the average number of months worked last year, the average number of days worked per working month, and the average number of hours worked per working day. In this case, we calculate the annual hours worked by the product of the above three measures. We obtain monthly hours worked by dividing annual hours by 10.7 and weekly hours worked by dividing monthly hours by 4.34.

[13]The share of salaried workers who have agricultural income or self-employed income is 5.85% in 2010, 26.65% in 2012, 12.33% in 2014, 22.79% in 2016, 9.64% in 2018, and 7.31% in 2020.

[14]In CFPS 2016, salaried workers who didn't change their primary jobs between 2014 and 2016 were not surveyed on the work module due to some technical mistakes in the questionnaire design. So, we can't observe the working hours for around one-third of the salaried employee sample (N=4,901/11,460). We impute their working hours by taking the average working hours in 2014 and 2018 if they were still in the CFPS 2018 sample and keep the same primary job of 2014 in 2018. This imputation adds around 1,811 observations in 2016.

data on individual-level labor incomes, the CFPS allows us to estimate the income process.

2.3 Empirical Facts

In this section, we first document that average hours worked increased in the last decade and take a look at both the extensive margin and intensive margin. We then show that the increasing working hours is not driven by different compositions of the population. Last, we show a not-increasing negative individual-level hours-wage elasticity, which suggests that the substitution effect might not play a dominant role in explaining the increasing hours.

2.3.1 Aggregate Hours Worked

We begin our analysis by comparing hours per person allocated to different categories between 2008 and 2018. Before this assessment, it's worth noting that the composition effect could confound our findings. The first is the age composition effect. One example is population aging, where people in 2018 are, on average, older than those in 2008, and older people tend to work less; then, an observation of increasing working hours from 2008 to 2018 could be underestimated due to the age composition effect. The second composition effect comes from education, where individuals were more educated in 2018 than in 2008, and people with higher education tend to work longer hours. This could overestimate the increase in working hours, if so. To account for the composition effects by age and by education, we divide individuals in the CTUS into 12 age groups and 6 education groups.[15] We calculate the population weight for 72 demographic groups in 2008 using the CTUS 2008 sample size, and then obtain weighted average hours in 2018 by summing the product between the raw average hours in 2018 in each demographic group and population weight in 2008 for this

[15]The 12 age groups are 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, and 70-74. The 6 education groups include no schooling or below the primary school, primary school completed, middle school completed, high school completed, tertiary school completed, and college and above.

demographic group.

Table 2.1 reports time allocation per person by area and gender between 2008 and 2018. Several findings emerge. First, there are huge rural/urban differences in hours allocated to various activities. For example, rural individuals have longer market hours than urban individuals. Urban males, on average, worked 18.7 hours per week less than rural males in 2008 and 8.1 hours per week less in 2018 than their rural counterparts. Rural females worked 7.3 hours per week more than urban females in 2008 and 5.2 hours per week more in 2018. Regarding child care, rural males spent less time than urban males, while rural females spent similar time as their urban peers. In addition, rural individuals, on average, spent less time on education and enjoyed less leisure than urban individuals for both genders.

Second, for urban individuals, we observe an increase in market hours. The average market hours per person among urban males were 33.0 hours per week in 2008 and increased to 39.6 hours per week in 2018. The average market hours per person among urban females increased from 25.0 to 27.9 hours per week from 2008 to 2018. Regarding the economic significance, the 6.6 (urban male) and 2.9 (urban female) higher weekly hours in 2018 correspond to 20 and 11.6 percent longer working time than in 2008. The increase in market hours among urban people is accompanied by a decline in home production and a decline in leisure. From 2008 to 2018, the average hours spent on home production decreased by 3.3 hours per week for urban males. Urban females experienced a much larger decline in hours spent on housework, from 23.6 hours per week to 17.3 hours per week, or a 6.3-hour decrease. Both urban males and urban females enjoyed a little less leisure between 2008 and 2018. The time spent on child care increased among urban individuals, from 2.1 hours per week to 4.0 hours per week for urban males and from 4.0 weekly hours to 8.7 weekly hours for urban females. Females, on average, spend more time on both housework and child care than males.

Third, for rural individuals, market hours per person decreased by 5 hours per week

for males and decreased by 4.2 for females. With the decline of market hours by rural individuals and the increase by their urban counterparts, the rural-urban gap in market hours has converged from 2008 to 2018. The decreasing working hours occurred with a slight decrease in time spent on housework, an increase in time spent on taking care of children, and an increase in leisure time. The only exception is that rural males spend 0.2 hours per week more on home production. While urban peers enjoyed less leisure in the decade, their rural counterparts consumed more leisure. The leisure time per week increased from 105.0 hours to 108.2 hours for rural males and from 100.8 hours to 104.4 hours for urban females between 2008 and 2018.

Table 2.1: Time Allocation by Area and Gender: Hours Per Person

	Urban male		Urban female		Rural male		Rural female	
	2008	2018	2008	2018	2008	2018	2008	2018
Market hours	33.0	39.6	25.0	27.9	51.7	47.7	37.3	33.1
Non-market hours								
Home production	10.7	7.4	23.6	17.3	6.9	7.1	22.5	19.2
Child care	2.1	4.0	4.0	8.7	1.2	1.6	4.4	7.9
Education	4.7	4.7	4.2	4.3	3.2	3.4	3.0	3.4
Leisure	117.0	112.8	111.2	109.8	105.0	108.2	100.8	104.4
Total	168	168	168	168	168	168	168	168

Notes: This table reports the average weekly hours spent on each broad-use category of activities. The rural-urban definition is based on where the individual lives at the time of the survey. The sample includes all individuals aged 15-74. All means are calculated using fixed demographic weights in 2008: 12 age groups $\times$ 6 education groups.
Data source: The Chinese Time Use Survey 2008 and 2018.

2.3.2 Extensive Margin vs. Intensive Margin

Differences in hours worked per person are shaped by two margins: the extensive margin – differences in employment rates, and the intensive margin – differences in average hours per worker. To investigate what explains the change in hours worked per person, we decompose market hours per person into the above-mentioned extensive margin and intensive margin.

Table 2.2 reports the average employment rates and hours per worker in 2008 and 2018. We highlight the following findings. First, there are great rural/urban differences and also gender differences in both years. For instance, urban individuals have a lower probability of being employed than rural peers for both genders. Urban employees worked less than rural employees, except for urban female employees in 2018 who worked longer time than rural female employees. Whether in rural areas or in urban areas, males are more likely to be employed, and male employees work much longer than females.

Second, for urban adults, we observe a slight increase in employment rate among males and a moderate decrease among females. Between 2008 and 2018, the employment rate increased from 69.5% to 71.6% for males while this declined from 55.3% to 54.4% for females. Along the intensive margin, urban employees experienced an increase in working time from 2008 to 2018 for both genders. Market hours per worker increased from 43.0 hours to 48.6 hours per week for males, and from 39.0 hours to 42.7 hours per week for females. The differences in hours per worker account for almost 76.5% (15.3%/20%) increase in hours worked per person for males and 81.8% (9.5%/11.6%) for females. Therefore, the increase in market hours per person is mainly driven by intensive margin for urban individuals, i.e., the increase in hours worked among employees.

Third, for rural adults, the proportion of those employed dropped for both males and females. The employment rate decreased by 3.3 percent for males and 6.5 percent for females. On the other hand, rural workers worked around 4.6 fewer hours per week in 2018 than in

2008 for both males and females. Thus, the decline in hours worked per person is driven by the decrease along both the intensive and the extensive margins. As the main puzzle is for urban individuals along the intensive margin, we mainly focus on wage workers from now on.

Table 2.2: Employment Rate and Hours Per Worker by Area and Gender

	Urban male		Urban female		Rural male		Rural female	
	2008	2018	2008	2018	2008	2018	2008	2018
Employment rate, %	69.5	71.6	55.3	54.4	89.9	86.6	82.0	75.5
Market hours per worker	43.0	48.6	39.0	42.7	55.6	50.9	42.4	38.8

Notes: This table reports the employment rate (the fraction of adults who report being employed) and average weekly hours worked per worker among employed. The rural-urban definition is based on where the individual lives at the time of the survey. The sample includes all individuals aged 15-74. All means are calculated using fixed demographic weights in 2008: 12 age groups $\times$ 6 education groups.
Data source: Chinese Time Use Survey 2008 and 2018.

2.3.3 Hours Worked for Different Demographic Groups

So far we document that aggregate hours worked increased in the last decade in China. In this subsection, we explore further heterogeneity related to this fact by leveraging CFPS data, which contains comprehensive demographic and socioeconomic characteristics. As introduced in Section 2.2, we focus on salaried employees using CFPS data from now on. Overall, we find that the increase in aggregate hours worked is prevalent across all demographic groups, such as age, gender, education, sector, and childbearing. These findings suggest that the increase in aggregate hours worked is not likely driven by different compositions of the population.

Age Table 2.3 first reports the average hours per worker by age groups between 2010 and 2020. In each year, the hours worked per worker decreases with age. Young people, on average, work more hours per week than older cohorts. Across different years, all age groups experienced an increase in hours worked per worker during the ten years from 2010 to 2020. The increase in hours worked per worker is especially larger for prime-age workers: roughly 4-6 hours per week for wage employees at their prime ages (16-55) from 2010 to 2020.

Gender The second heterogeneity turns to gender. Again, female workers work fewer hours per week than their male peers in all years. For example, in 2010, females average worked 49.3 hours per week while males worked 51.7 hours per week. The average hours per worker increased for both male and female workers between 2010 and 2020. In 2020, females worked 52.4 hours per week – 3.1 hours higher than in 2010; while males averaged 57.4 hours per week – 5.7 hours higher than in 2010. The magnitudes of increase in hours worked for both genders are similar to what we find using the CTUS in Table 2.2.[16]

Education To investigate the heterogeneity by education, we define four education groups: those educated with (i) at most primary school, (ii) middle school, (iii) high school, and (iv) at least college. Low-educated workers work more than their high-educated peers, but all education groups experienced an increase in working hours. Among individuals in the lowest education group, the average hours worked per week was 57.1 in 2010 compared to 60.6 in 2020. The difference between ten years amounts to 3.5 hours for the lowest education group. For individuals with middle school and high school, these two education groups experience a similar magnitude of increase in hours worked per week: 53.6 to 59.5 and 48.6 to 54.3. For the highest education group with at least college, the average hours per worker is 4.4 higher per week in 2020 compared to 2010.

[16]In Table 2.2, market hours per worker increased by 6.6 hours per week for urban male employees and increased by 3.7 hours per week for urban female employees.

Sector We obtain the industry information for each worker based on their primary job. We then compute the average hours per worker among employees working in two industries (manufacturing and services). Both manufacturing workers and service workers work longer hours than before. Between 2010 and 2018, the weekly working hours increased from 53.5 to 57.8 for manufacturing workers and increased from 49.0 to 54.2 for services workers.

Childbearing The last demographic composition we consider is childbearing, which is related to a pronounced demographic phenomenon in recent years in China – a rapid decline in the birth rate. If the share of adults in the population who are willing to give birth to a child decreases and their working hours are higher than their counterparts who spend more time on childbearing, then the increase in aggregate hours is driven mostly by the increase in the share of adults who have no child or don't need to childbearing. We group individuals by their gender, whether they had at least one child, and the age of their youngest child if they had ever given birth to any children before. For males and females with and without children, working hours have become longer for all of them ten years later since 2010. The magnitudes of increases in working hours for those who need childbearing or do not need are similar. This suggests that the increase in aggregate hours worked is unlikely to be associated with the fact that more young people choose no childbearing.

Table 2.3: Hours Per Worker: Heterogeneity

	2010	2012	2014	2016	2018	2020
Age						
16-25	54.5	55.2	57.1	61.5	60.1	59.5
26-35	50.2	50.8	53.5	55.4	54.9	55.3
36-45	50.7	52.1	53.6	53.3	55.1	55.2
46-55	48.9	50.7	52.5	51.7	54.6	55
56-65	50.7	47.3	51.1	51.6	53.4	50.1
Gender						
Male	51.7	53.2	55.2	56.0	57.1	57.4
Female	49.3	50.2	52.0	53.9	53.1	52.4
Education						
Primary and below	57.1	56.5	58.9	60.8	60.2	60.6
Middle school	53.6	54.9	56.7	58.2	59.3	59.5
High school	48.6	50.0	51.3	53.6	54.3	54.3
College and above	44.4	45.1	46.6	48.3	48.9	48.8
Sector						
Manufacturing	53.3	54.0	56.4	56.9	57.8	–
Services	49.0	50.2	52.5	54.2	54.2	–
Childbearing						
Female without child	51.4	51.2	52.2	56.0	53.8	52.1
Female with child at ages 0-6	48.7	50.8	52.9	54.8	51.8	51.9
Female with child at ages 7-12	49.8	49.7	52.7	54.1	54.1	53.7
Female with child at ages 13-18	48.9	49.9	50.2	51.4	53.5	53.8
Male without child	52.2	54.3	55.3	58.1	58.6	58.5
Male with child at ages 0-6	53.3	54.7	56.9	57.5	58.0	58.2
Male with child at ages 7-12	52.0	52.9	55.6	56.4	58.6	58.9
Male with child at ages 13-18	51.5	53.5	53.2	54.5	56.3	58.5

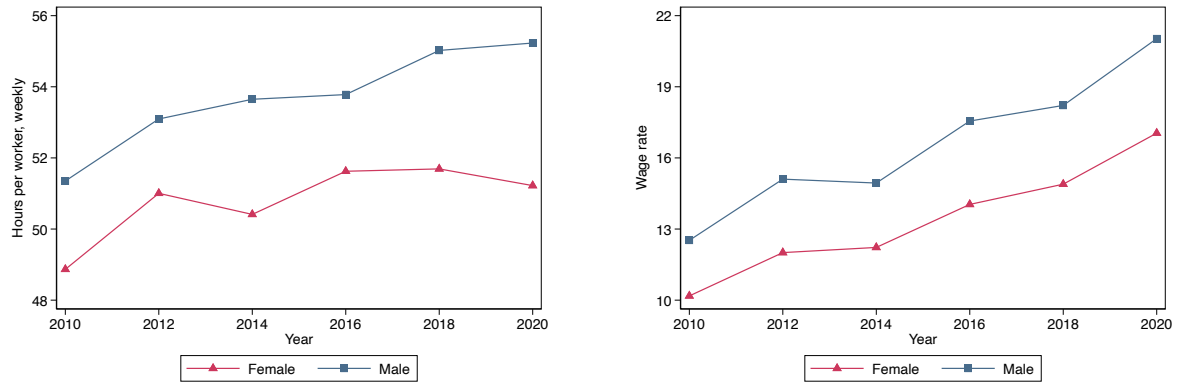
Notes: This table reports the average weekly hours worked per worker. The sample includes all salaried employees at ages 15+, and without farm or self-employment income. Industry information in 2020 is not publicly available. The age of the child is the age of the youngest child if the individual has two or more children.
Data source: CFPS 2010-2020.

2.3.4 Hours and Wages

In this section, we utilize individual-level wage data to study the relationship between hours and wages across sections and over time.

Figure 2.2 plots the trend in hours per worker and the trend in wage rate from 2010 to 2020. Within the sample of paid employees we focus on, we observe a rising trend in hours per worker for both male workers and female workers in Figure 2.2(a). Figure 2.2(b) shows the level of the wage rate for the same group of wage workers used in 2.2(a). The wage rate has experienced a surge coincident with the increasing working hours. Between 2010 and 2020, the average wage rate rose roughly 60% while market hours per worker increased around 6%. This is different from what happened in many developed countries, where increasing income levels come with declining working hours. One popular explanation for decreasing working hours in developed countries is the income effect, which assumes that people would work less if they received a higher income. However, we find that Chinese workers choose to work longer time even though they can earn a higher wage rate. This result implies that at the aggregate level, the other channels might be more important in explaining the longer working time in China instead of the income effect.

Figure 2.2: Hours Per Worker and Wage Rate, 2010-2020



(a) Hours Per Worker (b) Wage Rate

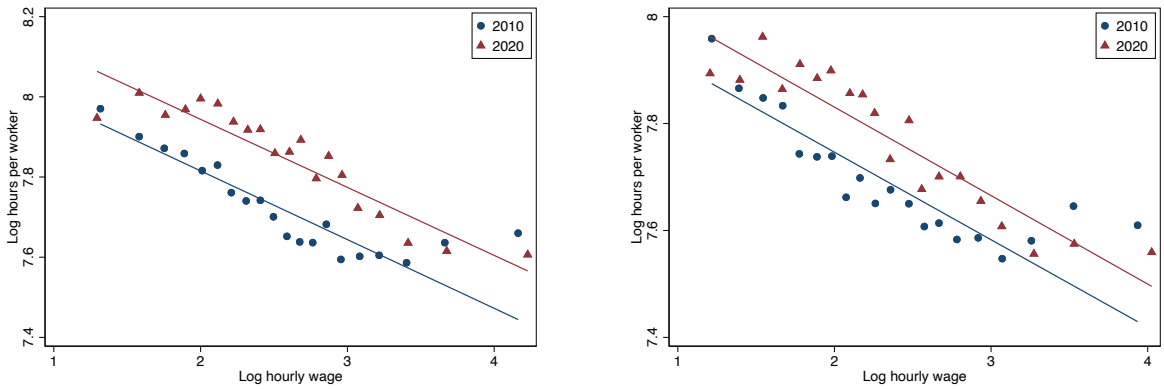
Notes: This figure plots the trend in hours per worker in Panel A and the trend in wage rate in Panel B. The wage rate is calculated by the total salary income divided by the total working hours during the same reference period. The sample is restricted to full-time workers (work at least 1200 hours per year) at prime ages (16-65).
Data source: CFPS 2010-2020.

We turn to the individual-level variations and ask whether the positive market hours-income relationship at the aggregate level over time is accounted for mostly by the individual elasticity. Figure 2.3 shows the bin scatter plots of log market hours per worker versus log wage rate in 2010 and 2020. From the cross-sectional perspective, market hours per worker are declining in the individual wage. Employees who earn a higher wage rate work fewer market hours, which points to a dominant role for preferences of income effects rather than substitution effects at the individual level. We run regressions in Table B.1 and find that the hours-wage elasticity is significantly negative. This is consistent with the evidence from the majority of countries in the world where Bick et al. (2018) found a decreasing relationship between hours per worker and the individual wage.

However, when we compare the individual-level hours-wage elasticity in 2010 and 2020, we find that it remains largely unchanged instead of being flatter or even positive. Via eyeballing check, the two fitted lines of log working hours against log hourly wage in 2010 and 2020 are almost parallel to each other. The fact holds true when we estimate the

elasticity by year as shown in Table B.1. This suggests that the increase in aggregate market hours worked is not a result of a flatter individual hours-wage elasticity, and indicates some other aggregate time-varying features that lead to higher hours. We discuss these possible forces resulting in longer market hours in the context of China in the next section.

Figure 2.3: Correlation of Hour and Wages: 2010 vs. 2020



(a) Male Sample (b) Female Sample

market hours have been rising sharply from 2008 to 2018, nonmarket hours contracts are almost at the same magnitude.l observations. To construct the binscatter plot, we divide log hourly wage into 20 equal-sized bins and plot the means of log hours in 2010 and 2020 within each bin. The solid line stands for the slope estimated using OLS regression on the 20 points. The sample is restricted to full-time workers (work at least 1200 hours per year) at prime ages (16-65).
Data source: CFPS 2010 and 2020.

2.4 Why Hours Still Increase While Wages Grow?

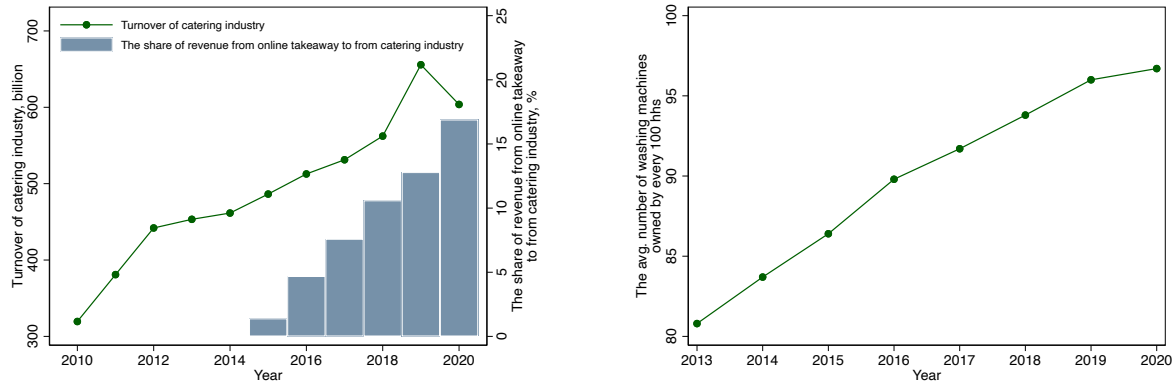
Thus far, we have shown that the average market hours worked have increased in China despite the concurrent growth of the wage rate. This increasing trend of market hours is unlikely to be driven by the composition of different demographic groups,, while market hours and wage rate are negatively correlated in cross-sections, suggesting the dominance of the income effect. In this section, we discuss potential time-varying factors that might explain why Chinese workers have higher market hours than before.

Substituting Non-Market Hours with Market Hours Though market hours have been rising sharply from 2008 to 2018, non market hours contracts almost at the same magnitude. Factors are required to explain this phenomenon shift in the structure of labor supply. This shift can be decomposed into two layers. First, households started to use market goods as substitutes for home goods. For example, a home away from home can substitute for food at home. As depicted in Panel A in Figure 2.4, turnover of the catering industry in China has been increasing by more than 100 percent from 2010 to 2019.

Secondly, home capital starts to substitute non-market hours in the production of home goods. For instance, even though dining out has become relatively cheaper, people would still prefer to have food at home sometimes as they are not perfect substitutes. To have food at home, one increasingly popular choice is ordering food delivery service. Panel A in Figure 2.4 shows that the share of online takeout in the catering industry grew from nothing to over 15 percent from 2014 to 2020. Panel B in Figure 2.4 also displays the universal usage of washing machines accomplished within the same period. The expenditure on food delivery services or purchasing washing machines can be viewed as an expenditure on home capital to produce home goods.

Intuitively, a higher wage rate implies a higher shadow cost of non-market hours and relatively cheaper market goods. It is optimal for households to substitute market goods for non-market goods and substitute home capital for non-market hours by allocating more time to market jobs if the two pairs are substitutes. Meanwhile, the price of home capital might become even cheaper relative to market goods due to productivity growth in producing home capital. This margin delivers an additional incentive to substitute home capital for non-market hours.

Figure 2.4: The Decline of Non-Market Hours: Food Production and Washing Machine



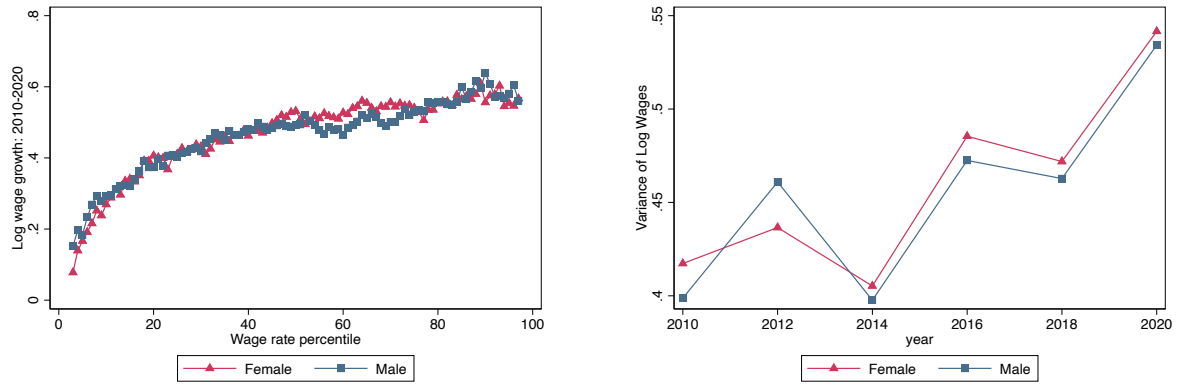
(a) Share of Online Takeout in Catering Industry (b) Numbers of Washing Machine Per 100 hhs.

Notes: This figure shows the trend of turnover of the catering industry and the share of the revenue from online takeaway from the catering industry in Panel A, and the trend of the average number of washing machines per 100 households in China
Data source: the turnover of the catering industry and the number of washing machines are from the National Bureau of Statistics, and the share of online takeaway is from The State Information Center.

Rising Wage Inequality and Uncertainty The second aggregate factor is wage inequality and income uncertainty. Rising income uncertainty might contribute to longer working hours via precautionary saving motives. Then a higher income uncertainty would induce people to save more by decreasing their consumption and increasing their labor supply.

Wage inequality in China has been rising in the last decade. Panel A in Figure 2.5 displays the wage growth between 2010 and 2020 by income percentile. We find that higher wage growth is achieved among those who have earned a high wage before. This demonstrates a rising wage inequality in the last ten years. Panel B further plots the variance of the log wage rate over time. For both genders, the variance of the log wage rate has been escalating, suggesting a growing wage inequality over time. To examine what contributes to the increasing wage inequality, we decompose the variances into deterministic and stochastic components in section 2.6. We estimate the income process by utilizing the individual-level panel data on wage residuals from CFPS. It allows us to estimate the contributions from persistent shocks and transitory shocks in explaining the rising inequality.

Figure 2.5: Wage Inequality Between 2010 and 2020



(a) Wage Growth By Income Percentile (b) Variance of Log (Wage) over Time

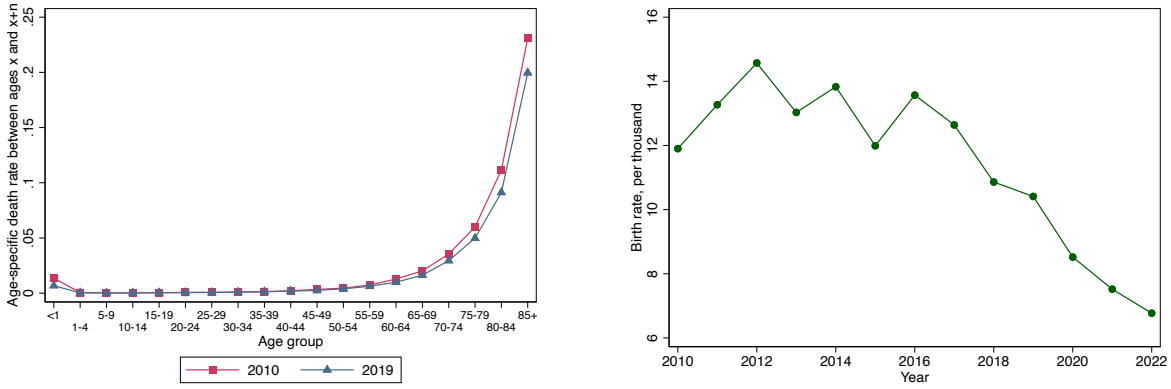
Notes: This figure shows the wage inequality by income percentile and age. Panel A plots the average wage rate growth between 2010 and 2020 by wage rate percentiles. We divide hourly wage into 100 equal-sized groups and obtain the median wage rate within each group in 2010 and 2020. The y-axis is the log difference of each group's median hourly wage rate. Panel B plots the variance of log individual wage rates in 2010 and 2020 by gender. The sample is restricted to full-time workers (work at least 1200 hours per year) at prime ages (16-65).
Data source: CFPS 2010 and 2020.

Aging and Public Pension The third change in institutional factors is the demographic changes. The age structure of the total population and life expectancy could affect individual decisions on working hours. An aging population and a decreasing number of newborn babies today together suggest a rise in the dependency ratio in the future. In other words, with the pay-as-you-go public pension system, fewer people working can support the dependent population in coming times. Additionally, with a prolonged life expectancy, people living in the present period are aware that they need to save more now for their future use in older ages. These demographic changes combined could be likely to explain why Chinese workers work longer now.

China is experiencing such demographic changes rapidly. Panel A of Figure 2.6 plots the age-specific death rate in 2010 and 2019. All age groups experienced a drop in death rate from 2010 to 2019. Furthermore, older cohorts saw a larger drop in death rate than younger cohorts at prime-working ages. Panel B shows the trend of birth rate since 2010. China's birth rate has been declining for years since the 2010s. Scrapping the one-child policy in

2016 only temporarily increase the birth rate to a very limited extent, largely leaving behind what people expected before. The combination of a decreasing birth rate and age-specific death rate means that there will be fewer young workers in the labor market in the future times. We will plug in parameters that capture this demographic shift in our quantitative model and investigate to what extent this relates to dynamics in labor supply moments we want to match.

Figure 2.6: Age-specific Death Rate and Birth Rate



(a) Death Rate

(b) Birth Rate

Notes: This figure plots the age-specific death rate in Panel A, and birth rate in Panel B. Data source: birth rate data from the National Bureau of Statistics and death rate data from the World Health Organization (WHO).

2.5 Model

The economies we investigate are extensions of Huggett (1996) with endogenous labor supply. The main characterization follows Heathcote et al. (2010b). Time is discrete and infinite. We first introduce demographics, production, and preferences. Next, we describe households' problems, firm's problems, and government policies. Finally, we define the steady state equilibrium in our economy.

2.5.1 Demographics

The economy is populated by a continuum of individuals, and there is no aggregate uncertainty. We consider the overlapping generation model, where age is indexed by j, $j \in \mathcal{J} \equiv \{1, 2, ..., J\}$. Individuals live a maximum of J periods and face a probability s_j of surviving up to j conditional on surviving up to $j-1$. Population is growing at an exogenous rate n. Let μ_j be the density of the population with age j:

$$\mu_j = \frac{s_j}{1+n}\mu_{j-1}$$

In an economy with a constant population growth rate and age-specific survival rate, age structure is stationary, so we normalize aggregate population density to be one.

$$\sum_{j=1}^{J} \mu_j = 1$$

Individuals enter the labor market at age $j = 1$ and work for Jw periods. They retire from $Jw+1$, starting to receive a pension, and die with a probability of 1 at age $j = J$. The retirement age is exogenous.

2.5.2 Production

The final good is produced by a representative firm that use aggregate capital K and aggregate market labor as inputs H with Cobb-Douglas technology:

$$Y = AK^{\alpha}H^{1-\alpha}$$

A is total factor productivity and α is capital share. Each period capital K depreciates at rate δ. Final goods can be used for market goods consumption, investment, and government

expenditure. We normalize the price of the final good to be one.

Final good can also be used to produce home capital K_h according to a linear technology:

$$K_h = A_h Y_h$$

where Y_h is the market good input, and A_h is the productivity in producing home capital. We assume the depreciation rate of home capital is 1[17]. We assume the market for producing home capital is competitive, and the implied price for home capital is $1/A_h$

2.5.3 Preferences

Following MaCurdy (1981), the period utility function is:

$$u(c, h) = \frac{c^{1-\gamma}}{1-\gamma} - \psi \frac{h^{1+\sigma}}{1+\sigma}$$

where $c \geq 0$ is final consumption and $h \in [0, 1]$ is the sum of market hours and non-market hours:

$$h = n_h + n_m$$

γ is the relative risk aversion and inverse of inter-temporal elasticity of substitution. σ is the inverse of Frisch elasticity of labor supply, measuring the elasticity of hours worked to the changes in wage rate, given a constant marginal utility of consumption.

We follow McGrattan et al. (1997) to characterize the preferences on home production. Final consumption is an aggregate over market goods c_m and home goods c_h where ω_2 is the weight on market goods and ξ_2 is the elasticity of substitution between market goods and home goods.

$$c = [\omega_2 c_m^{1-\frac{1}{\xi_2}} + (1-\omega_2) c_h^{1-\frac{1}{\xi_2}}]^{\frac{1}{1-\frac{1}{\xi_2}}}$$

[17]The household problem can be easily rewritten to allow depreciation in home capital as long as there is no adjustment cost in home capital.

Home goods is an aggregate over home capital k_h and non market hours n_h. ω_1 is the weight on home capital and ξ_1 is the elasticity of substitution between home capital and non-market hours.

$$c_h = [\omega_1 k_h^{1-\frac{1}{\xi_1}} + (1-\omega_1) n_h^{1-\frac{1}{\xi_1}}]^{\frac{1}{1-\frac{1}{\xi_1}}}$$

. Let us define the expenditure on home capital as d, $d = k_h/A_h$ and we can rewrite c_h as:

$$c_h = [\omega_1 (A_h d)^{1-\frac{1}{\xi_1}} + (1-\omega_1) n_h^{1-\frac{1}{\xi_1}}]^{\frac{1}{1-\frac{1}{\xi_1}}}$$

2.5.4 Households' Problem

Agents are born with identical preference at age $j = 1$:

$$\mathbb{E}[\sum_{j=1}^{J} \beta^j (\prod_{m=1}^{m=j} s_m) u(c_j, n_j)] \tag{2.1}$$

Agent's efficiency units per hour of market work (or individual labor productivity) depend on age(experience) and an idiosyncratic component labor productivity y_{ij} that follows the following stochastic process. Therefore, the hourly wage for an individual i of age j is:

$$p_{ij} = \underbrace{w}_{\text{common wage rate}} \times \underbrace{\exp[L(j) + y_{ij}]}_{\text{individual i's efficiency unit}} \times \underbrace{\frac{1}{\int_{\mathcal{S}} \exp[y_{ij}] d\lambda}}_{\text{normalization term}} \tag{2.2}$$

The normalization term aims to keep the expected level of the idiosyncratic component constant. Then, any changes in the stochastic process of y will not shift the aggregate labor productivity. Following income process estimation literature, We model y_{ij} as the sum of two orthogonal components: a persistent component $z_{ij} \in \mathcal{Z}$ and a transitory component $\varepsilon_{ij} \in \mathcal{E}$. The initial value of persistent component z_{i1} is drawn from an initial dispersion that

describes the labor productivity differentials when individuals enter the labor market.

$$y_{ij} = z_{ij} + \varepsilon_{ij}$$

$$z_{ij} = \rho z_{i,j-1} + \eta_{ij} \tag{2.3}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \quad \eta_{ij} \sim N(0, \sigma_\eta^2), \quad z_{i1} \sim N(0, \sigma_z^2)$$

To face idiosyncratic income risks, the financial market is incomplete. Agents can only save or borrow from a risk-free asset. Let $a \in \mathcal{A} = [\underline{a}, \infty]$ denotes the asset position where $\underline{a}$ is the exogenous borrowing limit. The net return on risk-free assets is r, and we assume there is no wealth tax given paying wealth tax is rare in our sample in China. Additionally, we assume that the assets of agents who die will be distributed proportionally to asset holdings within the same cohort for simplicity. Therefore, the total return of risk-free asset is $(1+r)/s_j$. Agents can also insure themselves with endogenous labor supply and earnings $p(w, z, j, \varepsilon)h$ is subject to tax rate τ^p. Notice that this tax refers to pension fund payment, which is used to finance pension expenditures. As discussed before, the average labor income tax rate individuals face is very low in China, and labor income tax is not a main source of government income. As a result, we do not model additional labor income tax in our model. Then we describe the agents' problem for working age individuals with age $j \in \{1, 2, .., J^w\}$ in the recursive form:

$$V(a, z, j, \varepsilon) = \max_{c_m, a', n_h, n_m, d} u(c, h) + \beta s_{j+1} \mathbb{E}[V(a', z', j+1, \varepsilon'|z)] \qquad s.t. \tag{2.4}$$

$$c_m + a' + d = \frac{1+r}{s_j} a + (1 - \tau^p) p(w, z, j, \varepsilon) n_m$$

$$a' \geq \underline{a}, c \geq 0, h \in [0, 1]$$

For individuals with age $j \in \{J^w + 1, .., J\}$, they are staying away from the labor market.

A fixed amount of pension comes from the social security fund b will be provided in each period. Though we still keep their productivity component variables in state variables for consistency, policy functions will only depend on (a, j).

$$V(a, z, j, \varepsilon) = \max_{c,a',n_h,d} u(c, h) + \beta s_{j+1} \mathbb{E}[V(a', z', j+1, \varepsilon'|z)] \qquad s.t. \tag{2.5}$$

$$c + a' + d = \frac{1+r}{s_j} a + b$$

$$a' \geq \underline{a}, c \geq 0, h = 0$$

2.5.5 Firm's Problem

There exists a representative firm that uses aggregate capital and labor to produce the final good. The firm's output is subject to a value-added tax τ^f, motivated by the fact that value-added tax is the largest single tax funding resource of the government in China. Given prices $\{w, r\}$ and tax rate, firms choose input to maximize profit.

$$\max_{K,H} (1 - \tau^f) A K^\alpha H^{1-\alpha} - wH - (r + \delta)K \tag{2.6}$$

The optimality conditions are:

$$w = (1 - \alpha)(1 - \tau^f) A K^\alpha H^{-\alpha}, \quad r + \delta = \alpha(1 - \tau^f) A K^{\alpha-1} H^{1-\alpha} \tag{2.7}$$

2.5.6 Government

The government has two independent budgets to balance. The pension system is Pay-as-you-go. A system in which retirement benefits are financed by contributions levied from current workers, as opposed to a funded system in which contributions are invested to pay for future benefits. Let τ^b be the replacement rate, which measures the ratio of pension benefit to

average labor earnings for the working age population. The pension system budget is:

$$\tau^p wH = b \sum_{j=j^w+1}^{J} \mu_j = \tau^b \frac{wH}{\sum_{j=1}^{J^w} \mu_j} \sum_{j=j^w+1}^{J} \mu_j \tag{2.8}$$

Government expenditure is financed by value-added tax.

$$\tau^f A K^\alpha H^{1-\alpha} = G \tag{2.9}$$

2.5.7 Competitive Equilibrium

Now, we describe the steady state recursive competitive equilibrium. The state space is denoted by $\mathcal{S} \equiv \mathcal{J} \times \mathcal{A} \times \mathcal{E} \times \mathcal{Z}$. Let $\Sigma_{\mathcal{S}}$ be the sigma algebra on $\mathcal{S}$ and $(\mathcal{S}, \Sigma_{\mathcal{S}})$ the corresponding measurable space. Denote the stationary distribution as λ.

A competitive equilibrium is a value function $V(s)$; decision rules $c(s), a'(s), h(s)$; firm choices H, K; prices r, w, tax rates τ^p, τ^f, retirement benefit b government expenditure G and and measures of agents λ, such that:

1. Given prices, retirement benefits and tax rates, the policy functions $c(s), a'(s), n_m(s), n_h(s), d$ solve the household's problem (4), (5) for working periods and retirement periods while $V(s)$ is the associated value function.

2. Given prices, the firm chooses optimally its capital K and its labor H, equation (7) is satisfied.

3. Labor market clears.

$$H = \int_{\mathcal{S}} n_m(s) d\lambda$$

4. Capital market clears. Government budget balances.

$$K(1+n) = \int_{\mathcal{S}} a'(s)d\lambda$$

5. Goods market clears

$$AK^{\alpha}H^{1-\alpha} = \int_{\mathcal{S}} c_m(s)d\lambda + (1+n)K - (1-\delta)K + G + \int_{\mathcal{S}} d(s)d\lambda$$

6. The government budget is balanced, and equations (8), and (9) are satisfied.

7. The invariant distribution λ is consistent with household decision rules. For all $s \in \mathcal{S}$ and $\mathbb{S} \in \Sigma_{\mathcal{S}}$, the invariant probability measure λ satisfies

$$\lambda(\mathbb{S}) = \int_{\mathcal{S}} Q(s, \mathbb{S})d\lambda$$

while the transition function $Q(s, \mathbb{S})$ is defined as:

$$Q(s, \mathbb{S}) = I\{j+1 \in \mathbb{J}\}I\{a(s) \in \mathbb{A}\}Pr(\varepsilon \in \mathbb{E}) \sum_{z' \in \mathbb{Z}} \pi(z', z)$$

2.6 Calibration to the Chinese Economy

With the quantitative model we describe in the last section, we now turn to calibrate it. The main target of our quantitative model is to reproduce rising market hours from 2010 to 2020 and conduct an accounting exercise to find what factors contribute to this phenomenon and by how much. The calibration exercises are conducted in three stages. In the first stage, we try to determine the value of externally calibrated parameters, including income process estimation. In the second stage, we try to choose a set of internally calibrated parameters

including the elasticity of between market goods and home goods as well as the elasticity of substitution between home capital and non-market hours to match a set of selected steady state moments in 2010. In the third stage, we choose time-varying internally calibrated parameters, the productivity growth in producing final goods and home capital, to match some moments in 2020.

2.6.1 First-Stage Calibration

Externally Calibrated Parameters

Table 2.6 summarizes the choices of externally calibrated parameters. We choose relative risk aversion to be 1.5, in the middle of the range that micro estimations suggest Attanasio (1999). This choice implies that the income effect dominates the substitution effect, in line with the long-run evidence documented by Boppart and Krusell (2020). The choice of relative risk aversion governs the elasticity of average market hours on total factor productivity. We choose Frisch elasticity of labor supply to be 1. Though the number is greater than some of the early micro estimations surveyed by Blundell and MaCurdy (1999), it is consistent with compensated elasticities at the macro level Keane and Rogerson (2015). Indeed, the financial frictions in our model imply why micro estimates might be subject to downward bias. We choose the length of the life cycle to be 70, reflecting individuals enter the labor market at age 20 and exit at age 90. Statutory retirement age varies by gender and occupation type in China. The statutory retirement age is 60 for male, 55 for female administration and research staff, and 50 for female manufacturing staff. We choose the length of working periods to be 40 as an upper bound. Age is the only demographics heterogeneity we consider in our model, we estimate experience profile from the regression we compute hourly wage rate dynamics. We chose the pension tax rate to be 20%, which corresponds to employer contribution in China's pension system. All employer contributions shall be deposited into the social pooling

funds as social security transfers, and it is consistent with our model setup. Notice that the model is isomorphic no matter whether this tax is imposed on the employer or employee. Since all employee contributions shall be deposited into his/her individual account, we omit this tool because it is a perfect substitute for individual savings in our model. We assume no borrowing is possible and set asset holding lower bound to be 0. α is chosen to match the capital share, and we borrow estimation of depreciation rate in China from Herd (2020).

The age-specific survival rates we plug in two steady states come from the WTO database, as shown in Panel A of Figure 2.6. Though the one-child policy was lifted at the beginning of 2016, the birth rate has been experiencing a dramatic decline since 2016. This quick shift in population growth rate is largely unexpected as most people believe the abolishment of the one-child policy can boost the birth rate for a longer period. Though the changing population growth rate will not affect the age structure and dependent ratio instantaneously, the effect on the pension system is fully anticipated. Forward-looking households will adjust their consumption, saving, and labor supply instantaneously. We assume that households form expectations on birth rate growth rate as the average birth rate growth rate from the last ten years. Under this belief, the population growth rate in the birth cohort is zero in the 2010 steady state and -0.03 in the 2020 steady state. Productivity in producing final goods in 2010 and productivity in producing home capital goods are normalized to be 1 in 2010 steady state.[18]

[18] A_h is not identifiable among ω_1 and ω_2.

Table 2.4: Summary of Externally Calibrated Parameters

Parameters	Description/Sources	Value
Invariant Parameters:		
γ	Micro-estimates of intertemporal elasticity of substitution	1.5
σ	Micro-estimates of elasticity of labor supply	1
J	Length of life cycle age 20-90	70
J^w	Length of working periods age 20-60	40
$L(j)$	Experience profile from equation	Equation 10
τ^p	Basic old-age insurance public fund tax rate	0.2
$\underline{a}$	No borrowing	0
α	Capital share	0.4
δ	Capital depreciation rate	0.05
ρ	Permanent shock	1
τ_f	Government expenditure to GDP ratio	0.25
A^{2010}	Normalization	1
A_h^{2010}	Normalization	1
Variant Parameters:		
$\sigma_{\varepsilon,2010}/\sigma_{\varepsilon,2020}$	Wage rate residuals dynamics from CFPS	0.155/ 0.143
$\sigma_{\eta,2010}/\sigma_{\eta,2020}$	Wage rate residuals dynamics from CFPS	0.0076/0.0182
$\sigma_{z,2010}/\sigma_{z,2020}$	Wage rate residuals dynamics from CFPS	0.1628/0.2400
$s_{j,2010}/s_{j,2020}$	Age specific survival rate	Figure 2.6 Panel A
n_{2010}/n_{2020}	Growth rate in birth rate	0/-0.03

Income Process Estimation

We utilize residuals in hourly wage rate dynamics from CFPS data estimating income process estimation that follows the model in the previous section. We restrict our sample to full-time workers who work at least 1200 hours per year at ages 20-60. Since the labor force participation rate for females is high in China, the selection is not very strong. Hence, we

do not restrict our sample only to male samples, and our estimation can cover the whole population. Let $w_{i,j,t}$ be the hourly wage rate for individual i, at age j and year t. We get residuals by regressing $w_{i,j,t}$ on a time dummy and a cubic polynomial in potential experience (age minus years of education minus six) $L(j)$.

$$ln(w_{i,j,t}) = \beta_t^0 + L(j) + y_{i,j,t} \tag{2.10}$$

The specification is consistent with equation (2) where we describe how the hourly wage at individual level is determined. Then, we are going to take residuals by period and education groups to estimate different sets of parameters, including the variance of transitory shocks, the variance of permanent shocks, and the variance of the initial distribution of permanent components. I follow the procedure of Heathcote et al. (2010a) in estimating the income process. Note that the variance of permanent and transitory shocks is time-variant, but we assume they do not depend on age profile. Intuitively, covariances give the same component between two periods, namely the permanent component. The variance of residuals gives the sum of the variance of a permanent component and a transitory component. Identification is achieved by the following two sets of identities.

$$var(y_{it}) - cov(y_{i,t+2}, y_{i,t}) = \sigma_{\varepsilon t}^2$$

$$var(y_{it}) - cov(y_{it}, y_{i,t-2}) = \sigma_{\varepsilon t}^2 + \sigma_{\eta,t-1}^2 + \sigma_{\eta,t-2}^2$$

The variance of initial dispersion is computed as the variance of log wage in age $j = 22$ minus the estimated variance of transitory shocks.

$$\sigma_{zt}^2 = var(y_{i,j=22,t}) - \sigma_{\varepsilon t}^2$$

Table 2.5: Income Process Estimation

	(1) σ_η^2	(2) σ_ε^2	(3) σ_z^2
2010		0.125	0.1886
		(0.0123)	
2012	0.0066	0.185	0.1478
	(0.0034)	(0.0145)	
2014	0.0086	0.147	0.2016
	(0.0044)	(0.0103)	
2016	0.0265	0.149	0.2177
	(0.0078)	(0.0085)	
2018	0.0171	0.118	0.2435
	(0.0040)	(0.0078)	
2020	0.0193	0.168	0.2364
	(0.0061)	(0.0132)	

Notes: In this table, I tabulate the estimation results for income process using CFPS sample from 2010 to 2020. Bootstrap standard error in parentheses.

Table 2.5 presents the estimation results. From 2010 to 2010, there is an increasing trend in the variance of permanent shocks and the variance of the initial distribution. The trend in variance of transitory shocks is less obvious, with a spike in 2012. Indeed, if we consider there exists a measurement error that follows mean zero variance σ_e^2 normal distribution and is independent of transitory shocks. The number we estimate here is a sum of the variance of transitory shocks and measurement errors. Hence, the main conclusion we draw from this income process estimation is increases in variance of persistent shocks and initial distribution mainly drive rising wage rate inequality. We do not model the micro foundation of initial dispersion, neither in structural model nor statistical model. But the idea is it potentially covers deterministic features like education premium, gender gap and geographic differentials. A rise in initial dispersion likely represents a rise in skill premium and compensation differentials across regions. To set parameters in the two steady states we consider, we take the average of the two earliest estimated values as parameters we plug in "2010 steady state"

and the average of the two latest estimation values in each column as parameters we plug in "2020 steady state".

2.6.2 Second and Third Stage Calibration

In the second stage, we jointly estimate the six parameters left in matching moments in the 2010 steady state. The model is complex and non-linear, and there is no precise one-to-one mapping that disciplines the model outcomes. However, we choose a selected set of moments to match, regulating the parameterization in intuitive ways. Within the production of home goods, a higher weight on home capital implies a higher share of expenditure on home capital relative to the expenditure on market goods. The static choice between non-market hours and expenditure on home capital is given by the following first-order conditions:

$$(1-\omega_1)n_h^{-\frac{1}{\xi_1}} = p(w,z,j,\varepsilon)\omega_1 A_h^{1-\frac{1}{\xi_1}} d^{-\frac{1}{\xi_1}} \tag{2.11}$$

Given $\xi_1 > 1$, home capital and non-market hours are substitutes, a higher wage rate or higher productivity in producing home capital implies a higher expenditure on home capital to non market hours ratio while the sensitivity is governed by ξ_1. Similarly, a higher weight on market goods implies a higher share of market hours in total hours. Agents with higher wages tends to consume a higher share of market goods while the sensitivity is affected by ξ_2. Our estimation results of weights and elasticities of substitution in home production are generally in line with McGrattan et al. (1997) and Dotsey et al. (2014).

Disuitability of labor affects total hours, and the discounting factor influences the interest rate and wealth-to-income ratio or capital-to-output ratio for the whole economy.

Table 2.6: Summary of Internally Calibrated Parameters

Parameters	Description/Moments to Match	Value	Relative Moments
Second-Stage			
ω_1	Weight on home capital	0.55	Average d/c_m
ξ_1	Sub. betw. n_h and k_h	1.52	Elas. of n_h to wage rate
ω_2	Weight on market goods	0.48	Average n_m/n_h
ξ_2	Sub. betw. market and home goods	2.16	Elas. of n_m/n_h to wage rate
ψ	Disutility of labor	4.29	Average total hours
β	Discounting factor	0.987	Average wealth to income ratio
Third-Stage			
A^{2020}	Productivity in producing final goods 2020	1.42	Change in wage rate
A_h^{2020}	Productivity in producing home capital 2020	1.45	Change in average n_m/n_h

In the third stage, we try to estimate the productivity growth in producing final good as well as home capital. The direct impact of productivity growth in producing final goods is reflected in the wage rate per efficiency unit. The indirect impact is households tend to use market hours to substitute non-market hours, as shown in equation 2.11. On the other hand, productivity growth in producing home capital also encourages households to substitute market hours for non-market hours. In the calibration, we first choose A^{2020} to match the growth in wage rate from 2010 to 2020 and then choose A_h^{2000} to recover the change in n_m/n_m which is not fully explained by A^{2020}.

2.7 Quantitative Results

With the calibrated model, we now turn to compute model outcomes in two steady states and see if the time-variant parameters and associated channels can explain the four main moments in labor supply we are interested in: the sharply rising market hours, the sharply declining

non-market hours, the flat total hours and the modestly declining negative cross-section correlation between hourly wage rate and market hours across individuals. The first three empirical moments are computed using urban sample in Chinese Time Use Survey while the fourth moment is computed by CFPS. As the model does not contain child care, we restrict non-market hours to core housework.[19]. With respect to model simulated moments, we restrict simulated data to agents with age 20-74 for the first three moments to be consistent with empirical moments. For the fourth moment, we restrict simulated data to agents with working ages 20-60.

From 2008 to 2018, weekly market hours per person increases from 29.0 to 33.8 hours according to the Chinese Time Use survey. The market hours per worker also increase the similar magnitude in our core CFPS sample. Interestingly, there is the same amount of rise in non-market hours, and the total hours do not change. The correlation between hourly wage rate and market hours drops around -0.034. From panel A in Table 2.7, not surprisingly, our quantitative model in 2010 matches moments in hours well since they are targeted moments. The model generates a negative correlation between market hours and wage rates across individuals. The level of the correlation between market hours and wages is an untargeted moment in our calibration, so we do not quite hit it. The performance to match data in 2020 can examine the success of our model since they are not targeted. Our model slightly undershoots total hours but reasonably replicates the diverging trend between market and non-market hours. Meanwhile, the model successfully generates a modest decline in correlation between wages and hours, as the data suggests.

To examine how much shift is led by which variant factors, we start with 2010 model parameters and only change one class of parameters to see the partial effect as presented in Panel B. The TFP growth leads to an 11.5% drop in total hours since the income effect

[19]We assume that the total available time is 112 hours per week. From table 1, we define empirical market hours as market hours/112, empirical non market hours as home production/112, leisure is defined by (112-market hours-home production)/112

Table 2.7: Model Outcomes

	h	n_m	n_h	$Corr_{p,n_m}$
Panel A: Model versus Data				
2010 Data	0.411	0.258	0.153	-0.384
2010 Model	0.411	0.258	0.153	-0.313
2020 Data	0.411	0.301	0.110	-0.418
	(0.0%)	(16.7%)	(-28.1%)	
2020 Model	0.402	0.296	0.106	-0.349
	(-2.2%)	(14.7%)	(-30.7%)	
Panel B: Model Partial effect				
TFP	0.364	0.235	0.129	-0.249
	(-11.5%)	(-8.9%)	(-15.7%)	
Productivity in home capital	0.393	0.282	0.111	-0.341
	(-3.9%)	(9.3%)	(-28.1%)	
Income Process	0.431	0.270	0.161	-0.344
	(4.9%)	(4.4%)	(5.2%)	
Demographics	0.448	0.283	0.165	-0.347
	(8.3%)	(9.7%)	(7.8%)	

Notes: Panel A reports empirical moments in two years and model simulated moments in two steady states separately. Panel B reports the partial effect of model generated moments through changing parameters from 2010 steady state, respectively.

dominates. As households substitute market hours for non-market hours, non-market hours experience a larger decline relative to market hours. TFP growth also generates an increase in correlation between wages and market hours, holding all else equal. This is consistent with cross-country evidence documented by Bick et al. (2018).

Secondly, productivity growth in producing home capital delivers contacting total hours due to a positive income effect. As home capital becomes relatively cheaper, the substitution effect on market hours overweighs the income effect and market hours expand. The substitution effect and income effect are in the same direction for non-market hours. This is the key channel to recovering a diverged trend between market and non-market hours.

Thirdly, we just modify the parameters in income process estimation to see a partial effect. Overall, rising initial dispersion and a larger variance of persistent shocks lead to higher hours and a stronger negative correlation between market hours and wage rates. An increase in initial dispersion expands the wage inequality directly. The existence of an incomplete market implies an unequal economy is associated with a larger density of households hit by borrowing constraints in a steady state. Individuals who hit the borrowing constraint are not Euler equation consumers and fail to smooth consumption. The inefficient low level of consumption would be associated with an inefficient high level of market and non-market hours implied by intertemporal substitution between consumption and leisure. An increase in the variance of permanent shocks would first increase ex-post wage inequality, and it first inherits all the effects of a higher initial dispersion. Furthermore, larger income uncertainty pushes up precautionary savings and reduces consumption as well as leisure. The magnitude of idiosyncratic shocks affects the correlation between market hours and wages in an intuitive way. If the shock is fully insurable and is independent of the marginal utility of consumption, only the substitution effect stands out. Workers work more when wages are high and less when wages are low. On the other hand, if the shock is uninsurable, the shock will affect the marginal utility of consumption, and therefore, the income effect may dominate, depending

on the strength between income and substitution effect and the pass-through from shocks to consumption. In our exercises, the main action is an increase in the variance of permanent shocks, which is highly uninsurable Blundell et al. (2008). Hence, the correlation between wages and hours goes down in response to the change in the idiosyncratic wage process.

Finally, the combination of a decline in the death rate and birth rate implies a worsening working population to whole population ratio. In our exercise, we assume the pension contribution rate to be constant, and therefore, the pension replacement ratio is endogenous. An aging population structure implies a lower pension replacement ratio in an expected way. Individuals have to build up savings because expected life expectancy is growing and expected transfer is declining. Since poor people rely more on pension transfer, the decline in the birth rate hurts the magnitude of social insurance endogenously by lowering the volume of transfer. Therefore, poor households have to work for even longer and this makes the correlation between wages and market hours smaller.

To conclude, not surprisingly, our model shows that pure TFP growth would lead to lower market hours and nonmarket hours as well as a higher correlation between market hours and wages, as suggested by cross-country evidence. However, all the mechanisms we propose in section 2.4, growing productivity in producing home capital, rising uninsurable risks as well as aging, contribute to reconciling the tension between rising hourly wage rate and rising market hours as a puzzling empirical finding in urban China given income effect dominates in the long run. Our quantitative model is only able to speak to the trend in the urban area. In the rural area, the income process could be very different from urban salary workers. Additionally, agriculture workers do not contribute to pension funds and will not link the changing demographic structures to social security benefits. In contrast, people in the rural area got large transfers in 2020 compared with 2010. Agriculture workers started to get a pure transfer as social pensions starting from 2009[20] and Huang and Zhang (2021)

[20]The pilot program started in 2009, and the program covered the whole country in 2014.

find the welfare gain from this program is substantial. The trend of time allocation in rural area is likely to be explained by a preference with subsistence level, an increase in TFP and a higher net transfer.

2.8 Conclusion

We conclude by answering the questions we ask in the first paragraph in section 2.1. Time allocation does shift dramatically in China in just ten years, but the direction and magnitude depend on where one lives. Urban workers experience an increase in market hours by around 3 to 6 hours per week, accompanied by a 60 percent growth in the average wage rate in the same period. Even though the trend in total hours is dampened by a substantial contract in non-market hours, the total hours are far from a declining trend. On the other hand, rural agriculture workers' lives become relatively easier with lower market hours and higher leisure. The secular trend of market hours for urban workers is very different from empirical evidence in other economies at the same development stage, making it difficult to rationalize the phenomenon that market hours and wage rates grow at the same time, but they are negatively correlated cross sections. To reconcile this conflict, we build a heterogeneous agent life cycle model with incomplete market and home production to conduct an accounting exercise. We find that rising productivity in producing home capital, income uncertainty, and worsening expected working population ratios generate a significant impact on market hours, as predicted by our calibrated quantitative model. The quantitative model successfully replicates the empirical observations across market hours, non-market hours, and the correlation between market hours and wage rates.

Chapter 3

Macroeconomic Implications of Heterogeneity in Residual Inequality

[1]

3.1 Introduction

Wage inequality has increased sharply in the past 40 years while wage differentials increase across and within education and occupation groups (Cutler and Katz (1992); Card and DiNardo (2002);Autor et al. (2008)). Results from Mincer regressions suggest changes in deterministic components, particularly skill premium, contribute to the rise in wage inequality. However, Mincer regressions can only explain less than half of the total variance, leaving more unexplained residuals. In particular, residual variance accounts for most of the increase in total variance according to Juhn et al. (1993). By investigating the idiosyncratic component of wages, one important consensus is this rising trend in residual inequality has

[1]I am highly indebted to Richard Rogerson and Nobuhiro Kiyotaki for all his guidance and support and to Mark Aguiar and Stephen Redding for their helpful comments, as well as all the participants in the Princeton Macro Student Workshops. Any errors that remain are solely the responsibility of the authors.

been interpreted as rising instability and transitory shocks of earnings using panel data. (Gottschalk et al. (1994); Gottschalk and Moffitt (2009)) However, one part that is under-explored in the literature is the heterogeneity in levels and secular trends of residual variance by skill groups and the macroeconomic implications of that, while previous studies usually assume the underlying idiosyncratic income or wage process is homogeneous across different skill groups. This study attempts to fill this gap. In other words, given there is plenty of literature focusing on the heterogeneity in deterministic profiles and the aggregate trend in the idiosyncratic component, we would like to investigate the heterogeneity in the idiosyncratic component of wage residuals by education groups.

On the other hand, Aiyagari-style models have become benchmark models in economics. Their popularity stems at least in part from a view that they capture important elements of reality. However, there has not been that much work assessing the quantitative predictions of these models, with important exceptions being the work of Heathcote et al. (2010b), Heathcote et al. (2014) and Kaplan and Violante (2010). I aim to assess the quantitative predictions of a standard incomplete market model in another setting, incorporating endogenous labor supply and leveraging the fact that the idiosyncratic shock processes have changed differentially across skill groups.

By utilizing data from both CPS and PSID samples, I find residual inequality by education groups disperses sharply from 1980 to 2018. Residual inequality increases more for high-skilled labor (with a college degree) than for low-skilled labor (without a college degree) in this period. Changes in within-group variances of wage residuals by experience-education cells strongly confirm the same fact. Meanwhile, this dispersion concentrates on the young cohort, suggesting that an increase in initial dispersion may play an important role in explaining rising residual inequality and its heterogeneity across different groups. Hence, I estimate the wage rate process by education groups and decompose the wage residual variance into the variances of permanent shocks, transitory shocks, and the initial dispersion

of permanent components. Using the male household heads sample in PSID, I find that in the aggregate, the rises in the variance of initial dispersion and transitory shocks contribute equally to the rising trend of residual inequality by around 0.025 log points, respectively. The gaps between the variance of transitory shocks, permanent shocks, and initial dispersion have all been widening by education groups from 1980 to 2010. In particular, for high-skill labor, the variances of transitory shocks, permanent shocks, and initial dispersion increase by 0.04, 0.001, and 0.03 log points. For low-skill labor, the variances of transitory shocks and initial dispersion of permanent components increase by 0.02 and 0.02 log points. In contrast, the variance of permanent shocks decreases by 0.003 log points for low-skill labor. I also estimate skill premium and experience premium accounting for the deterministic profile to calibrate my quantitative models. I find that the college premium has been rising much slower since 2000, while there has been a constant growth rate in relative labor supply from 1980 to 2010. After accounting for the education composition effect, I also find a decreasing trend in the experience profile for both high and low-skill labor.

Given four sets of wage rate process estimates by education groups and periods in hand, I am first interested in examining the extent to which differentials in the idiosyncratic wage process account for observed empirical differentials in macroeconomic moments by education groups and periods. Secondly, I am interested in the implications on aggregate variables by considering this heterogeneity in residual inequality by skill groups. Hence, I build a standard life cycle incomplete market heterogenous agent model with endogenous labor supply and two types of labor by education groups to study this problem quantitatively (Aiyagari (1994); Huggett (1993); Huggett (1996)). In a standard incomplete market model where individuals only face idiosyncratic shocks to human capital, the parameters we estimate directly regulate labor supply and consumption behaviors. I focus on three sets of empirical targets to match: the cross-sectional correlation between wage and hours, the variance of hours, and the variance of earnings. For each set of moments, we have four model-generated moments

to match their empirical counterparts: 1980 and 2010 steady states for high-skilled and low-skilled labor, respectively. The quantitative results show that our model can reasonably reproduce the levels and trends for the two types of labor in the three moments.

Essentially, the variance of log earnings can be decomposed into the variance of log wage rates, the variance of log hours, and the correlation between log wage rates and log hours. Larger initial dispersion, permanent shocks, and transitory shocks will increase variability in wage rate directly. The variability of labor supply will also increase through endogenous adjustments. On the other hand, an increase in the variance of transitory shocks has a positive effect on the cross-sectional correlation between wages and hours. Since transitory shocks are highly insurable, an increase in transitory shocks would have little impact on permanent income and have little income effect. When the substitution effect dominates, workers would like to supply more labor when the wage rate is temporarily high. In contrast, an increase in the variance of permanent shocks has a negative effect on the correlation between wages and hours since it is not very insurable. When the income effect dominates, workers would like to reduce labor supply when they receive a good draw in permanent human capital. Because the gaps between high-skilled and low-skilled labor in the variance of permanent and transitory shocks have all been widening, the two effects cancel out each other. Therefore, we see high and low-skilled labor experience around the same level of growth in the correlation between wages and hours. Furthermore, a softer borrowing constraint will increase the correlation between wages and hours since any idiosyncratic shocks are more insurable under a larger borrowing limit. These endogenous adjustments of labor hours significantly contributed to a rise in residual earnings inequality in addition to the exogenous rise of residual wage rate inequality.

Finally, the quantitative results show that considering the heterogeneity in residual wage rate inequality by skill groups has significant implications for aggregate variables, including wealth inequality and aggregate capital in equilibrium. Since high-skill labor faces larger

idiosyncratic risks, they have larger precautionary saving motives and relatively save more. Due to this positive correlation between relative saving motive and wage rate, total wealth inequality increases. The rise in aggregate savings also further suppresses the equilibrium interest rate.

This paper contributes to the literature in two main aspects. First, it documents a substantial heterogeneity in the secular trend in the idiosyncratic component of the income process by education groups since 1980. High-skilled labor experiences more growth in the variance of initial dispersion, the variance of permanent shocks, and transitory shocks. Second, it shows quantitatively that this heterogeneity in the idiosyncratic wage process is central to explaining the various levels and trends in the correlation between wages and hours and variability in labor supply and earnings. This heterogeneity in the idiosyncratic wage process also has implications for aggregate variables in general equilibrium.

The paper is organized as follows. I review the related literature in the next section. Section 3 introduces the data used and the process of selecting my sample. Section 4 describes the empirical trend in residual inequality and shows the estimates of the income process by education groups. I describe my model in section 5 and show quantitative results in section 6. Section 7 concludes.

3.2 Literature Review

This paper is related to at least two branches of literature. First, this paper relates to previous work that documents the rising residual variance and income instability. Juhn et al. (1993)conclude that the rise in residual inequality accounts for most of the total growth in wage inequality. In a later opposite paper, Lemieux (2006) finds that there is no uniform increase in residual inequality in all education-experience cells and concludes that the composition effect substantially contributes to the rise in residual inequality while

the working population becomes older and more educated. Using the covariance structure of wage residuals, a consensus has been reached that there is a growing trend in income instability, i.e., workers experience increasingly large fluctuations in earnings from year to year. (Gottschalk et al. (1994); Gottschalk and Moffitt (2009); Heathcote et al. (2010a)) This paper contributes by documenting a new fact that there is substantial heterogeneity in the idiosyncratic component of income process by education groups since 1980[2].

Second, this paper also relates to the literature that parameterizes idiosyncratic shocks in macroeconomic models. Heathcote et al. (2014) shows how changes in idiosyncratic shocks affect saving and labor supply decisions in a fully tractable framework. Krueger and Perri (2006) find a large increase in income inequality does not lead to a large increase in consumption inequality since increasing shocks are transitory and thus highly insurable. Heathcote et al. (2010b), probably the closest paper to mine, investigates comprehensively how changes in the wage structure, including rising transitory shocks, account for changes in empirical moments in consumption and labor supply. Our paper differs from theirs since high and low-skilled labor have various income dynamics in my quantitative model and the model does successfully match the correlation between wages and hours and variability in hours as well as earnings in both levels and trends for two types of labor.

[2]several papers estimate the income process by education groups. Guvenen (2009) does not find evident differences in variances of transitory shocks and permanent shocks using PSID from 1968 to 1993. Meghir and Pistaferri (2004) find that college graduates have larger permanent shocks compared with high school graduates. However, the transitory shock is much smaller for college graduates than for high school graduates. However, their data is a bit old; they used data from 1967 to 1992. Abbott et al. (2013) estimate the variance of persistent income shocks is around the same for college graduates and high school graduates using NLSY79. To the best of my knowledge, I am the first to document the heterogeneity in secular trends by education groups until the 2010s.

3.3 Data

3.3.1 CPS

We mainly use the CPS March file from 1980 to 2018.[3] CPS is a comprehensive database that contains rich demographic information, income, occupation, and working industry. The sample size is also big enough to describe the distribution of the whole population. The drawback of this data is that the income is top-coded so it is not very useful to recover the very top distribution, say top 1 percent. We follow Autor et al. (2008) to multiple top-coded incomes by 1.5 before 1996. After 1996, top-coded observations are reported as the average income in each top-coded cell. We restrict our main sample to full-time, full-year workers with more than 35 hours of working time a week and more than 40 weeks a year. We restrict our sample to those aged from 16 to 65. We also drop the very bottom distribution with income lower than 5,000 dollars a year and imputed hourly wage less than 60 percent of the minimum wage. The selected sample size is around 50,000 a year while the CPS sample size varies from 150,000 to 200,000 a year. The income measure is adjusted to 2018 dollars by price indices. In this paper, I use CPS data to generate most of the empirical secular trends and moments since its sample size is much larger than PSID.

3.3.2 PSID

PSID is the most commonly used panel data to estimate the income process. I use it mainly for estimating the income process to get sets of variances of permanent/transitory shocks and initial dispersion. I restrict PSID data to the SRC sample and only focus on male

[3]The data is provided by the Center for Economic and Policy Research. 2019. March CPS Uniform Extracts, Version 1.1. Washington, DC

household heads between 25 to 60 years old, which is consistent with our life cycle model to calibrate. Hourly wage is computed by total annual labor income divided by total annual working hours. I restrict the sample to those with hourly wages no lower than minimum wage. I use PSID data from 1968 to 2017 and drop observations that appear less than 5 times in the panel when we estimate the income process. The survey becomes biannually from annually after 1997.

3.4 Heterogeneity in Income Process by Education Groups

3.4.1 Heterogeneity in Residual Variance

We start by plotting the total variance and residual variance of log wages from 1980 to 2018 for men using CPS data in Figure 1. Residuals are obtained from regressing ln(wage) on a set of dummies accounting for educational attainment and an education group-specific quadratic form of experience profiles. Regressions are run by separate years. We will use this way to consistently define residuals in the rest of this paper. From 1980 to 2018, total variance increased by 12 log points while residual variance increased by 8 log points, explaining two-thirds of the growth. Figure 2 shows the secular trend of residual variance by college graduates and high school graduates using data from CPS and PSID respectively. Each sample shows a dispersion in residual inequality from the 1980s (around 1989 for CPS and 1982 for PSID). A few points can be summarized. First, overall, residual inequality increases for both of the two groups but more for college graduates. Second, the increasing trend for residual inequality has been slowed down after 2000 and the gap between the two groups remains relatively stable. Later we will see the trend for college premium also becomes relatively stable after 2000. Hence we refer to the trend after 2000 as "2010 steady-state". The trend around 1980 is not stable, but we also refer to the periods around 1980 as

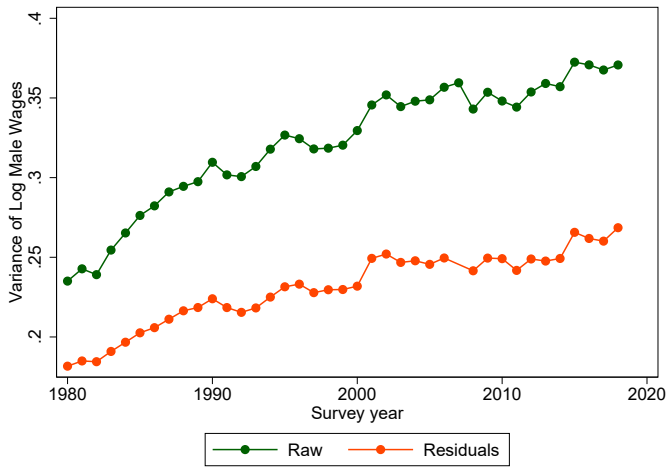


Figure 3.1: Varinace of Log Male Wage, CPS

Notes: In this figure, we plot the secular trend for the variance of log male wages and residuals of log male wages using the CPS sample from 1980 to 2018 separately. The attrition process is described in section 2. Additionally, I drop the top 1% percent in wages to plot this figure since the variance of wages in the top 1% highly depends on the top coding and imputation process. Nevertheless, our empirical findings and model do not focus on the top distribution. To get wage residuals, we regress log wage on a set of dummies accounting for education and education-specific quadratic forms of experience.

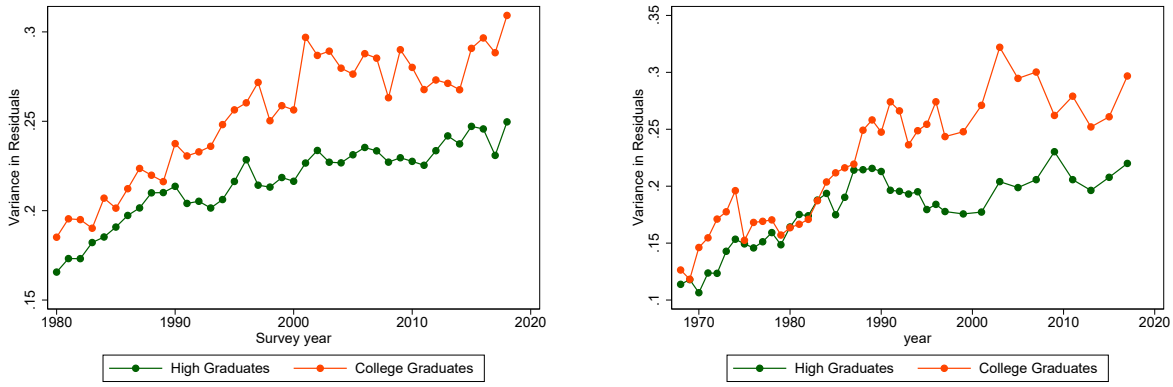
the "1980 steady state" for comparisons. In the part of the income process estimation and quantitative model, I will also focus on comparing the two types of labor in two steady states. To compare how residual inequality changes in each education-experience cell in two steady states, I show the levels and differences for each group in Table 1. To calculate variance in the 1980 steady-state, I pool residuals from 1980 to 1984 together for each cell. Similarly, I pool residuals from 2008 to 2012 to calculate variance in the 2010 steady state.

Several lessons can be learned from Table 1. First, given year and education level, residual variance increases over the life cycle. Second, given year and experience level, higher education is associated with higher within-group residual variance. Third, comparing two steady states, given education level and experience level, within-group variances increase for all cells, but by much more for college graduates. Fourth, comparing two steady states, within-group residual variance disperses sharply for young men by education groups. In 1980-1984, there was only a small dispersion in residual variance for the least experienced group.

Table 3.1: Within-Group Residual Variance of Wages by Experience-Education Cell for Men

	(1) 1980-1984	(2) 2008-2012	(3) Differences
Panel A. By Education and Experience			
Less Than High School:			
1-10	0.164	0.182	0.018
11-20	0.201	0.216	0.014
21-30	0.214	0.247	0.031
30+	0.216	0.265	0.049
High School Graduates:			
1-10	0.171	0.209	0.038
11-20	0.188	0.248	0.059
21-30	0.206	0.261	0.056
30+	0.204	0.267	0.063
Some College:			
1-10	0.179	0.222	0.043
11-20	0.192	0.247	0.056
21-30	0.225	0.271	0.046
30+	0.259	0.293	0.035
College Graduates:			
1-10	0.185	0.263	0.078
11-20	0.242	0.328	0.086
21-30	0.291	0.387	0.096
30+	0.335	0.406	0.071
Advanced:			
1-10	0.278	0.348	0.071
11-20	0.311	0.395	0.084
21-30	0.338	0.459	0.121
30+	0.384	0.481	0.097
Panel B. Weighted Average			
Actual Share	0.215	0.295	0.080
1980 Share	0.215	0.270	0.055
2010 Share	0.232	0.295	0.063

Notes: This table tabulates the wage residual inequality by education-experience cells in Panel A. All the changes in the variance are significantly different from zero at the 95-percent confidence level. In Panel B, I computed the weighted average by their labor share. For a discussion of the implications in this table, see section 3.



(a) CPS

(b) PSID

Figure 3.2: Heterogeneity in Residual Inequality by Skill Groups

Notes: In this figure, we plot the secular trend for variances of wage residuals by educational attainment using the CPS sample from 1980 to 2018 and the PSID sample from 1968 to 2017, separately. The attrition process is described in section 2. To get wage residuals, we regress log wage on a set of dummies accounting for education and education-specific quadratic forms of experience.

The gap is around 0.02 log points from "Less Than High School" to "College Graduates". But the same gap enlarges to 0.08 in 2008-2012. Panel B in Table 1 shows the extent to which the composition effect contributes to the aggregate growth in residual inequality. Composition effect contributes less than one-third to the total residual growth from 1980 to 2010, much less than what is calculated by Lemieux (2006) using data from 1975 to 2000.[4]

To sum up, there is a substantial heterogeneity in residual inequality across education groups in both levels and trends from 1980 to 2010. The composition effect caused by this heterogeneity does not contribute to the majority part of secular growth in residual inequality during the period. Potentially, this heterogeneity in cross-sectional residual inequality could be translated into heterogeneity in income processes. This heterogeneity by education groups and periods in the income process will have different predictions in macroeconomic moments.

[4]Table 1 is a replication of table 1A in Lemieux (2006) with different time coverage. Lemieux (2006) focuses more on explaining a large part of aggregate growth in residual inequality is due to a composition effect, and he also argues that CPS contains more measurement error than CPS/ORG. Instead of that, I focus on heterogeneity in increasing trends in within-group residual inequality. However, none of our findings fail even in Lemieux's original table.

3.4.2 Income Process Estimation: stochastic component

In this subsection, we will focus on translating the cross-sectional heterogeneity in residual inequality into four sets of parameters governing the income process through estimation with panel data. As I stated before, I will approximate parameters in two steady states: 1980 and 2010. To be consistent with the quantitative model, I estimate two types of labor respectively. Low-skill labor includes all individuals without a college degree and high-skill labor includes all individuals with a college degree. I think that it is sufficient and efficient to simplify the heterogeneity we observed in Figure 2 and Table 1 into the four types.

Labor economics has a long tradition of estimating the income process. See section 3 in Heathcote et al. (2009) for a good survey. Based on a long literature of empirical studies, it is believed that the stochastic part of the income process constitutes a permanent or a very persistent component and a pure transitory component. (Lillard and Willis (1978); Abowd and Card (1989)) We follow Heathcote et al. (2010b) to model the standard linear income process. We first model the income process as a sum of a deterministic component and a stochastic component where i denotes individual, j denotes age, e denotes education level and t denotes time. This is consistent with our regressions and y_{it} is just the residuals we obtain from regressions.

$$w_{it} = p_t^e \times exp(L^e(j) + y_{it})$$

$$ln(w_{it}) = ln(p_t^e) + L^e(j) + y_{it}$$

In terms of modeling the stochastic part, I follow Heathcote et al. (2010a) with a non-stationary linear income process with permanent and transitory shocks. ϵ_t, η_t are independent, serially uncorrelated, and normally distributed random variables with zero mean and variancesσ_ϵ, σ_η respectively. Note that there are two implications for this assumption. First, for a specific cohort, the variance of the permanent component would increase as they

become older. If the variance of transitory shocks is time-invariant, then we would expect to observe variance of y_{it} would increase as years of experience increase. Second, a life cycle version of the standard incomplete market model but a perpetual youth version just fit in this non-stationary income process.

$$y_{it} = z_{it} + \epsilon_{it}$$

$$z_{it} = z_{i,t-1} + \eta_{it}$$

$$and \quad \eta_{it} \sim N(0, \sigma_\eta) \quad z_{i0} \sim N(0, \sigma_{z0}) \quad \epsilon_{it} \sim N(0, \sigma_\epsilon)$$

Then we are going to take residuals by period and education groups to estimate different sets of parameters including the variance of transitory shocks, the variance of permanent shocks, and the variance of the initial distribution of permanent components. I follow the procedure of Heathcote et al. (2010a) in estimating income process but by different education groups. Note that the variance of permanent and transitory shocks is time-variant but we assume they do not depend on age profile. Intuitively, covariances give the same component between two periods, namely the permanent component. The variance of residuals gives the sum of the variance of a permanent component and a transitory component. Since the PSID data becomes biannual after 1997, we only use moments that can be computed using biannual data for consistency. Identification is achieved by the following two sets of identities.

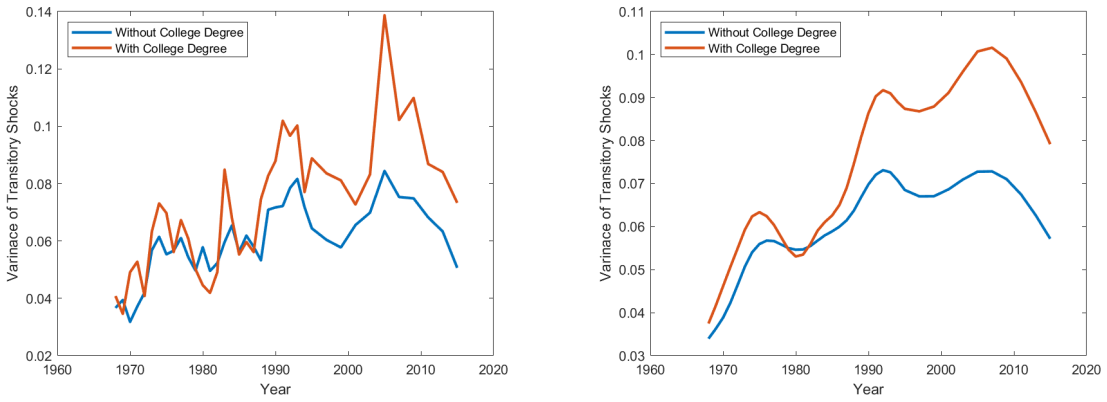
$$var(y_{it}) - cov(y_{i,t+2}, y_{i,t}) = \sigma_{\epsilon t}$$

$$var(y_{it}) - cov(y_{it}, y_{i,t-2}) = \sigma_{\epsilon t} + \sigma_{\eta,t-1} + \sigma_{\eta,t-2}$$

Figures 3 and 4 show the estimation results of variances of transitory and permanent shocks, respectively. Note that the system is just identified, so we compute the variance rather

than estimating that by minimizing moment distance. For transitory shocks, we are likely to conclude that the variance of transitory shocks has been increasing more for high-skill labor relative to low-skill labor, but there is also a long-run increasing trend in transitory shocks for low-skill labor. The volatility is much higher in our estimation of the variance of permanent shocks. In general, there is a slightly increasing long trend for high-skilled labor and a decreasing long-run trend for low-skilled labor.

We mainly focus on the comparison of two "steady states". I approximate the parameters in the initial 1980 steady state by averaging parameters from 1970 to 1984 and the final 2010 steady state from 2001 to 2015. The numbers are reported in Table 2. Estimates for all individuals without heterogeneity are also reported in Table 2. In aggregate, consistent with Heathcote et al. (2010a);Heathcote et al. (2010b); Krueger and Perri (2006); Moffitt and Gottschalk (2011) and Moffitt and Gottschalk (2012), we find an increasing trend for variance of transitory shocks. And we also find there has been a large increase in the variance of initial dispersion, which contributes 0.024 log points in the total growth of residual variance, roughly the same magnitude contributed by the increase in the variance of transitory shocks. Furthermore, I find heterogeneity in the trend of changes in the variances of transitory shocks by education groups. The variance of transitory shocks for high-skilled labor increases by more than two times that low skilled labor(0.037 and 0.018). The variance of permanent shocks remains stable in the aggregate, but it increases for high-skill labor and decreases in low-skill labor from 1980 to 2010. What accounts for rising transitory shocks and the heterogeneity in its secular trend? One explanation is business cycle will influence income stability (Moffitt and Gottschalk (2011)) and it is hard to reconcile with the heterogeneity by education groups. Another explanation is proposed by Violante (2002). It argues that technical acceleration reduces workers' capacity to transfer skills from old to new machines, generating a rise in the cross-sectional variance of wages. I follow the technological acceleration story to explain my empirical findings for two reasons. First, high-skill labor is supposed to



(a) Time series

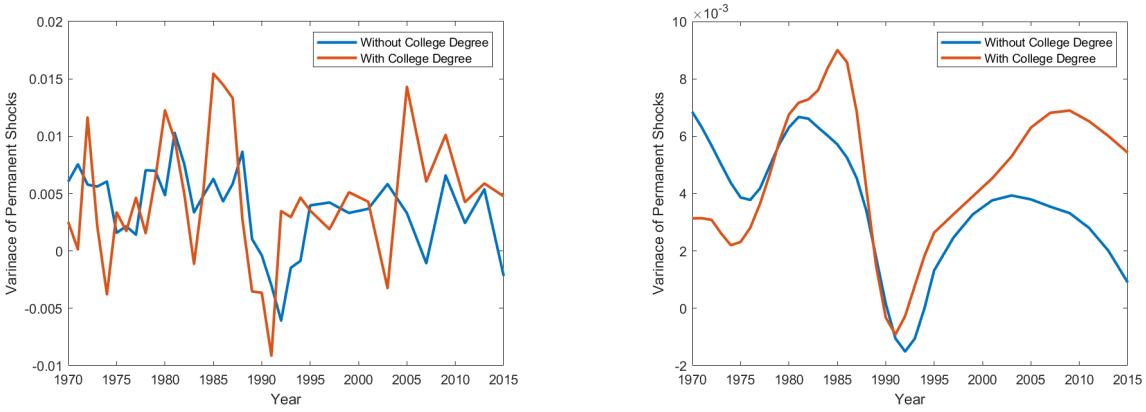
(b) HP-filtered with para=10

Figure 3.3: Variance of Transitory Shocks

Notes: In panel a, we plot the secular trend for estimated variances of transitory shocks by educational attainment by male wage residual dynamics using the PSID sample from 1968 to 2017. The attrition process is described in section 2. In panel b, we plot the HP-filtered data trend with parameter=10

have more exposure to technological acceleration than low-skill labor since their occupations are more likely to be involved with the technology frontier. This helps to understand why high-skilled labor experiences a larger increase in their variance of transitory shock. Second, it helps to understand why the variance of transitory shocks became relatively stable at high levels after the 2000s because we also observe less skill-biased technical change implied by Figures 6 and 7 around the same period.

The initial variance of the permanent component can not be computed from the previous two sets of identities directly. We compute the initial variance of the permanent component by calibrating based on 25 years old residual variance. We assume that 25 years old is the first year for every individual to enter labor markets, say j=0 at age=25. Therefore, the variance of residuals at age 25 can be decomposed into a sum of the variance of transitory shocks and the initial variance of permanent components. We find an increasing trend for dispersion in the initial distribution of permanent components for both high-skill and low-skill labor while more for high-skill labor. Figure 5 shows the secular trend of variance of male wage residuals at age 25 using CPS. It provides alternative evidence that the sum of the variances of



(a) Time series

(b) HP-filtered with para=10

Figure 3.4: Variance of Permanent Shocks

Notes: In panel a, we plot the secular trend for estimated variances of permanent shocks by educational attainment by male wage residual dynamics using the PSID sample from 1968 to 2017. The attrition process is described in section 2. Note that we identify a sum of variance in the adjacent two years., so we simply divide that by two. In panel b, we plot the HP-filtered data trend with parameter=10

Table 3.2: Income Process Estimation

	(1) σ_η	(2) σ_{z0}	(3) σ_ϵ
"1980" High Skill Labor	0.009	0.072	0.059
"1980" Low Skill Labor	0.006	0.073	0.051
"1980" All	0.007	0.073	0.053
"2010" High Skill Labor	0.010	0.101	0.096
"2010" Low Skill Labor	0.003	0.094	0.069
"2010" All	0.007	0.097	0.079

Notes: In this table, I tabulate the estimation results for income process by educational attainment using PSID sample from 1968 to 2017. I compute numbers in the 1980 steady state by averaging numbers from 1970 to 1984 and the 2010 steady state by averaging numbers from 2001 to 2015.

transitory shocks and variance of the initial distribution of permanent components increases by more for high-skill labor. Increasing heterogeneous returns to various majors is potentially helpful to explain that, i.e., individuals with a history or physics degree would make around the same compensation when they first enter the labor market in the past while college graduates with a CS degree may make significantly more than those with sociology degrees when they find their first job today.

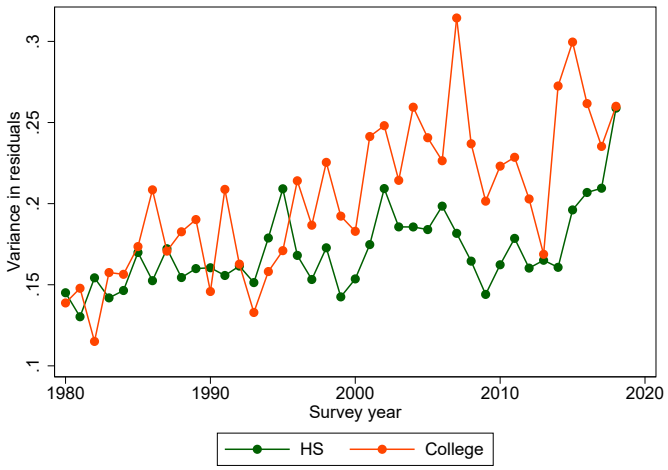


Figure 3.5: Variance of Log Male Wage Residuals at Age 25

Notes: In this figure, we plot the secular trend of variance in log male wage residuals at age 25 by educational attainment using the CPS sample from 1980 to 2018. The attrition process is described in section 2.

Reminding that only less than half of variation can be explained by deterministic profiles by Mincer regressions but Keane and Wolpin (1997) argued that 90% of a lifetime earning variance is predetermined at the time individuals enter the job market. That means there must be a large amount of "shocks" we identified as residuals are predetermined but not idiosyncratic. Therefore, I attempt to interpret the initial distribution of the permanent component as unobserved heterogeneity when individuals enter the labor market even though all individuals are ex-ante identical in our setting. This can be imagined as a predetermined "individual fixed effect" since they will carry this permanent component in their entire

career. In principle, one can build up a model with homogeneous deterministic profiles and permanent or very persistent components (RIP); or heterogeneous deterministic profiles with not very persistent components (HIP). Guvenen (2009)argues that it is hard to statistically distinguish the two models with typical panel data. Hence, we adopt RIP and interpret the initial distribution of the permanent component as predetermined unobserved heterogeneity.

3.4.3 Deterministic Component

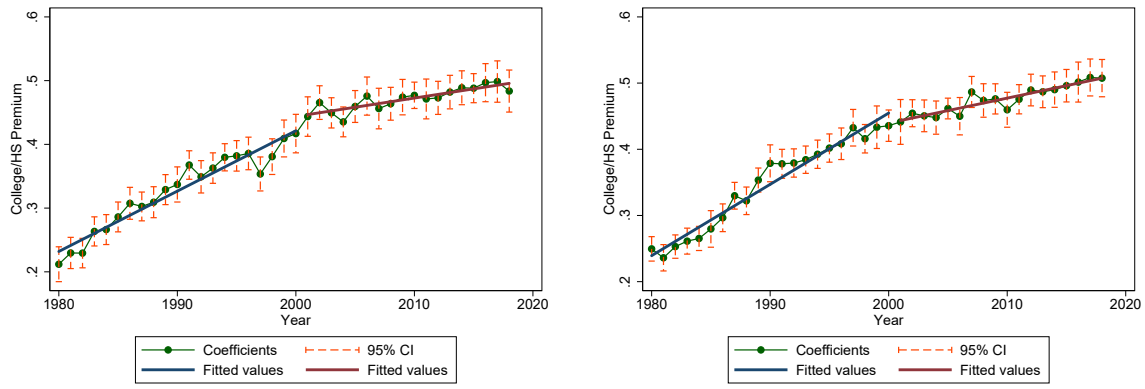
Though our primary focus is the heterogeneity and changes in the idiosyncratic component of the income process, we also estimate the corresponding deterministic component for calibration in the quantitative model.

College Premium

Increasing college premiums is one of the most striking wage structure changes in the past 40 years and it has been well documented(Autor et al. (2008); Katz and Murphy (1992)). Figure 6 shows the trend of skill premium by gender. I got the coefficient by regressing ln(wage) on a dummy account for college with cohort and state fixed effect for each gender in each year. (experience FE is not included since the current year and cohort with education level give the potential experience directly)

$$ln(y_i) = \beta_0 + \beta_1 \times college_i + \delta_{cohort} + \eta_{state} + \varepsilon_i$$

Note that the increasing trend of skill premium has been slowed down after the 2000s, especially for men. What does the aggregate labor supply look like? Figure 7 shows there is a constant rate of growth in the relative supply of college graduates to high school graduates. That leads to a conjecture that there is less skill-biased technological change after the 2000s based on the canonical model. We will not go deeply into this discussion but take coefficients



(a) Male (b) Female

Figure 3.6: College/High Graduate Wage Premium

Notes: In this figure, we plot the secular trend of college premium by genders using CPS sample from 1980 to 2018. The attrition process is described in section 2.

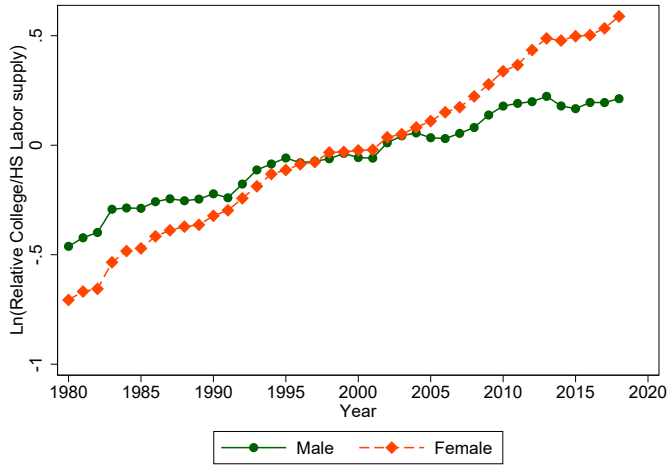


Figure 3.7: Growth Rate of Relative College/High Labor Supply

Notes: In this figure, we plot the relative labor supply by gender using the CPS sample from 1980 to 2018. The attrition process is described in section 2. We compute the relative labor share as college equivalent labor supply in hours divided by high school equivalent labor supply in hours. We treat individuals with some college experience as half high school graduates and half college graduates. We treat individuals with less than high school education experience as high school graduates and individuals with post-college degrees as college graduates, normalized by average wages, separately.

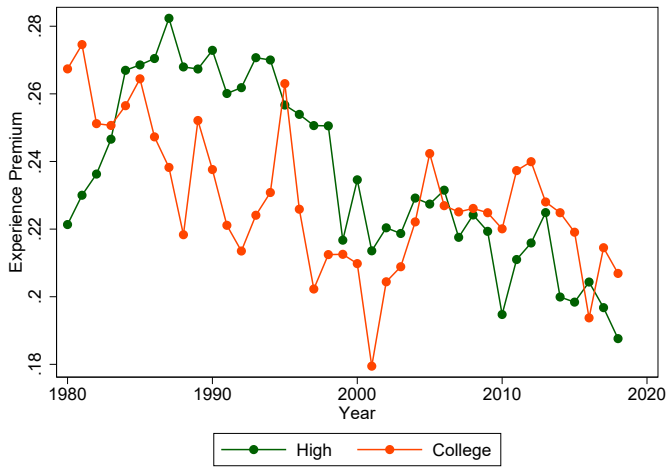


Figure 3.8: Secular Trend of Experience Premium

Notes: In this figure, we plot the secular trend of experience premium by educational attainment using CPS male sample from 1980 to 2018. The experience premium is defined by the wage gap in logs between an individual with 10 years of experience and 20 years of experience with the same education.

from regressions as prices in calibrating the deterministic part in the income process.

Experience Premium

Similarly, we also get coefficients in determining experience profiles from similar regressions. I regress log wage on a quadratic form of potential experiences by years and education groups. In Figure 8, I predict the wage gap between an individual with 10 years of experience and 20 years of experience by education groups. The experience premium increased in the 1980s and decreased afterward for low skill labor while it dropped continuously for high-skill labor from the 1980s. Our conclusion is contradicted with Heathcote et al. (2010a) since they do not take into account the composition effect of educational attainment. In 2010, there is a larger proportion of educated experienced labor than in 1980 because of the expansion in college education.

To sum up, we obtain four sets of parameters governing the income process for high and low-skill labor, around 1980 and 2010 steady states, respectively. In the "2010" steady-state, high-skill labor has a larger skill premium and faces larger idiosyncratic shocks compared

with low-skill labor relative to "the 1980" steady-state. Within the stochastic component, the relative gaps of variance in initial distribution, and the variance of transitory and permanent shocks have all been widening between the two education groups. In the rest of this paper, we will plug in those parameters and calibrate our quantitative model into the two steady states. We will then compute model predictions and compare them with empirical counterparts.

3.5 Model

My core question is to examine the extent to which differentials in income dynamics account for differentials in aggregate macroeconomic moments in wages and labor supply. To study this problem, a standard life cycle incomplete market heterogeneous agent model (Aiyagari (1994); Huggett (1993); Huggett (1996)) becomes a natural candidate for two reasons. First, in a standard incomplete market model where individuals only face idiosyncratic income shocks, the parameters we estimate regulate behaviors in labor supply and consumption directly. Second, it is straightforward to generate macroeconomic moments from simulated distributions for comparisons. In my model, there are two types of labor H and L facing different sets of idiosyncratic shocks and wage prices in population. There are two ways for individuals to insure themselves. Individuals can trade one-period risk-free bonds to smooth consumption and labor supply is endogenous.

3.5.1 Demographics

Time is discrete and continues forever. There is a continuum of individuals with measure 1 for the whole population. Age is indexed by j and works for 36 periods from $J = 0, 1, 2, 3, ...35$, which is equivalent to age 25 to 60 in the data. Then the agents retire for 20 periods and die with certainty at age $J = 55$ and a new cohort will enter at the same time. Hence the population structure is stationary and we calibrate the share of high and low-skill labor

externally.

3.5.2 Income process

Once they are born, they draw an initial realization of an idiosyncratic permanent component from $N(0, \sigma^e_{z0})$. The stochastic process is modeled as a unit root AR1 plus an i.i.d. transitory shocks:

$$y_{ij} = z_{ij} + \epsilon_{ij}$$

$$z_{ij} = z_{i,j-1} + \eta_{ij}$$

$$and \quad \eta_{ij} \sim N(0, \sigma_\eta) \quad z_{i0} \sim N(0, \sigma_{z0}) \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon)$$

The hourly wage for individual i at age j with education e is

$$w^e_{ij} = p^e \times exp(L^e(j) + y_{ij})$$

where $L^e(j)$ is the education specific experience profile and we call $exp(L^e(j) + y_{ij})$ as efficiency units $h^e(j, y_{ij})$

3.5.3 Production

The consumption good is produced by a representative firm. The functional form of production function is standard: $Y = AK^\alpha_t N^{\alpha-1}_t$ where K is aggregate capital and N is a CES aggregation of high skill and low skill labor with the elasticity of substitution θ, following Katz and Murphy (1992):

$$N = [\lambda H^{\frac{\theta-1}{\theta}} + (1-\lambda)L^{\frac{\theta-1}{\theta}}]^{\frac{\theta}{\theta-1}}$$

3.5.4 Household Problem

Agents make decisions on how much time to work and how much money to consume during working periods. Agents receive a fixed amount of benefit b when they retire and simply decide how much savings they want to carry to the next period. Working Period:

$$V^e(j,a,z,\epsilon) = \max_{c,a',n} u(c,n) + \beta E[V^e(j+1,a',z',\epsilon')] \;\; subject\ to$$

$$c + a' = (1+r)a + p^e h^e(j,y_j)n(1-\tau)$$

$$a' \geq \underline{a},\ c \geq 0,\ n \in [0,1]\ and\ a_0 = 0$$

Retirement Period:

$$V(j,a) = \max_{c,au(c,n)} +\beta E[V(j+1,a')] \;\; subject\ to$$

$$c + a' = (1+r)a + b$$

$$a' \geq \underline{a},\ c \geq 0$$

We set the functional form of utility function to be as follows where γ is risk aversion, σ is the inverse of Frisch elasticity and ψ is disutility of labor.

$$u(c_t, n_t) = \frac{c^{1-\gamma}}{1-\gamma} - \psi\frac{n^{1+\sigma}}{1+\sigma}$$

3.5.5 Definition of Stationary RCE

Let B_A, B_z and B_ϵ be the Borel sigma algebras of A, Z and ϵ. And P_E and P_J be the power sets of E and J. The state space is denoted by $S \equiv E \times J \times A \times \varepsilon \times Z$. Let Σ_S be the sigma algebra on S and (S, Σ_S) the corresponding measurable space. Denote the stationary

distribution as λ^*. Transition function Q is defined as:

$$Q((e,j,a,\epsilon,z),E\times J\times A\times\varepsilon\times Z)=I\{e\in E\}I\{j+1\in J\}I\{a(s)\in A\}Pr(\epsilon\in\varepsilon)\sum_{z'\in Z}\pi(z',z)$$

Given stationary distribution, a competitive equilibrium is a value function $V(s)$; decision rules $c(s), a_{t+1}(s), n_(s)$; firm choices H, L, K; prices r, p^e and transition function Q such that:

1. Given prices, the policy functions $c(s), a'(s), n(s)$ solve the household's problem and $V(s)$ is the associated value function.

2. Given prices, the firm chooses optimally its capital K and its labor H and L.

3. Labor market clears. Capital market clears. Government budget balances.

$$L=\int_{l\times J\times A\times\varepsilon\times Z}\gamma^l(j,y_{ij})n(s)d\lambda^*$$

$$H=\int_{h\times J\times A\times\varepsilon\times Z}\gamma^h(j,y_{ij})n(s)d\lambda^*$$

$$K=\int_{E\times J\times A\times\varepsilon\times Z}a_{t+1}(s)d\lambda^*$$

$$L\tau=b\times L_{retirement}$$

4. for all $(E\times J\times A\times\varepsilon\times Z)\in\Sigma_S$, the invariant probability measure λ^* satisfies

$$\lambda^*(E\times J\times A\times\varepsilon\times Z)=\int_{E\times J\times A\times\varepsilon\times Z}Q((e,j,a,\epsilon,z),E\times J\times A\times\varepsilon\times Z)d\lambda^*$$

3.6 Quantitative Results

Using a quantitative model, we are interested in the extent to which our model predictions be able to match their empirical counterparts. We target to match three set of moments to the empirical data: the correlation between wage and hours, the variance of hours and the

variance of earnings. For each set of moments, we have four empirical points to match 1980 and 2010 steady states for high skilled and low skilled labor. We are interested in examining if our quantitative model can reproduce both the level and trend for the two types of labor in the three moments. We also explore how do changes in the variance of unobserved initial dispersion, the variance of permanent shocks, the variance of transitory shocks and level of the borrowing limit affect the levels of three moments respectively to clarify the mechanisms.

3.6.1 Calibration

We need to calibrate and parameterize our model before solving it. We focus on partial equilibrium in the current stage since adding the full set of production and government sector does not provide more information in generating the moments we target here and computationally it is much more complicated. Essentially, solving the general equilibrium model would give us three endogenously decided prices: interest rate and wage rate for high and low-skilled labor. I set risk-free interest rate externally to be 3%, a typical number in the literature. I set the wage rate according to coefficients on education return from Mincer regressions using CPS in 1980 and 2010 for two steady states, respectively. If we do not set wage rates externally, we have to choose parameters in the production function but to match those numbers again. Hence, we think the moments generated from our partial equilibrium model should be very close to those from the general equilibrium model. In our life-cycle model, we also set deterministic age profiles according to corresponding coefficients from the same Mincer regressions. In terms of the stochastic component of the wage process, we simply plug in the four sets of parameters that we estimate from residual dynamics in section 3.

The level of risk aversion is one of the key parameters in explaining our empirical targets. It decides the level of correlation between wages and hours since it governs the relative strength between income and substitution effect. We set it to be 2, among the most accepted

ranges in the literature. In both our heterogeneous agent model and empirical targets, we target to examine the intensive margin in the variability of endogenous labor supply rather than the aggregate labor supply elasticity. Therefore, we choose to set labor supply elasticity to 0.5, a relatively small estimate from microdata. (Keane and Rogerson (2012)). The discounting factor is calibrated to match the ratio of median wealth to median labor income since the whole class of standard incomplete market models is not going to match the top distribution very well. The disutility of labor is calibrated to match the average market hours for males. The constant benefit level of the retirement period is going to match one-third of the average earning in working periods for each type of labor[5]. In our benchmark model, the borrowing limit is zero and agents can not borrow in principle. In the extended model, we allow for some degree of borrowing, setting a borrowing limit such that the fraction of agents with negative or zero wealth is 15%.

To solve the model quantitatively, we set 25 grids for assets, 25 grids for permanent components and 2 grids for transitory components and discretize the AR(1) process using the Tauchen method. Agents work for 36 periods (from age 25 to 60) and retire for 20 periods. Hence, there are 70k grids in the aggregate state space. After solving for policy functions at those grids using the EGM algorithm, we simulate the economy for 100K individuals and then generate the moments we choose to compare with empirical moments. In the rest of this section, I show model predictions in 3 models with some variations from Table 4 to Table 6. In column 3, the model does not contain the retirement period. In column 4, the model contains a retirement period and basically, we set the borrowing limit to be zero. In column 5, we set a negative borrowing limit to match the fraction of households with zero or negative assets in the data.

[5]In a general equilibrium solution, the retirement period benefit level is decided endogenously to satisfy the government budget constraint. In practice, the benefit level is pinned down given a predetermined tax rate when we guess a particular number of aggregate labor supply. Here, we solve the model in a partial equilibrium framework. To put it simply, we calibrate the benefit level in a reduced form way.

Table 3.3: Calibration Parameters

Parameters	Target/Resources	Value
γ	micro-estimates of intertemporal elasticity of substitution	2
σ	inverse of labor supply elasticity	2
r	risk-free interest rate	0.03
L_j^e	male education specific wage life-cycle profile (CPS)	see figure 8
p^e	male college premium (CPS)	see figure 6
$\sigma_{\eta 0}, \sigma_\eta, \sigma_\epsilon$	male residuals dynamics (PSID)	see table 2
β	ratio of median wealth to median labor income	0.965
$\underline{a}$	no borrowing/15% with negative or zero wealth	0/-2
ψ	average hours	2.5
b	0.33 average earnings	3.1, 4.0, 2.8, 4.2

3.6.2 Correlation between Log wage and Log hour



(a) CPS

(b) PSID

Figure 3.9: Correlation between Log (wage) and Log (hour)

Notes: In this figure, we plot the secular trend of correlation between log wages and log hours for a male using CPS and PSID samples from 1980 to 2018. The attrition process is described in section 2.

Figure 9 shows the secular trend for correlation between Log(wage) and Log(hour) using CPS and PSID, respectively. We compute empirical moments from two datasets to take full advantage of them. CPS has many more observations and could be more precise and representative to serve as a benchmark while we estimate income dynamics directly from

Table 3.4: Correlation between Log (wage) and Log (hour)

	(1) CPS	(2) PSID	(3) W/O R	(4) W/ R; $\underline{a} = 0$	(5) W/ R; $\underline{a} = -2$
"1980" Low Skilled Labor	-0.142	-0.152	-0.588	-0.253	-0.208
"1980" High Skilled Labor	-0.169	-0.162	-0.638	-0.243	-0.193
"2010" Low Skilled Labor	-0.040	-0.083	-0.487	-0.130	-0.070
"2010" High Skilled Labor	-0.059	-0.073	-0.454	-0.153	-0.128

Notes: In this table, we tabulate the empirical and model-predicted correlation between wages and hours from different data sources and various models. We compute empirical moments using CPS and PSID in columns 1 and 2. I compute the numbers in the 1980 steady state by averaging moments from 1980 to 1984 and 2010 steady states by averaging moments from 2008 to 2012. This is consistent with time coverage table 1. From columns 3 to 5, we compute model predicted moments by simulation. In column 3, there is no retirement period in the model. In column 4, there is a retirement period but with zero borrowing constraints. In column 5, there is a negative borrowing constraint.

PSID, so empirical moments computed from PSID should be closer to our model predictions in some sense. Both data from CPS and PSID show a shoulder-by-shoulder rising trend in the correlation between wages and hours for two types of labor. In Table 4, we compute numbers for 1980 steady states by averaging from 1980 to 1984 and 2010 steady-state by averaging from 2008 to 2012.

In Column (1) of Table 4, the correlation between log wage and log hour increased from -0.142 to -0.040 from the 1980-period to the 2010-period for low-skilled labor, according to CPS. In Column (4), the correlation between log wage and log hour increases from -0.253 to -0.130 in the model with retirement and no-borrowing. The comparable correlation for high-skilled labor increased from -0.169 to -0.059 in CPS while from -0.243 to -0.130 in the model. Although the model-generated correlation is generally lower[6], the magnitudes of increases in the correlation between log wage and log hour are comparable between data and model. The model performs better by including retirement from columns (3) to (4) and by allowing some borrowing from (4) to (5) in terms of matching the level. The related empirical moments

[6]One can always adjust this correlation to a higher level by choosing a smaller risk aversion, then the income effect would be smaller relative to the substitution effect.

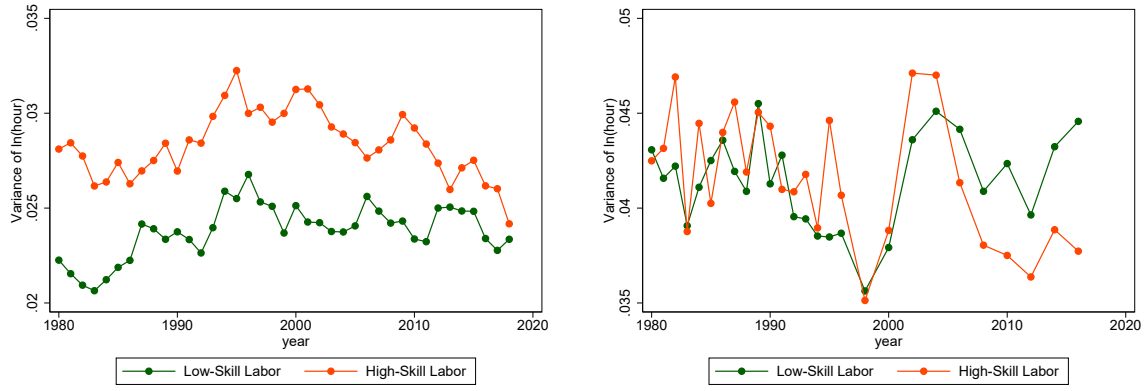
are comparable in levels and trends between the CPS and PSID samples.

This correlation is mostly sensitive to the size of initial dispersion, permanent shocks, transitory shocks, and the borrowing limit. Our decomposition shows that in our model, a larger transitory shock leads to an increase in the correlation between wages and hours. In contrast, an increase in permanent shocks decreases the correlation between wages and hours. To discuss that, we first take one step back to discuss insurable idiosyncratic shocks and uninsurable idiosyncratic shocks. Insurable shocks do not affect consumption but positively affect labor supply. Uninsurable shocks positively affect consumption, and the degree of risk aversion regulates how uninsurable shocks affect labor supply. If the income effect dominates the substitution effect, individuals will choose to work less after positive, uninsurable shocks. Therefore, an increase in insurable shocks would raise the correlation between wage and hour. In contrast, an increase in uninsurable shocks would reduce this correlation when risk aversion is greater than one. Then the question is how much idiosyncratic shocks are insurable? In standard incomplete market models, individuals can only trade a one-period risk-free non-contingent bond. Hence, no shocks are perfectly insured or fully uninsured. It is well documented that transitory shocks are highly insurable (with an insurance coefficient of 0.95), and permanent or highly persistent components are not very insurable. (Kaplan and Violante (2010); Blundell et al. (2008); Blundell and Preston (1998)) Therefore, an increase in transitory shocks is very likely to increase the correlation between wage and hour, while an increase in permanent shocks is likely to decrease the correlation between wage and hour. The latter link[7] would become looser and even could reverse if relative risk aversion becomes smaller because the effect of uninsurable shocks would become smaller as substitution effect canceling out a large part of income effect. Relatively speaking, high-skill

[7]Decomposition exercise in Heathcote et al. (2010b) shows that an increasing persistent shock has a small positive effect on this correlation since they choose a relatively smaller risk aversion, and persistent shocks could be more insurable than permanent shocks. But they end up with a much higher level in this correlation compared with data through their calibration, this could be a signal that they chose a small number for risk aversion.

labor has experienced more growth in both transitory and permanent shocks than low-skill labor from 1980 steady-state to 2010 steady-state. Therefore, the two effects cancel out each other, leading to a rising trend at the same pace. From column 5, it is easy to notice that the level of this correlation increases in all cases after we set a loose borrowing limit. For agents in any group, a looser borrowing limit would increase insurance coefficients for both permanent and transitory shocks, raising the correlation between wages and hours. In particular, poorer groups will benefit more from this loose borrowing limit since they are more likely to be around the borrowing constraint. We set the uniform borrowing constraint for all types of labor. If we set a looser borrowing constraint for high-skilled labor, the level of correlation for two types of labor would be closer, mimicking the empirical moments in a better way. We do not allow the borrowing limit to vary from the 1980 steady-state to the 2010 steady-state without additional empirical support, but a more complete financial market or larger borrowing limit would increase the correlation between hours and wages even if there is no change in variance of idiosyncratic shocks. Our model confirms this point quantitatively.

3.6.3 Variance of Log (hour)



(a) CPS (b) PSID

Figure 3.10: Variance of Log (hour)

Notes: In this figure, we plot the secular trend of variance of log hours for males using CPS and PSID samples from 1980 to 2018. The attrition process is described in section 2.

Table 3.5: Variance of Log (hour)

	(1) CPS	(2) PSID	(3) W/O R	(4) W/ R; $\underline{a} = 0$	(5) W/ R; $\underline{a} = -2$
"1980" Low Skilled Labor	0.021	0.041	0.024	0.017	0.017
"1980" High Skilled Labor	0.027	0.043	0.031	0.018	0.017
"2010" Low Skilled Labor	0.024	0.045	0.027	0.020	0.020
"2010" High Skilled Labor	0.028	0.044	0.041	0.027	0.026

Notes: In this table, we tabulate the empirical and model-predicted variance of log hours from different data sources and various models. See notes in Table 4 for details.

The second empirical target we examine is the variability of labor supply. We explore the intensive margin, so we restrict our sample to workers who work at least 1400 market hours annually(40 weeks times 35 hours per week). The empirical trend for variance of labor supply is relatively stable from 1980 to 2010, roughly speaking. In detail, no clear trend

is consistently supported by both CPS and PSID. Empirical moments generated from CPS show that there was an increasing trend from 1980 to 2000, and then the trend reversed. From Column 1 of Table 5, the 1980 steady-state to the 2010 steady-state, the log(hour) variance rises from 0.021 to 0.024 for low-skilled labor and from 0.027 to 0.028 for high-skilled labor, according to the CPS sample. In Column 2 of Table 5, the variance of labor supply computed using PSID is much larger than the size of using CPS. In the same period, low-skilled labor experienced a 0.004 growth in the variance of labor supply, and high-skilled labor experienced a growth of 0.001. This is generally consistent with what CPS suggests.

Back to our quantitative model, a larger level of initial dispersion, the variance of permanent shocks, and transitory shocks lead to a larger dispersion in labor supply. Hence, qualitatively, it is straightforward to understand that high-skill labor should have a higher level of labor supply variability since they have larger initial dispersion and higher variance of both types of idiosyncratic shocks. Meanwhile, high-skilled labor should experience a larger growth in variability in labor supply since the gaps of initial dispersion and the variance of permanent and transitory shocks between high and low-skilled labor have been enlarged. Quantitative results from column 3 to column 5 in Table 5 confirm these predictions. Our predicted level generally reproduces what we observe in CPS.[8]. In Column 4 of Table 5, the model without borrowing possibility predicts a growth in the variance of labor supply from 0.017 to 0.020 for low-skilled labor and from 0.018 to 0.027 for high-skilled labor. Allowing some degree of borrowing does not change the results by much, as shown in Column 5.

To decompose the aggregate effect, I find that larger initial dispersion and larger permanent and transitory shocks increase the variance in labor supply. Higher benefit levels, flatter experience profiles, and looser borrowing constraints reduce the labor supply variance.

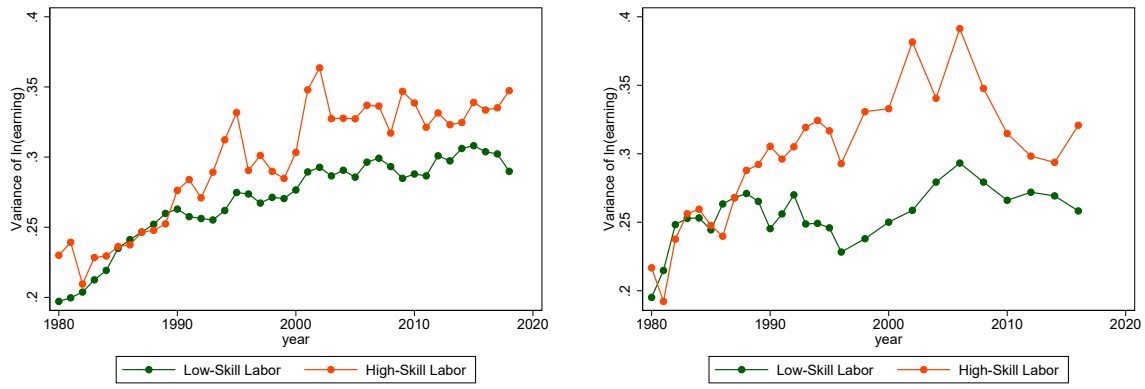
Regarding the levels, the model-generated moments are closer to the data counterparts

[8]The empirical counterpart of the variance of labor supply highly depends on how we restrict our sample. Here, we restrict our sample to "full-time" workers with working weeks over 40 and hours more than 35 per week, following Autor et al. (2008)

provided by CPS. First, the model-generated moments successfully reproduce the rising trend in the variance of hours for both high-skilled and low-skilled groups. However, though the model suggests that high-skilled labor should experience a higher growth in the variance of hours, the data suggests the opposite.

Two arguments could help to alleviate the results. First, high-skilled labor could have a higher fraction of retirement benefit level than low-skilled labor, for which we set the same here. Second, just as what we show in column 5 in table 5, a looser borrowing limit also reduces the variance of labor supply. Since saving and elastic labor supply are the two ways to insure themselves in our model, the agent could adjust the labor supply less elastically as one channel strengthens. If high-skilled labor has a looser borrowing constraint than low-skilled labor, they tend to have a smaller variance of hours.

3.6.4 Variance of Log (earning)



(a) CPS (b) PSID

Figure 3.11: Variance of Log (earning)

Notes: In this figure, we plot the secular trend of variance of log earnings for male using CPS and PSID samples from 1980 to 2018. The attrition process is described in section 2.

Table 3.6: Variance of Log (earning)

	(1) CPS	(2) PSID	(3) W/O R	(4) W/ R; $\underline{a} = 0$	(5) W/ R; $\underline{a} = -2$
"1980" Low Skilled Labor	0.214	0.226	0.210	0.249	0.257
"1980" High Skilled Labor	0.219	0.236	0.268	0.263	0.267
"2010" Low Skilled Labor	0.280	0.243	0.240	0.268	0.276
"2010" High Skilled Labor	0.311	0.348	0.353	0.357	0.362

Notes: In this table, we tabulate the empirical and model predicted variance of log earnings from different data sources and various models. See notes in table 4 for details.

The third moment we look for is the variance of earnings. Since earning is a product of wage and hour, simple variance decomposition tells us the aggregate growth in earnings inequality can be divided into three parts. First, the changes in wage inequality are captured in our wage process estimation. Second, the changes in labor supply variability that we just discussed in the previous section. Third, the changes in covariance between hours and wages are discussed in section 6.2.

In Figure 11, both CPS and PSID reflect a growing secular trend for earnings inequality while high-skilled labor increases by more. The difference is that high-skilled labor increases more, and low-skilled labor increases less in PSID regarding earnings inequality. From the 1980 steady-state to the 2010 steady-state, the log(earning) variance rises from 0.214 to 0.280 for low-skilled labor and 0.219 to 0.311 for high-skilled labor, according to the CPS sample in Column 1 of Table 6. The PSID sample suggests similar growth for high-skilled labor and smaller growth for low-skilled labor.

Regarding model-generated moments, the quantitative model predicts a rise in the earnings variance from 0.249 to 0.268 for low-skilled labor in Column 4 of Table 6, a number close to what PSID suggests from 0.226 to 0.243 according to Column 2 of Table 6. It is natural because we estimate the wage process directly from wage residuals in PSID. The growth is

relatively small because the decline in the variance of permanent shocks offsets a positive contribution from larger initial dispersion and transitory shocks. The model also predicts a growth in the variance of earnings from 0.263 to 0.357 for high-skilled labor in Column 4, a number close to what is suggested by CPS as well as PSID in Column 1 and Column 2 of Table 6. As Table 6 shows, it is unsurprising that our model also matches the level in four steady states very well.

Hence, our model-generated moments successfully match the empirical counterparts in level and the trend we observe in the data for correlation between wages and hours and variability in labor supply and earning. Essentially, larger initial dispersion, permanent shocks, and transitory shocks contribute to larger variability in labor supply and earnings, so we see highly skilled labor experiences a larger growth in those two moments in both model predictions and data. Meanwhile, increasing permanent shocks has a negative effect on the correlation between wages and hours. Since the gaps between high-skilled and low-skilled labor in the variance of permanent and transitory shocks have all been widening, the two effects cancel out each other. Therefore, we see high and low-skilled labor experience the same level of growth in the correlation between wages and hours.

3.6.5 Implications on Aggregate Moments

Our objective is to assess the quantitative predictions of a standard incomplete market model focusing on labor market outcomes, leveraging the fact that the idiosyncratic shock processes have changed differently across skill groups. In this subsection, we try to argue that heterogeneity in idiosyncratic shock processes across skill groups has important quantitative implications on aggregate moments, suggesting the need to consider this heterogeneity in future studies.

To fulfill this goal, we first assume the two education groups have the same idiosyncratic shock process and estimate the two sets of parameters that govern income processes in 1980

Table 3.7: Implications on Aggregate Moments

	(1) Var(Asset)	(2) $\frac{Mean(Asset_{high\ \ skill})}{Mean(Asset_{low\ \ skill})}$	(3) Var(wage)	(4) Mean(Asset)
"1980 S.S."	-2.7%	-1.2%	-4.7%	0.0%
"2010 S.S."	10.8%	10.7%	16.6%	1.7%

Notes: This table documents the deviation of a set of simulated moments of a model with heterogeneous idiosyncratic shock processes from the simulated moments of a model with homogeneous idiosyncratic profile across education groups.

S.S. and 2010 S.S. separately, following the same procedure in section 4.2. Then, we simulate a set of key moments as the benchmark results. Next, we simulate the same moments from models where we assume the idiosyncratic shock process differs across two education groups and calculate the deviation from the benchmark case.

Not surprisingly, the deviation from the benchmark results in 1980 S.S. is small because the idiosyncratic shock processes are not very different between low-skill and high-skill groups in 1980. In contrast, in 2010, the model allowing heterogeneous idiosyncratic shock processes generates significantly larger aggregate dispersion in wages and assets. In particular, the high-skill group has more precautionary saving motives compared with the benchmark case, and therefore, the first and second moments of assets go up. These results justify the significance of considering this heterogeneity and helping generate larger inequality, given that the standard incomplete market model makes it hard to generate enough inequality compared with data on both levels and trends.

3.7 Conclusion

Wage patterns have been changing dramatically in the US since 1980. Most of previous studies focus on the heterogeneity in the deterministic component by education groups and the aggregate trend in the idiosyncratic component. This paper provides a new dimension to examine the changing wage pattern: the heterogeneity in the stochastic component, namely

wage residual inequality, by education groups and its macroeconomic implications. First, using data from CPS and PSID, I find substantial heterogeneity in the secular trend of wage residual inequality by education groups. High-skilled labor experiences larger growth in residual variance. Second, under a standard income process statistical model with an AR(1) unit root permanent component and an i.i.d. transitory component, I decompose the residual inequality into variances of initial dispersion, permanent shocks, and transitory shocks. From 1980 to 2010, high-skilled labor experienced larger growth in all of the three parts compared to low-skilled labor. Third, I examine the extent to which the heterogeneity in income processes can explain the heterogeneity in empirical moments. In a standard life-cycle heterogeneous agent incomplete market model, I find moments generated from the model successfully match their empirical counterparts in levels as well as trends we observe in the data for the correlation between wages and hours, variance in labor supply, and variance in earnings.

To summarize, this paper documents a substantial heterogeneity in wage residual inequality by education groups and translates that into heterogeneity in income process through wage process estimation. Our quantitative model confirms that this documented heterogeneity is at the center of explaining various empirical moments in wages and labor supply. At least two themes related to our research can be investigated in future research. First, we want to understand what drives the dispersion in residual inequality by education groups. Answers to this question can illuminate what accounts for rising residual inequality in the aggregate, given rising residual equality accounts for most of the recent growth in total wage inequality. Specifically, examining heterogeneous returns to various majors is potentially helpful to explain the increase in the variance of initial dispersion for high-skilled labor, and looking at the job turnover rate by education groups could be useful to provide empirical evidence in explaining the increasing trends for the variance of transitory shocks. Second, we have shown how powerful our framework is in explaining macroeconomic moments related

to labor supply decisions, it would also be interesting to investigate the heterogeneity in saving decisions and the various pass-through under fiscal and monetary policies by education groups.

Appendix A

Appendix to Chapter 1

A.1 Data Construction

This subsection describes the firm-level variables used in the empirical analysis of the paper, based on China Annual Survey of Industrial Firms.

A.1.1 Summary Statistics

Table A.1: Average of Key Variables by Firm Size

Size	Employment	Fixed Assets	Output	Exit Rate	$\frac{\text{Debt}}{\text{Assets}}$	Borrowing rate
1	28.28	5.16	16.79	0.17	0.542	0.047
2	63.40	6.21	22.31	0.12	0.552	0.043
3	110.52	9.82	31.44	0.10	0.571	0.040
4	205.52	18.43	50.34	0.09	0.587	0.037
5	1039.25	154.67	288.01	0.07	0.606	0.031

Notes: We calculate the average number of several key variables by 5 groups indexed by employment size. We include all the firms from 1998 to 2007. The unit for fixed assets and output is million RMB.

Sector Level Capital Recovery Rate The sector-level capital recovery rate is taken from Kermani and Ma (2023). I match the US 2-digit sector and China 2-digit sector within the manufacturing sector. Since I focus on fixed capital, the recovery rate I use is the type of PPE.

Figure A.1: Sector Level Capital Recovery Rate

SIC2	PPE	Inventory	Receivable	Intangible	Intan (ex. goodwill)
10 Metal Mining	0.24	0.44	0.44	0.02	0.03
12 Coal Mining	0.22	0.56	0.79	0.36	0.62
13 Oil/Gas Extraction	0.47	0.41	0.73	0.23	0.63
14 Quarrying-Nonmetals	0.45	0.50	0.73	0.00	0.00
15 Building Construction	0.28	0.32	0.65	0.00	0.00
17 Construction Contractors	0.37	0.20	0.29	0.00	0.00
20 Food Products	0.37	0.46	0.74	0.09	0.19
22 Textile Products	0.43	0.52	0.64	0.25	0.71
23 Apparel Products	0.19	0.47	0.66	1.09	2.15
24 Wood Products	0.58	0.32	0.79	0.04	0.20
25 Furniture and Fixtures	0.33	0.35	0.71	0.06	0.18
26 Paper Products	0.33	0.59	0.64	0.05	0.14
27 Printing and Publishing	0.31	0.34	0.63	0.14	0.35
28 Chemical Products	0.25	0.50	0.62	0.38	0.59
30 Rubber and Plastics Products	0.45	0.48	0.63	0.10	0.25
32 Stone, Clay, Glass, and Concrete	0.42	0.43	0.71	0.00	0.00
33 Primary Metal	0.44	0.62	0.66	0.24	0.60
34 Fabricated Metal Products	0.38	0.46	0.69	0.02	0.04
35 Machinery	0.40	0.45	0.48	0.05	0.13
36 Electronic Equipment	0.28	0.29	0.62	0.50	0.99
37 Transportation Equipment	0.36	0.59	0.61	0.21	0.56
38 Analytical Instruments	0.40	0.37	0.89	0.32	0.63
39 Misc. Manufacturing	0.33	0.67	0.67	0.28	0.58

A.2 More on Empirical Results

The below table tests the correlation between the tightness of financing constraints and MRPK as well as MRPL.

Table A.2: The Correlation between Tightness of Financing Constraint and Marginal Revenue of Inputs

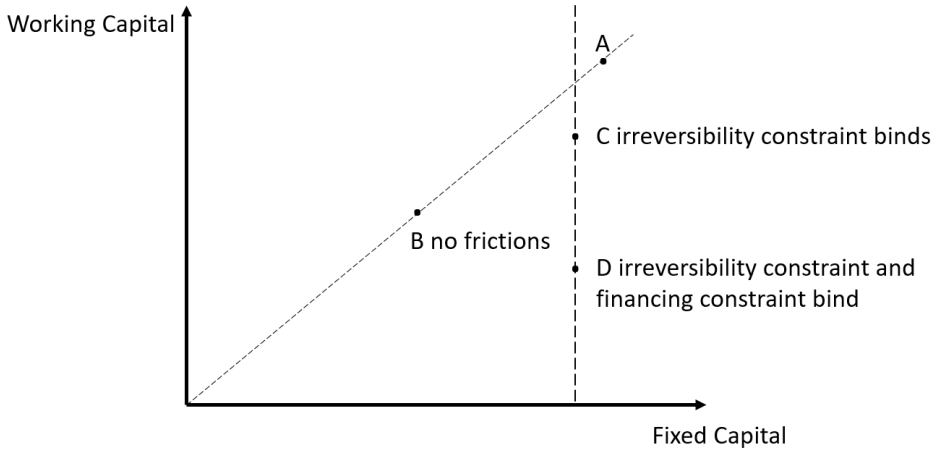
	(1) Log(MRPK)	(2) Log(MRPK)	(3) Log(MRPL)	(4) Log(MRPL)
Log(Debt/Asset)	0.054***		0.028***	
	(0.002)		(0.001)	
D(Investment=0)× Log(Debt/Asset)	**-0.074****		0.011(***)	
	(0.002)		(0.001)	
Log (Borrowing Rate)		0.033***		0.035***
		(0.001)		(0.001)
D(Investment=0) × Log(Borrowing Rate)		**-0.017****		0.005***
		(0.001)		(0.001)
Constant	0.160***	0.072***	1.502***	1.538***
	(0.001)	(0.004)	(0.001)	(0.001)
Year × Sector FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	2875653	2140368	2875653	2140368
R-Square	0.839	0.855	0.804	0.825

Notes: Results from estimating (1.32). α_i is a firm i fixed effect, α_{st} is a year-two digit sector fixed effect, $\mathbb{I}\{i_t = 0\}$ is an indicator specifies whether one makes a positive adjustment in fixed capital or not. ARPK y_{it} is defined as value-added over fixed capital to measure MRPK. The tightness of financing constraint is proxied by two variables $x_{it} \in \{\text{leverage ratio}, \text{borrowing rate}\}$. The leverage ratio is defined as total liabilities over total assets, and the borrowing rate is defined as total interest payment over total liabilities. Data source: Chinese Annual Survey of Industrial Firms, 1998-2007. Standard errors are clustered at the firm level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.3 More on Theoretical Results

A.3.1 Toy Example

Figure A.2: Toy Example of Investment Decisions



This figure illustrates the investment decisions one firm makes in different scenarios. Suppose the firm uses fixed capital and working capital to produce final goods while the elasticity of substitution between fixed capital and working capital is one. When the productivity is high, the firm will choose the inputs at point A. Then, a negative productivity shock hits the firm. If there are no frictions, then the firm would like to reduce the size of fixed capital and working capital instantly and simultaneously. The choice point B will be on the same line since it is a static choice.

However, if the fixed capital is partially irreversible, the firm might find it too costly to sell the fixed capital immediately because restocking fixed capital is expensive. To enjoy the option value, the firm might want to keep the fixed capital. To make the best use of the working capital, the firm would like to finance enough working capital to match with that due to the complementarity between the two inputs. Since there is no financial friction, it is still efficient, and the firm will be located at point C.

If there exists financial friction, then the borrowing constraint on working capital tends to be tighter because compared with point B: (i) no cash can be liquidated from selling fixed capital, (ii) relative funding demand for working capital is higher. If the borrowing constraint starts to bind to working capital, the firm would have too much fixed capital and too little working capital and end up with point D. The production will be very unbalanced and unprofitable. The default risk will rise in the next period. When the creditor recognizes this risk, they must charge a higher default premium to compensate for the higher default risk. This will further tighten the borrowing constraint on working capital and make production even more unbalanced and unprofitable. This self-reinforced lock pushes the firm into a trap.

A.3.2 Proof

Proof of Proposition 1 Let us consider the continuation problem. Depending on whether negative fixed capital adjustment or non-negative capital adjustment is used, it can be di-

vided into two subproblems. V_+^{cont} denotes the maximum value of the firm chooses non-negative investment $k' \geq (1-\delta)k$, and V_-^{cont} denotes the maximum value that can be reached if the firm chooses negative investment $k' < (1-\delta)k$.

$$V^{cont}(n,k,z) = \max\{V_+^{cont}, V_-^{cont}\} \tag{A.1}$$

Let us write the Bellman equation for the two sub-optimal problems.

$$V_+^{cont}(n,k,z) = \max_{d,k',b',l'} d + \mathbb{E}[\Lambda' V(n',k',z')|z] \tag{A.2}$$

Subject to:

$$d = n + F(b',k',l',z) - \xi - wl' - qk' \tag{A.3}$$

$$d \geq 0,\ k' \geq (1-\delta)k \tag{A.4}$$

$$V_-^{cont}(n,k,z) = \max_{d,k',b',l'} d + \mathbb{E}[\Lambda' V(n',k',z')|z] \tag{A.5}$$

Subject to:

$$d = n + F(b',k',l',z) - \xi - wl' - q((1-\eta)(1-\delta)k + \eta k') \tag{A.6}$$

$$d \geq 0,\ k' < (1-\delta)k \tag{A.7}$$

Then, it is useful to derive the default threshold by two sub optimal problems. Not all continuation firms' problems yield a solution since some of them can't satisfy non-negative dividend payment constraints no matter what choices are made. We can define default threshold $\underline{n}_+(z,k)$ and $\underline{n}_-(z,k)$ as the minimum starting net worth such that the maximum of dividends payment one can make is zero in the two sub-optimal problems.

It is easy to see the default threshold for upward adjustment only depends on z but not

k.

$$\underline{n}_+(z,k) \equiv -\max_{b',k',l'}\left\{F(b',k',l',z) - wl' - \xi - qk'\right\} \tag{A.8}$$

It is easy to see $\underline{n}_-(z,k)$ is increasing in k is fixed capital is partially irreversible $\eta < 1$.

$$\underline{n}_-(z,k) \equiv -\max_{b',k',l'}\left\{F(b',k',l',z) - wl' - \xi - q((1-\eta)(1-\delta)k + \eta k')\right\} \tag{A.9}$$

As a result, under normal regularity conditions, there exists $k(z)$ such that $\underline{n}_+(z,k(z)) = \underline{n}_-(z,k(z))$. While the default threshold $\underline{z}(n,k)$ is the maximum of the two sub-optional default thresholds.

$$\underline{z}(n,k) = \max\{\underline{z}_+(n,k), \underline{z}_-(n,k)\} \tag{A.10}$$

Hence, there exists a threshold $\underline{n}(k,z)$, such that the firm defaults when $n < \underline{n}(k,z)$. $\underline{n}(z,k)$ is weakly increasing in k. And there exists $\underline{k}(z)$ such that $\underline{n}(z,k) = \underline{n}(z,0)$ all for $k \leq \underline{k}(z)$. $\underline{k}(z)$ is increasing in z. $V(n,k,z) = V^{exit}$ for $n < \underline{n}(k,z)$.

Proof of Proposition 2 and Proposition 3 To see this, it is more convenient to rewrite the dynamic problem as a sequential problem. For simplicity, I assume $\pi_d = 0$. The general case that allows a positive exogenous death rate follows a similar procedure but with more complicated notations.

$$V_{i0} = \mathbb{E}_0 \sum_{t\geq 0} \Lambda_t d_{it} \tag{A.11}$$

subject to:

$$\begin{aligned} d_{it} = n_{it} + F(b_{i,t+1}, k_{i,t+1}, l_{i,t+1}, z_{it}) - \xi - wl_{i,t+1} - \\ \mathbb{I}_{k_{i,t+1}\geq(1-\delta)k_{i,t}} \cdot qk_{i,t+1} - \mathbb{I}_{k_{i,t+1}<(1-\delta)k_{it}} \cdot q((1-\eta)(1-\delta)k_{it} + \eta k_{i,t+1}) \end{aligned} \tag{A.12}$$

$$n_{it} = (A_t z_{it}\varepsilon_{it} k_{it}^{\alpha} \ell_{it}^{1-\alpha})^{1-\frac{1}{\theta}} Y_t^{\frac{1}{\theta}} + (1-\delta)q_t k_{it} - b_{it} \tag{A.13}$$

$$d_{it} \geq 0 \tag{A.14}$$

Let λ_{it} be the multiplier on the flow of cash constraint and $\phi_i t$ be the multiplier on the flow of cash constraint. First-order condition yields:

$$\lambda_{it} = 1 + \phi_{it} \geq 1 \tag{A.15}$$

Since agents are risk-neutral, the marginal value of consumption is one. λ_{it} is the marginal value of liquidity valued at period t, which is larger than one if ϕ_{it}>0. This implies the non-negative dividends constraint is binding. We define firms as financially constrained if their marginal value of liquidity is larger than 1. How does a firm value today's marginal liquidity versus tomorrow's marginal value of liquidity? The first order condition with respect to the borrowing bridges the trade of one face over the two periods.

$$\underbrace{\frac{\partial F(b,k',l',z)}{\partial b'} - \mathbb{E}[\Lambda']}_{\text{<0: price wedge in risky bond over safe bond}} = \mathbb{E}\left[\Lambda' \underbrace{\left(\frac{\frac{\partial V(n',k',z')}{\partial n'}}{\frac{\partial V^{cont}(n,k,z)}{\partial n}} - 1\right)}_{\text{<0: tightness of the financing constraint at date t}} |z\right] \tag{A.16}$$

$$\lambda_t = \mathbb{E}[\lambda_{t+1}] \times \frac{\Lambda_{t+1}}{\frac{\partial F}{\partial b_{t+1}}} \tag{A.17}$$

It is easy to see λ_{it} is a supermartingale as $\lambda_t \geq \mathbb{E}[\lambda_{t+1}]$. In economic terms, the firm accumulates wealth to save out of financial constraints constantly. Suppose a firm saves out of financial constraint at time j with $\lambda_j = 1$. It implies $\lambda_t = 1$ for $j \geq 1$ since the lower bound of the marginal value of liquidity is 1.

Then, there must exist a threshold $\overline{n}(z,k)$ such that the firm saves out of financial constraint. The definition comes from a similar fixed point problem to define the default threshold. The intuition is this threshold is the minimum starting net worth such that no matter what productivity is drawn in the next period, the net worth in the next period is still higher than the $\overline{n}(z',k'|k)$. As fixed capital is irreversible, $\overline{n}(z,k)$ is decreasing in k when

downward fixed capital adjustment maximizes the firm value.

When the firm saves out of financing constraints, the policy function will not depend on the financial position, and the state variables will collapse into two-dimensional space. Since the firm is indifferent between paying dividends from zero to the maximum payment such that the net worth after paying dividends meets the threshold. The firm value can be expressed as:

$$V^{cont}(n,k,z) = n - \overline{n}(k,z) + V^{cont}(\overline{n}(k,z),k,z) \tag{A.18}$$

Proof of Proposition 4 This proposition specifies the default zone expands in the transition path in response to a negative TFP shock.

The proof follows three steps. First, following the definition of book value of the net worth, $n'(k',\ell',b'|z)$ and $n^{e\prime}(k',\ell',b'|z)$ are increasing in A'.

$$n'(k',\ell',b') = (A'z\varepsilon k'^{\alpha}\ell'^{1-\alpha})^{1-\frac{1}{\theta}}Y^{\frac{1}{\theta}} + (1-\delta)qk' - b' \tag{A.19}$$

Second, the debt one can borrow contingent on decisions is also increasing in A' as the default probability is decreasing in A'

$$F(b',k',l',b') = \left(\Lambda \begin{Bmatrix} b' \cdot [\pi_d \mathbb{I}_{n^{e\prime}>0} + (1-\pi_d)(1-\mathbb{I}_{n'<\underline{n}'}\mathbb{I}_{n^{e\prime}<0})] + \\ \min(b', \eta q(1-\delta)k) \cdot [\pi_d \mathbb{I}_{n^{e\prime}<0} + (1-\pi_d)\mathbb{I}_{n'<\underline{n}'}\mathbb{I}_{n^{e\prime}<0}] \end{Bmatrix}\right) \tag{A.20}$$

Thirdly, following the definition of the default threshold, $\underline{n}(z,k)$ is increasing in $F(b',k',l',b')$ so it is also increasing in A'. As a result, along the transition dynamics with $\varepsilon^A_{t_0} < 0$, the default threshold is higher than in the steady state

$$\{\underline{n}_t(k,z)\}_{t=t_0}^{\infty} > \underline{n}(k,z)$$

Meanwhile, it is straightforward to see the policy function k'(n,k,z) is increasing in A'. The

default threshold for upward adjustment is decreasing in A' since $F(b', k', l', z)$ is increasing in A'.

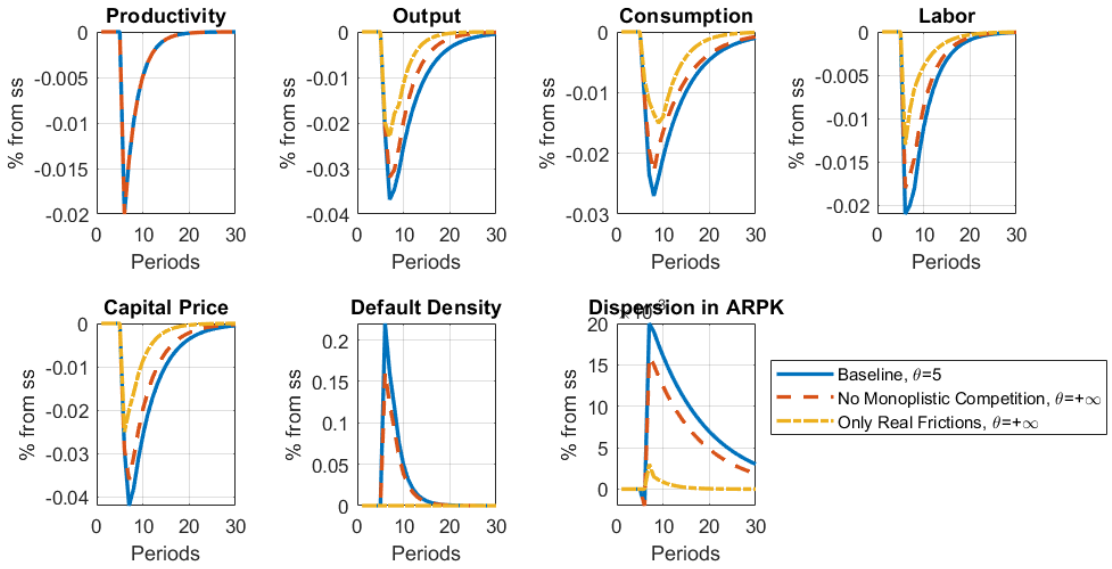
$$\underline{n}_+(z,k) \equiv -\max_{b',k',l'}\left\{F(b',k',l',z) - wl' - \xi - qk'\right\} \tag{A.21}$$

As a result, the threshold specifies the default threshold dependency on k is increasing in A' and

$$\{\underline{k}_t(z)\}_{t=t_0}^{\infty} < \underline{k}(z)$$

A.4 More on Quantitative Results

Figure A.3: The Role of Monopolistic Competition



This figure shows the role of monopolistic competition in amplifying the recession. The elasticity of substitution is 5 in the baseline results. Then we shut down the monopolistic competition by assuming the differentiated goods are perfect substitutes. The production function is assumed to be decreasing return to scale to match the coefficient in the production function estimation.

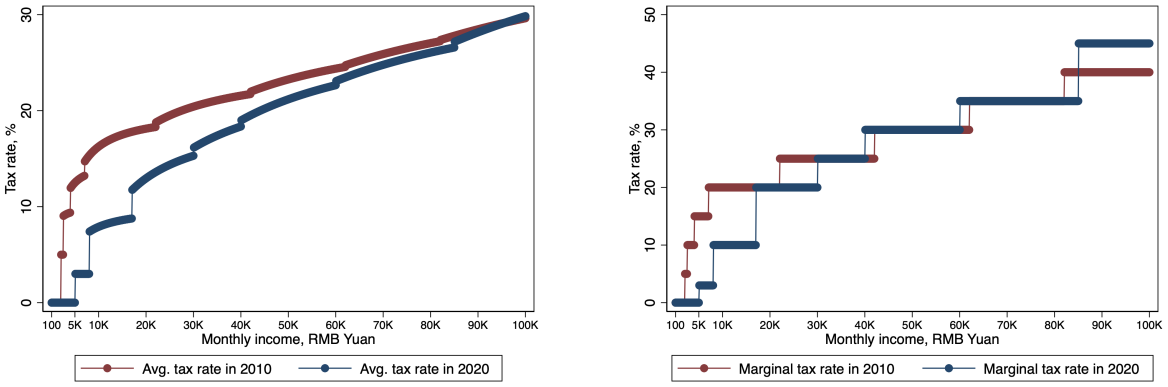
The results suggest that adding monopolistic competition accounts for around one-third of the amplification effect in terms of aggregate output. Due to the general equilibrium feedback, more default and misallocation are generated.

Appendix B

Appendix to Chapter 2

B.1 More on Empirical Results

Figure B.1: Tax Rate, 2010 vs. 2020



(a) Average Tax Rate

(b) Marginal Tax Rate

Notes: We plot these two figures according to two versions of the Personal Income Tax Law in China revised in 2007 and 2018. The law files can be found in the following link: `https://law.pkulaw.com/falv/71ad23cd41c523c2bdfb.html`.

Table B.1: Elasticity of Hours to Individual Income

	Dependent variable: ln(hours per worker)					
	2010	2012	2014	2016	2018	2020
Panel A. All Gender						
ln(hourly wage)	-0.161^a	-0.170^a	-0.166^a	-0.187^a	-0.177^a	-0.163^a
	(0.006)	(0.006)	(0.005)	(0.006)	(0.004)	(0.004)
R^2	0.151	0.158	0.162	0.207	0.192	0.189
Observations	4,588	4,614	6,771	4,786	7,470	6,615
Panel B. Male						
ln(hourly wage)	-0.175^a	-0.184^a	-0.190^a	-0.194^a	-0.197^a	-0.181^a
	(0.008)	(0.009)	(0.006)	(0.008)	(0.006)	(0.006)
R^2	0.165	0.182	0.208	0.228	0.233	0.230
Observations	2,827	2,706	4,107	2,761	4,398	3,876
Panel C. Female						
ln(hourly wage)	-0.165^a	-0.177^a	-0.170^a	-0.211^a	-0.184^a	-0.172^a
	(0.009)	(0.010)	(0.008)	(0.009)	(0.006)	(0.006)
R^2	0.170	0.168	0.173	0.240	0.209	0.212
Observations	1,761	1,908	2,664	2,025	3,072	2,739

Notes: This table reports the estimated OLS coefficients of equation **??** by regressing individual-level log hours on log wage rate. The estimation equation is

$$\log(h_i) = \alpha + \beta \log(w_i) + \delta_1 age_i + \delta_2 age_i^2 + \epsilon_i$$

where i is for individual and w is the individual wage. The sample is restricted to full-time salaried employees (who work at least 1200 hours per year) at prime ages (16-65). The dependent variable is the logarithm of individual hours worked per worker. Age and age squared are controlled in all regressions. Standard errors in parentheses are robustly clustered. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$.

Bibliography

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante, "Education policy and intergenerational transfers in equilibrium," Technical Report, National Bureau of Economic Research 2013.

Abel, Andrew B and Janice C Eberly, "A unified model of investment under uncertainty," *The American Economic Review*, 1994, *84* (5), 1369.

_ **and** _ , "Optimal investment with costly reversibility," *The Review of Economic Studies*, 1996, *63* (4), 581–593.

Abowd, John M and David Card, "On the covariance structure of earnings and hours changes," *Econometrica: Journal of the Econometric Society*, 1989, pp. 411–445.

Aguiar, Mark and Erik Hurst, "Measuring trends in leisure: The allocation of time over five decades," *The quarterly journal of economics*, 2007, *122* (3), 969–1006.

Aiyagari, S Rao, "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 1994, *109* (3), 659–684.

Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote, "Work and leisure in the United States and Europe: why so different?," *NBER macroeconomics annual*, 2005, *20*, 1–64.

Arellano, Cristina, Yan Bai, and Patrick J Kehoe, "Financial frictions and fluctuations in volatility," *Journal of Political Economy*, 2019, *127* (5), 2049–2103.

Attanasio, Orazio P, "Consumption," *Handbook of macroeconomics*, 1999, *1*, 741–812.

Autor, David H, Lawrence F Katz, and Melissa S Kearney, "Trends in US wage inequality: Revising the revisionists," *The Review of economics and statistics*, 2008, *90* (2), 300–323.

Baley, Isaac, Andrés Blanco et al., "The Macroeconomics of Partial Irreversibility," Technical Report 2022.

Baqaee, David, Emmanuel Farhi, and Kunal Sangani, "The Supply-Side Effects of Monetary Policy," *Journal of Political Economy*, 2023, *accepted.*

Baqaee, David Rezza and Emmanuel Farhi, "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 2020, *135* (1), 105–163.

Becker, Gary S, "A Theory of the Allocation of Time," *The economic journal*, 1965, *75* (299), 493–517.

Benhabib, Jess, Richard Rogerson, and Randall Wright, "Homework in macroeconomics: Household production and aggregate fluctuations," *Journal of Political economy*, 1991, *99* (6), 1166–1187.

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

Bertola, Guiseppe and Ricardo J Caballero, "Irreversibility and aggregate investment," *The Review of Economic Studies*, 1994, *61* (2), 223–246.

Bick, Alexander, Nicola Fuchs-Schündeln, and David Lagakos, "How do hours worked vary with income? Cross-country evidence and implications," *American Economic Review*, 2018, *108* (1), 170–199.

Bils, Mark, Peter J Klenow, and Cian Ruane, "Misallocation or mismeasurement?," *Journal of Monetary Economics*, 2021, *124*, S39–S56.

Blanchard, Olivier Jean and Nobuhiro Kiyotaki, "Monopolistic competition and the effects of aggregate demand," *The American Economic Review*, 1987, pp. 647–666.

Bloom, Nicholas, "The impact of uncertainty shocks," *econometrica*, 2009, *77* (3), 623–685.

_ , Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry, "Really uncertain business cycles," *Econometrica*, 2018, *86* (3), 1031–1065.

Bloom, Nick, Stephen Bond, and John Van Reenen, "Uncertainty and investment dynamics," *The review of economic studies*, 2007, *74* (2), 391–415.

Blundell, Richard and Ian Preston, "Consumption inequality and income uncertainty," *The Quarterly Journal of Economics*, 1998, *113* (2), 603–640.

_ and Thomas MaCurdy, "Labor supply: A review of alternative approaches," *Handbook of labor economics*, 1999, *3*, 1559–1695.

_ , Luigi Pistaferri, and Ian Preston, "Consumption inequality and partial insurance," *American Economic Review*, 2008, *98* (5), 1887–1921.

Bolton, Patrick, Hui Chen, and Neng Wang, "A unified theory of Tobin's q, corporate investment, financing, and risk management," *The journal of Finance*, 2011, *66* (5), 1545–1578.

Boppart, Timo and Per Krusell, "Labor supply in the past, present, and future: a balanced-growth perspective," *Journal of Political Economy*, 2020, *128* (1), 118–157.

Buera, Francisco J and Yongseok Shin, "Financial frictions and the persistence of history: A quantitative exploration," *Journal of Political Economy*, 2013, *121* (2), 221–272.

_ , Joseph P Kaboski, and Yongseok Shin, "Finance and development: A tale of two sectors," *American economic review*, 2011, *101* (5), 1964–2002.

Caggese, Andrea, "Financing constraints, irreversibility, and investment dynamics," *Journal of Monetary Economics*, 2007, *54* (7), 2102–2130.

Cao, Dan, Guido Lorenzoni, and Karl Walentin, "Financial frictions, investment, and Tobin'sq," *Journal of Monetary Economics*, 2019, *103*, 105–122.

Card, David and John E DiNardo, "Skill-biased technological change and rising wage inequality: Some problems and puzzles," *Journal of labor economics*, 2002, *20* (4), 733–783.

Chamon, Marcos, Kai Liu, and Eswar Prasad, "Income uncertainty and household savings in China," *Journal of Development Economics*, 2013, *105*, 164–177.

Chen, Zhao, Xian Jiang, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu, "Tax policy and lumpy investment behaviour: Evidence from China's VAT reform," *The Review of Economic Studies*, 2023, *90* (2), 634–674.

Cooper, Russell, John Haltiwanger, and Laura Power, "Machine replacement and the business cycle: lumps and bumps," *American Economic Review*, 1999, *89* (4), 921–946.

Cooper, Russell W and John C Haltiwanger, "On the nature of capital adjustment costs," *The Review of Economic Studies*, 2006, *73* (3), 611–633.

Cui, Wei, "Macroeconomic effects of delayed capital liquidation," *Journal of the European Economic Association*, 2022, *20* (4), 1683–1742.

Cutler, David M and Lawrence F Katz, "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's," *The American Economic Review*, 1992, *82* (2), 546–551.

Dotsey, Michael, Wenli Li, and Fang Yang, "Consumption and time use over the life cycle," *International Economic Review*, 2014, *55* (3), 665–692.

Eisfeldt, Andrea L and Adriano A Rampini, "Capital reallocation and liquidity," *Journal of monetary Economics*, 2006, *53* (3), 369–399.

Erickson, Timothy and Toni M Whited, "Measurement error and the relationship between investment and q," *Journal of political economy*, 2000, *108* (5), 1027–1057.

Farhi, Emmanuel and Iván Werning, "A theory of macroprudential policies in the presence of nominal rigidities," *Econometrica*, 2016, *84* (5), 1645–1704.

Fazzari, Steven M and Bruce C Petersen, "Working capital and fixed investment: new evidence on financing constraints," *The RAND Journal of Economics*, 1993, pp. 328–342.

Ghez, Gilbert and Gary S Becker, *The allocation of time and goods over the life cycle* number ghez75-1, National Bureau of Economic Research, 1975.

Gilchrist, Simon, Jae W Sim, and Egon Zakrajšek, "Uncertainty, financial frictions, and investment dynamics," Technical Report, National Bureau of Economic Research 2014.

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez, "Capital allocation and productivity in South Europe," *The Quarterly Journal of Economics*, 2017, *132* (4), 1915–1967.

Gottschalk, Peter and Robert Moffitt, "The rising instability of US earnings," *Journal of Economic Perspectives*, 2009, *23* (4), 3–24.

_ , _ , Lawrence F Katz, and William T Dickens, "The growth of earnings instability in the US labor market," *Brookings Papers on Economic Activity*, 1994, *1994* (2), 217–272.

Greenwood, Jeremy and Zvi Hercowitz, "The allocation of capital and time over the business cycle," *Journal of political Economy*, 1991, *99* (6), 1188–1214.

_ , _ , and Gregory W Huffman, "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, 1988, pp. 402–417.

Guntin, Rafael, "Firms' rollover risk and macroeconomic dynamics," Technical Report, Tech. rep., Working paper, New York University 2022.

Guvenen, Fatih, "An empirical investigation of labor income processes," *Review of Economic dynamics*, 2009, *12* (1), 58–79.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L Violante, "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic dynamics*, 2010, *13* (1), 15–51.

_ , Kjetil Storesletten, and Giovanni L Violante, "Quantitative macroeconomics with heterogeneous households," *Annu. Rev. Econ.*, 2009, *1* (1), 319–354.

_ , _ , and _ , "The macroeconomic implications of rising wage inequality in the United States," *Journal of political economy*, 2010, *118* (4), 681–722.

_ , _ , **and** _ , "Consumption and labor supply with partial insurance: An analytical framework," *American Economic Review*, 2014, *104* (7), 2075–2126.

Hennessy, Christopher A, Amnon Levy, and Toni M Whited, "Testing Q theory with financing frictions," *Journal of financial economics*, 2007, *83* (3), 691–717.

Herd, Richard, "Estimating Capital Formation and Capital Stock by Economic Sector in China," 2020.

House, Christopher L, "Fixed costs and long-lived investments," *Journal of Monetary Economics*, 2014, *68*, 86–100.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Huang, Wei and Chuanchuan Zhang, "The power of social pensions: Evidence from China's new rural pension scheme," *American Economic Journal: Applied Economics*, 2021, *13* (2), 179–205.

Huggett, Mark, "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of economic Dynamics and Control*, 1993, *17* (5-6), 953–969.

_ , "Wealth distribution in life-cycle economies," *Journal of Monetary Economics*, 1996, *38* (3), 469–494.

Jones, Charles I and Peter J Klenow, "Beyond GDP? Welfare across countries and time," *American Economic Review*, 2016, *106* (9), 2426–2457.

Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce, "Wage inequality and the rise in returns to skill," *Journal of political Economy*, 1993, *101* (3), 410–442.

Kaplan, Greg and Giovanni L Violante, "How much consumption insurance beyond self-insurance?," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 53–87.

_ **, Benjamin Moll, and Giovanni L Violante**, "Monetary policy according to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

Katz, Lawrence F and Kevin M Murphy, "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 1992, *107* (1), 35–78.

Keane, Michael and Richard Rogerson, "Micro and macro labor supply elasticities: A reassessment of conventional wisdom," *Journal of Economic Literature*, 2012, *50* (2), 464–76.

_ **and** _ , "Reconciling micro and macro labor supply elasticities: A structural perspective," *Annu. Rev. Econ.*, 2015, *7* (1), 89–117.

Keane, Michael P and Kenneth I Wolpin, "The career decisions of young men," *Journal of political Economy*, 1997, *105* (3), 473–522.

Kermani, Amir and Yueran Ma, "Asset Specificity of Nonfinancial Firms," *The Quarterly Journal of Economics*, 2023, *138* (1), 205–264.

Khan, Aubhik and Julia K Thomas, "Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics," *Econometrica*, 2008, *76* (2), 395–436.

_ **and** _ , "Credit shocks and aggregate fluctuations in an economy with production heterogeneity," *Journal of Political Economy*, 2013, *121* (6), 1055–1107.

_ **, Tatsuro Senga, and Julia K Thomas**, "Default risk and aggregate fluctuations in an economy with production heterogeneity," *Unpublished manuscript*, 2014, *5*.

Kiyotaki, Nobuhiro, "A perspective on modern business cycle theory," *FRB Richmond*

Economic Quarterly, 2011, *97* (3), 195–208.

_ **and John Moore**, "Credit cycles," *Journal of political economy*, 1997, *105* (2), 211–248.

Kocherlakota, Narayana et al., "Creating business cycles through credit constraints," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2000, *24* (3), 2–10.

Krueger, Dirk and Fabrizio Perri, "Does income inequality lead to consumption inequality? Evidence and theory," *The Review of Economic Studies*, 2006, *73* (1), 163–193.

Lanteri, Andrea, "The market for used capital: Endogenous irreversibility and reallocation over the business cycle," *American Economic Review*, 2018, *108* (9), 2383–2419.

_ **and Adriano A Rampini**, "Constrained-efficient capital reallocation," *American Economic Review*, 2023, *113* (2), 354–395.

Lemieux, Thomas, "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?," *American Economic Review*, 2006, *96* (3), 461–498.

Lillard, Lee A and Robert J Willis, "Dynamic aspects of earning mobility," *Econometrica: Journal of the Econometric Society*, 1978, pp. 985–1012.

MaCurdy, Thomas E, "An empirical model of labor supply in a life-cycle setting," *Journal of political Economy*, 1981, *89* (6), 1059–1085.

McGrattan, Ellen R, Richard Rogerson, and Randall Wright, "An equilibrium model of the business cycle with household production and fiscal policy," *International Economic Review*, 1997, pp. 267–290.

_ **,** _ **et al.**, "Changes in hours worked, 1950-2000," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2004, *28* (1), 14–33.

Meghir, Costas and Luigi Pistaferri, "Income variance dynamics and heterogeneity," *Econometrica*, 2004, *72* (1), 1–32.

Mendoza, Enrique G, "Sudden stops, financial crises, and leverage," *American Economic Review*, 2010, *100* (5), 1941–1966.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and misallocation: Evidence from plant-level data," *American economic review*, 2014, *104* (2), 422–458.

Moffitt, Robert A and Peter Gottschalk, "Trends in the Transitory Variance of Male Earnings in the US, 1970-2004," Technical Report, National Bureau of Economic Research 2011.

_ **and** _ , "Trends in the transitory variance of male earnings methods and evidence," *Journal of Human Resources*, 2012, *47* (1), 204–236.

Moffitt, Robert and Peter Gottschalk, "The growth of earnings instability in the US labor market," *Brookings Papers on Economic Activity*, 1994, *2*, 217–72.

Moll, Benjamin, "Productivity losses from financial frictions: Can self-financing undo capital misallocation?," *American Economic Review*, 2014, *104* (10), 3186–3221.

Oberfield, Ezra, "Productivity and misallocation during a crisis: Evidence from the Chilean crisis of 1982," *Review of Economic Dynamics*, 2013, *16* (1), 100–119.

Ohanian, Lee, Andrea Raffo, and Richard Rogerson, "Long-term changes in labor supply and taxes: Evidence from OECD countries, 1956–2004," *Journal of Monetary Economics*, 2008, *55* (8), 1353–1362.

Ottonello, Pablo, "Capital unemployment," *Work. Pap., Dep. Econ., Columbia Univ., New York Google Scholar Article Locations: Article Location A*, 2017.

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

Pijoan-Mas, Josep, "Precautionary savings or working longer hours?," *Review of Economic dynamics*, 2006, *9* (2), 326–352.

Prescott, Edward C, "Why do Americans work so much more than Europeans?," 2004.

Restuccia, Diego and Richard Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 2008, *11* (4), 707–720.

Riddick, Leigh A and Toni M Whited, "The corporate propensity to save," *The Journal of Finance*, 2009, *64* (4), 1729–1766.

Rogerson, Richard, "Understanding differences in hours worked," *Review of Economic dynamics*, 2006, *9* (3), 365–409.

_ , "The Impact of Labor Taxes on Labor Supply," Technical Report, American Enterprise Institute 2010.

Sandleris, Guido and Mark LJ Wright, "The costs of financial crises: Resource misallocation, productivity, and welfare in the 2001 argentine crisis," *The Scandinavian Journal of Economics*, 2014, *116* (1), 87–127.

Santaeulalia-Llopis, Raül and Yu Zheng, "The price of growth: consumption insurance in China 1989–2009," *American Economic Journal: Macroeconomics*, 2018, *10* (4), 1–35.

Thomas, Julia K, "Is lumpy investment relevant for the business cycle?," *Journal of political Economy*, 2002, *110* (3), 508–534.

Violante, Giovanni L, "Technological acceleration, skill transferability, and the rise in residual inequality," *The Quarterly Journal of Economics*, 2002, *117* (1), 297–338.

Whited, Toni M, "Debt, liquidity constraints, and corporate investment: Evidence from panel data," *The journal of finance*, 1992, *47* (4), 1425–1460.

Winberry, Thomas, "Lumpy investment, business cycles, and stimulus policy," *American Economic Review*, 2021, *111* (1), 364–396.