UCLA
UCLA Electronic Theses and Dissertations

Title

# Essays on Macroeconomic Implications of Technological Change

Permalink
https://escholarship.org/uc/item/0rw616vv

Author
Shen, Shihan

Publication Date
2023

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Macroeconomic Implications of Technological Change

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Shihan Shen

2023

© Copyright by

Shihan Shen

2023

## ABSTRACT OF THE DISSERTATION

Essays on Macroeconomic Implications of Technological Change

by

Shihan Shen

Doctor of Philosophy in Economics

University of California, Los Angeles, 2023

Professor Lee Ohanian, Co-Chair

Professor Pierre-Olivier Weill, Co-Chair

This dissertation contributes toward our understanding of how technological changes in recent decades affect firm behaviors and various aspect of the aggregate economy, including changes in market concentration, productivity, labor share and credit allocation.

In Chapter 1, "Customer Acquisition, Rising Concentration and US Productivity Dynamics", I document that the cost of marketing and advertising has declined enormously in recent decades due to the advance of digital technologies. This chapter then studies the macroeconomic consequences of lower marketing cost, and finds that it is a critical driving force of several striking macroeconomic trends, including rising market concentration and productivity growth slowdown since the 1990s. I develop an endogenous growth model with product market search frictions. Firms invest in innovation and marketing to build customer base, which is a long-term asset. Then I exogenously feed in the observed large drop of marketing cost into the quantitative model and find that it accounts for 83% of the rise in market concentration, measured by the largest firm's market share. Cheaper marketing generates a positive level effect and a negative growth effect on productivity. These two effects together explain around 1/3 of the decline in productivity growth rate and successfully capture its hump-shaped pattern over time. Finally, I conduct

a welfare analysis and find that firms tend to over-invest in marketing compared to the socially optimal allocation, which implies that welfare can be improved by a marketing tax.

In Chapter 2, "Revisiting Capital-Skill Complementarity, Inequality, and Labor Share", jointly written with Lee Ohanian and Musa Orak, we analyze the quantitative contribution of capital-skill complementarity in accounting for rising wage inequality, as in Krusell et al. (2000). We study how well the KORV framework accounts for more recent data, including the large changes in labor's share of income that occurred after the KORV estimation period ended. We also study how using information and communications technology (ICT) capital as the complementary capital stock affects the model's implications for inequality and overall model fit. We find significant evidence for continued capital-skill complementarity across all model permutations we analyze. Despite nearly 30 years of additional data, we find very little change to the original KORV estimated substitution elasticity estimates when the total stock of capital equipment is used as the complementary capital stock. We find much more capital-skill complementarity when ICT capital is used. The KORV framework continues to closely account for rising wage inequality through 2019, though it misses the three-percentage point decline in labor's share of income that has occurred since 2000.

In Chapter 3, "Private Information, Adverse Selection and Small Business Financing", I develop a competitive search model with asymmetric information and search frictions in bank-borrower relationship. The development of data-driven technologies increases the enforceability of lenders when the borrowers fail to meet the standard repayment requirement. In a credit market where borrowers can easily escape debt obligations, banks tend to use market tightness for screening and post too few contracts for safe borrowers to deter imitation from riskier borrowers. When lender enforceability rises, however, the required repayment becomes a better screening device. To stand out from the risky ones, the safe borrowers accept the contract in which they are over-charged and over-offered the loans.

The dissertation of Shihan Shen is approved.

Hugo Andres Hopenhayn

Andrew Granger Atkeson

Matthew Saki Bigio Luks

Lee Ohanian, Committee Co-Chair

Pierre-Olivier Weill, Committee Co-Chair

University of California, Los Angeles

2023

# DEDICATIONS

*To my beloved family*

# Contents

# List of Figures

# List of Tables

## ACKNOWLEDGMENT

I am deeply grateful to my Committee Chairs Lee Ohanian and Pierre-Olivier Weill for constantly believing in me, guiding me and encouraging me every step of the way throughout my PhD studies.

I am especially grateful to my committee member Hugo Hopenhayn, who has devoted tremendous time and effort to helping me become a real researcher - I am forever indebted to you.

I am beholden for the support and advice from my committee members Andrew Atkeson and Saki Bigio.

I am thankful to my great mentor and friend Steven N. S. Cheung for his care and support.

I thank my coauthor Musa Orak, from whom I have learned a great deal.

I thank Oleg Itskhoki, David Baqaee, Martin Hackman, Bruno Pellegrino and Huifeng Chang, and other participants at the UCLA macro proseminar and UCLA IO proseminar for insightful discussion.

I thank participants in various conferences for helpful comments.

I thank UCLA for giving me the chance and financial support to thrive in a fantastic environment.

I thank my parents and my fiancé Peiqi Wang - I could not have made this journey without your unconditional love and support, and I dedicate this milestone to you.

# VITA

## Education

### University of California, Los Angeles

M.A. in Economics 2019

### Peking University

M.A. in Economics 2017

B.A. in Economics 2014

## Honors and Awards

Best Paper Award, Macroeconomics Proseminar, UCLA 2022

Dissertation Year Fellowship, UCLA 2022–2023

Lewis L. Clarke Graduate Fellowship Fund, UCLA 2021

Graduate Student Fellowship 2017-2018

Outstanding Graduate Student, Peking University 2014

# Chapter 1

## Customer Acquisition, Rising Concentration and US Productivity Dynamics

### 1.1 Introduction

With the advance of digital technologies in the past few decades, the cost of marketing and advertising has declined enormously due to a dramatic change in the composition of advertising delivery. The resulting rise of marketing/advertising intensity reshapes the way firms build customer base and further affects firm dynamics. This paper studies the macro consequences of lower marketing cost and finds that it well explains two recent macroeconomic trends: rising market concentration, measured by market share of the largest firms, and slowdown in aggregate productivity growth. To my knowledge, this paper is the first study to connect macroeconomic trends to the evolution of marketing landscape and customer base accumulation.

To do this, I develop an endogenous growth model with product market search frictions, where firms choose how much to innovate and how much to advertise. The informational frictions between firms and buyers render customer relationships long-term in nature, and turn the customer base into a form of capital for firms. Then I conduct an experiment in which I exogenously feed in the observed large drop of marketing cost into the quantitative model to study how this may have contributed to changes in the aggregate economy.

I find that the decline of marketing cost is a critical driving force of the recent rise in aggregate market concentration. From 2016 to 2019, the average market share of the largest firm within 3-digit SIC industries has increased from 24% to almost 30%. Quantitative analysis shows that cheaper marketing can account for 83% of this change. There are always concerns among economists and policymakers that the considerable rise in concentration may depress long-run economic performance by squeezing out competitors[1]. After all, during the same period of rising concentration, the TFP growth rate dropped from around 1.5% to 0.6%, despite a temporary boom in the late 1990s. My paper contributes to this debate by studying how lower marketing cost, as a key mechanism of increasing concentration, affects long-run productivity growth and economic welfare.

I show that on the transitional dynamics, the declining marketing cost generates a positive "level effect" on productivity due to higher output quantity brought about by more efficient matching, and a negative "growth effect" driven by lower innovation intensity due to a shift towards marketing investment. The combination of these two effects explains around 1/3 of the decline in productivity growth and successfully captures the "first-rise-then-fall" pattern in the TFP growth rate over time. In a welfare analysis, I find that as marketing becomes cheaper, the level effect dominates the growth effect and results in an increase of total welfare. However, compared to the socially optimal allocation, firms tend to over-invest in marketing because one firm's marketing generates a negative externality on other firm's customer acquisition.

This paper highlights an essential part of firm operations: building and maintaining customer base. In fact, the rise of many large firms' market share is coincided with their huge and fast-growing customer base. Almost all the superstar firms emerged in the past 20

[1] Policymakers are concerned about the negative effects of superstar firms on competition. For instance, the Executive Order on Promoting Competition in the American Economy in 2020 addresses concerns on the potential unfair competition between big firms and small businesses. (For more details, check link). The recent bill of "American Innovation and Choice Online Act" (S. 2992) considers limiting the "big tech" companies (e.g., Facebook, Google, Amazon) from using their website to promote their own products over similar products by other companies (For details, check link).

years, such as Amazon, Uber, etc., own a large number of users or subscribers[2]. Conventional wisdom focuses on the contribution of innovation to attract customers. Specifically, firms invest in R&D to develop better product quality, which increases their relative market size by diverting demand away from firms selling low-quality goods. However, in addition to R&D, firms also invest substantially in marketing, advertising and the like, to propagate their products and build customer base.

Between 1996 and 2019, firms spend as much as 8% of GDP on marketing related activities, with advertising alone taking up around 2.2%. This is comparable to the share of R&D expenditures over GDP (2.4% to 3%) in the same period. The heavy spending on marketing and advertising provide evidence of frictions in product markets, which require firms to spend resources on customer acquisition. The importance of marketing is also elaborated in the industrial organization and marketing literature. Argente et al. (2021a) shows that firms build market share by adding new customers through advertising rather than manipulating markups. Thomas (2001) also points out that many firms are taking a customer-oriented approach as compared to a product-oriented approach in the new economy where more and more firms sell services and customer relationship becomes critical.

The landscape of marketing has changed dramatically in recent years: the development of digital technologies greatly reshapes the delivery of advertising. From late 1990s, online marketing gradually becomes prevalent and accounts for almost half of total advertising in today's world. This raises marketing efficiency and thus reduces the cost of advertising. Using the composition of U.S. advertising spending and the price index of information and communications technologies (ICT) investment, I find that the average advertising price index has plummeted by more than 40% from 1996 to 2019. Meanwhile, aggregate firm marketing expenditures have increased enormously. After taking into account the decline

[2] For instance, from 2014 to 2021, Amazon's sales share in the U.S. had doubled from 28.1% to 56.7%, and in the meantime, their Prime member user base expanded by four times, from 40 million to 157.4 million. Take Uber as another example. Founded in 2009, Uber grew from a small startup into a giant that owns 71% of sales in the ride-sharing market in 2022. Their customer base today has reached 118 million, which is almost 1/3 of total U.S. population.

of price index, real advertising spending of U.S. firms has risen by 5 times over the past 20 years. Interestingly, the gap of marketing expenditures between large and small firms (by sales) has widened, with the large firms increasing their marketing investment and small firms downsizing it. This implies that the technological change in marketing might potentially explain the enlarging gap in customer base and market share across firm size.

To capture the effect of marketing cost on customer base accumulation, firm dynamics and further on the aggregate economy, I develop an endogenous growth model where a continuum of firms sell goods to customers in a market hampered by informational frictions. My model features two key aspects: (i) firms innovate to raise quality through R&D investment, and (ii) firms invest in marketing and advertising to become more visible to customers who search for new products to buy. Search frictions in the product market renders the customer base a valuable long-term asset to firms. Those in the customer base of a firm will buy again at the firm if their lifetime utility of continuing the customer relationship outweighs that of searching.

Marketing investment increases product exposure to buyers and enables the firm to match with more customers in the search process. Upon meeting, a customer discovers the quality, price, as well as her preference on the good, and decides whether to buy based on the revealed information. Holding other factors constant, the better the product quality, the more likely a customer makes the purchase.

Marketing and R&D investments are complementary, because by performing R&D, a firm improves product quality, which increases a customer's probability of purchase and thus raises the return to advertising. In general, firms with a larger customer base invest more in both R&D and marketing. Due to the scale effect of innovation, the cost of innovation does not scale up with firm customer base, but profits do because non-rivalrous technologies/ideas can be replicated at close-to-zero marginal cost (Haskel and Westlake (2017), De Ridder (2019)). Hence, large firms tend to invest more in innovation than their smaller counterparts to take advantage of their large customer base. By complementarity, large firms also invest more intensively in marketing.

Then I use the model to analyze impact of declining marketing cost on firm dynamics. As marketing becomes less costly, every firm has incentives to invest more in marketing. With the increasing presence of advertisement, it becomes easier for customers to learn the existence or characteristics of a product, which generates more firm-customer matches for all firms. However, as large firms invest more in R&D and are more likely to provide better quality, they are able to convert a larger proportion of matched customers into actual buyers relative to their smaller counterparts. Therefore, although marketing becomes equally cheaper for all firms, large firms disproportionately obtain more new buyers in the search process, which enlarges the difference in customer base across firm size. As customer base directly affects firm decisions on R&D and marketing investment, the gap between large and small firms is further amplified in a positive feedback loop, and eventually leads to a surge of market concentration and the rise of superstar firms.

Despite reallocating economic activities towards a smaller set of firms, rising concentration caused by lower marketing cost is not necessarily harmful to the aggregate productivity. In fact, cheaper marketing generates mixed effects on productivity growth by differently affecting production and innovation incentives for large and small firms. As marketing becomes cheaper, the aggregate level of marketing increases, which alleviates search frictions and facilitates more matches and consumption. The resulting rise in aggregate output quantity has a positive effect on productivity growth.

Nevertheless, due to enlarging difference in customer base, large firms increase their marketing investment by disproportionately more than the smaller ones, which implies that small firms now have a lower share of advertisement and are thus less likely to be encountered by customers than before. Despite a rise in aggregate marketing, the capacity of consumption is limited by the total measure of customers (population), creating a congestion effect among firms. Consequently, the customer base of small firms might even shrink due to fewer matches with new customers, which further reduces their long-term incentives to innovate and/or advertise.

On the other hand, large firms expand their customer base by obtaining more new customers during the search process. However, in an economy with high aggregate level of marketing, the option value of search is also larger because customers are now more likely to get matched. Therefore, in order to retain their existing customers, firms now have to offer more competitive prices, which reduces profits earned on their current customer base. In this sense, declining marketing cost reduces firms' market power to charge high prices or markups, and thus lowers the expected return of innovation investment. This constitutes a force towards lower productivity. These offsetting factors exert an ambiguous overall effect on aggregate productivity growth and this is why the quantitative analysis is of paramount importance.

I calibrate the model on the balanced growth path by matching various moments for the U.S. economy in the late 1990s (1996-1999), using data on all U.S. listed firms from Compustat as well as aggregate moments. With this initial calibration I describe firms' innovation and marketing strategies to develop intuition about the model. I then calibrate the reduction in marketing cost to fully capture the five-fold increase in aggregate real marketing investment observed in the data.

Using the calibrated model, I conduct an experiment where I exogenously feed in the observed drop of marketing cost to infer how this may have changed the economy from late 1990s to today. First, I quantify the aggregate impact of marketing shock on business concentration and productivity growth on the balanced growth path. I find that my model can explain 83% of the rise in the largest firm's market share within 3-digit SIC industries, and around 1/3 of the decline in aggregate productivity growth. To understand the role of changing customer base in rising concentration, I conduct a counterfactual exercise to decompose the effect on concentration into changes in quantity and changes in prices. It shows that the market concentration is mainly driven by the enlarging gap in customer base across firm size rather than large firms charging higher markups, supporting the findings in

Baqaee and Farhi (2020) and De Loecker et al. (2020).

Second, I demonstrate the dynamics of transition from the low to high marketing equilibrium. Over the transition path, market leader's sales share increases rapidly at the beginning and then gradually grows to its new steady state. The aggregate productivity growth experienced an initial surge of 37 basis point for 5 years and then it declines by 68 basis point over time to reach the new balanced growth path. This is consistent with the "first-rise-then-fall" pattern in data. A decomposition analysis indicates that the temporary boom at the beginning is driven by the increase in output quantity due to more matches in the product market. This level effect gradually fades out as the economy converges to the new steady state. Additionally, marketing also affects firm R&D incentives and, therefore, has an indirect effect on growth (the "growth effect"). With an increasing customer base, large firms tend to increase their R&D investment, although the lower prices on existing customers generate opposite incentives. Small firms decrease their R&D investment due to both deteriorate marketing share and lower prices. In the calibrated model, the average innovation growth rate decreases from 1.51% in the first period to 1.19% in the new steady state, implying that the negative effect of small firms and declining price dominates the rise of customer base in large firms.

Last but not the least, I study the consequences of lower marketing cost on welfare. Following the classic literature Nelson (1974), Butters (1977), Grossman and Shapiro (1984) and Milgrom and Roberts (1986), informative advertising weakens the negative effect of information frictions, and is usually associated with welfare improving. Indeed, I find that as marketing becomes cheaper, the level effect dominates the growth effect, and the overall welfare increases. However, advertising is also a taste shifter that firms use to steal customers from each other and maintain their own market shares ( Dixit and Norman (1978), Becker and Murphy (1993), Benhabib and Bisin (2002), Benhabib and Bisin (2011) and Molinari and Turino (2009)). In this case, the combative marketing expenditures might result in a waste of resources. Therefore, I derive the social planner's problem and find that compared to the socially optimal allocation, firms in the decentralized economy over-invest in marketing as they ignore the negative impact of their advertisement in other firms' customer acquisition. Finally, I discuss the policy implications and find that there exists an optimal level of tax on marketing that maximizes social welfare.

**Related Literature.** This paper is closely related to a new literature that embeds customer acquisition and marketing into a macroeconomic framework. Relevant work including Fishman and Rob (2003), Gourio and Rudanko (2014) and Paciello et al. (2019) explore the impact of long-term customer relationship in a product market with search frictions. My paper builds on their idea of customer capital but focuses on its implication for market concentration and productivity growth. Another two closely related papers are Cavenaile and Roldan-Blanco (2021) and Cavenaile et al. (2021). These two papers also develop an endogenous growth model to analyze the implications of advertising for firm dynamics and economic growth through its interaction with R&D. The key difference between their papers and mine is the role of advertising. In their papers, advertising benefits firms by either influencing consumer tastes, while my paper emphasizes the role of advertising to mitigate informational frictions in a search environment. Although we all talk about the effects on productivity, I propose a mechanism of positive feedback loop of customer base accumulation to explain increasing market concentration. The long-term customer relationship is the key to explain the sharp rise in large firm's market share.

Rachel (2021) claims that the rise of leisure goods provided by advertising has a negative effect on welfare because less labor are allocated to R&D. This is consistent with the "growth effect" in my model. However, advertising in my paper has an additional information-based role as opposed to purely shifting customer tastes, and the resulting "level effect" explains the different welfare effect in the two papers. Greenwood et al. (2021) distinguishes traditional and digital advertising and argues that digital advertising increases welfare significantly and is disproportionately financed by better-off consumers.

Perla (2019) develops a model in which consumers learn about firms through a network of connections that endogenously evolves over the life cycle. Other relevant papers include papers that use pricing to accumulate a customer base (e.g. Drozd and Nosal (2012) Foster et al. (2016), Roldan-Blanco and Gilbukh (2021)), and papers that study marketing and advertising activities in international economics (Fitzgerald et al. (2017)).

This paper also contributes a novel mechanism to the large and growing literature linking trends in concentration, productivity growth, and business dynamism using models of endogenous growth. The existing literature highlights the importance of technological change and intangible assets to explain macroeconomic trends. For example, De Ridder (2019) shows that the rise of intangible inputs such as software can jointly explain the slowdown of productivity growth and rising market power. Olmstead-Rumsey (2019) explains these trends using the declining innovativeness of smaller firms, and Aghion et al. (2019) emphasizes the importance of less technology spillover. In addition, there are also non-technological explanations including demographic changes (Hopenhayn et al. (2018) , Jones (2020), Peters and Walsh (2021), Karahan et al. (2019), Eggertsson et al. (2019), Bornstein et al. (2018)) or declining real interest rates (Kroen et al. (2021), Chatterjee and Eyigungor (2020)). Specifically, Bornstein et al. (2018) argues that population aging affects consumer inertia and the evolution of customer base. Rather than emphasizing the product competition, my paper provides a novel idea of the importance of marketing and customer base accumulation to firms. I show that marketing and advertising also plays a key role in firms' pursuit of profits, and potentially affects competition, which further explains the enlarging gap among large and small firms and the potential threat to productivity growth.

More broadly, this paper relates to the literature that provides micro-foundations to how advertising works. The two-sided market is first proposed by Caillaud and Jullien (2003) and Rochet and Tirole (2003). These papers are built upon the idea that advertising platform serves a two-sided market between firms and consumers. The platform is a monopoly that maximizes its profit by choosing the number of ads it sells. In my paper, I model advertising as a type of investment for firms to build customer base, and do not explicitly study the relationship between ad agencies and firms. Another related strand of literature explores the role of intangibles for firm and industry dynamics (Arkolakis (2016), Dinlersoz and Yorukoglu (2012), Perla (2017) and Bhandari et al. (2021)). I contribute to this literature by highlighting the importance of marketing investment and exploring its interaction with firm R&D investment. My paper models firm innovation with an endogenous growth model, which inherits the classical quality-ladder framework of Klette and Kortum (2004) and Lentz and Mortensen (2008).

**Outline.** The paper proceeds as follows. Section 1.2 shows the motivating facts in marketing landscape and firm investment. Section 1.3 presents the endogenous growth model with customer acquisition and Section 1.4 discusses the mechanism. I structurally estimate the model in Section 1.5, and discuss results in Section 1.6. Section 1.7 concludes.

## 1.2 Empirical Motivation

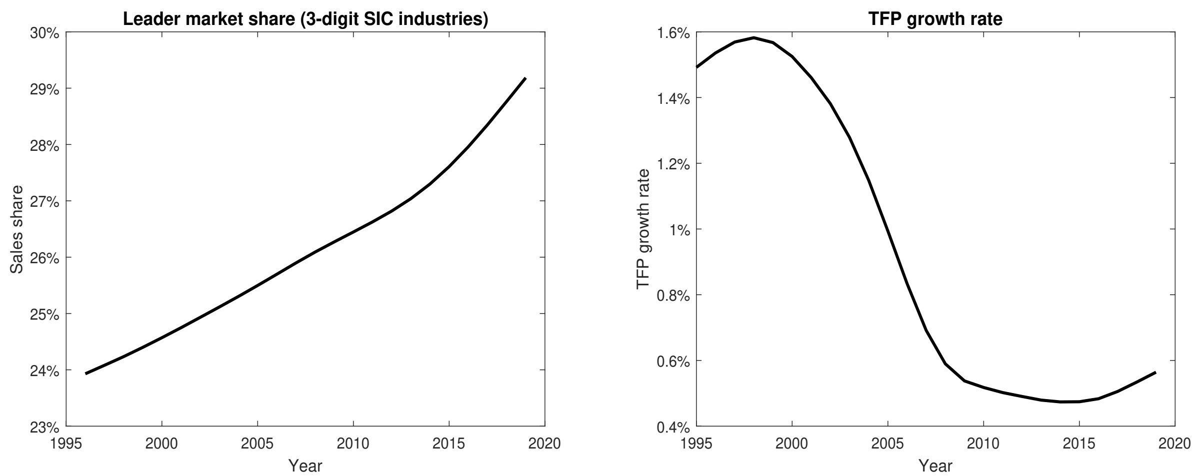
In this section, I first display recent macroeconomic trends of increasing concentration and slowdown in productivity growth in the past twenty years. Then I describe the stylized facts on marketing landscape during the same period, including the composition of advertising spending, the evolution of marketing cost and trends in firm marketing investment.

### 1.2.1 Market Concentration and Productivity Growth

Figure 1.1 plots the evolution of average market share of industry leader for all U.S. public companies (left panel), and the total factor productivity growth rate (right panel). Using Compustat data, I calculate the largest firm's sales share for all 3-digit Standard Industrial Classification (SIC) industries, and take the average weighted by industry sales as market concentration. Over the past 25 years, average market concentration has risen from around

24% in the l996 to 29.2% in 2019. Annual total factor productivity (TFP) growth is taken from the updated series of TFP in Fernald (2014). Based on their accounting method, the average TFP was about 1.5% in the late 1990s, slightly increased to 1.6% in the late 1990s, and has declined all the way to 0.6% in the following 20 years.

Figure 1.1: Macroeconomics Trends of TFP and Concentration




Notes: This figure presents the evolution of concentration and TFP growth rate from 1996 to 2019. The calculation of average market share of the largest firm (by sale) in 3-digit SIC industries uses firm level data from Compustat, weighted by industry sales. Total factor productivity (TFP) estimates are taken from FRBSF Working Paper 2012-19. The plots are smoothed using an HP filter with an annual smoothing parameter of 20 and 200, respectively.

### 1.2.2 Customer Base

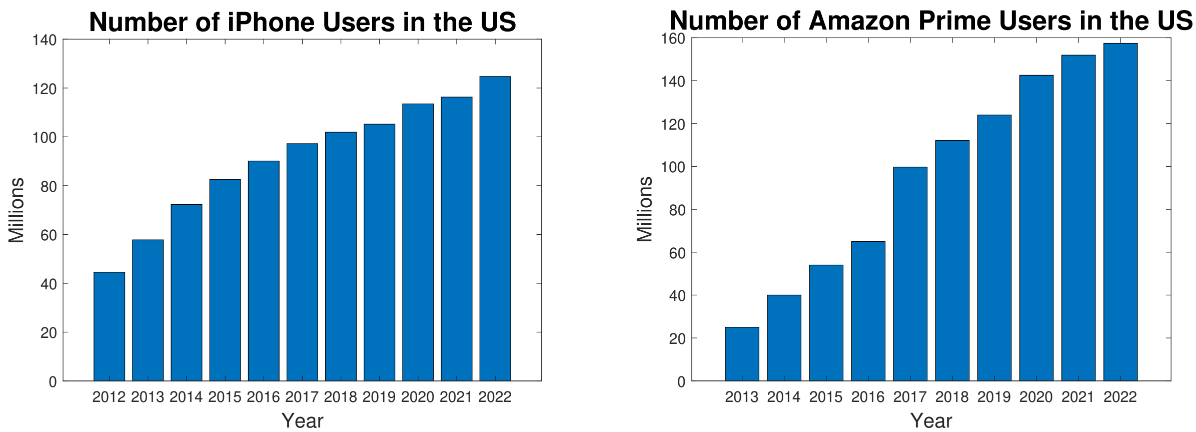
#### 1.2.2.1 Sales and Customers

A phenomenon that coincides with increasing concentration is the rise of superstar firms with exceptionally high revenues and market value. According to Fortune 500[3], the ten largest companies by revenue in 2022 are Walmart, Amazon, Apple, CVS Health, UnitedHealth Group, Exxon Mobil, Bershire hathaway, Alphabet, McKesson Corporation and AmerisourceBergen. As many papers have pointed out (Olmstead-Rumsey (2019),

[3] Fortune 500 is an annual list compiled and published by Fortune magazine that ranks 500 of the largest United States corporations by total revenue for their respective fiscal years. The list includes publicly held companies, along with privately held companies for which revenues are publicly available.

Aghion et al. (2019)), successful innovation plays a key role in the rapid growth of these superstar firms, but on top of that, another common feature of these superstar firms is that they all have an enormous and fast-growing customer base regardless of the industries. Take the technology sector as an example. There are various types of tech companies with complex networks, but the biggest ones (Walmart, Amazon Apple, Alphabet and Microsoft) all have a huge number of customers, users or subscribers. Figure 1.2 shows the number of iPhone users and Amazon Prime subscribers over the past decade.

Figure 1.2: Number of Customers for Superstar Firms (Apple and Amazon)




Notes: Data source: annual reports of Apple and Amazon.

The relationship between customer base and sales is also documented by the work of Afrouzi et al. (2020), where they show that differences in the sizes of firms' customer bases account for three-quarters of the variation in firms' sales use Nielsen Homescan product data and Compustat firm data. Using a different data set on Visa card transactions, Einav et al. (2021) show that about 80% of sales variation can be traced to the number of customers. Bernard et al. (2022) document the importance of customers for Belgian inter-firm transactions.

#### 1.2.2.2 Building Customer Base

Given the importance of customers in boosting revenues, firms invest a lot in building customer base. In reality, firm come up with a critical statistic "customer lifetime value"

(CLV) as a metric that measures how much a business can plan to earn from the average customer over the course of the relationship. A customer not only generates revenue in one period, but also brings cash flows in the long run thanks to consumer inertia.

In canonical quality-ladder models where firms engage in Bertrand competition, the firm with the highest productivity (equivalently, lowest production cost) is the only producer in a product line, which implies that it obtains all the customers for that variety. In other words, firms invest in innovation to build customer base in such frameworks. However, the meeting between firms and customers is a highly frictional process, and firms invest substantially in marketing and advertising to attract the attention of customers. Since 1990s, the share of US R&D expenditures over GDP fluctuated between 2.27% and 2.82%[4], while at the same time firms spent on average around 2.2% of GDP on advertising each year[5], which is comparable to the expenditures on R&D.

### 1.2.3 Changes in Marketing Landscape

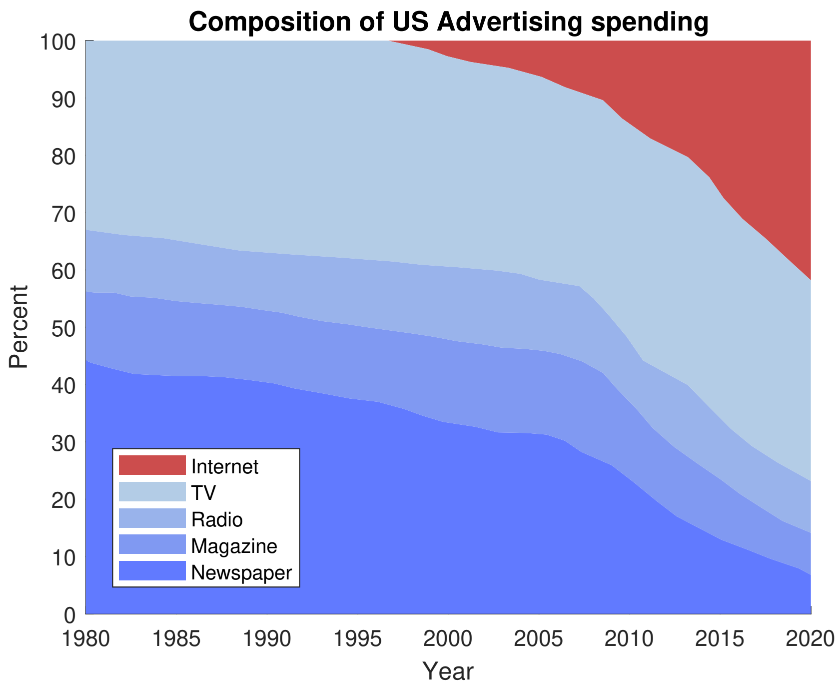
Although the share of expenditures on advertising do not display a significant rise over time, the advance of digital technologies has greatly reshaped the delivery of advertising, and thus affects firms' accumulation of customer base. Digital marketing, also known as online advertising, evolves quickly because of Internet technologies, with the online channels now accounting for almost half of total advertising, and strong growth in social media, search engines, mobile apps and e-commerce. Figure 1.3 shows the composition of various marketing/advertising platforms. Internet advertising, which was almost absent in the late 1990s, grows quickly in recent years and becomes a major marketing platform that accounts for more than 40% of total US advertising today. Advertising on print-based media and radio declines rapidly, while the share of TV advertising remains relatively stable over time.

Even traditional media is deeply influenced by the development of internet and

[4] Source: OECD data, available at http://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm

[5] Source: Coen Structured Advertising Expenditure Dataset, extracted from the McCann Erikson advertising agency (available at http://www.purplemotes.net/2008/09/14/us-advertising-expenditure-data/).

Figure 1.3: Composition of US Advertising Spending



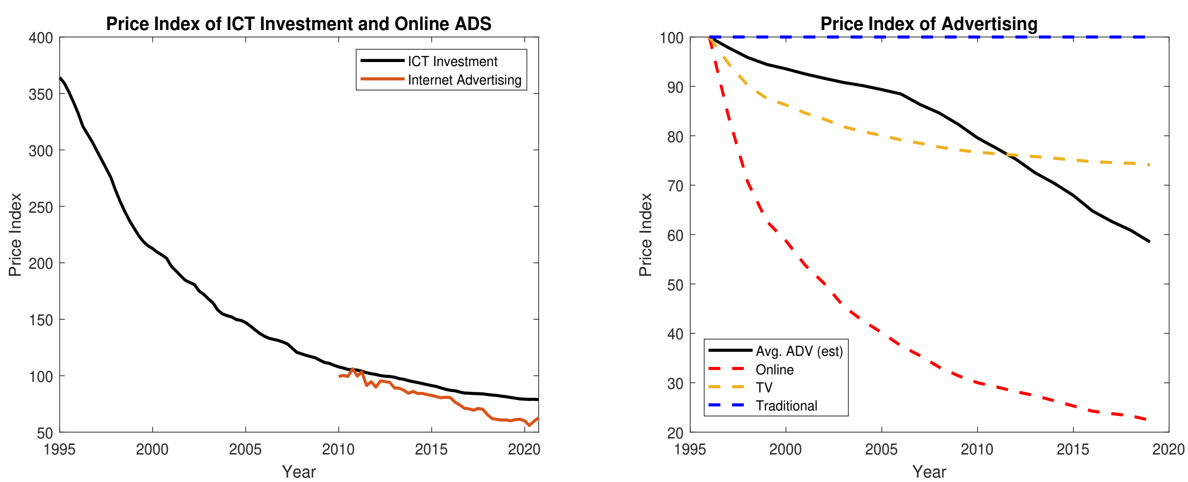

Notes: This figure depicts the percentage of US advertising spending on various platforms, including Internet, TV, Radio, Magazine and Newspaper. The data is obtained from Zenith and McCann.

communications technologies (ICT) capital, with the digital technologies reshaping the display of advertising and consequently affecting the cost of marketing for businesses. For example, in recent years, there has been a shift from traditional to connected TV (CTV). With CTV advertising, ad buys are not based on air times or channels as in traditional TV advertising. Instead, CTV ads are delivered one at a time based on the specific viewer watching a program. This saves the cost of ad buyers because they no longer have to guess which shows their target audiences are watching. Instead, they can build a target audience based on demographic and behavioral signals, then serve ads to specific viewers.

Goldfarb (2014) points out that the fundamental difference between traditional and digital advertising is a substantial reduction in the cost of targeting for digital channels. With online advertising, advertisers can target the keywords customers use in search engines, demographic characteristics, and a consumer's past online behavior at relatively low cost. On the macro level, this technological change leads to a large decline of the cost of marketing/advertising over time. The left panel of Figure 1.5 displays the price index of information processing equipment (black curve) taken from NIPA Table 5.3.4. The price index has dropped by 2/3 from 318.5 in late 1990s to 80.4 in 2019. The red curve is the price index of Internet advertising, a main platform of advertising as aforementioned. The statistic is provided by BEA and starts from 2010. In the past 10 years, the price of Internet advertising has declined by 40%. Although the price of Internet advertising is not available for years before 2009, the downward trends between the two price indices are quite similar, with the price of ICT investment declining even slightly faster.

In order to obtain an average price index of advertising, but in the mean time not over-estimate the decline, I assume 1/3 of the TV advertising uses Internet technologies, and none of the other traditional platforms uses ICT. Next, I use the price index of ICT investment as the proxy for Internet advertising over the past 20 years. Then taking the share of different advertising platforms shown in Figure 1.3 as the weight, I calculate the average marketing price index, and plot it in the right panel of Figure 1.5 (black curve). It plummeted by 40% from 1996 to 2019.

Figure 1.4: Price Indexes of Marketing Input




Notes: This figure depicts the price indexes of advertising/marketing. The black curve on the left shows the price index of information processing equipment taken from the website of National Institute of Pension Administrators (NIPA). The red curve on the right shows the estimated price index of Internet advertising sales, calculated as is described in the main text.

### 1.2.4 Firm Marketing Investment

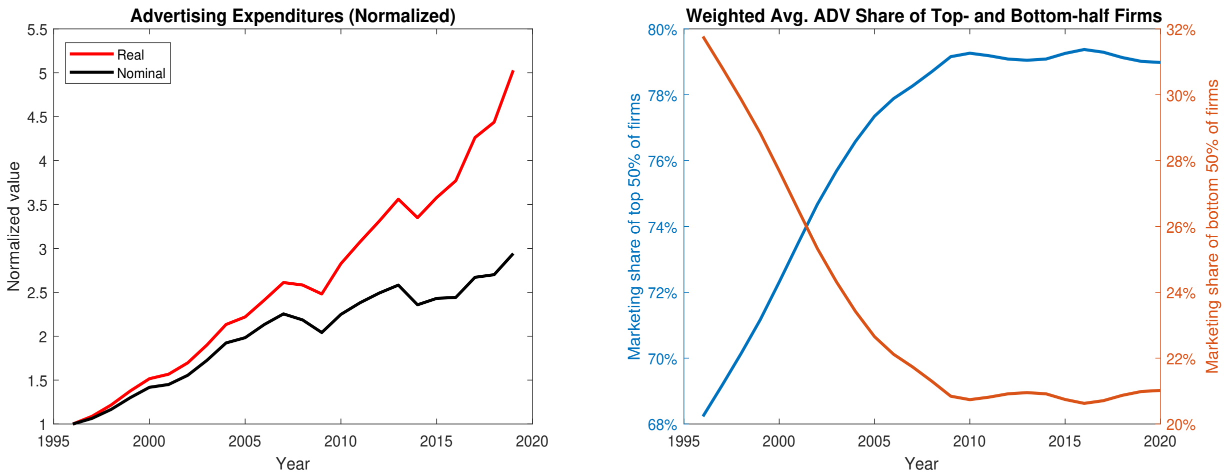
With the shift of the landscape of marketing, firm marketing investments also demonstrate dramatic changes. Similarly to Chiavari (2022), I use two approaches to compute firm-level expenditures devoted to the acquiring customers using Compustat data. The first measure is XAD, which reports the firm-level expenditure in advertisements. This is the only available item in Compustat that measures only (and somehow cleanly) the expenditures of interest; however, this measure suffers from two drawbacks: (i) it reports the cost of advertising media (radio, television, newspapers, and periodicals) and promotional expenses but excludes selling and marketing expenses, and (ii) approximately half of the observations are missing. As an alternative measure to the above, I use Selling General and Administrative (XSGA). This item in Compustat has been the focus of many recent studies such as Gourio and Rudanko (2014), Ptok, Jindal and Reinartz (2018), Afrouzi et al. (2020), and Morlacco and Zeke (2021). However, despite the acknowledged ability of Selling General and Administrative to capture firm-level selling expenditures, it is well known that this item reports many expenditures that are not directly related to the firm's selling efforts, such as bad debt expenses, expenditures in pensions and retirement, rents, and expenditures in research and development. Therefore, I follow the method in Cavenaile and Roldan-Blanco (2021) to construct an adjusted measure of marketing $a_{it}$:

$$a_{it} = XSGA_{it} - XRENT_{it} - XPR_{it} - RECD_{it} - XRD_{it} \tag{1.1}$$

During the same period, the aggregate investment of marketing has increased (Figure 1.5). The total nominal spending of advertising of Compustat firms increased by 3 times. And taking into account the decline of advertising price, the real expenditures increased by 5 times. Digital technologies lowers the cost of advertising and induces firms to invest more in marketing and selling activities.

Although aggregate amount of marketing investment has risen, large firms increase their

Figure 1.5: Evolution of Firm Advertising Investment




Notes: This figure depicts the evolution of firm advertising investment. Left: Aggregate firm advertising expenditure. Black: nominal spending. Red: real spending by dividing the nominal spending with the average price index of advertising. Data source: Compustat variable "XAD". Right: Weighted average difference of advertising share between large and small firms within industries. Use industry sales as weights.

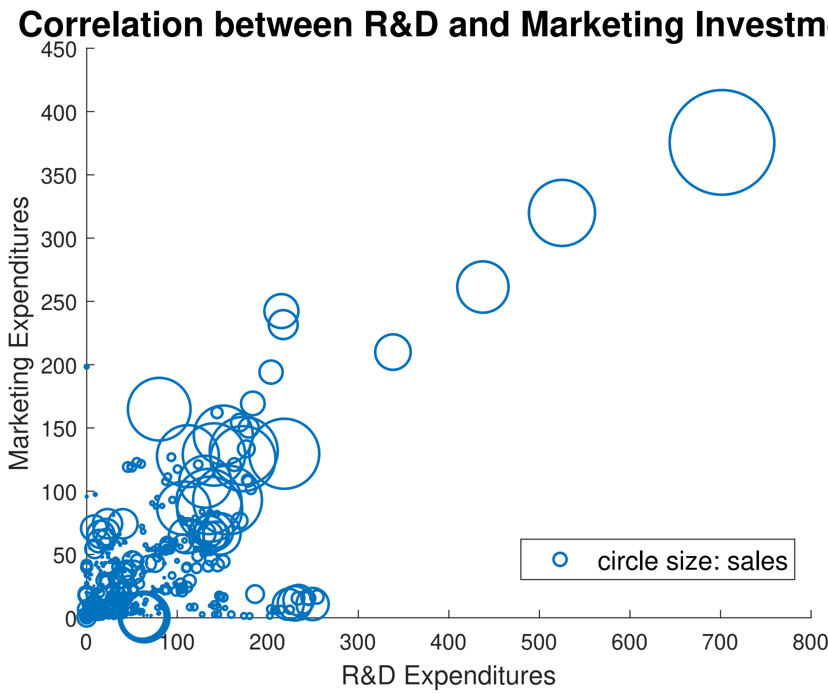
investment by more than the small ones. The right panel of Figure 1.5 shows the average share of marketing investment of the top-half and bottom-half firms in Compustat, weighted by their industry sales. This ratio increases from 68% in 1996 to 80% in 2020.

### 1.2.5 Relationship between Marketing and R&D

The coincidence between the increase of firm real marketing expenditures and the rapid growth of concentration suggests that changes in marketing may be relevant to the trends of market concentration. In this subsection, I investigate the relationship between marketing and R&D investment. As Figure 1.6 shows, big firms invest more in both R&D and marketing.

Relatedly, I show that R&D and marketing complements each other in driving the growth of sales (see Table 1.1).

Figure 1.6: Correlation between R&D and Marketing Investment




Notes: This figure depicts the correlation between R&D and marketing investment. Firm size refers to sales. The measure of R&D is taken from Compustat variable "XRD".

Table 1.1: Relationship between Sales, Marketing and R&D

| | Sale | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| R&D | 0.511*** | 0.432*** | 0.439*** | 0.415*** |
| | (0.00463) | (0.00461) | (0.00460) | (0.00457) |
| marketing | 0.229*** | 0.236*** | 0.233*** | 0.216*** |
| | (0.00340) | (0.00322) | (0.00322) | (0.00318) |
| R&D × marketing | 0.0256*** | 0.0204*** | 0.0203*** | 0.0209*** |
| | (0.0000713) | (0.0000677) | (0.0000676) | (0.0000666) |
| Observations | 24858 | 24858 | 24858 | 24858 |
| $R^2$ | 0.723 | 0.762 | 0.762 | 0.763 |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Year FE | | ✓ | ✓ | ✓ |
| SIC FE | | ✓ | ✓ | |
| Firm FE | | | ✓ | |

Notes: This table presents the relationship between sales, marketing and R&D investment from 1996 to 2019. Control variables include quarterly return on assets, log of tangibility ratio, log of Tobin's q (all lagged by one quarter). Coefficients are reported with t-statistics in parentheses. ***, **, and * indicate significance at 1%, 5% and 10% levels. Standard errors are clustered at the firm level.

## 1.3 Model

In this section, I investigate the effect of declining marketing cost, or equivalently, increasing marketing efficiency, on firm dynamics and the macroeconomy. To do this, I develop an endogenous growth model in which the product market is frictional, and customers are a long-term asset to firms.

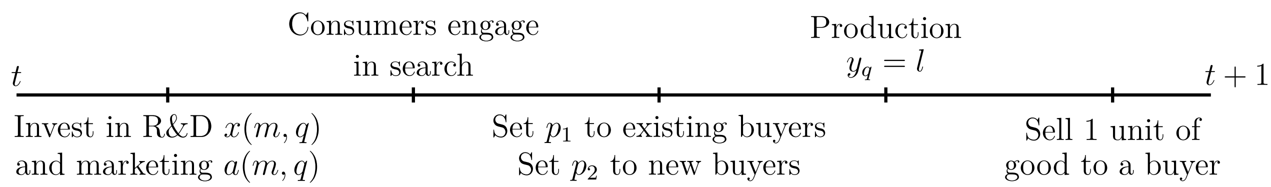
### 1.3.1 Model Setup and Timeline

**Firms.** Time is discrete and lasts forever. There is a measure-one continuum of firms that are risk-neutral and discount the future at rate $\beta$. Firms are heterogeneous along two dimensions, their customer base $m$, which is defined as the mass of customers who bought from that firm in the previous period, and their product quality $q$, which reflects the accumulated effect of R&D. The timeline of firms is shown in the figure below. At the beginning of period $t$, firms invest in R&D and marketing, which are denoted by $x(m,q)$ and $a(m,q)$, respectively. Next, the outcome of R&D is realized, with a probability $x(m,q)$ of achieving a successful innovation. Based on quality of products, firms decide on the optimal prices that they will charge to the existing buyers, who bought from the firm last period, $p_1$, and the price to the new buyers, $p_2$. Next, customers make decisions on where to buy goods, and firms lose and gain customers. After customers are settled, firms produce using labor with a linear technology $y = l$. Finally, firms sell 1 unit of good to each buyer, and the period ends.

Assume a firm has a constant probability $\tau$ of exiting the market each period. Once a firm exits the market it loses all of its customers and has a value of zero. An exiting firm is replaced by a new entrant that starts with a customer base $m_0 = 0$.

**Customers.** The economy is populated by a continuum of mass $N$ consumers, who are risk-neutral and discount the future at rate $\beta$. The customers are ex-ante homogeneous, but they

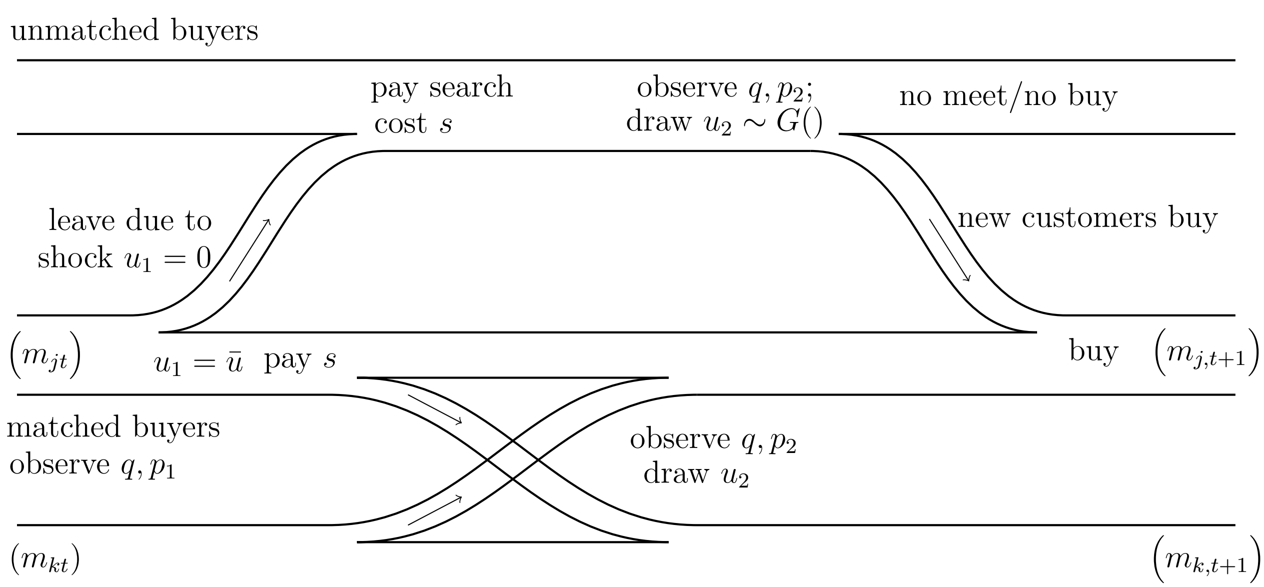
Figure 1.7: Model Timeline - Firm




Notes: This timeline shows a firm's decision and activities at period $t$. Time is discrete. At the beginning, firms invest in R&D and marketing. Next, consumers incur a search cost for a chance to meet a firm. Firms charge prices $p_1$ to the returning buyers and $p_2$ to the new buyers. After customers are settled, firms produce using labor with a linear technology. Finally, firms sell 1 unit of good to each buyer, and the period ends.

are heterogeneous in their preferences over products , which they draw from a distribution $G$ each period. At the beginning of a period, some of the customers (called matched buyers) are in the customer base of the firm they bought from last period, while others are unmatched buyers, who can only meet a firm through searching. The flow of customers in each period is illustrated as follows:

Figure 1.8: Model Timeline - Customers




Notes: This timeline shows a customer's activities within period $t$. The matched customers know the current firm's goods without searching. All customers have a chance to meet a firm. Upon meeting a new firm, the customer observes the good's characteristics, quality and price, and decide whether to buy or not.

Both the unmatched and matched consumers pay a search cost $s$ for a chance to meet a firm in the search process. Since the matched customers consumed at the firm last period, they know the product characteristics, and thus their per-quality utility $u_1$ on the product is a constant, $\bar{u}$ or 0.[6] After innovation has taken place, the customer observes perfectly the firm's product quality $q$ and price $p_1$, and decide whether to consume at the newly matched firm or continue their relationship with the current firm.

After paying the search cost, some of the consumers successfully get matched with a firm, where they observe the quality $q$ and price $p_2$. They also draw their per-quality utility $u_2$ from a known distribution $G$, where $\mathbb{E}(u_2) = \bar{u}$. It is reasonable to assume $u_1$ is less disperse than $u_2$ because experienced customers know the product well and have smaller variance on utility than customers who first bump into it. Based on the information discovered upon meeting, the customer decides whether to buy the good or not. Those who buy will become part of the customer base of the newly matched firm. A customer can meet at most one firm in a period, and thus those who do not meet a firm, or who meet a firm but decide not to buy will remain unmatched and search again in the next period.

### 1.3.2 Investment in Innovation

Firms improve the quality of their current product by investing in R&D. In each period, firms innovate upon a baseline product quality $q$, which is the average product quality of all firms from last period $\bar{q}_{t-1} = \int q_{j,t-1} dj$. This assumption is based on the idea of "learning and forgetting" (See Benkard (2000)). Firms that used to take lead in previous innovations do not necessarily ace in the new generations of technology development. On the other hand, laggard firms may also benefit from the spillover of knowledge. Therefore, it is reasonable to assume firms innovate based on the average level of technology in the economy. Moreover, if we assume firms innovate based on their own quality $q_j$, the large firms even have more advantage compared to their smaller counterparts, because the large firms are usually the most productive ones. The main conclusions about the impact of concentration

[6] The details will be discussed in the "Customer Problem" subsection later.

and productivity growth in my model will still hold, and become even stronger due to this additional benefit of being large.

Upon successful innovation, a firm increases product quality from $q$ to $q(1+\lambda)$, where $\lambda$ is the quality increment in each round of innovation. When investing, firms choose the probability of R&D success $x \in (0,1)$. To achieve innovation rate of $x$, firms employ

$$R(x) = \phi \frac{x^{\eta_x}}{1-x} q \tag{1.2}$$

researchers, where $\phi > 0$ and $\eta_x > 0$. So, the total R&D investment cost is convex in the rate of innovation $x$. In this economy, labor is the numeraire, and therefore the detrended wage $w$ is normalized to 1.

The decision to model R&D as a process of own-product quality improvement is consistent with the findings of Garcia-Macia et al. (2019) that growth mainly occurs through quality improvements rather than new varieties. The specification of cost function reflects the scale effect of R&D. This is built upon the innovation literature that argues ideas are non-rivalrous (Bloom et al. (2020), Jones (1995)) and that intangible inputs (R&D) can be duplicated at close to zero marginal costs (Haskel and Westlake (2017), Hsieh and Rossi-Hansberg (2019)).

### 1.3.3 Product Market and Investment in Marketing

Another essential part of firm operations involves building and maintaining customer base, because it usually costs less to keep existing customers than it does to acquire new ones. Firms that care about customer loyalty has a key measure to track: customer lifetime value (CLV). CLV is the value of a customer to a company, not just on a purchase-by-purchase basis but over the entire customer relationship.

To capture this long-term value of customer relationship, I introduce search frictions into the product market of my model, in which it takes time for an unmatched customer to find a firm. Therefore firms also invest substantially in marketing[7], which helps to increase product exposure to potential buyers. An unmatched customer does not know the existence of a product until they see an advertisement, meet a salesperson, or run into the product in a store of the firm. As a result, marketing investment allows firms to enhance demand at the extensive margin. The advertisement describes the product's characteristics, quality and price, which are important factors for customers to decide whether to buy or not after seeing the ads, and in this way marketing also influences demand at the intensive margin.

Marketing investment also uses labor. Assume decreasing returns in advertising: to post $a$ units of advertisement requires

$$I(a) = \psi a^{\eta_a} q \tag{1.3}$$

of labor, where $\psi > 0$, and $\eta_a > 1$ to have cost convexity. $a$ refers to the quantity of ads received and viewed by consumers. Multiplying the number of ads will not bring in the same amount of views. As consumers have limited attention, reaching additional consumers is more difficult than attracting the first set of consumers. As a result, the number of successful ads that receive potential customers' attention has diminishing returns, which is equivalent to a convex cost structure.

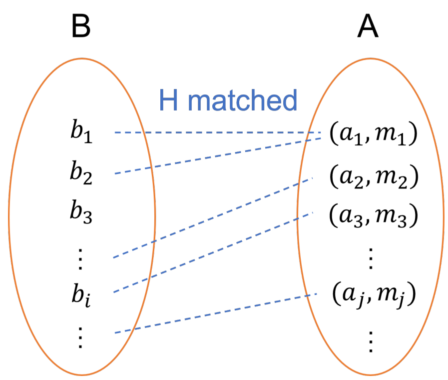
Figure 1.9 depicts the matching process between unmatched customers and firms. Firms are indexed by $j$. Let $A = \int a_j dj$ denote the aggregate amount of advertising from all firms, and $B$ the measure of potential buyers (unmatched), then the measure of total match $H$ is given by

$$H(B, A) = \xi B^{\gamma} A^{1-\gamma}, \tag{1.4}$$

[7] The marketing investment in this paper refers to the broadly defined investment to promote product exposure, including expenditures on purchasing product space, expanding distributional channels, hiring sales people, and posting advertisement. Many of these promotional activities can be done online or advanced by Internet technologies, and therefore experiences a cost reduction.

where $\xi > 0$ and $\gamma \in (0, 1)$. It is convenient to define the average queue length $\theta = B/A$. This allows writing the probability a customer succeeds in matching as a decreasing function of the queue length $H(1, 1/\theta)$. The more aggregate ads are there, the more likely that an unmatched customer is exposed to an ad and become aware of a product she did not know before.

Figure 1.9: Illustration of the Formation of Firm-Customer Relationship.




At the firm level, let $h_j$ be the measure of customers that firm $j$ is matched with, and $\kappa_j$ the share of newly matched customers visiting the firm ($\sum \kappa_j = 1$). Apparently, the more ads a firm posts, the more likely that customers see its ads and miss other firms' ads. Thus, it is reasonable to assume $\kappa_j = a_j/A$, then

$$h_j = \kappa_j H = \frac{\xi B^\gamma}{A^\gamma} \cdot a_j. \tag{1.5}$$

In this specification, $h_j$ is increasing in (i) the aggregate amount of ads $A$, which increases the total measure of matched customers; and (ii) the individual amount of ads posted by the firm relative to other firms.

### 1.3.4 Customer Problem

#### 1.3.4.1 Matched Customers

In the benchmark model, I assume the matched customers do not do "on-the-job" search to gain some tractability. Assume in each period, a fraction $\delta$ of the customer base leaves the matched firm for random reasons, such as moving to a different city. For the rest of matched customers, searching another firm is an outside option of continuing the customer relationship. Assume a fraction $(1-\alpha_1)$ of customers care about quality improvement, and will leave the firm if it cannot upgrade the product because of R&D failure (Think about the "techies" who look for fancy technologies especially when they purchasing e-products). These people have $u_1 = 0$ if R&D fails. The rest $\alpha_1$ share of customers are "loyal" ones who would consider continuing their customer relationship even if the firm fails to make innovations. After innovation is realized, the "techies" in R&D success and the loyal customers have per-quality uility $u_1 = \bar{u}$, and will compare the outside option (expected value of search) $U_s$ and the value of continuing the customer relationship at the firm $U_0$. The latter depends on product quality $q$, the price charged by the current firm $p_1$, and the utility $u_1$.

Due to frictions in the product market, customer relationships become long-term in nature. Unless there is random separation (with probability $\delta$) or R&D failure, the relationship lasts as long as $U_0 \geq U_s$. Since product appeal $u_1$ is the same for all the existing customers, they either all stay ($U_0 \geq U_s$) or all leave ($U_0 < U_s$). To maximize profits on existing customers, firms will charge $p_1$ to make them just indifferent between staying and searching, $U_0 = U_s$, extracting their entire surplus of the match. A customer's value of staying with the firm $U_0$ is given by

$$U_0 = \bar{u}q_j - p_1 + \beta U_0', \quad \text{where } q_j = \begin{cases} \bar{q}(1+\lambda), & \text{if R\&D succeeds} \\ \bar{q}, & \text{if R\&D fails} \end{cases} \tag{1.6}$$

and the price is $p_1 = \bar{u}q_j - U_0 + \beta U_0'$.

#### 1.3.4.2 Unmatched Customers

Unmatched customers pay a search cost $s$ for the chance of meeting a firm. Upon matching with the firm, she observes the quality $q$ and price $p_2$. Similarly to the previous subsection, I assume a fraction $(1 - \alpha_2)$ of customers care about quality improvement, and will have $u_2 = 0$ if the firm fails in their R&D effort. The newly matched customers draw a per-quality utility $u_2$ from a distribution $G(\cdot)$. Firms know the distribution $G$, but they do not know each customer's utility $u_2$, and thus make a take-it-or-leave-it offer $p_2$ to all the customers.

If an unmatched customer meets a firm of size-$m_j$ after searching, her utility of consuming the product will be $u_2 q_j$. She will buy the good only if $u_2 q_j \geq p_2(m_j)$. Thus, a customer's value function of meeting with a size-$m$ firm is

$$U_s(m_j, q_j) = \int_{\frac{p_2(m,q)}{q_j}} \Big[u_2 q_j - p_2(m_j, q_j)\Big] \, dG(u_2) + \beta U_0' \tag{1.7}$$

Thus, the expected value of searching is given by

$$U_s = -s + \int_j \left[\frac{h(m_j, q_j)}{B} U_s(m_j, q_j)\right] dF\left(m_j, q_j\right) + \left(1 - \frac{H}{B}\right) \beta U_0' \tag{1.8}$$

Indifference condition requires $U_0 = U_s$.

### 1.3.5 Firm Problem

#### 1.3.5.1 Value Function

Each period, firms make decisions on the R&D investment $x$, marketing investment $a$, prices to existing customers upon R&D success $p_1^{rd}$ and failure $p_1^{nrd}$, and prices to new customers

$p_2$, to maximize their value function, which reads

$$
\begin{aligned}
V(m,q) = \max_{p_i^{(n)rd},a,x} \; & x\left\{\left(p_1^{rd}-q^{rd}\right)y_e^{rd}+\left(p_2^{rd}-q^{rd}\right)y_n^{rd}+\beta(1-\tau)V\left(m'^{rd},q^{rd}\right)\right\} \\
& +(1-x)\left\{\left(p_1^{nrd}-q\right)y_e^{nrd}+\left(p_2^{nrd}-q\right)y_n^{nrd}+\beta(1-\tau)V\left(m'^{nrd},q\right)\right\} \\
& -\psi a^{\eta_a}q-\phi\frac{x^{\eta_x}}{1-x}q
\end{aligned} \tag{1.9}
$$

$$
\begin{aligned}
s.t.\;\; y_e^{(n)rd} &= (1-\alpha)\left[1-\delta\left(m,q^{(n)rd}\right)\right]m \\
y_n^{(n)rd} &= \int\left[1-G\left(u_2^*\left(m,q^{(n)rd};\tilde{m},\tilde{q}\right)\right)\right]dF(\tilde{m},\tilde{q})\,\frac{\xi B^\gamma}{A^\gamma}a \\
m'^{(n)rd} &= y_e^{(n)rd}+y_n^{(n)rd} \\
g = \frac{\bar{q}'}{\bar{q}} &= \lambda\int x(m,q)dF(m,q)
\end{aligned}
$$

#### 1.3.5.2 Entry and Exit

Entrepreneurs randomly draw an entry cost $\kappa \sim U[0,\bar{E}]$, and start with customer base $m=0$ and average quality $\bar{q}$

$$
\begin{aligned}
V(0,\bar{q};\kappa) = \max_{p_2^{(n)rd},a,x} \; & x\left\{\left(p_2^{rd}-\bar{q}^{rd}\right)y_n^{rd}+\beta(1-\tau)V(m'^{rd},\bar{q}^{rd})\right\} \\
& +(1-x)\left\{\left(p_2^{nrd}-\bar{q}\right)y_n^{nrd}+\beta(1-\tau)V(m'^{nrd},\bar{q})\right\} \\
& -\psi a^{\eta_a}\bar{q}-\phi\frac{x^{\eta_x}}{1-x}\bar{q}-\kappa\bar{q}
\end{aligned} \tag{1.10}
$$

Exiting firms are replaced with entering firms $(0,\bar{q})$. The exiting rate of incumbents is $\tau$, which is equal to the measure of entrants to ensure the existence of a balanced growth path.

### 1.3.6 Equilibrium

To close the model, the total measure of consumers who search is equal to the total population $N$:

$$\int [x(m,q)m + (1 - x(m,q))\alpha_1 m] (1-\delta) dF(m,q) + B = N \tag{1.11}$$

Firms take the aggregate advertising level $A$ and TFP growth rate $g$ as given, and in equilibrium $A$ is equal to the sum of advertising intensity of each firm, and $g$ is the average of R&D intensity multiplied by $\lambda$:

$$\int a(m,q) dF(m,q) = A \tag{1.12}$$

$$\lambda \int x(m,q) dF(m,q) = g \tag{1.13}$$

**Definition 1**. A **balanced growth path equilibrium** is for every $t$, the allocations $\{x(m,q), a(m,q)\}$, the prices $\{p_i^{rd}(m,q), p_i^{nrd}(m,q)\}$, firm distribution $F(m,q)$, the aggregates $\{A, B\}$, and the growth rate $g$ such that

(i) The decision rules of incumbents and entrants solve (1.9) and (1.10);

(ii) Customers choose where to buy to maximize lifetime utility;

(iii) Firm distribution $F(m,q)$ is stationary;

(iv) The individual choices add up to the aggregate $A, g, N$.

## 1.4 Model Mechanism

This section analyzes the main mechanisms of the model. I show in Section 1.3 that firms with more customers invest more in both R&D and marketing. Based on these findings, I discuss in this section how and why the large and small firms react differently to a higher marketing efficiency, and further explain how the distinct responses lead to rising market concentration and lower productivity growth rate.

### 1.4.1 Optimal choices

Detrend the value function by dividing baseline quality $q$ on both sides. Denote $\tilde{V}(m) = \frac{V(m;q)}{q}$, $\tilde{p}_i = \frac{p_i}{q}$, $\tilde{w} = \frac{w}{q}$, and $\tilde{\beta} = \beta(1-\tau)(1+g)$.

Firms set prices on new customers $p_2$ according to

$$1 - G\left(\frac{\tilde{p}_2}{1+\lambda}\right) = \left[\tilde{p}_2 - \tilde{w} + \tilde{\beta}(1-\delta)\tilde{V}'(m'^{rd})\right] G'\left(\frac{\tilde{p}_2}{1+\lambda}\right)\frac{1}{1+\lambda} \tag{1.14}$$

$$1 - G\left(\tilde{p}_2\right) = \left[\tilde{p}_2 - \tilde{w} + \tilde{\beta}(1-\delta)\tilde{V}'(m'^{nrd})\right] G'\left(\tilde{p}_2\right) \tag{1.15}$$

Increasing price $\tilde{p_2}$ increases the per unit profit by the same amount, and thus the marginal benefit is equal to the measure of new customers (recall that each customer buys at most one unit of good). The marginal cost is that it reduces the measure of new customers willing to consume, which has both a direct effect of lowering profits in the current period, and also a long-term effect through the slower accumulation of customer base.

**Proposition 1.** *Firms charge more when they succeed in R&D than when they fail.* $p_1^{rd} > p_1^{nrd}$

**Proof:** Appendix 1.8. Proposition 1 is intuitive because R&D success brings higher product quality, and increases customer's utility of consuming the good. Therefore, customers are willing to pay more for higher quality.

The firm's optimal innovation rate is the solution to the first order condition of the value function with respect to $x$,

$$
\begin{aligned}
\phi\eta_x x^{\eta_x-1}\tilde{w} = & \left\{(\tilde{p}_1^{rd}-\tilde{w})-(\tilde{p}_1^{nrd}-\tilde{w})\alpha_1\right\}(1-\delta)m \\
& + \left\{(\tilde{p}_2-\tilde{w})\left[1-G\left(\frac{\tilde{p}_2}{1+\lambda}\right)\right]-(\tilde{p}_2-\tilde{w})\left[1-G\left(\tilde{p}_2\right)\right]\alpha_2\right\}\frac{\xi B^\gamma}{A^\gamma}a \\
& + \tilde{\beta}\tilde{V}(m'^{rd})-\tilde{\beta}\tilde{V}(m'^{nrd})
\end{aligned}
\tag{1.16}
$$

The marginal benefit of innovation is composed of three parts. The first line of equation (1.16) is the increment in the profits on existing customers due to higher R&D intensity. When R&D is successful, firms charge higher unit price ($\tilde{p}_1^{rd}$ vs. $\tilde{p}_1^{nrd}$) and also sell to more customers ($m$ vs. $\alpha_1 m$). The second line is the increment in profits on new customers. The last term is the difference in the future value of a firm between R&D success and failure. Better product quality is critical to expanding the customer base, for it not only prevents more customers from leaving but also turns more visiting customers into actual buyers, and therefore is likely to generates a higher customer base $m'^{rd} > m'^{nrd}$.

As for the choice of advertising, the first order condition with respect to advertising intensity $a$ reads

$$
\begin{aligned}
\psi\eta_a a^{\eta_a-1} = & x\left[\tilde{p}_2-\tilde{w}+\tilde{\beta}(1-\delta)\tilde{V}'(m'^{rd})\right]\left[1-G\left(\frac{\tilde{p}_2}{1+\lambda}\right)\right]\frac{\xi B^\gamma}{A^\gamma} \\
& + (1-x)\left[(\tilde{p}_2-\tilde{w})+\tilde{\beta}(1-\delta)\tilde{V}'(m'^{nrd})\right]\alpha_2\left[1-G\left(\tilde{p}_2\right)\right]\frac{\xi B^\gamma}{A^\gamma}
\end{aligned}
\tag{1.17}
$$

One additional unit of advertising helps acquire a mass of $\left[1-G\left(\frac{\tilde{p}_2}{1+\lambda}\right)\right]\frac{\xi B^\gamma}{A^\gamma}$ customers upon R&D success and $\alpha_2\left[1-G\left(\tilde{p}_2\right)\right]\frac{\xi B^\gamma}{A^\gamma}$ customers upon R&D failure. The new customers bring more profits in the current period and increase long-term value by enlarging the customer base.

Denote the right hand side of the value function as $H(m, x, a)$.

**Proposition 2. (Supermodularity)** *Firms with a large customer base $m$ invest in* *more innovation $x$ and more marketing $a$.*

**Proof:** Appendix 1.8. Supermodularity is easily checked numerically and appears to hold in all the calibrations. Intuitively, the profit on existing customers is proportional to the customer base. A successful R&D allows the firm to charge a higher price and retain more existing customers than in the case of R&D failure. Therefore, the marginal benefit of innovation is increasing in the customer base $m$. However, due to the scale effect of R&D, the innovation outcome can be applied to all goods at close to zero marginal cost. Therefore, a larger firm chooses higher innovation investment to take advantage of their large customer base.

As for marketing investment, the marginal benefit of marketing is to obtain $\frac{\xi B^\gamma}{A^\gamma}$ measure of new customers. Each of them generates a current profit for the firm, and increases the lifetime value of a larger customer base $m'$. The convexity of function $V$ leads to larger marginal benefit of marketing for large firms. Moreover, large firm's higher likelihood of high quality also enables the firm to charge higher prices and convert more matched customers to really buying the good, than in the case of R&D failure. In this sense, innovation and marketing is complementary. This explains why large firms also invest more in marketing $a$.

### 1.4.2 Impact on Market Concentration

The process of customer base building falls into a positive feedback loop as follows:

Figure 1.10: Positive Feedback Loop of Customer Base Building

$$
\begin{array}{ccccc}
\text{large customer base} & \Rightarrow & \text{high R\&D} & \Rightarrow & \text{high ADV} \\
\Uparrow \; m' = (1-\delta)m + y & & & & \Downarrow \; h = \frac{\xi B^\gamma}{A^\gamma} a \\
\text{more eventual purchases} & & \overset{y = [1 - G(u_2^*)]dF \cdot h}{\Leftarrow} & & \text{more matches}
\end{array}
$$

Firms with a large customer base $m$ invest more in both R&D $x$ and marketing $a$. and more marketing allows large firms to acquire a larger share of customers during the search process. This advantage of large firms is further amplified by their higher R&D, because better quality can convert more matches into purchases. The new customers add to the customer base next period, and the loop starts again from a larger customer base next period. This explains why there is distinction between large and small firm's customer base.

Then consider the effect of a shock on the cost of marketing, reflected by a reduction in $\psi$, which means firms now are able to produce the same amount of ads as before at a lower cost. On a balanced growth path, all firms equate their marginal cost of marketing $MC_a = \psi \eta_a a^{\eta_a - 1} q$ and the marginal value of marketing,

$$\begin{aligned} MV_a =& x \left[ p_2^{rd} - q^{rd} + \beta(1-\tau) V_m(m'^{rd}, q^{rd}) \right] \int \left[1 - G\left(u_2^*\right)\right] dF \frac{\xi B^\gamma}{A^\gamma} \\ &+ (1-x) \left[ p_2^{nrd} - q + \beta(1-\tau) V_m(m'^{nrd}, q) \right] \int \left[1 - G\left(u_2^*\right)\right] dF \frac{\xi B^\gamma}{A^\gamma} \end{aligned} \tag{1.18}$$

which is larger for firms with more customers because of their higher innovation intensity. When $\psi$ declines, there is a sudden drop in $MC_a$ but the marginal value $MV_a$ does not change immediately because it is not directly affected by $\psi$. At this point, all firms have incentives to raise their marketing intensity $a$ to match $MC_a$ with $MV_a$. Assume $\psi$ declines from $\psi_1$ to $\psi_2$, and firm $j$ increases its marketing intensity from $a_{1j}$ to $a_{2j}$. Then $\psi_1 \eta_a a_{1j}^{\eta_a - 1} = MV_{a,j} = \psi_2 \eta_a a_{2j}^{\eta_a - 1}$, which implies $\frac{a_{2j}}{a_{1j}} = (\frac{\psi_1}{\psi_2})^{\frac{1}{\eta_a - 1}}$. So, every firm increases its marketing intensity by the same proportion. This increases the aggregate advertising intensity $A = \int a_j dj$, and therefore enables more customers to get matched with firms in the search. However, the relative share of customers that each firm meets with remains the same as before ($a_{1j}/A_1 = a_{2j}/A_2$), because all firms increase their marketing intensity by the same proportion.

Although each firm obtains the same measure of more matched customers during search and matching, large firms turn more visiting customers into eventual buyers with their advantage in R&D. Their higher likelihood of R&D success attracts more techies to stay with them upon meeting, and the higher product quality also appeals to a wider range of customers. As a result, the variation of innovation across firm size generates different pass-through rate from potential buyers to actual new customers, and thus widens the gap in customer base between large and small firms.

As was discussed in Section 1.3, the innovation investment is positively related to firm size. Although all firms have incentives to increase R&D investment, the enlarging difference in customer base $m$ will result in a larger gap between large and small firms in terms of innovation intensity $x$ as well as prices $p_2$. According to equation (1.18), all of these factors are important determinants of the marginal value of marketing $MV_a$, and thus exert an indirect effect on marketing intensity $a$. With a larger customer base, higher innovation intensity and higher prices, large firms elevate their $MV_a$ by more than their smaller counterparts, and this further amplifies the difference in marketing intensity between large and small firms. This unequal indirect effect on marketing reallocates more customers to large firms during the search, and creates a positive feedback loop on the gap in customer base, R&D and prices through equation (1.17). The spiral is further amplified through the long-term customer relationship, and eventually gives rise to superstar firms that take up a substantial market share.

### 1.4.3 Impact on Productivity Growth

The impact on aggregate productivity growth rate is more complicated because there are offsetting effects on firm innovation decisions. In response to a rise in marketing efficiency (lower $\psi$), each firm increases its marketing intensity which brings the firm more new customers in the search. With a larger customer base, all firms have incentives to increase their innovation intensity $x$. Nevertheless, due to the enlarging gap of marketing intensity across firm size as aforementioned, relatively more new customers are exposed to large firms' advertisement than in the higher-$\psi$ economy. In other words, the largest firm's share of new customers in the match is higher while that of the smallest firm is lower than before. Moreover, there is a limited capacity of consumption because of the decreasing return of advertising in the matching technology $Y = \xi B^{\gamma} A^{1-\gamma}$. Each customer purchases at most one unit of product in a period. Even if every unmatched customer gets matched during the search, the total purchase would be $N$ units. Under this circumstance, the uneven increase in marketing creates congestion in the product market, where the smallest firms might end up matching with fewer new customers than before. A smaller customer base discourages firms from innovation, and the resulting lower product quality further reduces the incentives of advertising and slows down the accumulation of customer base, which puts the small firms at a long-term disadvantage compared to the large ones. The lower R&D incentives by small firms negatively affects the productivity growth rate.

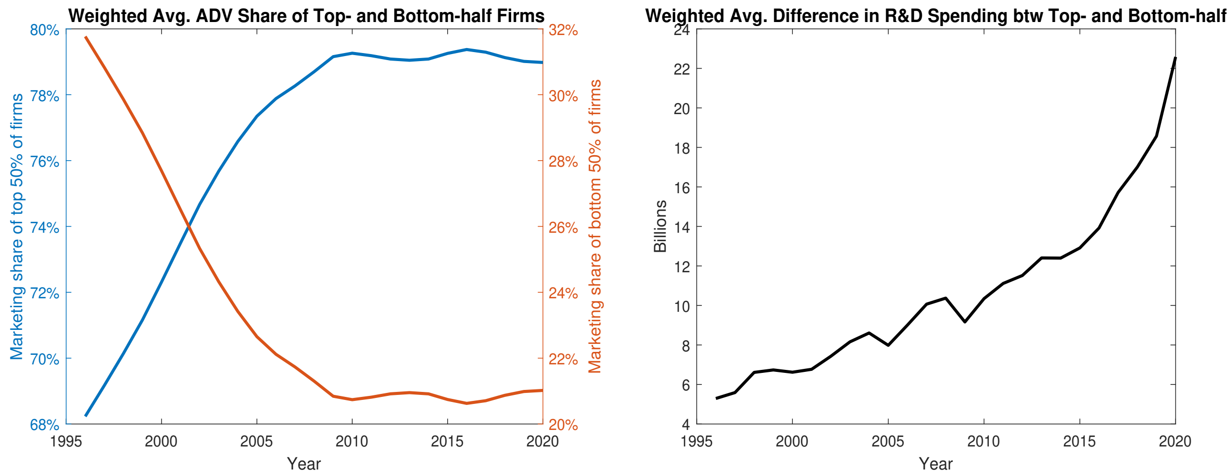
The analysis above indicates opposite forces on productivity growth: large firms increase their innovation intensity while small firms tend to invest less in R&D. In addition, there is also a third effect on innovation decisions through the labor market. Higher marketing efficiency facilitates matching during product search and generate more purchases. The boosted labor demand for production increases wages and also shifts more labor from R&D to production, both of which depress firm R&D investment. Due to these offsetting factors, the overall impact on productivity growth is ambiguous. In the calibrated model, the latter forces dominate and firms on average choose a lower rate of innovation.

### 1.4.4 Empirical Validations

The model predictions of enlarging gap in marketing and innovations are consistent with the observed trends in firm investment data. The left panel of Figure 1.11 shows the average share of the marketing investment of the top-half and bottom-half firms in Compustat, weighted by their industry sales. This ratio increases from 68% in 1996 to 80% in 2020.

The right panel shows the difference between large and small firms' R&D expenditures is also widened in the past 20 years, validating the predictions that large firms disproportionately increase their innovation and marketing investment compared to smaller ones.

Figure 1.11: Marketing and Innovation Investment of Small and Large Firms




Notes: This figure demonstrates the evolution of marketing and innovation investment of small and big firms from 1996 to 2019.

## 1.5 Quantification

In this section, I calibrate the model on the balanced growth path by matching various moments for the U.S. economy in the late 1990s (1996-1999), using data on all U.S. non-financial listed firms from Compustat as well as aggregate moments. Using this initial calibration I describe firms' pricing and innovation strategies to develop intuition about the model. I then calibrate the rise in marketing efficiency (the reduction in $\psi$) to fully capture the change in aggregate real marketing investment in the data.

Using the calibrated model, I conduct three exercises to infer changes to the economy between late 1990s and today. The first analysis displays how firm investments in R&D and marketing change in respond to the rise in marketing efficiency. Specifically, I show a sharp contrast between large and small firms. Second, I quantify the aggregate effects of advertising shock on innovation, economic growth and market concentration on the balanced growth path. Finally, I demonstrate the dynamics of transition from the low to high marketing equilibrium.

### 1.5.1 Calibration

The model is calibrated at an annual frequency. There are three parameters calibrated externally. I calibrate the curvature of R&D ($\eta_x$) to 2 following the literature of Acemoglu et al. (2018) and Akcigit and Kerr (2018). I calibrate the cost elasticity of advertising ($\eta_a$) to 2. This is a key parameter because it determines the concavity of the return to advertising. If marketing activities concentrate among fewer firms, the fact that $\eta_a > 1$ implies that the average effect of these investments on growth is lower. I will do a robustness check of using other values of $\eta_a$ in the Appendix. The discount rate $\beta$ is set to 0.95.

The statistics for firm entry rate and residents moving rate can be found directly. They are 9.6% and 11.2% respectively. The remaining seven parameters are estimated using indirect inference by matching moments from the U.S. Compustat data. I use the Genetic Algorithm to choose combinations of parameters within broad bounds on their possible values. Using the equilibrium values for innovation and entry rates, the firm-size distribution, rates of creative destruction and aggregate quantities such as the efficiency wedge, wages and output, I simulate the economy for 20,000 firms until the distribution of has converged, and simulate data for five more years to collect moments on the simulated sample. The Genetic Algorithm then updates the combinations of parameters based on a comparison of the theoretical and data moments along the following objective function:

$$\min \sum_{k=1}^{7} \frac{|\text{model}_k - \text{data}_k|}{(|\text{model}_k| + |\text{data}_k|)\, 0.5} \Omega_k \tag{1.19}$$

where the weights $\Omega = 2$ for TFP growth rate and leader market share, and $\Omega = 1$ for the rest of targeted moments.

Table 1.2 display the targeted moments in model and data. The generalized model does not yield an analytical solution, and thus we cannot express the targeted moments in this form. Table 1.3 presents an overview of the calibrated parameters. In fact, each targeted moments is jointly affected by multiple parameters.

Table 1.2: Model Fit for Targeted Moments

| Targeted Moments | Model Value | Data Value |
|---|---|---|
| TFP growth rate | 1.52% | 1.51% |
| Leader market share (3-digit SIC) | 24.04% | 24.04% |
| Avg. R&D/Advertising | 4.03 | 3.85 |
| Avg. R&D/Sales | 4.79% | 4.28% |
| Avg. customer turnover rate | 14.6% | 10%-20% |
| Avg. profit share | 5.24% | 5.31% |
| Labor share of income | 0.69 | 0.67 |

Notes: This table displays the model fit for targeted moments from calibration of seven parameters for the late 1990s.

Table 1.3: Overview of Estimated Parameters

| Parameter | Description | Method | Value |
|---|---|---|---|
| $\beta$ | Discount factor | External | 0.95 |
| $\eta_a$ | Cost elasticity of advertising | External | 2 |
| $\eta_x$ | Cost elasticity of R&D | External | 2 |
| $\bar{u}$ | Product appeal | TFP growth rate | 2.5 |
| $\xi$ | Matching function coefficient | Leader market share | 0.95 |
| $\psi$ | Cost of advertising | Avg. R&D/Advertising | 4.98 |
| $\phi$ | Cost of R&D | Avg. R&D/Sales | 2.0 |
| $\alpha$ | Customer loyalty | Customer turnover rate | 0.4 |
| $N$ | Total population | Labor share of income | 0.499 |
| $\gamma$ | Matching function elasticity | Avg. profit share | 0.5 |
| $\delta$ | customer depreciation rate | Share of residents moving | 0.008 |
| $\tau$ | Firm exit rate | BDS firm entry rate | 0.096 |

Next, the parameters $\gamma$ and $\xi$ of the matching function are determined based on evidence on the share of time spent in buying, selling and production activities at the aggregate level. Our target for time spent in selling relative to production is 7.15%, and our target for time spent in buying relative to selling 25%. (Note that time devoted to production well exceeds that devoted to buying and selling.) To arrive at these targets, we use data on the share of the labour force in sales-related occupations from the Occupational Employment Statistics (OES) survey, and the amount of time consumers spend shopping from the American Time Use Survey (ATUS). According to the OES survey, 11% of U.S. workers are employed in sales-related occupations. Examples of such occupations include sales representatives, advertising agents, retail salespersons, cashiers, and real estate brokers. Because workers in other occupations are likely to spend a share of their time in selling activities also, we attribute 10% of their time to selling as well. Examples of other occupations with a significant selling component are advertising and promotions managers, marketing and sales managers, as well as waiters. Overall, this implies that 20% of working time is spent in selling activities. Finally, in reality not all of this time is spent on new customers. To take this into account, we attribute a third of selling time to new customer acquisition. With 6.67% of market labour devoted to selling, time spent on selling relative to production amounts to 7.15%. Turning to resources spent on buying, time-use data document that Americans spend on average 0.4 hours per day shopping. If we again attribute a third of this time to the new-customer margin, our target for buying time becomes 0.56% of total time. With a third of total time spent on market labour (production and sales), buying relative to selling becomes 25%.

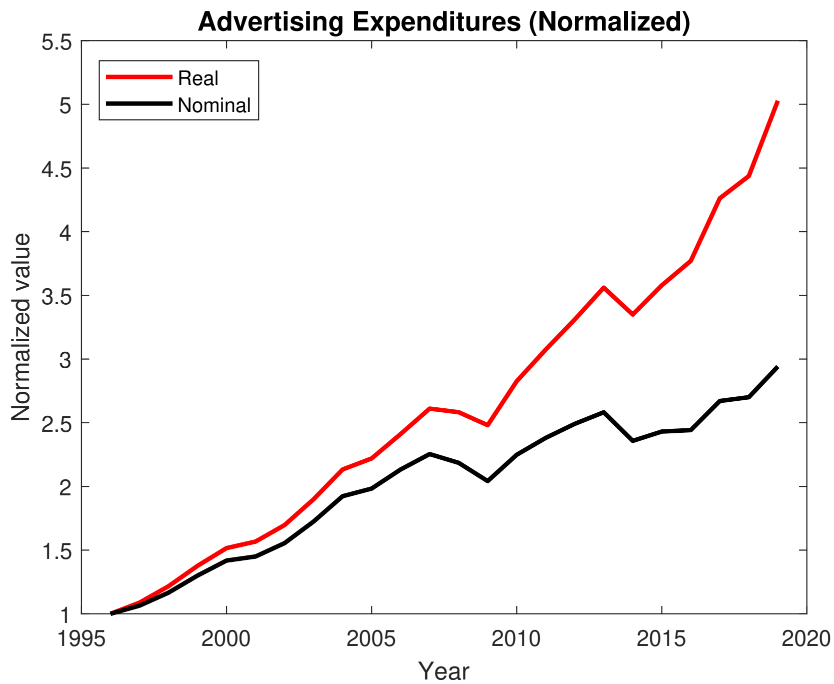
Then I calibrate the decline in marketing cost $\psi$ from 1996 to 2019 to fully capture the increase in real aggregate marketing expenditures.

As aforementioned in Section 1.2, digital technologies lowers the cost of advertising and induces firms to invest more in marketing and selling activities. From 1996 to 2019, the total nominal spending of advertising of Compustat firms increased by 3 times. And taking into account the decline of advertising price, the real expenditures increased by 5 times, as is shown in Figure 1.12. This requires efficiency parameter of advertising $\psi$ to decrease from 4.98 to 0.0218 in the quantitative model.

### 1.5.2 Results: on the Balanced Growth Path

The effect of increasing marketing efficiency as reflected by the reduction of $\psi$ is summarized in Table 1.4. It presents the variables of interest in differences from the original balanced growth path. The change in real expenditures on advertising is targeted. I compute the

Figure 1.12: Firm Aggregate Advertising Investment




Notes: This figure depicts the evolution of firm aggregate advertising investment. Black: nominal spending. Red: real spending by dividing the nominal spending with the average price index of advertising. Data source: Compustat variable "XAD".

changes in data that are explained by a change in $\psi$ as follows:

$$M_j\left(\boldsymbol{\theta}_{1990s}, \psi_{low}\right) - M_j\left(\boldsymbol{\theta}_{1990s}, \psi_{high}\right) \tag{1.20}$$

where $M_j$ is moment $j$ in the model BGP with the other parameters $\theta$ held fixed at their estimated 1990s values.

Table 1.4: Balanced Growth Path Change due to Plummet of Marketing Cost

| | Targeted | $\Delta$ Data | $\Delta$ Model |
|---|---|---|---|
| Expenditures on advertising | Yes | 5.08 | 5.08 |
| Productivity growth rate | No | -0.98 pp | -0.31 pp |
| Leader market share | No | 5.15 pp | 4.25 pp |

The comparison between the new and old balances growth path shows that the change in marketing cost predicts a 0.37 percentage-point drop in productivity growth, and an increase of 4.25pp for the top 1% of firms' market share[8]. As a result, the model can explain around 1/3 of the decline in productivity growth and about 83% of rise in aggregate concentration. It therefore seems that technological change in marketing are not responsible for most of the slowdown of growth in the US. This is reasonable because the technological progress in marketing increases production and consumption, which boosts firms' incentives to make innovations.

### 1.5.3 Results: Composition Effect in Rising Concentration

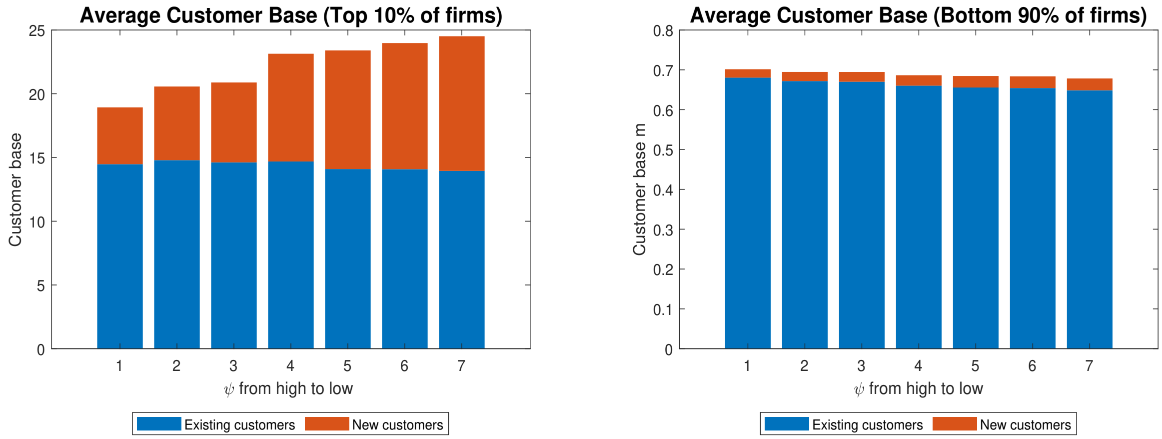
#### 1.5.3.1 Distribution of Customer Base

Each period, the customer base is accumulated by adding new customers to the original customer base, adjusted for random separations. Lower $\psi$ saves advertising cost for all firms, but more so for large firms because of their higher innovation investment and the complementarity between product quality and measure of new matches. Therefore, more new customers are reallocated to the large firms. This is graphically shown in the red components of the bar plot in Figure 1.13, where as $\psi$ drops, both the top 10% and bottom 90% of firms have an increasing number of new buyers, but the former has a much higher fraction than the smaller ones.

The blue components in Figure 1.13 display the measure of existing customers in customer base formation. A lower $\psi$ has two opposite effects on top firms' existing customers. On one hand, easier matches due to higher marketing intensity increases the option value of searching and thus attract more matched customers to leave the previously matched firms. This reduces large firms' measure of existing customers. On the other hand, as $\psi$ drops, large firms have an increasing number of new customers each period and are more likely to reach a larger customer base. The left panel of Figure 1.13 indicates that in the calibrated model, the former dominates the latter and the size of existing customers

[8] There is around 100 firms within each SIC-3 industry, so the largest firm is approximately the top 1% of firms.

Figure 1.13: Changes in Customer Base of Large and Small Firms



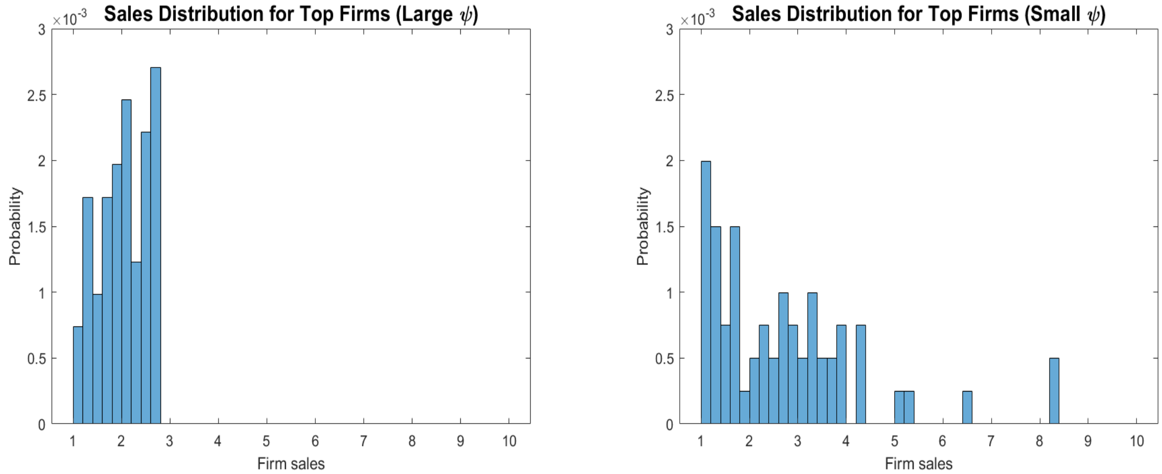

shrinks a little bit. This makes sense because the initial tech shock happens to advertising which directly shifts customers from staying to searching. In spite of this, the massive increase in new customers outweigh the minor drop in existing ones, so the total customer base of large firms are getting bigger with the enhanced marketing efficiency. As for the small firms, the two effects are in the same direction: more customers choose to search rather than continue the customer relationship, and fewer new customers bump into their advertisement, both of which lead to shrinking customer base.

Figure 1.14 demonstrate the distributions of firms' sales revenue for the top 10% of firms on initial (left panel) and new balanced growth paths (right panel)[9]. The efficiency change in marketing gives rise to a fatter-tailed distribution. The economy now is populated with more superstar firms with larger number of customers and higher revenues.

In addition to the customer base, changes in prices can affect firms' market share as well. However, a lot of studies show that the rising market share of large firms is not driven by a larger gap in prices or markups between large and small firms. Rather, it is the composition effect (reallocation of resources towards large productive firms) that explains most of the rising superstar firms. For example, Baqaee and Farhi (2020) and De Loecker et al. (2020)

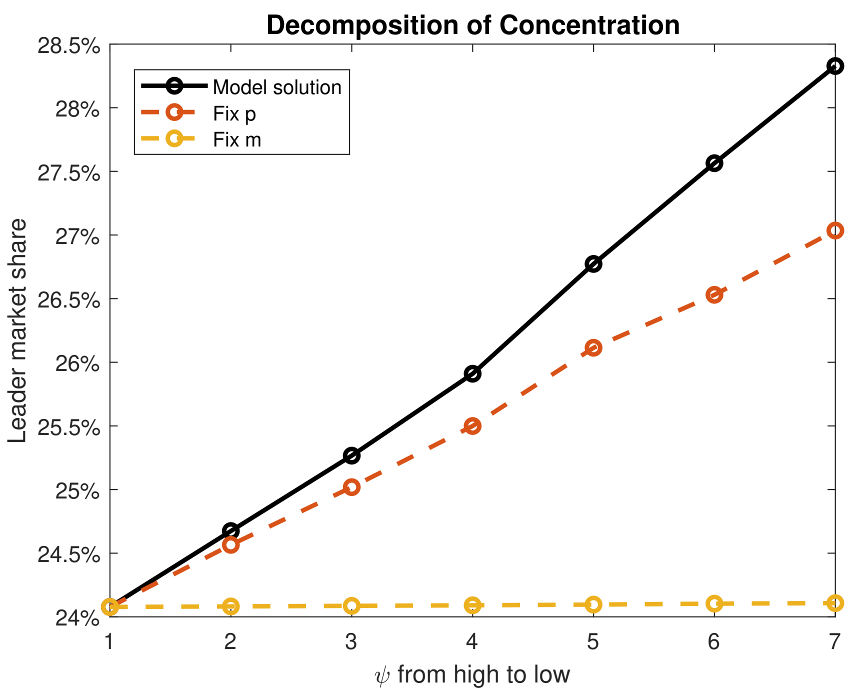
[9] It is not easy to compare the two BGP by viewing the entire distribution because the distribution is highly fat-tailed and the main difference is on the right tail. The entire distribution can be seen in Appendix A.

Figure 1.14: Sales Distribution for Initial and New Steady States




find that average markups have been increasing primarily due to a between-firm composition effect, whereby firms with high markups have been getting larger, and not to a within-firm increase in markups.

Figure 1.15: Decomposition of Concentration



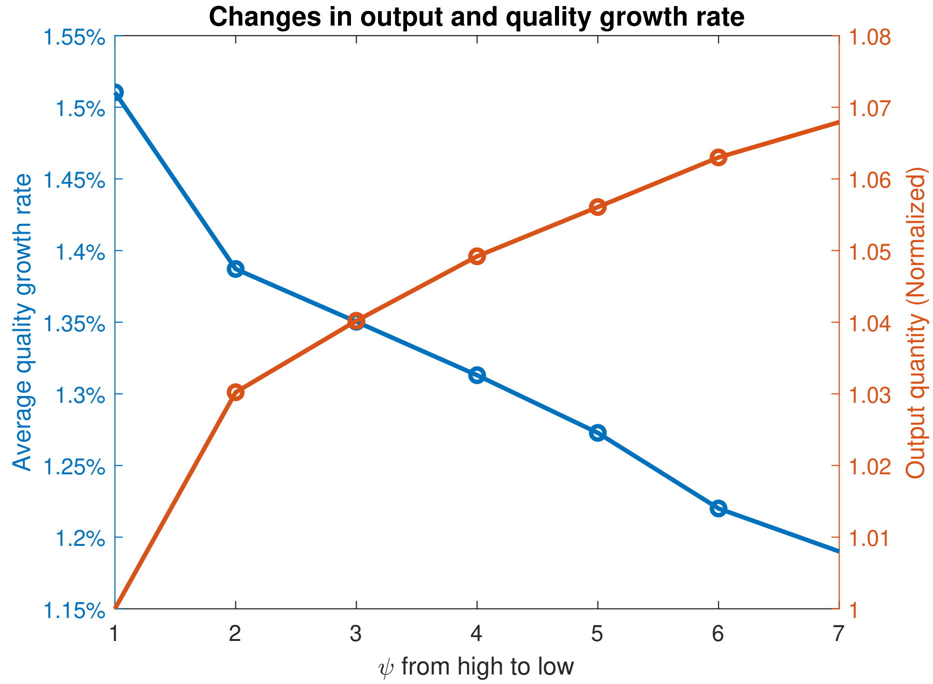

Notes: This figure presents the changes in market concentration when marketing cost $\psi$ declines. The black curve depicts the model solution. The red curve is obtained by fixing firms' price choices at the initial level but letting firms accumulate customer base as the model implies. The yellow curve is obtained by fixing firms' customer base at the initial level but letting firms choose the price levels.

To understand whether the enlarging gap between large and small firms' customer base is a main driver of rising concentration, I conduct a counterfactual exercise to decompose the effect on rising leader market share into customer base ($m$) and prices ($p_1$, $p_2$). I study the effect for seven different values of marketing cost $\psi$. $\psi = 1$ corresponds to the initial balanced growth path where $\psi = 4.98$; and $\psi = 7$ corresponds to the new balanced growth path with $\psi = 0.0218$. Figure 1.15 shows the decomposition. The red dashed curve is obtained by fixing the prices charged to the same level as on the $\psi = 1$ balanced growth path, while allow firm customer base to change as on the new balanced growth paths. As $\psi$ declines, the enlarging gap in customer base raises the market share of large firms. On the new balanced growth path, it explains 2.96 percentage-point rise in top 1% firms' sales share, which accounts for 70% of the total rise predicted by the model (the black curve). Next, fix the customer base as the $\psi = 1$ BGP and change prices, and the results are shown by the almost flat yellow dashed curve, which only explains a rise of 0.3 percentage-point out of the total 4.25 percentage-point. This suggests that the market concentration is mainly driven by the enlarging gap in customer base between large and small firms.

### 1.5.4 Results: Productivity Growth

Next, I study the quantitative effect of lower marketing cost on productivity growth rate. As was discussed in Section 1.4, the increase in marketing activities facilitates more matches, which raises the aggregate output quantity and has a positive effect on productivity growth. This is the "level effect". On the other hand, firm investment decisions on R&D have also been modified. Small firms invest less in innovations because of a lower price and smaller customer base. Large firms now obtain more customers in total. However, as marketing becomes cheaper, demand from new customers become more "elastic" because they can more easily find another firm in the economy, firms have to lower prices to retain more customers. Hence, there is ambiguous effect on their innovation. The changes in the rate of quality improvement is the "growth effect".

Figure 1.16: Level Effect and Growth Effect in Productivity Growth



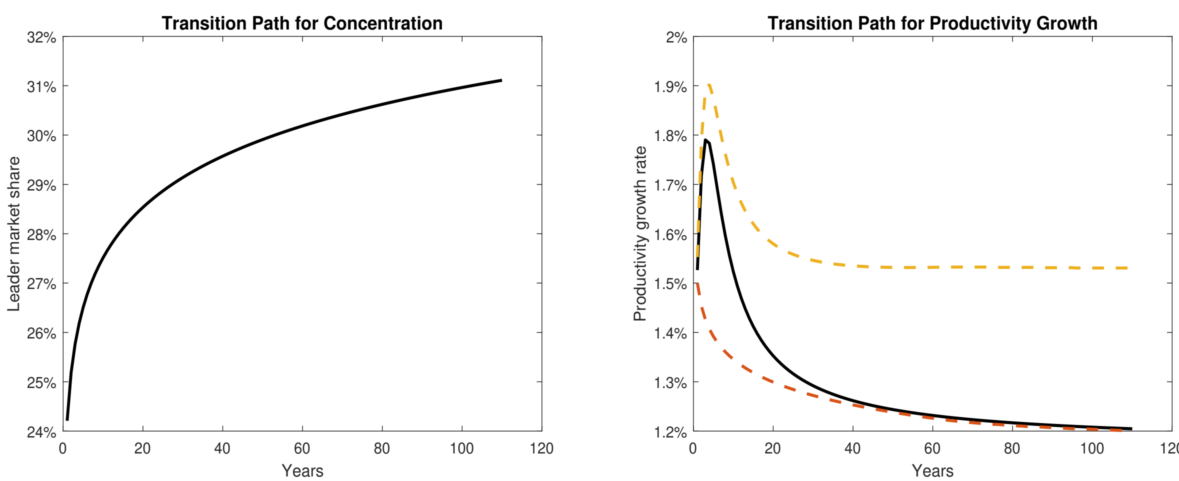

In Figure 1.16, the red curve accounts for a "level effect" of marketing: as marketing becomes cheaper ($\psi$ decreases), firms expand demand contemporaneously, which increases the perceived utility derived from aggregate consumption. Compared to the low-marketing equilibrium, output in the new economy increases by 6.79%. Additionally, marketing critically shapes firm R&D incentives and, therefore, has an indirect effect on growth (the "growth effect"). Although it is theoretically ambiguous whether the large firms will increase or decrease innovation, the calibrated model demonstrates a declining average innovation growth rate from 1.51% to 1.19%, implying the the negative effect of small firms and price drop dominates the rise of customer base in large firms.

### 1.5.5 Results: Transition Dynamics

The analysis thus far has studied the effect of a lower marketing cost along the balanced growth path. This section shows that short-term dynamics are substantially different. To quantify the transition path, I numerically solve for the path of top 1% firms' market share and productivity growth[10]. To understand the transition dynamics from the low-marketing to high-marketing equilibrium, I consider a permanent decrease in the cost parameter $\psi$ as documented in the previous section. I assume the transition takes 120 years, with the marketing efficiency $1/\psi$ rising at a steady pace.

Over the transition path, the market leader share (left panel of Figure 1.17) increases rapidly at the beginning and gradually grows to the new balanced growth path. As for aggregate productivity growth (right panel of Figure 1.17), there is an initial surge in the growth rate for 5 years, and then it gradually declines to a growth rate lower than the initial value as it reaches the new balanced growth path. This "first-rise-then-fall" pattern is consistent with the trend in the right panel of Figure 1.1, which shows the TFP growth rate experienced a temporary boom during the late 1990s and then declined all the way down to today.

Figure 1.17: Transition Dynamics for Marketing and R&D



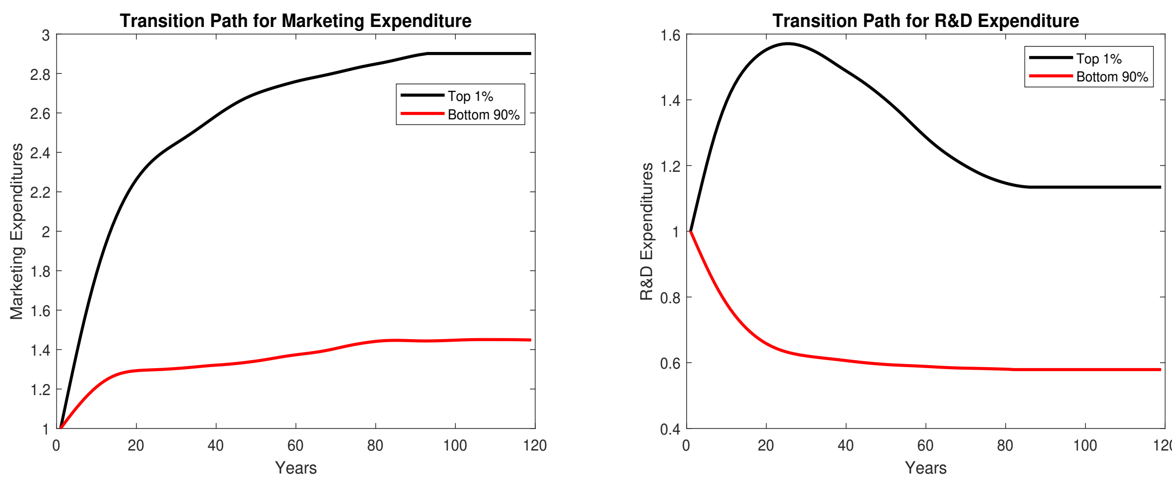

Notes: This figure depicts the model-implied transition paths of weighted average concentration and TFP growth rates from 1996 to 2019.

To understand the mechanism behind the hump shape immediately after the marketing shock, I decompose the change in productivity growth into the level effect and the growth effect. In the right panel of Figure 1.17, the yellow dashed curve is obtained by keeping fixed the average quality growth rate $g$ to its initial level on the high-$\psi$ balanced growth path,

[10] The computational algorithm is described in Appendix B.

and allowing the output quality to change as it is on the transition path. The productivity growth increases by 37 basis point in the first few years right after marketing becomes cheaper. Then the positive effect on growth gradually fades out as the economy converges to the new balanced growth path. Next, I keep the output quantity fixed and let R&D investment change on the transition path and find that the average innovation growth rate monotonically declines from 1.51% to 1.19%. The long-run TFP growth on the new balanced growth path is the same as the innovation growth rate as the output will eventually become stable over time. This implies that the increase in overall TFP growth at the beginning is all driven by the increase in output quantity due to more matches in the product market.

Figure 1.18: Transition Dynamics for Marketing and R&D




Notes: This figure depicts the model-implied transition paths of marketing and innovation investment from 1996 to 2019.

Figure 1.18 demonstrates the transition dynamics of marketing and R&D expenditures for large and small firms, respectively. The left panel indicates that large firms increase their marketing expenditures by much more than small firms, which is consistent with the trends in data. The right panel shows that large firms increase innovation investment in the first 20 years or so, because of their large customer base, but the will gradually downsize their R&D investment as they have to offer lower prices to retain customers. As for small firms, there is always an adverse effect, leading to less innovation efforts.

## 1.6 Welfare Analysis

### 1.6.1 Welfare Effect of Marketing

In Section 1.5, I pointed out that there are two offsetting effect of marketing: (i) overcomes search frictions and allow more customers to buy products they like; (ii) discourages firms from innovation and thus harm the growth in the calibrated model. In this section, I study the consequences of lower marketing cost on welfare. There has been a long-lasting debate in economics about the welfare implications of advertising.

In this paper, following the classic literature of Nelson (1974), Butters (1977), Grossman and Shapiro (1984) and Milgrom and Roberts (1986), I focus on the informative role of marketing/advertising, where advertisement helps remove information frictions by providing relevant information about product characteristics and quality, or simply about the existence of the product. Informative advertising weakens the negative effect of information frictions and is usually associated with welfare improving.

On the other hand, advertising is also a taste shifter that firms use to steal customers from each other and maintain their own market shares (See e.g. Dixit and Norman (1978), Becker and Murphy (1993), Benhabib and Bisin (2002), Benhabib and Bisin (2011) and Molinari and Turino (2009)). In this case, marketing expenditures are a pure combative tool for a limited number of customers and thus might result in a waste of resources. Moreover, in my model, the decline of marketing cost also affects firm's R&D decisions, which further create another side effect from the perspective of productivity growth rate. Given both the pros and cons, it is important to quantify the welfare implications of the declining cost of marketing.

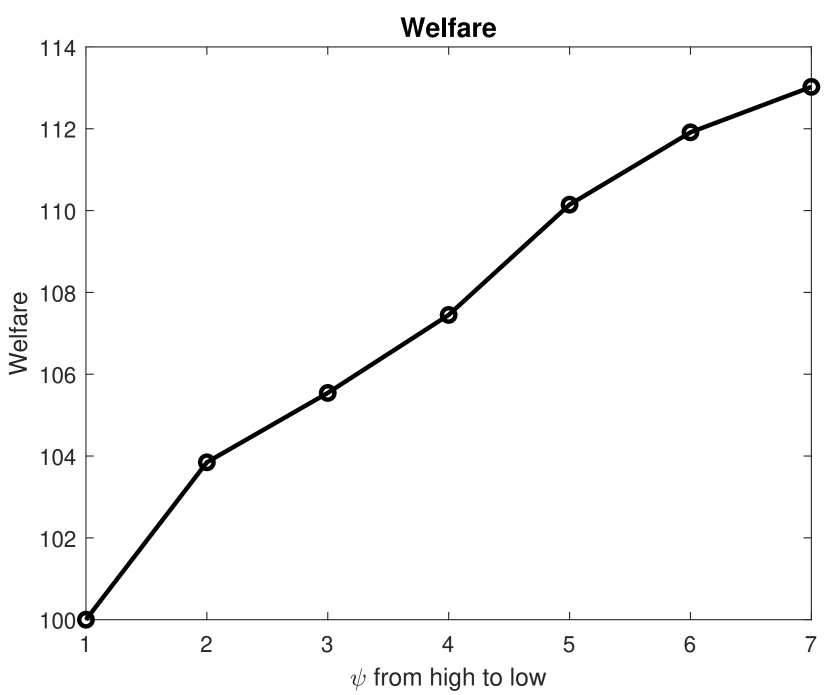
In this economy, labor is the numeraire and customers have a quasi-linear utility function. The welfare effect of the decline of marketing cost is given by the change in the discounted sum of consumption utility. The total welfare in the economy can be written as

$$\mathbb{E}\sum_{t=0}^{\infty}[\beta(1-\tau)(1+g)]^t \int \Bigg\{ x_{jt}\left[\bar{u}(1+\lambda)m_{jt} + \int_{u_{2j}^{*}{}^{rd}} u_2(1+\lambda)dG(u_2)\frac{\xi B_t^{\gamma}}{A_t^{\gamma}}a_{jt}\right] + (1-x_{jt})\left[\alpha_1\bar{u}m_{jt} + \alpha_2\int_{u_{2j}^{*}{}^{nrd}} u_2 dG(u_2)\frac{\xi B_t^{\gamma}}{A_t^{\gamma}}a_{jt}\right]\Bigg\} dj - sB_t \tag{1.21}$$

where $u_{2j}^{*}$ stands for the threshold of utility above which the matched customers will buy. As $\psi$ decreases, the trade-off is between a decline of long-run growth rate $g$ and a rise in the output driven by more matched new customers $a$.

Figure 1.19 shows that the level effect dominates the growth effect, and the overall welfare increases as marketing cost drops. This is reasonable considering that the shock is a technological progress that overcomes information frictions in the market. Despite the negative side effects on growth and business stealing, the first-order effect of facilitating matches dominates and raises total welfare.

Figure 1.19: Change in Welfare with the Decline in $\psi$



### 1.6.2 Constrained Efficiency

Despite the increase in welfare, trade-off between the level effect and the growth effect is still present. Is the heavy spending in advertising socially efficient? Does the decentralized equilibrium optimally allocate resources between innovation and marketing? To answer these questions, I derive the social planner's problem.

The goal of the social planner is to maximize lifetime utility of all customers subject to the search frictions and technological constraints of the economy. The planner generally does not set prices to customers, but to maintain the market structure where existing and new customers co-exist, I assume there are imaginary prices that make the existing customers indifferent between searching and continuing the relationship. This implies that there is still a threshold on product utility $u_2^{*(n)rd}$ for the new buyers[11]. With this restriction in the market, the social planner chooses the optimal marketing $a(m)$, innovation $x(m)$ and the thresholds $u_2^{*(n)rd}$ to solve the following maximization problem:

$$\max_{a_{jt},x_{jt},u_{2j}^{*}{}^{(n)rd}} \mathbb{E}\sum_{t=0}^{\infty}[\beta(1-\tau)(1+g)]^t \int \left\{ x_{jt}\left[\bar{u}(1+\lambda)(1-\delta)m_{jt} + \int_{u_{2j}^{*}{}^{rd}} u_2(1+\lambda)dG(u_2)\frac{\xi B_t^{\gamma}}{A_t^{\gamma}}a_{jt}\right]\right.$$
$$\left. +(1-x_{jt})\left[\alpha_1\bar{u}(1-\delta)m_{jt} + \alpha_2\int_{u_{2j}^{*}{}^{nrd}} u_2 dG(u_2)\frac{\xi B_t^{\gamma}}{A_t^{\gamma}}a_{jt}\right]\right\} dj - sB_t \quad (1.22)$$

$$s.t. \quad \int a_t(m_{jt})dj = A_t \quad (1.23)$$

$$B_t + \int m_{jt}dj = N \quad (1.24)$$

$$\int \left\{\phi x_{jt}^{\eta_x} + \psi a_{jt}^{\eta_a} + x_{jt}m_{j,t+1}^{rd} + (1-x_{jt})m_{j,t+1}^{nrd}\right\} dj = N \quad (1.25)$$

$$m_{j,t+1}^{rd} = (1-\delta)m_{jt} + y_{jt}^{rd},\ y_{jt}^{rd} = \left[1-G\left(u_2^{*rd}\right)\right]\frac{\xi B_t^{\gamma}}{A_t^{\gamma}}a_{jt} \quad (1.26)$$

$$m_{j,t+1}^{nrd} = (1-\delta)\alpha_1 m_{jt} + \alpha_2 y_{jt}^{nrd},\ y_{jt}^{nrd} = \left[1-G\left(u_2^{*nrd}\right)\right]\frac{\xi B_t^{\gamma}}{A_t^{\gamma}}a_{jt} \quad (1.27)$$

[11] Absent this assumption, the planner would always want the matched customers to buy because their $u_2 \geq 0$ and there is not price $p_2$.

**Proposition 3.** *The constrained efficient allocation specifies lower marketing investment and higher innovation investment compared to the decentralized equilibrium: $a_j^{sp} < a_j$, $x_j^{sp} > x_j$.*

**Proof:** Appendix 1.8. Intuitively, marketing creates a negative "congestion externality", where one firm's marketing reduces all other firms' chance of meeting customers $\frac{\xi B^\gamma}{(A_{-j}+a_j)^\gamma}$, and therefore in the decentralized economy firms over-invest in advertising.

### 1.6.3 Policy Implications

The results so far raise some questions in terms of policy implications, especially regarding advertising. In this section, I study whether a tax on marketing or a subsidy on firm R&D activities could be welfare-improving. In particular, I focus on linear taxes and subsidies. The revenues from taxes are rebated back to the consumers, and subsidies are financed through lump-sum taxes.

With the quasi-linear utility function, the welfare function with marketing tax is

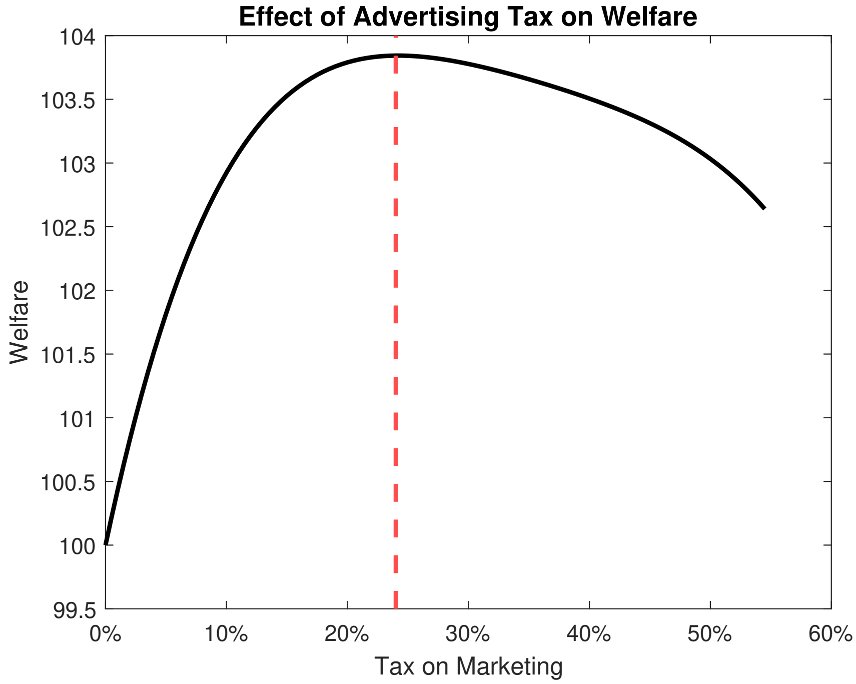
$$\mathbb{E}\sum_{t=0}^{\infty}[\beta(1-\tau)(1+g)]^t \int \left\{ x_{jt}\left[\bar{u}(1+\lambda)m_{jt} + \int_{u_{2j}^{*}{}^{rd}} u_2(1+\lambda)dG(u_2)\frac{\xi B_t^\gamma}{A_t^\gamma}a_{jt}\right] \right.$$
$$\left. +(1-x_{jt})\left[\alpha_1\bar{u}m_{jt} + \alpha_2\int_{u_{2j}^{*}{}^{nrd}} u_2 dG(u_2)\frac{\xi B_t^\gamma}{A_t^\gamma}a_{jt}\right]\right\} dj + T_t - sB_t \tag{1.28}$$

where the proceeds of marketing tax is rebated to consumers. Assume the linear tax rate on marketing is $\varepsilon$, then

$$T_t = \varepsilon \int \psi a_j^{\eta_a} dj \tag{1.29}$$

There are several forces associated with taxation of advertising that go in opposite directions regarding welfare. I find that there exists an optimal level of tax on marketing that maximizes welfare equal to 24% (see Figure 1.20). This tax is associated to a 0.24% increase in growth and a 0.54% reduction in quantity of output, and an overall increase in

Figure 1.20: Effect of Marketing Tax on Welfare




Notes: This figure depicts the effect of a marketing tax on welfare.

welfare of 3.8%.

## 1.7 Conclusion

This paper documents a large decline in the cost of marketing and advertising due to the development of digital technologies and the rise of online marketing platforms. The average price index of advertising has plummeted by 40% in the past 20 years. In response to this technological change, firms invest more in marketing and advertising, with the large firms increasing their investment by much more than their smaller counterparts.

To understand the macroeconomic consequences of lower marketing cost, I develop an endogenous growth model in which the product market is frictional, and customers are a valuable long-term asset to firms. Firms invest in marketing and advertising to increase the visibility of their products to customers, and invest in R&D to increase product quality. Large firms invest more in both R&D and advertising due to the complementarity between these two types of investments.

As marketing becomes cheaper, it is less costly for all firms to make advertisement and match with new customers. However, since large firms invest more in R&D and generate better product quality, they convert a higher proportion of customers to buyers, and gain a higher return of advertising investment. Therefore, more new customers are reallocated to the large firms. This widens the gap in customer base between large and small firms, which further affects future R&D and marketing investment, giving rise to increasing concentration and the rise of superstar firms.

Rising concentration caused by cheaper marketing does not necessarily do harm to aggregate productivity. Instead, it generates a positive level effect due to more efficient matching in the product market, and a negative growth effect mainly due to lower innovation incentives of small firms. These offsetting factors make effects on productivity growth ambiguous.

The quantitative analysis implies that the declining marketing cost is a critical driving force of rising marketing concentration: it explains 83% of the rise in the largest firm's market share within 3-digit SIC industries. Moreover, consistent with other classical studies, rising market concentration is mainly driven by the enlarging gap in customer base across firm size rather than large firms charging higher markups.

Nevertheless, lower marketing cost generates mixed effects on productivity growth. On average, it only accounts for 1/3 of the decline in TFP growth rate. The initial surge of productivity growth at the beginning of transition path is also consistent with the temporary boom in TFP growth rate observed in data. This contributes to the heated discussion of whether rising concentration depresses long-run economic performance by squeezing out competitors. The findings in my paper suggest that increasing concentration caused by lower marketing cost is not responsible for the majority of productivity growth slowdown. It could be driven by other competing factors, such as the fact that ideas are getting harder to find or an increasing effort for firms to protect their intellectual properties. Further work should investigate these potential mechanism that account for the productivity growth slowdown.

## 1.8 Appendix: Mathematical Proofs

Each period, firms make decisions on the R&D investment $x$, marketing investment $a$, as well as the prices upon R&D success $p_i^{rd}$ ($i = 1$ for existing customers, $i = 2$ for new), and prices upon R&D failure $p_i^{nrd}$, to maximize their value function, which reads

$$\begin{aligned} V(m;q) = \max_{p_i^{rd}, p_i^{nrd}, a, x} \; & x\left\{(p_1^{rd} - w)m + (p_2^{rd} - w)y^{rd} + \beta(1-\tau)V(m'^{rd}; (1+g)q)\right\} \\ & + (1-x)\left\{(p_1^{nrd} - w)\alpha_1 m + (p_2^{nrd} - w)\alpha_2 y^{nrd} + \beta(1-\tau)V(m'^{nrd}; (1+g)q)\right\} \\ & - \psi a^{\eta_a} w - \phi \frac{x^{\eta_x}}{1-x} w \qquad (1.30) \\ s.t. \;\; & p_1^{rd} = \bar{u}(1+\lambda) - U_0 + \beta U_0' \qquad (1.31) \\ & p_1^{nrd} = \bar{u} - U_0 + \beta U_0' \qquad (1.32) \\ & y^{rd} = \left[1 - G\left(\frac{p_2^{rd}}{(1+\lambda)q}\right)\right] \frac{\xi B^\gamma}{A^\gamma} a \qquad (1.33) \\ & m'^{rd} = (1-\delta)m + y^{rd} \qquad (1.34) \\ & y^{nrd} = \left[1 - G\left(\frac{p_2^{nrd}}{q}\right)\right] \frac{\xi B^\gamma}{A^\gamma} a \qquad (1.35) \\ & m'^{nrd} = (1-\delta)\alpha_1 m + \alpha_2 y^{nrd} \qquad (1.36) \end{aligned}$$

There are two types of customers. Existing customer's value of continuing relationship with the firm

$$U_0 = \bar{u} q_j - p_1 + \beta U_0' \qquad (1.37)$$

The value function of an unmatched customer to meet with a size-$m$ firm is

$$U_s(m_j) = \int_{\frac{p_2(m)}{q_j}} \left[u_2 q_j - p_2(m_j)\right] dG(u_2) + \beta U_0' \qquad (1.38)$$

Thus, the expected value of searching is given by

$$U_s = -s + \sum_j \left[ \frac{h(m_j)}{B} U_s(m_j) \right] + \left(1 - \frac{Y}{B}\right) \beta U_0' \tag{1.39}$$

Indifference condition requires $U_0 = U_s$.

### 1.8.1 Proof of Proposition 1

Firms charge $p_1^{rd}$ (when they succeed in R&D) and $p_1^{nrd}$ (when they fail in R&D) to the existing customers:

$$p_1^{rd} = \bar{u}(1 + \lambda) - \hat{U}_0 \tag{1.40}$$

$$p_1^{nrd} = \bar{u} - \hat{U}_0 \tag{1.41}$$

Apparently, $p_1^{rd} > p_1^{nrd}$ because $\lambda > 0$.

### 1.8.2 Proof of Proposition 2

Assume the value function $V(m)$ is convex.

The value function can be written as form:

$$V(m) = \max_{x,a} R(m, x, a) + \tilde{\beta} \mathbb{E} V(g(m, x, a)) \tag{1.42}$$

where $R(m, x, a)$ is the profit function once prices are maximized out:

$$\begin{aligned} R(m, x, a) =& x \left[ \left(p_1^{rd} - w\right) (1 - \delta) m + (p_2 - w) y^{rd} \right] \\ &+ (1 - x) \left[ \left(p_1^{nrd} - w\right) \alpha_1 (1 - \delta) m + (p_2 - w) \alpha_2 y^{nrd} \right] \\ &- \psi a^{\eta_a} w - \phi \frac{x^{\eta_x}}{1 - x} w \end{aligned} \tag{1.43}$$

and the FOC of $p_2$ reads

$$x\left[1-G\left(\frac{p_2}{1+\lambda}\right)\right]+(1-x)\alpha_2\left[1-G\left(p_2\right)\right]$$
$$=x\left[p_2-w+\tilde{\beta}V'\left(m'^{rd}\right)\right]\frac{1}{2\bar{u}(1+\lambda)}+(1-x)\left[\alpha_2(p_2-w)+\tilde{\beta}V'\left(m'^{nrd}\right)\right]\frac{1}{2\bar{u}} \tag{1.44}$$

$m'=g(m,x,a)$ is the law of motion for customer base.

$$m'^{rd}=g^{rd}(m,x,a)=(1-\delta)\,m+y^{rd}\quad \text{w.p. } x \tag{1.45}$$
$$m'^{nrd}=g^{nrd}(m,x,a)=\alpha_1\,(1-\delta)\,m+\alpha_2 y^{nrd}\quad \text{w.p. } (1-x) \tag{1.46}$$

#### 1.8.2.1 Proof of the supermodularity

In this section, I prove the supermodularity of $R+\tilde{\beta}\mathbb{E}V(g)$. Note that here we do not require supermodularity of $R$ or $g$. In fact, $R$ may not be supermodular.

Denote

$$H(m,x,a)=R(m,x,a)+\tilde{\beta}\left[xV(g^{rd}(m,x,a))+(1-x)V(g^{nrd}(m,x,a))\right] \tag{1.47}$$

1. $\frac{\partial^2 H}{\partial m\partial a}$

$$\frac{\partial H}{\partial m}=\frac{\partial R}{\partial m}+\tilde{\beta}\left[xV'(m'^{rd})\frac{\partial g^{rd}}{\partial m}+(1-x)V'(m'^{nrd})\frac{\partial g^{nrd}}{\partial m}\right] \tag{1.48}$$

Second derivative

$$\begin{aligned}\frac{\partial^2 H}{\partial m\partial a}=&\frac{\partial^2 R}{\partial m\partial a}+\tilde{\beta}x\left[V''(m'^{rd})\frac{\partial g^{rd}}{\partial m}\frac{\partial g^{rd}}{\partial a}+V'(m'^{rd})\frac{\partial^2 g^{rd}}{\partial m\partial a}\right]\\&+\tilde{\beta}(1-x)\left[V''(m'^{nrd})\frac{\partial g^{nrd}}{\partial m}\frac{\partial g^{nrd}}{\partial a}+V'(m'^{nrd})\frac{\partial^2 g^{nrd}}{\partial m\partial a}\right]\end{aligned} \tag{1.49}$$

where

$$
\begin{aligned}
\frac{\partial^2 R}{\partial m \partial a} =& x\frac{\partial p_2}{\partial m}\left\{\left[1-G\left(\frac{p_2}{1+\lambda}\right)\right]-(p_2-w)\frac{1}{2\bar{u}(1+\lambda)}\right\}\frac{\xi B^\gamma}{A^\gamma} \\
&+(1-x)\alpha_2\frac{\partial p_2}{\partial m}\left\{\left[1-G\left(p_2\right)\right]-(p_2-w)\frac{1}{2\bar{u}}\right\}\frac{\xi B^\gamma}{A^\gamma} \qquad (1.50)\\
=&\left\{x\left[\left(1-G\left(\frac{p_2}{1+\lambda}\right)\right)-\frac{p_2-w}{2\bar{u}(1+\lambda)}\right]+(1-x)\alpha_2\left[\left[1-G\left(p_2\right)\right]-\frac{p_2-w}{2\bar{u}}\right]\right\}\frac{\xi B^\gamma}{A^\gamma}\frac{\partial p_2}{\partial m} \qquad (1.51)
\end{aligned}
$$

From equation (1.44), we have

$$
\begin{aligned}
&x\left[\left(1-G\left(\frac{p_2}{1+\lambda}\right)\right)-\frac{p_2-w}{2\bar{u}(1+\lambda)}\right]+(1-x)\alpha_2\left[\left[1-G\left(p_2\right)\right]-\frac{p_2-w}{2\bar{u}}\right] \\
=&x\tilde{\beta}V'\left(m'^{rd}\right)\frac{1}{2\bar{u}(1+\lambda)}+(1-x)\tilde{\beta}V'\left(m'^{nrd}\right)\frac{1}{2\bar{u}} \qquad (1.52)
\end{aligned}
$$

So,

$$
\frac{\partial^2 R}{\partial m \partial a} =\tilde{\beta}\left[xV'\left(m'^{rd}\right)\frac{1}{2\bar{u}(1+\lambda)}+(1-x)V'\left(m'^{nrd}\right)\frac{1}{2\bar{u}}\right]\frac{\xi B^\gamma}{A^\gamma}\frac{\partial p_2}{\partial m} \qquad (1.53)
$$

Next, we also have

$$
\begin{aligned}
\frac{\partial^2 g^{rd}}{\partial m \partial a} &= -\frac{1}{2\bar{u}}\frac{1}{1+\lambda}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma} \qquad (1.54)\\
\frac{\partial^2 g^{nrd}}{\partial m \partial a} &= -\frac{1}{2\bar{u}}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma} \qquad (1.55)
\end{aligned}
$$

Part of the second derivatives can be replaced with

$$\tilde{\beta}xV'(m'^{rd})\frac{\partial^2 g^{rd}}{\partial m\partial a}+\tilde{\beta}(1-x)V'(m'^{nrd})\frac{\partial^2 g^{nrd}}{\partial m\partial a} \tag{1.56}$$

$$=-\tilde{\beta}xV'(m'^{rd})\frac{1}{2\bar{u}(1+\lambda)}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}-\tilde{\beta}(1-x)V'(m'^{nrd})\frac{1}{2\bar{u}}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma} \tag{1.57}$$

$$=-\tilde{\beta}\left[xV'\left(m'^{rd}\right)\frac{1}{2\bar{u}(1+\lambda)}+(1-x)V'\left(m'^{nrd}\right)\frac{1}{2\bar{u}}\right]\frac{\xi B^\gamma}{A^\gamma}\frac{\partial p_2}{\partial m} \tag{1.58}$$

which exactly offset $\frac{\partial^2 R}{\partial m\partial a}$.

Therefore, the second derivative is left with

$$\frac{\partial^2 H}{\partial m\partial a}=\tilde{\beta}xV''(m'^{rd})\frac{\partial g^{rd}}{\partial m}\frac{\partial g^{rd}}{\partial a}+\tilde{\beta}(1-x)V''(m'^{nrd})\frac{\partial g^{nrd}}{\partial m}\frac{\partial g^{nrd}}{\partial a} \tag{1.59}$$

By convexity of $V$, $V''()>0$. Notice that

$$\frac{\partial g^{rd}}{\partial a}=\left[1-G\left(\frac{p_2}{1+\lambda}\right)\right]\frac{\xi B^\gamma}{A^\gamma}>0 \tag{1.60}$$

$$\frac{\partial g^{nrd}}{\partial a}=\alpha_2\left[1-G\left(p_2\right)\right]\frac{\xi B^\gamma}{A^\gamma}>0 \tag{1.61}$$

As for $\frac{\partial g^{(n)rd}}{\partial m}$,

$$\frac{\partial g^{rd}}{\partial m}=1-\delta-\frac{1}{2\bar{u}(1+\lambda)}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a \tag{1.62}$$

$$\frac{\partial g^{nrd}}{\partial m}=\alpha_1(1-\delta)-\alpha_2\frac{1}{2\bar{u}}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a \tag{1.63}$$

A sufficient condition for these two terms to be positive is $\mid\frac{\partial p_2}{\partial m}\mid$ is quite small. This holds when the distribution $G$ is narrow (in general, we do not require $G$ to be uniform).

The economic intuition for this is that the price elasticity of demand is quite small (new customers are relatively price inelastic). Here we don't want the effect of price to overturn the effect of $a$ or $x$. In other words, we do not want to see that if firms lower their prices, they can attract a lot of customers even if they invest in very little $a$ (or $x$). Empirical evidence from Argente et al. (2021) have shown empirical evidence for this: firms build market share mainly through advertising rather than manipulating prices.

If the condition on price (in)elasticity holds, we have (1.59) to be positive.

2. $\frac{\partial^2 H}{\partial x \partial a}$

$$\frac{\partial H}{\partial x} = \frac{\partial R}{\partial x} + \tilde{\beta}\left\{V(g^{rd}) - V(g^{nrd})\right\} \tag{1.64}$$

Second derivative

$$\frac{\partial^2 H}{\partial x \partial a} = \frac{\partial^2 R}{\partial x \partial a} + \tilde{\beta}\left[V'(g^{rd})\frac{\partial g^{rd}}{\partial a} - V'(g^{nrd})\frac{\partial g^{nrd}}{\partial a}\right] \tag{1.65}$$

where

$$\frac{\partial^2 R}{\partial x \partial a} = \left\{(p_2 - w)\left[1 - G\left(\frac{p_2}{1+\lambda}\right)\right] - (p_2 - w)\,\alpha_2\left[1 - G\left(p_2\right)\right]\right\}\frac{\xi B^\gamma}{A^\gamma} \tag{1.66}$$

which is obviously positive because $\alpha_2 < 1$ and $\lambda > 0$.

Also, since firms are charging the same $p_2$ regardless of innovation outcome, they always hire more new customers in R&D success compared to R&D failure, due to higher quality: $1 - G\left(\frac{p_2}{1+\lambda}\right) > 1 - G\left(p_2\right)$. Therefore,

$$\frac{\partial g^{rd}}{\partial a} = \left[1 - G\left(\frac{p_2}{1+\lambda}\right)\right]\frac{\xi B^\gamma}{A^\gamma} > \alpha_2\left[1 - G\left(p_2\right)\right]\frac{\xi B^\gamma}{A^\gamma} = \frac{\partial g^{nrd}}{\partial a} \tag{1.67}$$

Similarly,

$$g^{rd} =(1-\delta)m + \left[1 - G\left(\frac{p_2}{1+\lambda}\right)\right]\frac{\xi B^\gamma}{A^\gamma}a > \alpha_1(1-\delta)m + \alpha_2\left[1 - G\left(p_2\right)\right]\frac{\xi B^\gamma}{A^\gamma}a = g^{nrd} \tag{1.68}$$

By convexity of $V$, we have $V'(g^{rd}) > V'(g^{nrd})$. So, the second term of (1.65) is also positive. We have shown (1.65) is positive.

3. $\frac{\partial^2 H}{\partial x \partial m}$

$$\frac{\partial^2 H}{\partial x \partial m} = \frac{\partial^2 R}{\partial x \partial m} + \tilde{\beta}\left[V'(g^{rd})\frac{\partial g^{rd}}{\partial m} - V'(g^{nrd})\frac{\partial g^{nrd}}{\partial m}\right] \tag{1.69}$$

where

$$\frac{\partial^2 R}{\partial x \partial m} = \left[\left(p_1^{rd} - w\right) - \left(p_1^{nrd} - w\right)\alpha_1\right](1-\delta) \tag{1.70}$$

$$+ \left\{\left[1 - G\left(\frac{p_2}{1+\lambda}\right)\right] - (p_2 - w)\frac{1}{2\bar{u}(1+\lambda)}\right\}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a \tag{1.71}$$

$$- \alpha_2\left\{[1 - G(p_2)] - (p_2 - w)\frac{1}{2\bar{u}}\right\}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a \tag{1.72}$$

The second term of (1.69) can be written as

$$\tilde{\beta}\left\{V'\left(g^{rd}\right)\left[1 - \delta - \frac{1}{2\bar{u}(1+\lambda)}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a\right] - V'\left(g^{nrd}\right)\left[1 - \delta - \frac{1}{2\bar{u}(1+\lambda)}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a\right]\right\} \tag{1.73}$$

Combining the above two equations yields

$$\begin{aligned}\frac{\partial^2 H}{\partial x \partial m} =& \left[\left(p_1^{rd} - w\right) - \left(p_1^{nrd} - w\right)\alpha_1 + \tilde{\beta}\left(V'\left(g^{rd}\right) - V'\left(g^{nrd}\right)\right)\right](1-\delta) \\ &+ \left\{\left[1 - G\left(\frac{p_2}{1+\lambda}\right)\right] - \frac{p_2 - w}{2\bar{u}(1+\lambda)} - \frac{\tilde{\beta}V'\left(g^{rd}\right)}{2\bar{u}(1+\lambda)}\right\}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a \\ &- \alpha_2\left\{[1 - G(p_2)] - \frac{p_2 - w}{2\bar{u}} - \frac{\tilde{\beta}V'\left(g^{nrd}\right)}{2\bar{u}}\right\}\frac{\partial p_2}{\partial m}\frac{\xi B^\gamma}{A^\gamma}a\end{aligned} \tag{1.74}$$

When $\mid \frac{\partial p_2}{\partial m} \mid$ is small, the second and third line of (1.74) is small, and the sign of $\frac{\partial^2 H}{\partial m \partial x}$ is dominated by the first line. We have shown that $g^{rd} > g^{nrd}$, combined with convexity of $V$, we have $V'(g^{rd}) - V'(g^{nrd}) > 0$. Apparently, the first line is positive.

#### 1.8.2.2 Proof of convexity

In this section, I prove the convexity $V(m) = H(m, x, a)$.

According to the envelope condition,

$$\frac{dV}{dm} = \frac{\partial H}{\partial m} \tag{1.75}$$

and

$$\frac{d^2V}{dm^2} = \frac{\partial^2 H}{\partial m^2} + \frac{\partial^2 H}{\partial m \partial a}\frac{da}{dm} + \frac{\partial^2 H}{\partial m \partial x}\frac{dx}{dm} \tag{1.76}$$

We have shown the supermodularity of $H$, so $\frac{\partial^2 H}{\partial m \partial a} > 0$, so $\frac{\partial^2 H}{\partial m \partial x} > 0$, $\frac{da}{dm} > 0$, $\frac{dx}{dm} > 0$.

To show convexity of $V$, it remains to be shown that $\frac{\partial^2 H}{\partial m^2} > 0$.

$$\begin{aligned}
\frac{\partial^2 H}{\partial m^2} =&(1-\delta)^2\tilde{\beta}\left[xV''\left(m'^{rd}\right) + (1-x)\alpha_2 V''\left(m'^{nrd}\right)\right] \\
&- (1-\delta)\tilde{\beta}\xi a\frac{1}{2\bar{u}}\left[\frac{xV''\left(m'^{rd}\right)}{1+\lambda} + (1-x)\alpha_2 V''\left(m'^{nrd}\right)\right]\frac{\partial p_2}{\partial m} \\
&- 2\xi a\frac{1}{2\bar{u}}\left(\frac{x}{1+\lambda} + (1-x)\alpha_2\right)\cdot\left(\frac{\partial p_2}{\partial m}\right)^2 \\
&+ \xi a\left[x\left(1 - \frac{2p_2 - w}{2\bar{u}(1+\lambda)} - \frac{\tilde{\beta}V'\left(m'^{rd}\right)}{2\bar{u}(1+\lambda)}\right) + \alpha_2(1-x)\left(1 - \frac{2p_2 - w}{2\bar{u}} - \frac{\tilde{\beta}V'\left(m'^{nrd}\right)}{2\bar{u}}\right)\right]\frac{\partial^2 p_2}{\partial m^2}
\end{aligned} \tag{1.77}$$

When $\mid \frac{\partial p_2}{\partial m} \mid$ and $\mid \frac{\partial^2 p_2}{\partial m^2} \mid$ are small (price inelasticity), the second to fourth lines of (1.74) are very small, and the sign of $\frac{\partial^2 H}{\partial m^2}$ is dominated by the first term, which is obviously positive.

Thus, we have shown that $\frac{\partial^2 H}{\partial m^2} > 0$.

## 1.9 Appendix: Computational Algorithm

Computational approach to solving the stationary equilibrium.

(i) Guess the values of $g, B, A, U_0$. ($Y = \xi B^{\gamma} A^{1-\gamma}$)

(ii) Guess the functional form of value function $V(m)$ for the case of R&D success and failure, respectively.

(iii) Calculate firm's optimal choices of $p_2^{rd}(m), p_2^{nrd}(m)$ for R&D success and failure, using equations (1.14) and (1.15), and firm's innovation decision $x(m)$ and advertising decision $a(m)$.

(iv) Plug the results from Step 4 into equations (1.9) and update the guesses of value functions $V(m)$ until they converge.

(v) Perform firm simulation to obtain firm size distribution $F(m)$.

(vi) Plug the results in Step 4 and 5 and the simulated $F(m)$ into the closing model conditions, and update the guesses of $g, B, A, U_0$ until the model converges.

To calculate comparative statics, there is one more outer loop on the value of $\psi$.

To calculate transition path, there is one more outer loop of guessing the value functions $V(m)$ over the transition path.

## 1.10 Appendix: Comparison across Firm Size

### 1.10.1 Advertising

As was analyzed in Section 1.4, there is both a direct effect and indirect effect on firm's advertising investment. Although all firms have incentives to raise marketing intensity at first, the uneven pass-through rate amplifies the gap between large and small firms in customer base, innovation and marketing. Small firms thus get a smaller fraction of customers in the matches. Due to the limited consumption capacity, the measure of customers exposed to small firms could even shrink as $\psi$ decreases.

Figure 1.21: Changes in the Share of Matched Customers for Large and Small Firms

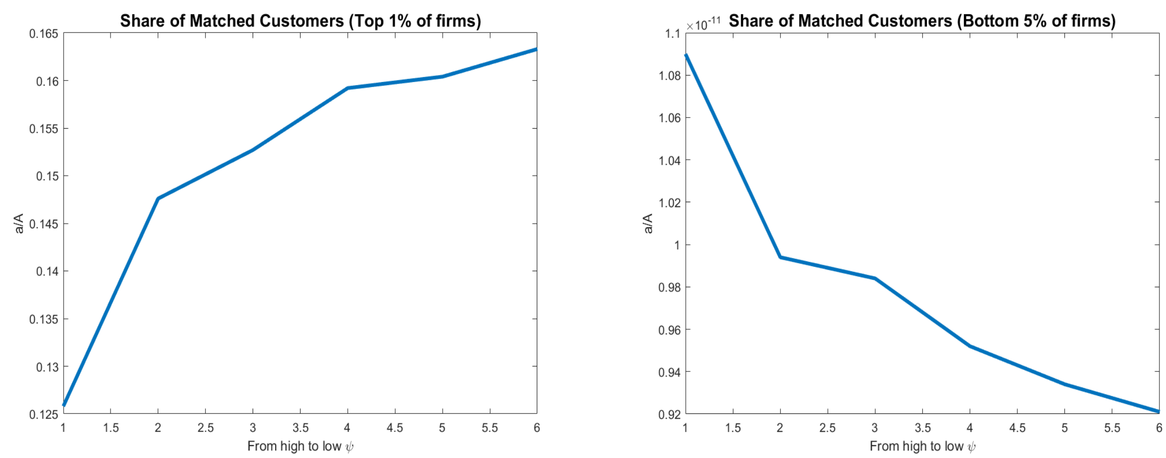
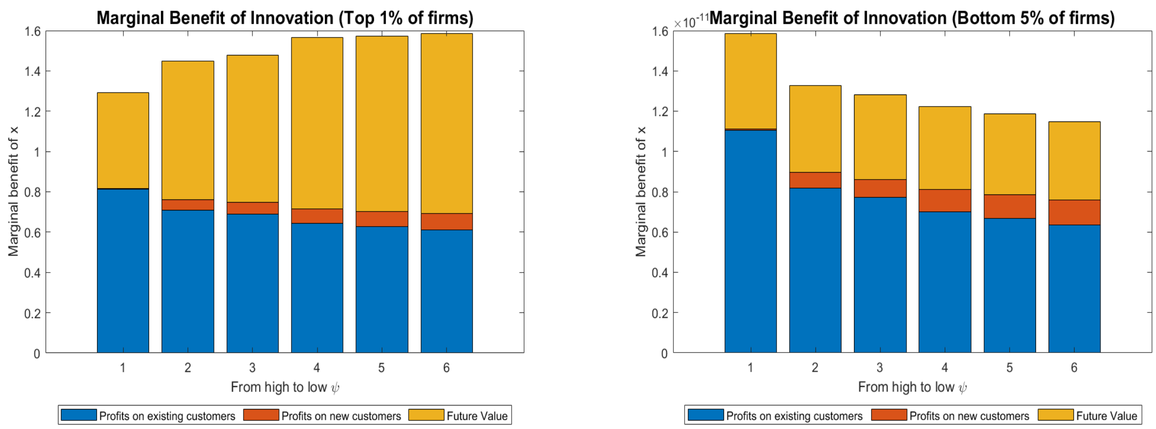



Figure 1.21 shows in the calibrated model, how the share of matched customers change with $\psi$ for the top 1% of firms and the bottom 5% of firms. At the initial level of marketing efficiency ($\psi = 4.98$), the largest firms attract 12.5% of the searching customers. Then as efficiency rises ($\psi$ decreases), the share surges to 16.5%. On the other hand, the already tiny fraction of visitors for small firms has deteriorated from 1.1E-11 to 9E-12. This means the small firms are having less exposure to the searching customers via advertising when marketing efficiency increases, because their advertisement become shaded in the outburst of marketing by larger competitors.

### 1.10.2 Innovation

There is also a sharp contrast between large and small firms in terms of the impact of a lower $\psi$ on firm R&D investment. As equation 1.16 indicates, the marginal benefit of innovation can be divided into three parts: profits on existing customers, profits on new customers and the increase in future value.

Figure 1.22: Changes in Marginal Benefit of Innovation for Large and Small Firms




With higher marketing efficiency, searching becomes more attractive and fewer customers choose to be loyal to their original firm. This reduces firms' incentives of extracting surplus from loyal customers by improving product quality. On the other hand, the increasing number of new customers matched through searching boost firm innovation incentives. Regardless of the firm size, an additional unit of innovation brings less profit on existing customers and more on new customers. The third important component in innovation is the difference in future value between a successful and failed innovation. As I have discussed in Section 1.4, the advantage of more advertising from large firms is weaved into the period-by-period accumulation of customer capital. To take advantage of the richer future profits, large firms have incentives to make more innovations. In contrast, the shrink of customer base for small firms makes them less willing to invest in innovations.

# Chapter 2

## Revisiting Capital-Skill Complementarity, Inequality, and Labor Share

### with Lee Ohanian and Musa Orak

## 2.1 Introduction

Krusell et al. (2000) (KORV, henceforth) found that empirically plausible differences in substitution elasticities between skilled labor and capital equipment, and unskilled labor and capital equipment, coupled with rapid growth in the stock of quality-adjusted capital equipment, can largely account for changes in the U.S. skill premium from 1963 to 1992. Krusell et al. (2000) thus provided a theory of skill-biased technological change, showed how to measure that change, and quantified its importance in understanding wage inequality.

KORV's production technology and substitution elasticity estimates continue to be used by other researchers studying inequality and related labor market topics, and the conclusions of Krusell et al. (2000) regarding the importance of capital-skill complementarity continue to be cited in the literature. However, there have been a number of important changes since 1992 that may affect the estimated elasticities of KORV and the quantitative importance of capital-skill complementarity in accounting for wage inequality.

One key change is that information, communications, and other advanced technologies that in part motivated the conceptual basis for the KORV production function, have advanced enormously since 1992.[1] To put some of these changes in perspective, we note that in 1992, the last year of the KORV estimation period, Lotus 1-2-3 was one of the most popular and sophisticated business software programs, and smart phones, online commerce, cloud computing, 3-D printing, and the portable document formatting technology (PDF), among many other technologies used today had not yet arrived.

Another important change is that labor's share of income, which is a moment condition in the Krusell et al. (2000) estimation, and which was quite stable in their 1963-1992 estimation period, has declined significantly since 1992.

Yet another change is that depreciation rates have increased and have become more volatile since 1992. Depreciation not only affects the accumulation of capital stocks, but also affects one of the moment conditions in the Krusell et al. (2000) estimation, which in turn may affect the estimated elasticities.

This paper studies the KORV framework in light of these changes. To our knowledge, this paper provides the most comprehensive assessment of this framework's empirical performance, with a focus on it's ability to account for the skill premium and labor's share of income.

We update the KORV dataset through 2019, estimate the model parameters, and analyze the model fit and its implications for the skill premium.

To address changes in depreciation rates since 1992, we conduct the analysis with both time-varying and constant depreciation rates. To address the decline in the labor share, we use labor income measured using gross output and using output net of depreciation, as well as using income from the non-farm business sector. To address the remarkable technological advances in some forms of capital equipment, we specify the complementary capital stock to skilled labor as Information-Communications-Technology (ICT) capital, as well as use the KORV baseline specification of total equipment capital.

[1] The growth rate of the relative price of equipment capital, a proxy for the inverse of equipment-specific technological progress, has averaged negative 6.4 percent per year after 1992, compared to negative 4.2 percent observed during 1963–1992.

As in Krusell et al. (2000), the model's skill premium is not part of the model's estimation, but rather is an outcome of the estimated parameters and data. We find that the Krusell et al. (2000) estimated elasticities change little since 1992, are robust to changes in depreciation measurement, measurement of labor share, and the type of capital that is most complementary to skilled labor. We find that KORV framework continues to account for much of the U.S. skill premium through 2019.

However, the model's fit for the labor share deteriorates when more recent data are applied in the analysis. This arises because capital-skill complementarity, combined with even faster capital-biased technological change measured in recent data, tends to increase the model's labor share by driving up the productivity of skilled labor.

We therefore frame the three modifications to Krusell et al. (2000) that we study - ICT capital as the complementary capital stock, different measures of labor's share of income, and changes in capital depreciation rates - as potential avenues for understanding the model's labor share deviation. We also use generalized method of moments (GMM) as an alternative estimation method to KORV's full-information likelihood-based method to assess whether KORV's estimator is affecting the model's ability to account for the labor share.

We find that changing the concept of capital-skill complementarity from equipment capital, as in Krusell et al. (2000), to ICT capital captures much of the skill premium over the entire period and modestly improves the model's fit of the labor share, accounting for about half of the drop in labor's share between the early 1960s and 2019. However, the model does not capture the much more recent declines in labor's share that occurred in the last 15 years.

The rest of the paper proceeds as follows. Section 2.2 summarizes related literature. Section 2.3 summarizes the data and its construction. Section 2.4 presents the theoretical model, and Section 2.5 presents the quantitative analysis. Section 2.6 presents a summary and conclusion.

## 2.2 Related Literature

Our paper contributes to several strands of literature with Krusell et al. (2000), whose elasticity estimates have been used widely, being the most obvious one. Polgreen and Silos (2008) analyzed the KORV study using data through 2004, and found that capital-skill complementarity is robust to alternative approaches to constructing the real stock of capital equipment. An earlier version of this research (Ohanian and Orak (2016b)), which has evolved into this paper, re-estimated the KORV framework and studied its fit through 2013, finding continued evidence of capital-skill complementarity. It also noted the deviation of KORV framework in accounting for the labor share after 1992. Building off of our 2016 analysis, this paper focuses on how the KORV framework can confront the post-1992 decline in labor's share, using alternative definitions of output, depreciation, labor's share, estimation techniques, and the conceptual measure of the complementarty capital stock.

Maliar et al. (2020) and Maliar et al. (2022) re-estimated the KORV model and studied its fit with data through 2017, also confirming capital-skill complementarity. They used their estimates to predict the future evolution of wage inequality. They forecast that the skill premium will continue to grow up to 2037, albeit at a slower rate. However, their paper does not focus on the decline in labor's share within the KORV framework. Additionally, their estimation generates model rates of return to capital investment that are much too high, ranging between 30 to 50 percent over the full period (see Maliar et al. (2022), Figure 4). These returns are much higher than observed returns to capital investment over this period (see, for example Marx et al. (2019) and Òscar Jordà et al. (2019)), and are also much higher than the average model-generated returns in Krusell et al. (2000) (around 4 percent) and in this paper (ranging between 8 to 12 percent in the baseline estimation).[2] The size of

[2] Marx et al. (2019) report that return on U.S. productive capital has increased from 6 percent in 1980s to around 10 percent in late 1990s, before falling back to around 8 percent by 2010. Òscar Jordà et al. (2019) find the post-1980 real return on U.S. risky assets to be near 7 percent.

investment returns has important implications for the skill premium because such excessively high returns would be expected empirically to lead to much more investment, which in turn would significantly widen the skill premium through capital-skill complementarity.

One factor that is contributing to substantially higher returns in their paper is that they estimate a higher value for the parameter governing the income share of structures capital in the model, nearly 0.2, compared to our estimate of 0.11, which is similar to the estimate in Krusell et al. (2000), and to the income share of structures capital in Greenwood et al. (1997). However, this difference alone cannot account for the very high returns their model generates, and we are unaware of other differences in their analysis that would deliver such high returns.

The paper also relates to the literature on factor income shares. Labor's share of gross income in the United States has been declining in recent years (see, for instance, Karabarbounis and Neiman (2014b) and Armenter (2015)). Karabarbounis and Neiman (2014b) study changes in the labor share across a panel dataset including 56 countries, and analyze this dataset using a model in which rapid productivity growth in capital goods, which in turn decreases the relative price of capital goods, incentivizes producers to substitute away from labor input into capital input within a production technology in which labor and capital are more substitutable than Cobb-Douglas. This substitution away from labor to capital decreases labor's share of income. Quantitatively they find that the observed long-run decline in investment goods prices accounts for about half of the change in labor share, even after allowing for other mechanisms, including changes in monopoly rents and changes in the skill composition of workers. Other factors studied within this literature include, trade and offshoring (Elsby et al. (2013)), foreign direct investment in inflows and mechanization (Guerriero (2012)), structural change and heterogeneity (Alvarez-Cuadrado et al. (2015)), a global productivity slowdown (Grossman et al. (2018)), and increasing concentration within industries (Dorn et al. (2017)). This paper connects the declining labor share to technological change, including Orak (2017), who links the decline in labor's share to technological change and the resulting shift in the occupational composition of the workforce; vom Lehn (2018), who explains the decline with replacement of workers engaged in routine (repetitive) occupations (job polarization); Eden and Gaggl (2018), who attribute half of the decline to the rise in the income share of ICT capital, using a framework distinguishing between ICT and non-ICT capital; and Eden and Gaggl (2019), who show that more than one quarter of the global decline in the labor share can be explained by a change in capital composition that works through automation. Analyzing the KORV framework after 1992 allows the model to confront these observations and analyze their quantitative importance in estimating the production function parameters.

There are also several studies suggesting that the decline in labor's share is less significant once some factors, such as a significant rise in housing capital (Rognlie (2015)); capitalization of intellectual property products (Koh et al. (2015)); a substantial rise in equity-based compensation (Eisfeldt et al. (2022)); and depreciation and taxes (Bridgman (2018)), are netted out.[3] Sherk (2016) argues that the decline in labor's share reflects how increased depreciation of capital and the income of the self-employed are accounted for. Given these issues regarding gross and net income, we construct a measure of net labor share to use in the analysis, which is indeed more stable than the gross labor share.[4]

## 2.3 Data

We construct capital stocks and labor inputs between 1963 and 2019 along the lines of Krusell et al. (2000). We collected equipment (including intellectual property products) and structures investment series from the National Income and Product Accounts (NIPA) and

[3] Note that Karabarbounis and Neiman (2014a) also analyze the labor share of gross income and of income net of depreciation, but they conclude that there is a declining trend in both labor share series on a global scale. However, Bridgman (2018) and our study focus solely on U.S. data.

[4] KORV's measure of the gross labor share already excluded self-employment.

used the perpetual inventory method to construct capital stocks following the formula below:

$$Final\ inventory = Beginning\ inventory * (1 - Depreciation\ rate) + investment.$$

We obtain structures investment from NIPA Table 5.2.5, then use the implicit GDP price deflator to generate real investment levels for each year.[5] The quarterly data are transformed into an annual series via simple averaging.

As Krusell et al. (2000) point out, many economists (see, for example, Gordon (1990)) argue that, despite the BEA's best efforts, NIPA substantially understates the increases in quality of durable goods over time, including capital equipment, which in turn overstates the rate of price inflation among these goods. We follow Krusell et al. (2000) and use the deflator of the equipment investment series provided by DiCecio (2009), which are constructed following the procedure pioneered by Gordon (1990) and Cummins and Violante (2002).[6,7]

When constructing the capital stock series, we use time-varying depreciation rates, which we calculate from the NIPA tables by dividing the current cost capital consumption series by current cost capital stock series. This choice is motivated by the fact that the depreciation rates of equipment capital have risen significantly since the original KORV study (see Figure 2.3), particularly during the technological boom period of late 1990s and early 2000s. Thus, time-varying depreciation rates may give a more accurate measure of the actual capital stock in any particular year than using an average rate. Alternatively, we use constant depreciation rates as in KORV, but calculate the average values using the most recent data.

[5] U.S. Bureau of Economic Analysis, Gross Domestic Product: Implicit Price Deflator [GDPDEF], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/GDPDEF`.

[6] DiCecio, Riccardo, Equipment Deflator [EDEF], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/EDEF`.

[7] Figure 2.11 in the Appendix 2.9 compares the NIPA equipment price series with that of DiCecio (2009).

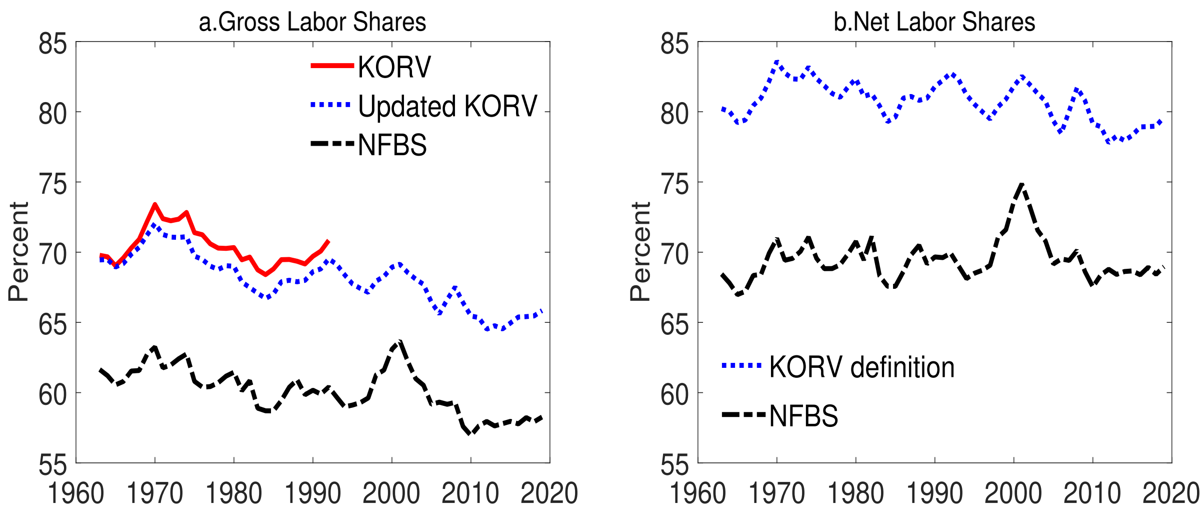
### 2.3.1 Labor Inputs and Wage Rates

As in Krusell et al. (2000), we specify skilled and unskilled labor based on educational attainment. The data are drawn mainly from the Current Population Survey (CPS) March Supplement, now called the Annual Social and Economic (ASEC) Supplement, integrated by IPUMS (Flood et al. (2015)) for the years 1963 through 2019. We use all of the person-level data, excluding those who are younger than 16 or older than 70, unpaid family workers, and those working in the military (together with other institutionalized population) to maintain comparability with earlier studies, including Katz and Murphy (1992) and Krusell et al. (2000). Although we drop the self-employed from our wage sample, we include them when constructing labor input series. We also drop the observations of those who report working less than 40 weeks or 35 hours a week or both from the wage sample as it is standard in the related literature. Finally, we exclude individuals with allocated income, those with hourly wages below half of the minimum federal wage rate, and those whose weekly pay was less than $62 in 1980 dollars to remove outliers and misreporting. A detailed description of the construction of labor input and hourly wages is in the Appendix 2.7.1.

### 2.3.2 Labor Share

We construct our labor share series using the BEA NIPA tables. To facilitate comparison with Krusell et al. (2000), we begin by using their definition of labor share, which is constructed in a manner similar to what Cooley and Prescott (1995) describe. As such, we define labor share as the ratio of labor income (wages, salaries, and benefits) to the sum of labor income plus capital income (depreciation, corporate profits, net interest, and rental income of persons). This is our benchmark definition and is called the “KORV definition." As an alternative, we also use the nonfarm business sector (NFBS) labor share, which is the most commonly used definition in the labor share literature. The data construction is in Appendix 2.7.2, and we also report some of the findings with this alternative definition in Appendix 2.9. As shown in the top left panel of Figure 2.1, although the (gross) labor share was nearly flat in KORV's data, it has been trending down since that time. Apart from level differences, both the KORV and NFBS labor shares show this pattern, though the decline is less pronounced for the NFBS labor share.

Figure 2.1: Labor Shares

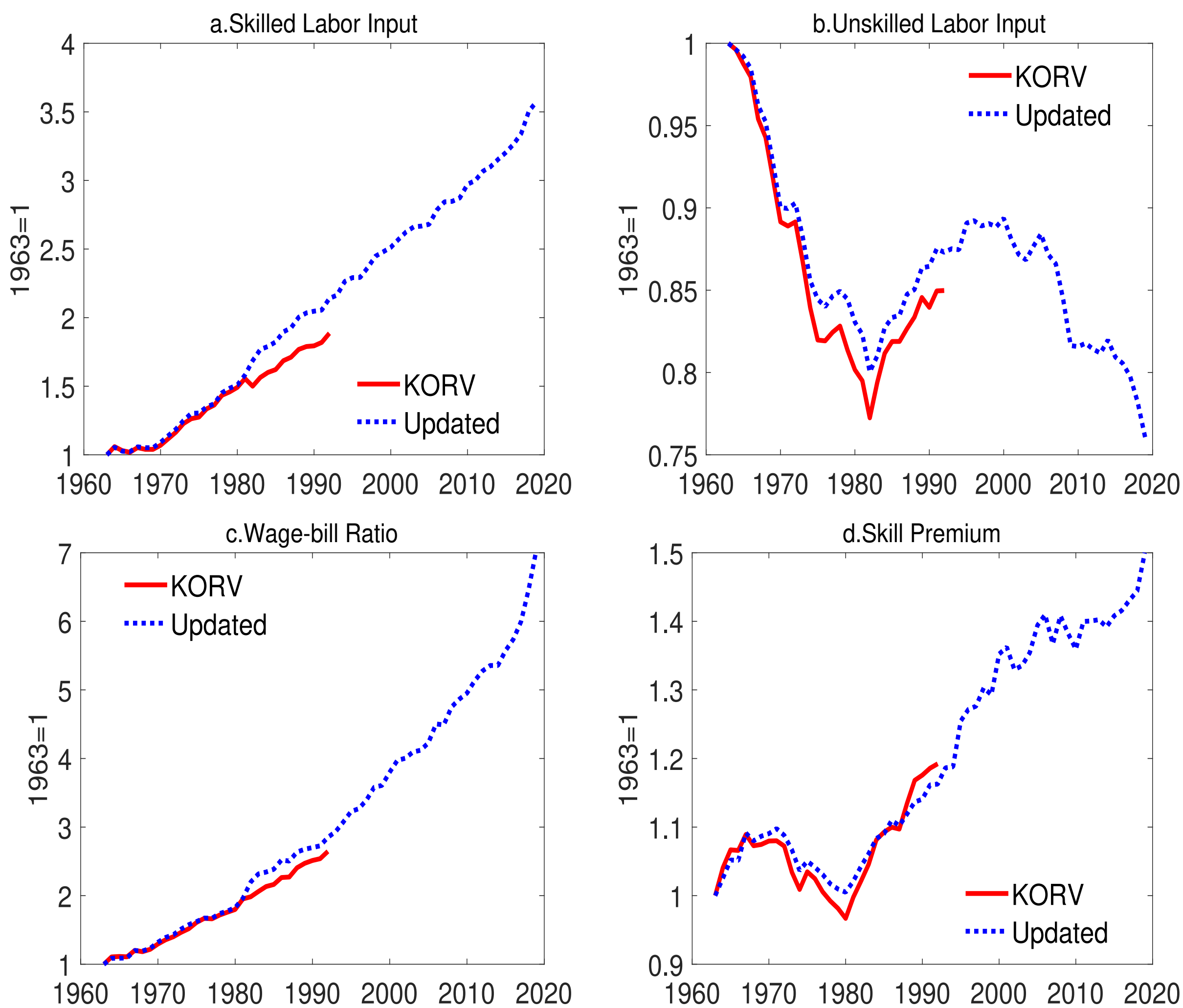



While the declining labor share has been considered one of the most striking and puzzling features of the recent U.S. economy, some claim that the decline reflects increased depreciation of capital (see, for example, Sherk (2016)). To analyze how netting out depreciation affects our findings, we construct an alternative measure of labor share that subtracts depreciation from gross income, which we then use to construct the labor share of net income. As seen in the right panel of Figure 2.1, these net measures of labor share do not exhibit a significant trend decline, though they are very volatile.

### 2.3.3 Summary of the Data

Figure 2.2 presents the evolution of the labor data from 1963 through 2019, along with a comparison to the original KORV data. Skilled labor input (panel a) has been continuously rising since the early 1970s, while unskilled labor input (panel b) declined by almost 25 percent over the 1963-2019 period. These patterns are largely in line with what KORV documented for the 1963-1992 period, though there is a level shift in skilled labor input (reflecting data revisions) beginning in 1982 relative to KORV.[8] The wage-bill ratio, which is the ratio of the labor income of skilled labor to that of unskilled labor, has continued to increase (panel c). More specifically, we construct the wage-bill ratio as the ratio of the product of average wage and total hours of the skilled labor (as constructed in Appendix 2.7) to that of the unskilled. Finally, income inequality has continued to widen since KORV's study, with the skill premium rising from a normalized level of about 1.2 in 1992 (the final year of KORV) to about 1.5 in 2019 (panel d).

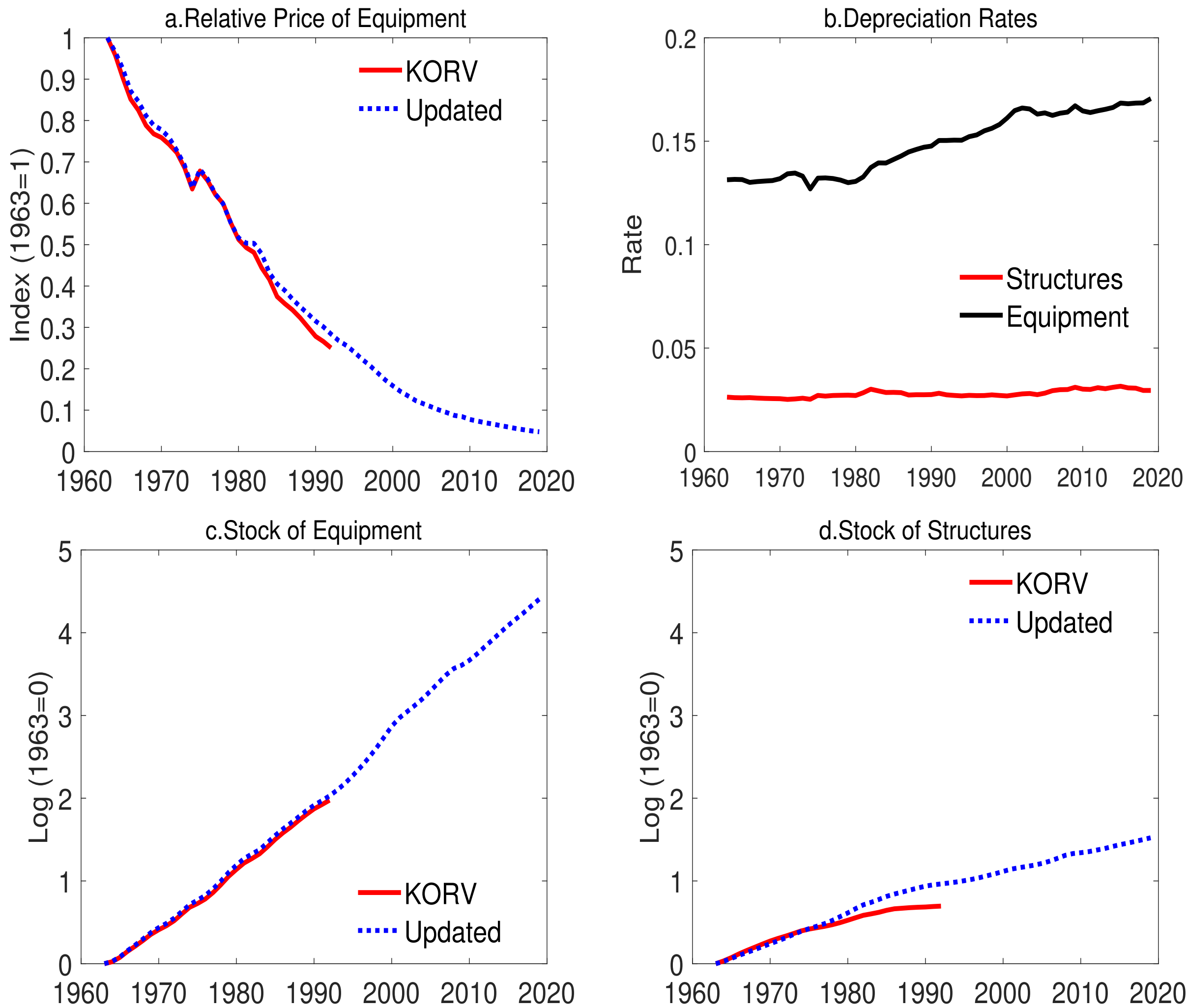
Figure 2.2: Comparison of the Original KORV Data with Updated Labor Data




[8] Other studies replicating KORV data, including Polgreen and Silos (2008) and Maliar et al. (2020), have documented a similar level shift.

Figure 2.3 presents the evolution of the capital stock data over the 1963-2019 period,with a comparison to the original KORV data when applicable. Consider first panel a, which shows the growth rate of the relative price of equipment capital, which in the model is equal to the inverse of the equipment-specific technology parameter. The decline in the relative price of equipment capital accelerated during the late 1990s and early 2000s, which is commonly cited as the “IT Boom" period. Since 1992, the growth rate of the relative price of equipment capital has averaged about negative 6.4 percent per year, compared with the negative 4.2 percent observed during the period of the original Krusell et al. (2000) study. This indicates faster technological progress in equipment after 1992, which coincides with a rising depreciation rate (panel b) and an acceleration of the increase in the stock of equipment capital (panel c). By 2019, the stock of equipment capital is more than eighty times larger than its 1963 level, whereas the stock of structures is only about 4.7 times larger in 2019 than in 1963 (panel d).

Figure 2.3: Comparison of Original KORV Data with the Updated Capital Data




## 2.4 Model

### 2.4.1 Model Environment

We use the same theoretical framework as Krusell et al. (2000). There are four factors of production: structures ($k_{st}$) and equipment ($k_{eq}$) capital; and skilled ($s$) and unskilled ($u$) labor inputs. These inputs are combined using a nested CES aggregate production function that allows different substitution elasticities between unskilled labor input and the composite output of equipment capital and skilled labor input, and between equipment capital and skilled labor input.

There are three final goods in the economy: consumption ($c$), investment in structures capital ($i_{st}$), and investment in equipment capital ($i_{eq}$). Consumption and structures capital are produced using the same constant returns to scale technology, and prices of both are normalized to one, as is standard in the literature. Of note, the growth rates of prices of consumption and structures are almost perfectly correlated in the data, as shown in Figure 2.12 in the Appendix 2.9.

There is equipment-specific technological change in which one unit of the final good that is invested in equipment produces $q_t$ units of equipment capital, where $q_t$ is equipment-specific productivity.

Perfect competition guarantees that

$$p_{eq,t} = \frac{1}{q_t}, \tag{2.1}$$

where $p_{eq,t}$ is the relative price of equipment capital, and its inverse is used as the proxy for equipment-specific technological progress.

Given competition and constant returns to scale, the aggregate resource constraint for this economy is as follows:

$$Y_t = c_t + i_{st,t} + \frac{i_{eq,t}}{q_t} = A_t G\left(k_{st,t}, k_{eq,t}, h_{u,t}, h_{s,t}, \delta_{eq,t}, \delta_{st,t}; \Upsilon\right), \tag{2.2}$$

where $Y$ is the final good, $h_u$, and $h_s$ are raw unskilled and skilled labor units, respectively. $A$ denotes neutral technological change. Finally, $\Upsilon$ is the set of model parameters, which is detailed below.

### 2.4.2 Production Technology

Following Krusell et al. (2000), we use the CES aggregate production function below:

$$G\left(k_{st,t}, k_{eq,t}, h_{u,t}, h_{s,t}, \delta_{eq,t}, \delta_{st,t}; \Upsilon\right) = k_{st,t}^{\alpha}\left(\mu u_t^{\sigma} + (1-\mu)\left[\lambda k_{eq,t}^{\rho} + (1-\lambda) s_t^{\rho}\right]^{\frac{\sigma}{\rho}}\right)^{\frac{1-\alpha}{\sigma}}. \quad (2.3)$$

[9]

In equation (2.3), $u_t$ and $s_t$, efficiency hours for the respective skill groups, are defined as follows:

$$u_t = e^{\varphi_{u,t}} h_{u,t} \quad (2.4)$$

$$s_t = e^{\varphi_{s,t}} h_{s,t}. \quad (2.5)$$

The model has the following set of parameters to be estimated: $\Upsilon \in \{\sigma, \rho, \alpha, \mu, \lambda\, \varphi_u, \varphi_s\}$. Here, $\mu$ and $\lambda$ are parameters governing the factor shares. The parameter $\alpha$ is the income share of structures capital. $\varphi_u$ and $\varphi_s$ are the efficiencies of unskilled and skilled labor, respectively. The parameters $\sigma$ and $\rho$ govern the elasticities of substitution between equipment capital and the two types of labor input. Following Krusell et al. (2000), the elasticity of substitution between equipment capital and skilled labor input is $\frac{1}{1-\rho}$. The elasticity of substitution between unskilled labor input and the CES composite of equipment capital and skilled labor is $\frac{1}{1-\sigma}$. Holding other factors constant, the elasticity of substitution between unskilled labor and equipment is also $\frac{1}{1-\sigma}$.[10]

Krusell et al. (2000) show that capital-skill complementarity requires $\sigma > \rho$ (equivalently, $\frac{1}{1-\sigma} > \frac{1}{1-\rho}$). This implies that equipment-specific technological progress increases the relative demand for skilled labor input, and depresses the relative demand for unskilled labor.

[10] Note that the KORV substitution elasticity definition assumes that all other factors are held constant. There are alternative definitions of the substitution elasticity between two factors in a production function that has more than three or more factors, including the Allen and Moroshima elasticities. Polgreen and Silos (2008) report that capital-skill complementarity also holds in the KORV framework for the Allen and Morishima elasticities, though the magnitudes are different. However, these distinctions in defining the substitution elasticity do not play a role in understanding the impact of capital deepening on wage inequality in the model. As we note immediately below, this depends on whether the value of $\sigma > \rho$.

### 2.4.3 The Model Skill Premium

Given perfect competition, the firm's problem is

$$\Pi_t = Y_t - r_{st,t}k_{st,t} - r_{eq,t}k_{eq,t} - w_{u,t}h_{u,t} - w_{s,t}h_{s,t}, \tag{2.6}$$

where $r_{st,t}$ and $r_{eq,t}$ are the rental rates of structures and equipment capital, respectively. Similarly, $w_{u,t}$ and $w_{s,t}$ denote the wage rates of unskilled and skilled labor at time $t$.

The skill premium at time $t$ is:

$$\pi_t = \frac{w_{s_t}}{w_{u,t}} = \frac{MPL_{s,t}}{MPL_{u,t}} = \frac{A_t G_{s,t}}{A_t G_{u,t}},$$

where $MPL_{s,t}$ and $MPL_{u,t}$ are marginal products of skilled and unskilled labor inputs, respectively.

As presented in Krusell et al. (2000), the skill premium in this model is as follows:

$$\pi_t = \frac{(1-\mu)(1-\lambda)}{\mu}\left[\lambda\left(\frac{k_{eq,t}}{s_t}\right)^{\rho} + (1-\lambda)\right]^{\frac{\sigma-\rho}{\rho}}\left(\frac{h_{u,t}}{h_{s,t}}\right)^{1-\sigma}\left(\frac{\varphi_{s,t}}{\varphi_{u,t}}\right)^{\sigma}. \tag{2.7}$$

When equation 2.7 is log-linearized and differentiated with respect to time, one obtains:

$$g_{\pi_t} \approx \overbrace{(1-\sigma)(g_{h_{u,t}} - g_{h_{s,t}})}^{\text{relative quantity effect}} + \overbrace{\sigma(g_{\varphi_{s,t}} - g_{\varphi_{u,t}})}^{\text{relative efficiency effect}} + \overbrace{(\sigma-\rho)\lambda\left(\frac{k_{eq,t}}{s_t}\right)^{\rho}(g_{k_{eq,t}} - g_{h_{s,t}} - g_{\varphi_{s,t}})}^{\text{capital-skill complementarity effect}}, \tag{2.8}$$

where $g_{j,t}$ denotes growth rate of variable $j$ at time $t$.

As shown in equation 2.8, Krusell et al. (2000) decompose the growth in the skill premium into three components. The first component, the *relative quantity effect*, shows that when $\sigma < 1$, faster growth in skilled labor supply reduces the skill premium. The *relative efficiency effect* depends on the sign of $\sigma$. When $\sigma > 0$ ($\sigma < 0$), relatively faster growth of skilled labor efficiency drives the skill premium higher (lower). Finally, when there is *capital-skill complementarity effect*, meaning that $\sigma - \rho > 0$, faster growth in equipment capital relative to the supply of skilled labor input increases the skill premium. This effect would get smaller (larger) over time if $\rho < 0$ ($\rho > 0$).

## 2.5 Quantitative Analyses

### 2.5.1 Estimation Strategy

We estimate the model from 1963 through 2019 and we also estimate the model in sub-samples so we can compare the results to those from Krusell et al. (2000), who estimated the model from 1963 to 1992. As in Krusell et al. (2000), our baseline framework uses equipment capital as the capital stock that is complementary with skilled labor. As an alternative, we estimate the model using Information and Communication capital (ICT) as the complementary capital stock. We examine the model using this alternative measure of capital given the very rapid technological advances in this category of capital goods (e.g. computer hardware, software, telecommunications devices, etc.).

As noted above, we use the same empirical methodology as Krusell et al. (2000). This includes a two-stage simulated pseudo maximum likelihood estimation (SPMLE) procedure to estimate most of the model parameters. Appendix 2.8 describes this in detail. Since we later will find that the KORV framework will be challenged in capturing labor's share of income, we also estimate the parameters with generalized method of moments (GMM) methodology, which is more widely used than SPMLE and does not require the full likelihood of the model. Details are described in Appendix 2.8.

For the SPMLE methodology, there are two stochastic elements to close the model and ensure that the likelihood is non-singular. This involves introducing stochastic components into the two labor inputs. Following Krusell et al. (2000), we specify the stochastic process as:

$$\varphi_t = \varphi_0 + \omega_t, \tag{2.9}$$

where $\varphi_t$ is a $2 \times 1$ vector of the log of labor efficiencies for skilled and unskilled labor at time $t$, $\varphi_0$ is a $2 \times 1$ vector of constants that correspond to the average levels of efficiencies (efficiency levels in the absence of efficiency shocks), and $\omega_t$ is a $2 \times 1$ vector of labor efficiency shocks, which are assumed to have a multivariate normal distribution with zero mean and covariance matrix $\Omega = \begin{bmatrix} \eta_\omega^2 & 0 \\ 0 & \eta_\omega^2 \end{bmatrix}$, where $\eta_\omega^2$ is the common variance of efficiency shocks. Note that, for comparability, we rule out trend growth for labor efficiencies as in Krusell et al. (2000).

The relative price of equipment capital is the second stochastic element. This price affects the rate of return to investment in equipment capital. Krusell et al. (2000) motivated this condition by hypothesizing a risk neutral investor, for whom arbitrage would equate the ex-ante expected returns on the two investments. Krusell et al. (2000) called this a “No Arbitrage" condition, and it is given as follows:

$$\underbrace{q_t A_{t+1} G_{eq,t+1} + (1 - \delta_{eq,t+1}) E\left(\frac{q_t}{q_{t+1}}\right)}_{\text{ex-ante expected return on equipment capital}} = \underbrace{A_{t+1} G_{st,t+1} + (1 - \delta_{st,t+1})}_{\text{ex-ante expected return on structures}}, \tag{2.10}$$

where $G_{eq,t+1}$ and $G_{st,t+1}$ are derivatives of function $G\left(k_{st,t}, k_{eq,t}, h_{u,t}, h_{s,t}, \delta_{eq,t}, \delta_{st,t}; \Upsilon\right)$ with respect to equipment and structures capital stocks at time $t+1$, respectively, and $\delta_{st,t}$ and $\delta_{eq,t}$ are the corresponding depreciation rates at time $t$. The first term on the left-hand side is the marginal product of equipment investment, and the second term is undepreciated equipment capital, adjusted by the expected change in its market price. The right-hand side terms are analogues for capital structures.

Equation 2.10 is one of the three equations used in the estimation. Krusell et al. (2000) developed this equation based on the idea of a risk neutral investor choosing between investing at the margin in equipment or structures, where both types of capital have the same tax treatment, and $(1-\delta_{eq,t+1})E\left(\frac{q_t}{q_{t+1}}\right) = (1-\delta_{eq,t+1})\frac{q_t}{q_{t+1}} + \epsilon_t$ with $\epsilon_t$ is assumed to be normally distributed with mean zero and variance $\eta_\epsilon^2$. As such, we use the following form of equation 2.10 in the estimation:

$$0 = q_t A_{t+1} G_{eq,t+1} + (1-\delta_{eq,t+1})E\left(\frac{q_t}{q_{t+1}}\right) - A_{t+1}G_{st,t+1} - (1-\delta_{st,t+1}) + \epsilon_t. \tag{2.11}$$

The other two equations used in the estimation are as follows:

$$\frac{w_{s,t}h_{s,t}}{w_{u,t}h_{u,t}} = wbr_t(X_t, \varphi_{u,t}, \varphi_{s,t}; \Upsilon) \tag{2.12}$$

$$\frac{w_{s,t}h_{s,t} + w_{u,t}h_{u,t}}{Y_t} = lshare_t(X_t, \varphi_{u,t}, \varphi_{s,t}; \Upsilon), \tag{2.13}$$

where the $wbr_t$ is the wage-bill ratio and $lshare_t$ is the labor share at time $t$, both of which are obtained from model-implied marginal products of the skilled and unskilled labor inputs ($w_{s,t}$ and $w_{u,t}$, respectively). Each of these marginal products is a function of observable factor inputs $X_t = \left\{k_{st,t}, k_{eq,t}, h_{u,t}, h_{s,t}, \delta_{eq,t}, \delta_{st,t}\right\}$, unobservable labor efficiencies $\varphi_{u,t}$ and $\varphi_{s,t}$, and a set of parameters $\Upsilon = \{\sigma, \rho, \alpha, \mu, \lambda, \varphi_{u0}, \varphi_{s0}, \eta_\epsilon, \eta_\omega\}$.

The system is a nonlinear state space model with three observation equations $Z_t = f(X_t, \varphi_{u,t}, \varphi_{u,t}, \varepsilon_t; \Upsilon)$, and two state equations $\varphi_t = \varphi_0 + \omega_t$ (one for skilled and one for unskilled labor efficiency). Here, $Z$ is a $3 \times 1$ vector of observables corresponding to equations 2.11 to 2.13. The right-hand sides of the equations are the model counterparts of the corresponding series implied by marginal products of inputs, observables, and estimated parameters and efficiencies.[11]

As in Krusell et al. (2000), we calibrate some of the parameters. Depreciation rates are obtained from the NIPA capital stock and capital consumption tables. Unlike Krusell et al.

[11] The left-hand sides are: a vector of zeros for the no-arbitrage condition, the ratio of the product of average wage and hour of the skilled labor to that of the unskilled for the wage-bill ratio, and labor shares as constructed from the NIPA tables.

(2000), we use time varying depreciation rates in our benchmark analysis, as these depreciation rate means and volatilities have increased since the Krusell et al. (2000) estimation period.

The depreciation rate of structural capital averages 0.0278, while the depreciation rate of equipment capital averages 0.1483 over the 1963-2019 period, beginning with 0.1311 in 1963 and rising up to 0.1706 in 2019. For simplicity, we assume the future depreciation rates in the case of time-varying depreciation are known. We will compare the results from time-varying depreciation to those with fixed depreciation below.

To calibrate $\eta_\epsilon$, we estimate an ARIMA model for the growth rate of the relative price of equipment capital: $\hat{q}_t = \frac{q_{t-1}}{q_t}$. The estimated ARIMA model is $\hat{q}_t = 2.12 - 0.001t - 0.48\hat{q}_{t-1} + 0.58\epsilon_{t-1} + \epsilon_t$ with an estimated standard deviation of white-noise disturbance ($\sigma_\epsilon$) of $= 0.0233$. As Krusell et al. (2000) do, $\eta_\epsilon$ is then calibrated as $(1 - \bar{\delta}_{eq})$ times $\sigma_\epsilon$, where $\bar{\delta}_{eq}$ is the mean of the equipment capital depreciation rate. Using this calibration method, we set $\eta_\epsilon$ to $= 0.020$.[12]

Furthermore, because $\mu$, $\lambda$, $\varphi_{u0}$, and $\varphi_{s0}$ are scaling parameters, we normalize $\varphi_{s0}$ as Krusell et al. (2000) do, setting it equal to 2.

The remaining parameters are estimated using the two-stage SPMLE method of Krusell et al. (2000), which is discussed in more detail in Appendix 2.8. In the first stage, to allow for the possible dependence of labor input on shocks, we follow Krusell et al. (2000) and regress the labor inputs on the set of instruments they used: current and lagged stocks of both types of capital, the lagged relative equipment capital price, a time trend, and the lagged value of the index of leading business cycle indicators of the Conference Board. The fitted (instrumented) values are then used in the SPMLE stage.

As in KORV, we choose the value of $\eta_\omega$ that minimizes the joint sum of squared deviations between the skill premium and its model counterpart and between the ex-post returns on structures and equipment investments. Note that we do not view this parameter

[12] When we estimate the ARIMA model for the 1963–1992 period, we obtain: $\hat{q}_t = 2.85 - 0.001t + 0.55\hat{q}_{t-1} - \epsilon_{t-1} + \epsilon_t$ with $\sigma_\epsilon = 0.0243$. By the same calibration strategy, we obtain $\eta_\epsilon = 0.021$ for this period.

as having specific economic interest within the scope of this analysis; rather it is introduced to ensure a non-singular likelihood.[13]

The Krusell et al. (2000) estimation is complex, which in part is due to a number of latent exogenous variables within the likelihood function. We therefore also estimate the parameters using GMM, which does not use the full likelihood of the model.

To implement GMM, we use the same instruments as those used in the SPMLE estimation described above. We estimated the production function parameters $\alpha$, $\sigma$, $\mu$, $\rho$, and $\lambda$; and efficiency level of unskilled labor $\varphi_u$ using the three moment conditions consisting of the wage-bill ratio, the labor share, and the no-arbitrage condition. As with SPMLE, we normalize $\varphi_s$ at 2. Details of the GMM estimation are discussed in Appendix 2.8.

### 2.5.2 Findings

This subsection presents the estimated parameters and model fit for the full 1963-2019 period as well as 1963-1992, the period analyzed by Krusell et al. (2000). Given recent changes in labor share and the large literature which has studied this change, we also estimate the model using two alternatives to Krusell et al. (2000), who use the standard measure of labor share of gross output. The alternatives are the share of labor income net of depreciation (as discussed in Bridgman (2018)) and nonfarm business sector (NFBS) labor income share, based on both gross output and output net of depreciation.

Because the findings are very similar to our baseline results when we use non-farm business sector output rather than real output of the business sector, we report parameter estimates and model fits for this case only in the Appendix (see Tables 2.7 and 2.8 and Figures 2.14, 2.15 and 2.16).[14] Although the parameter estimates change slightly in this

[13] As the working paper version of Krusell et al. (2000) (Krusell et al. (1997)) notes, if $\eta_\omega$ is estimated jointly with the rest of the parameters, the algorithm chooses a large value as it helps fit the difference between the two rates of return to capital in the mid-1970s when the relative price of equipment is extremely volatile and exhibits a very large spike in 1975. A very high value of $\eta_\omega$, however, worsens the fit of the skill premium.

[14] When using the nonfarm business sector labor share in estimation, we drop farm, households and government

case, capital-skill complementarity remains sizable and significant, as in the baseline case, which uses KORV's definition of the labor share.[15] When the nonfarm business sector labor share is used in the estimation, the model fit improves slightly for the skill premium and for the ex-ante no-arbitrage condition, though ex-post rates are somewhat larger than their empirical counterparts. In terms of the labor share, using the nonfarm business sector labor share brings no improvement (see Table 2.5).

Our baseline estimation involves time-varying depreciation rates to ensure consistency of data construction and model assumptions. However, because almost all macroeconomic models assume constant depreciation rates, we also estimated the baseline model using constant depreciation rates to check applicability of our results to general equilibrium models with standard assumptions for depreciation rates. The findings, which are reported in Appendix 2.9 (see Table 2.9 and Figure 2.17), are very similar to the baseline findings, suggesting that time-varying depreciation rates assumption are not impacting the results.

#### 2.5.2.1 Baseline results with KORV's definition of gross labor share

Table 2.1 presents the estimated parameters for 1963 through 1992, which is the original time period analyzed by KORV, and using the original KORV data.

Row I reports the estimates of KORV and row II is our replication using the original KORV data and their SPMLE methodology. Figure 2.13, which is presented in the Appendix, should be compared with figures 5 through 8 in Krusell et al. (2000), as it shows the model fit with the same data and estimation methodology.

The SPMLE estimated parameters and the model's fit are both nearly identical to KORV, confirming that any changes in our results in the following discussion will be data-driven. GMM estimates are presented in row III, and are also similar to the SMPLE

sectors when constructing wage and labor input data. The resulting skill premium series are highly correlated for both cases, reflecting the fact that the excluded sectors account for a small part of business sector output. We also subtracted farm sector investment when constructing corresponding capital inputs, which already excluded government capital as in Krusell et al. (2000).

[15] The only parameter that is considerably different from the baseline case is $\alpha$, which is the income share of structures capital. This difference is driven by the level difference between KORV and NFBS labor shares.

estimates. GMM confirms significant capital-skill complementarity, though indicating somewhat larger substitution elasticities between equipment capital and skilled labor and between their composite output and unskilled labor compared with the original and replicated KORV elasticities (rows I and II).

Table 2.1: Parameter Estimates for the 1963-1992 Period with Original KORV Data

| **Model** | **Methodology** | $\sigma$ | $\rho$ | $\alpha$ | $\eta_\omega$ |
|---|---|---|---|---|---|
| I. KORV (2000) | SPMLE | 0.401 | $-0.495$ | 0.117 | 0.043 |
| | | (0.234) | (0.048) | (0.007) | (0.003) |
| II. KORV Replication | SPMLE | 0.411 | $-0.505$ | 0.108 | 0.043 |
| | | (0.014) | (0.055) | (0.003) | (0.005) |
| III. KORV Replication | GMM | 0.471 | $-0.396$ | 0.117 | – |
| | | (0.021) | (0.043) | (0.002) | – |

Notes: The values in parentheses are standard errors.

Table 2.2 presents the parameter estimates when we use our revised and updated data in estimation for both 1963-1992 and 1963-2019 periods, for KORV's definition of both gross and net labor shares, and with the two estimation methodologies. In all cases, SPMLE and GMM parameter estimates and model fits are very similar. First consider the results for original period of KORV's study with KORV's definition of the gross labor share and SPMLE estimation. In this case, $\rho$, the parameter governing the elasticity of substitution between equipment capital and skilled labor input, and $\alpha$, the share of structures capital in production, are little changed from the original KORV results and from our replication with original KORV data in Table 2.1.

Our estimate of $\sigma$, the parameter governing the elasticity of substitution between equipment capital and unskilled labor input, is only slightly different with the revised data. KORV's estimated elasticity was 1.67, compared to our estimate of 1.70 using their original data (row II of Table 2.1), and 1.77 with revised data, indicating even more substituability between equipment and unskilled labor in the revised data. The estimated parameters change only slightly with GMM, and the degree of capital-skill complementarity is largely unchanged.

Table 2.2: Parameter Estimates with Updated Data

| **Labor Share** | **Period** | **Methodology** | $\sigma$ | $\rho$ | $\alpha$ | $\eta_\omega$ |
|---|---|---|---|---|---|---|
| KORV Gross | 1963-1992 | SPMLE | 0.438 | $-0.520$ | 0.105 | 0.083 |
| | | | (0.020) | (0.043) | (0.002) | (0.007) |
| | | GMM | 0.467 | $-0.478$ | 0.106 | – |
| | | | (0.018) | (0.035) | (0.002) | – |
| | 1963-2019 | SPMLE | 0.431 | $-0.309$ | 0.109 | 0.085 |
| | | | (0.013) | (0.026) | (0.002) | (0.005) |
| | | GMM | 0.461 | $-0.298$ | 0.112 | – |
| | | | (0.007) | (0.013) | (0.002) | – |
| KORV Net | 1963-1992 | SPMLE | 0.412 | $-0.606$ | 0.098 | 0.111 |
| | | | (0.024) | (0.047) | (0.002) | (0.015) |
| | | GMM | 0.428 | $-0.592$ | 0.098 | – |
| | | | (0.022) | (0.041) | (0.002) | – |
| | 1963-2019 | SPMLE | 0.422 | $-0.381$ | 0.097 | 0.090 |
| | | | (0.016) | (0.031) | (0.002) | (0.006) |
| | | GMM | 0.460 | $-0.339$ | 0.098 | – |
| | | | (0.008) | (0.014) | (0.002) | – |

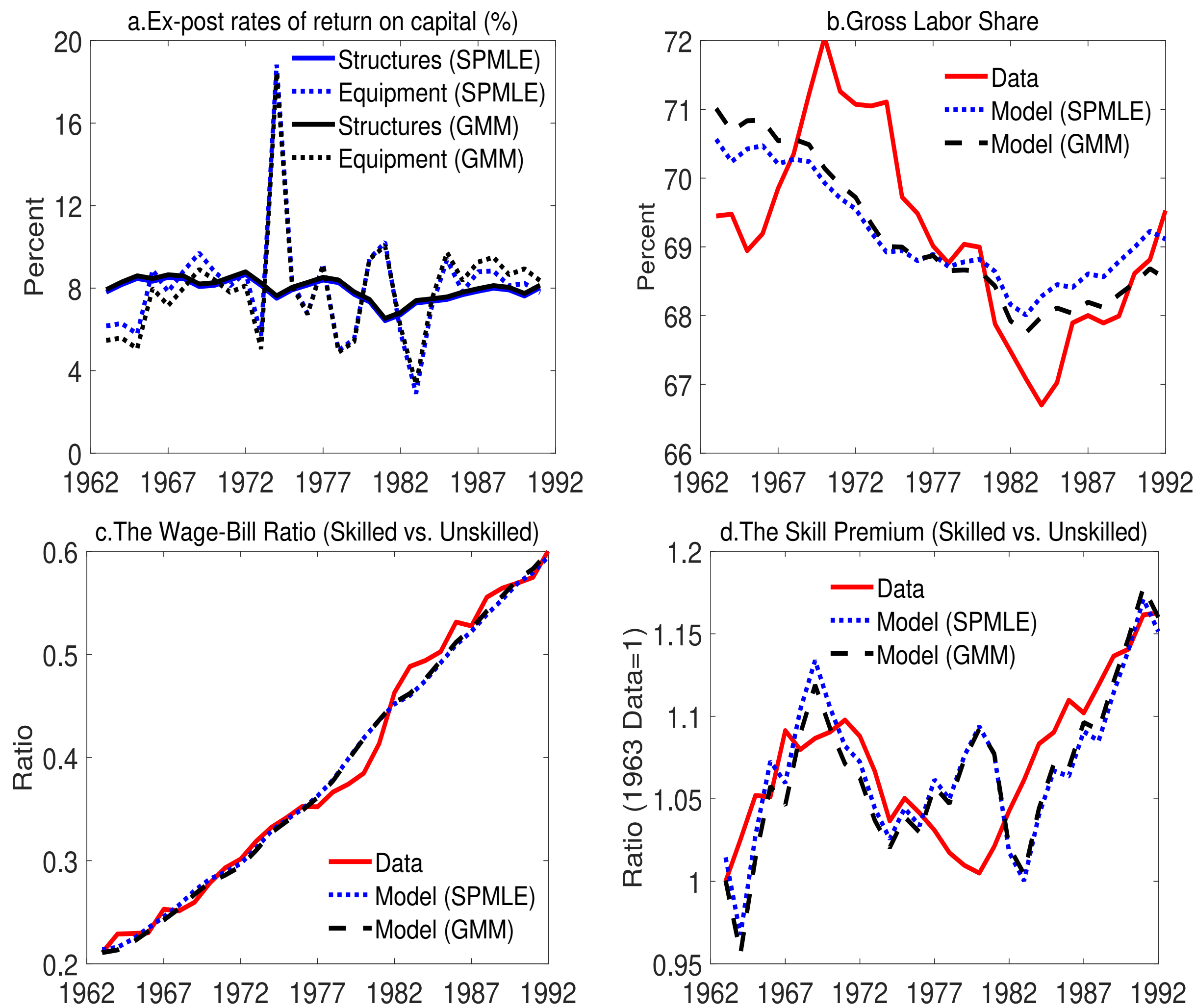
Notes: The values in parentheses are standard errors.

Figure 2.4 presents the model's fit for 1963 through 1992 using the revised data with both SPMLE and GMM estimation and with KORV's definition of gross labor share. Panels a through d shows that the model predictions are broadly in line with the data for both SPMLE and GMM. Panel d shows that the model captures the rise in the skill premium in the late 1960s, the decline until the 1980s, except for an increase in the early 1980s, and the large rise thereafter. Regarding the labor share, the model generates a labor share that is too smooth (see panel b), but the model captures a sizable component of long-run changes in labor share, predicting a decline until the early 1980s and a slight increase afterwards. Both the data and the model have the same average labor share of 69.2 percent.

The elasticity parameters estimated by Krusell et al. (2000) have been used extensively in the inequality literature. To evaluate the empirical fit of the model after 1992, we kept these parameters constant from row II in Table 2.1, which are obtained using original KORV data, but projected the model through 2019, extending the original KORV data with the growth rates of variables since 1992. Figure 2.5 shows these results with SPMLE estimation.[16]

[16] Because the model fits are very similar for two methodologies, we do not report the GMM fits in this figure.

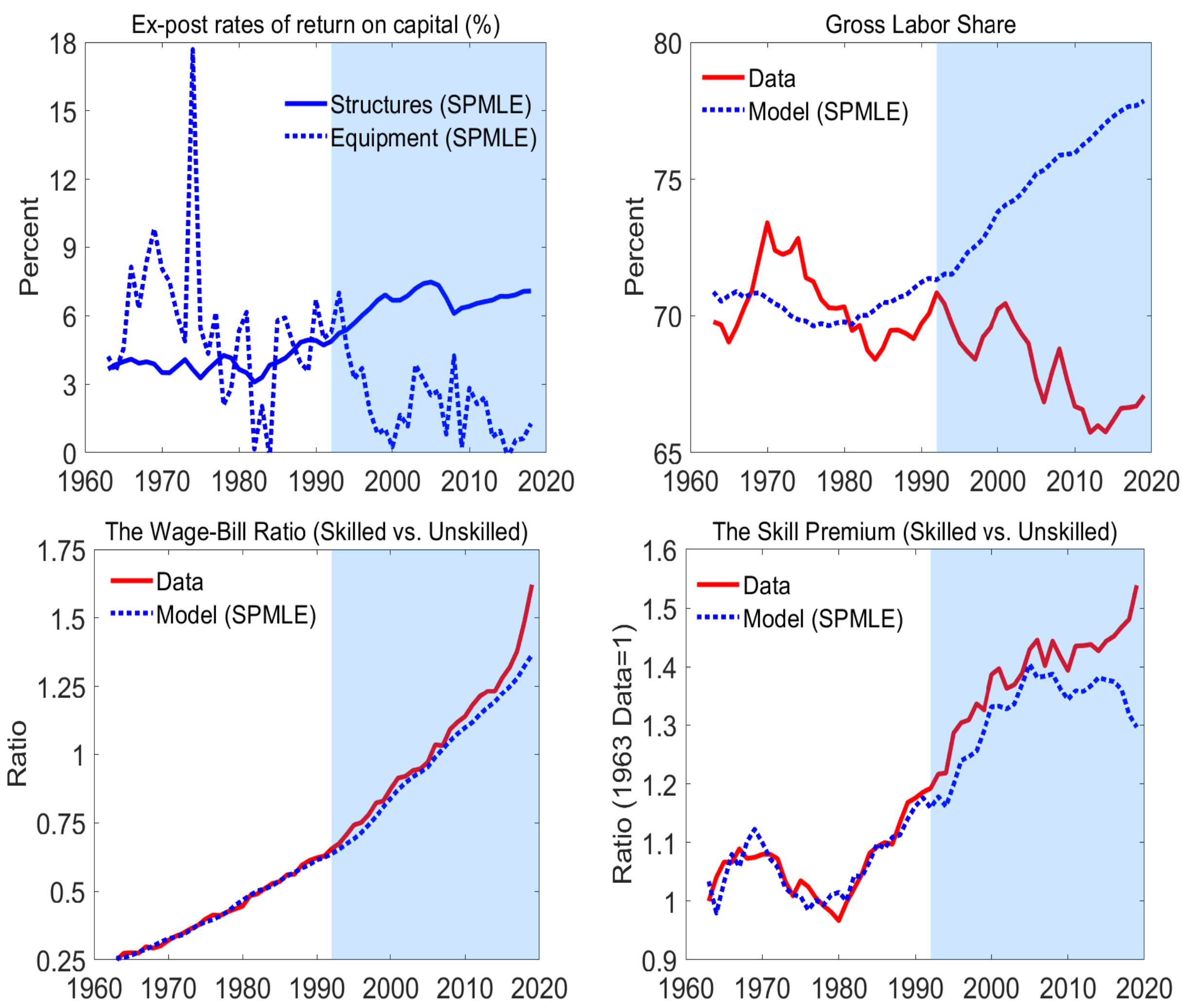
Figure 2.4: Model Fit for the 1963-1992 Period with Updated Data




Notes: These charts are produced using the observed factor inputs and the parameters estimated using data for 1963-1992. KORV's definition of gross labor share is used in the estimation. While panel a runs through 1991, the rest of the panels plot the data and the model fit for the entire 1963-1992 period.

As seen in panels c and d, the model does remarkably well regarding the wage-bill ratio and the skill premium until recently. Consider the skill premium, as shown in panel d. Although the parameters are obtained with data until 1992, the model predicts the rise in the skill premium until the early 2000s, as well as the slowdown in its growth rate until around 2014. However, the model with the original KORV parameter estimates fails to capture the increase in the skill premium in the past few years.

Figure 2.5: Model Out-of-sample Predictions for the 1993-2019 Period



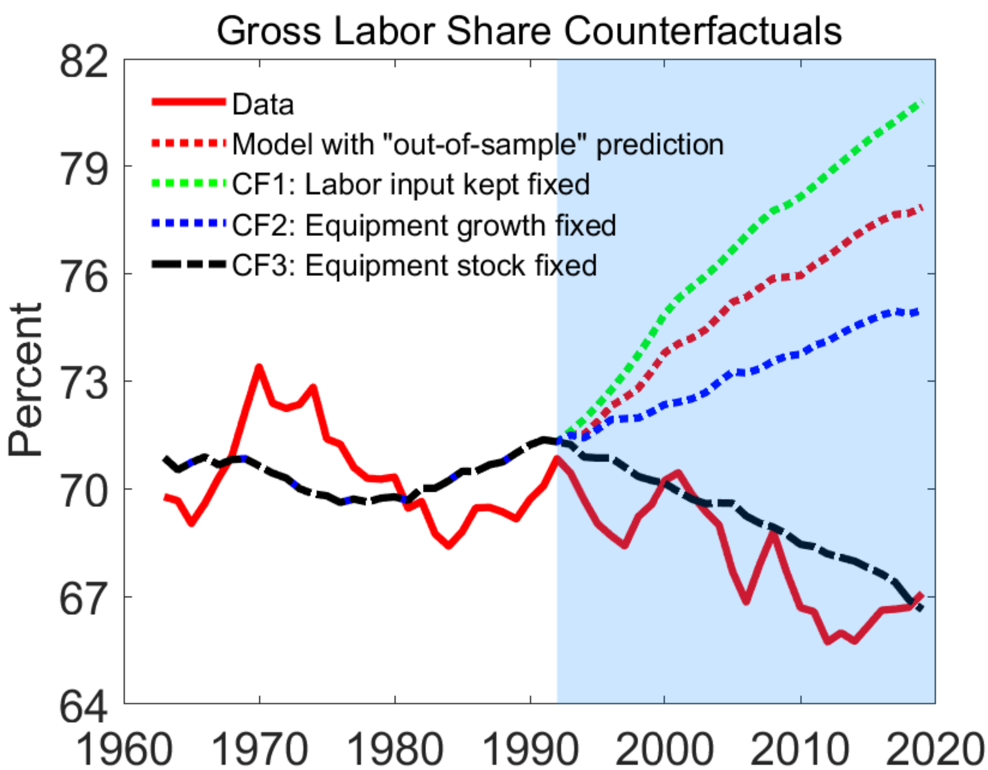

Notes: These charts are produced using the observed factor inputs and the parameters estimated with the original KORV data until 1992 and with the SPMLE methodology. KORV's definition of gross labor share is used in the estimation. The blue area represents the out of sample prediction. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963-2019 period.

The KORV framework does not capture the ongoing decline in the (gross) labor share.

In contrast, it predicts a counterfactual rise of the labor's share up to about 78 percent by 2019, from an average value of 69.2 percent. Similarly, the model tends to miss equating the ex-ante expected rates of returns of structures and equipment capital after the 1990s.

To understand why the original KORV parameters predict a large rise in the labor share, we conduct several counterfactuals, which are presented in Figure 2.6.

Figure 2.6: Gross Labor Share Counterfactuals




Notes: The chart is produced using parameters estimated using the original KORV data until 1992 with the SPMLE methodology. The blue area represents the out-of-sample dates. The solid line is the original KORV data up to 1992, and we then extend it afterwards. The red dashed line is produced using the observed factor inputs. In counterfactual 1, we keep skilled and unskilled labor inputs constant at their 1992 levels. In counterfactual 2, we keep the growth rate of equipment capital during the post-1992 period at its average rate over the 1963-1992 period. Counterfactual 3 keeps the stock of equipment capital fixed at its 1992 level.

Consider first the green dotted line, which is generated by keeping the skilled and unskilled labor inputs constant at their 1992 levels. The figure shows that keeping these inputs fixed generates an even larger increase in the labor share, as keeping these inputs fixed increases the scarcity of labor as a factor of production.

Consider next the blue and black lines, which are generated by keeping the growth rate of equipment capital during the post-1992 period at its average rate over the 1963-1992 period, and by keeping the stock of equipment capital fixed at its 1992 level, respectively. These counterfactuals show that the model's failure to track the labor share results from the enormous rise in the stock of equipment capital.

Given the large post-1992 changes in the data, it is natural to re-estimate the model and these key elasticities with data through 2019. Table 2.2 reports the parameters estimated from 1963 through 2019 for our baseline case with KORV's definition of gross labor share for the two estimation methodology we used.

As seen in Table 2.2, the parameters $\sigma$ and $\alpha$ do not change significantly from those estimated by Krusell et al. (2000) and our estimates for the 1963-1992 period (see first row of Table 2.1 and first two rows of Table 2.2, respectively). However, the estimated value of $\rho$, the parameter governing the elasticity of substitution between equipment capital and the skilled labor input shows somewhat less complementarity than in Krusell et al. (2000) when estimated through 2019, with an estimated value of -0.309 and -0.299 with the SPMLE and GMM methodologies, respectively, compared with -0.517 and -0.477, when estimated using revised data from 1963 through 1992.

From a technological perspective, the finding that equipment capital and skilled labor became somewhat less complementary since the early 1990s may reflect the notion that equipment-specific technology is now replacing some jobs involving skilled labor. For example, artificial intelligence and machine learning are now being used as a substitute for skilled labor in some industries.

From a goodness-of-fit perspective, the optimization algorithm chooses a somewhat lower complementarity between skilled labor and equipment capital to attenuate the model's prediction of a higher labor's share of income, as a very high complementarity increases skilled labor's productivity and thus labor's share.

Nonetheless, there is still significant capital-skill complementarity when the model is estimated through 2019, with an estimated elasticity between unskilled labor and equipment of 1.77, and an estimated elasticity of 0.76 between skilled labor and capital equipment with SPMLE methodology. The difference in these two elasticities is very similar to the differnce reported in KORV. Table 2.3 compares our elasticity estimates with SPMLE methodology reported in Table 2.2 with those of Krusell et al. (2000) (row I of Table 2.1).

Table 2.3: Estimated elasticities of substitution(using SPMLE methodology)

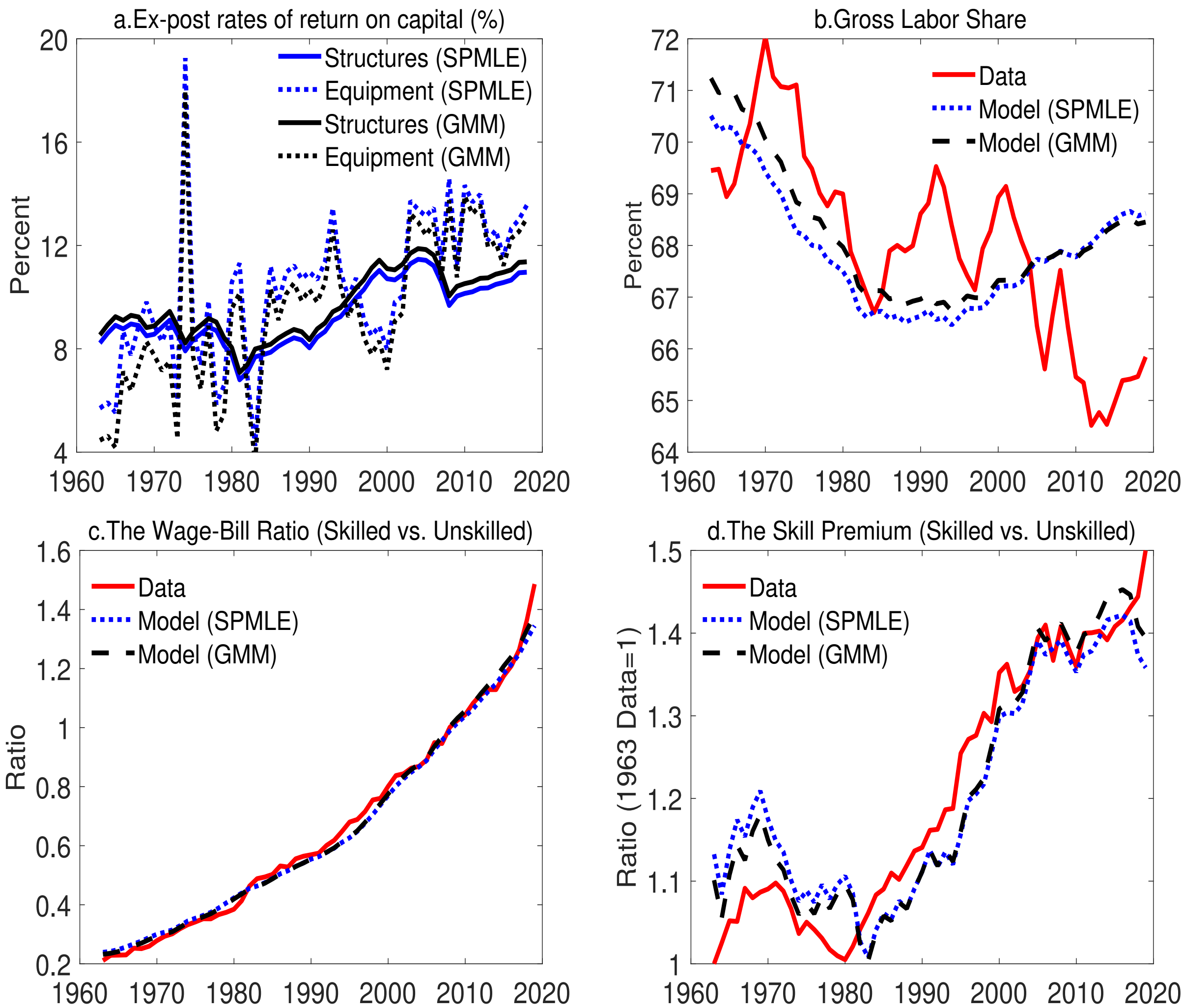
| | I. KORV (2000) (1963-1992) | II. Updated (1963-1992) | III. Updated (1963-2019) |
|---|---|---|---|
| $\frac{1}{1-\sigma}$ | 1.67 | 1.77 | 1.76 |
| $\frac{1}{1-\rho}$ | 0.67 | 0.66 | 0.76 |

Figure 2.7 presents the model's fit for 1963 to 2019 with the two estimation procedures and with KORV's definition of gross labor share. Compared to using parameters estimated for the original KORV period (Figure 2.5), the model estimated through 2019 gives a much better fit of the labor share (panel b) as the counterfactual rise in the labor share is attenuated. Although the model misses the volatility of labor's share, particularly the fall since the early 2000s, the average labor share in the model and data are both about 68 percent.

The model estimated over 1963-2019 also improves the fit of the no-arbitrage condition, as ex-post rates of return move together for both types of capital (see panel a), though the model's predicted ex-post rates of return are a bit higher than what empirical studies suggest. To compare, Marx et al. (2019) report that return on U.S. productive capital increased from 6 percent in 1980s to around 10 percent in late 1990s, before falling back to around 8 percent by 2010. In contrast, our model predicts a return on capital just above 10 percent since early 2000s.

The model continues to capture the large changes in the skill premium, which include the rise until the early 1970s, the fall until the early 1980s, and the rise thereafter, together with a slowdown in the rise since the early 2000s up until recently. This indicates that

Figure 2.7: Model Fit for the 1963-2019 Period




Notes: These charts are produced using the observed factor inputs and the parameters estimated employing data for the 1963-2019 period. KORV's definition of gross labor share is used in the estimation. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963-2019 period.

the KORV framework, and the hypothesis of capital-skill complementarity more broadly, remains quantitatively important from 1963 through 2019, a period with remarkable growth in the relative supplies of skilled and unskilled workers and a period featuring enormous technological change.

The results with constant depreciation rates are almost unchanged from the baseline estimation (see Table 2.9), suggesting that our assumption on depreciation rates has no substantive effect on our findings. This can also be clearly seen in Figure 2.17, which shows that the resulting model fit is nearly identical when either time-varying or constant depreciation rates are used for constructing the capital stocks and in the construction of capital stock series and in the estimation.

#### 2.5.2.2 Estimation with net labor share

This section discusses estimation results when the labor share is measured using income net of depreciation. As discussed earlier, gross labor share shows around a 5 percentage points decrease (our baseline), while net labor share does not have any obvious trend change, given higher deprecation.

When using net labor share, we change the labor share equation as:

$$\widetilde{lshare}_t = \frac{A_t G_{s,t} h_{s,t} + A_t G_{u,t} h_{u,t}}{A_t G_t - p_{eq,t} \delta_{eq,t} k_{eq,t} - \delta_{st,t} k_{st,t}}, \tag{2.14}$$

while the wage-bill-ratio equation and the no-arbitrage condition remain unchanged.

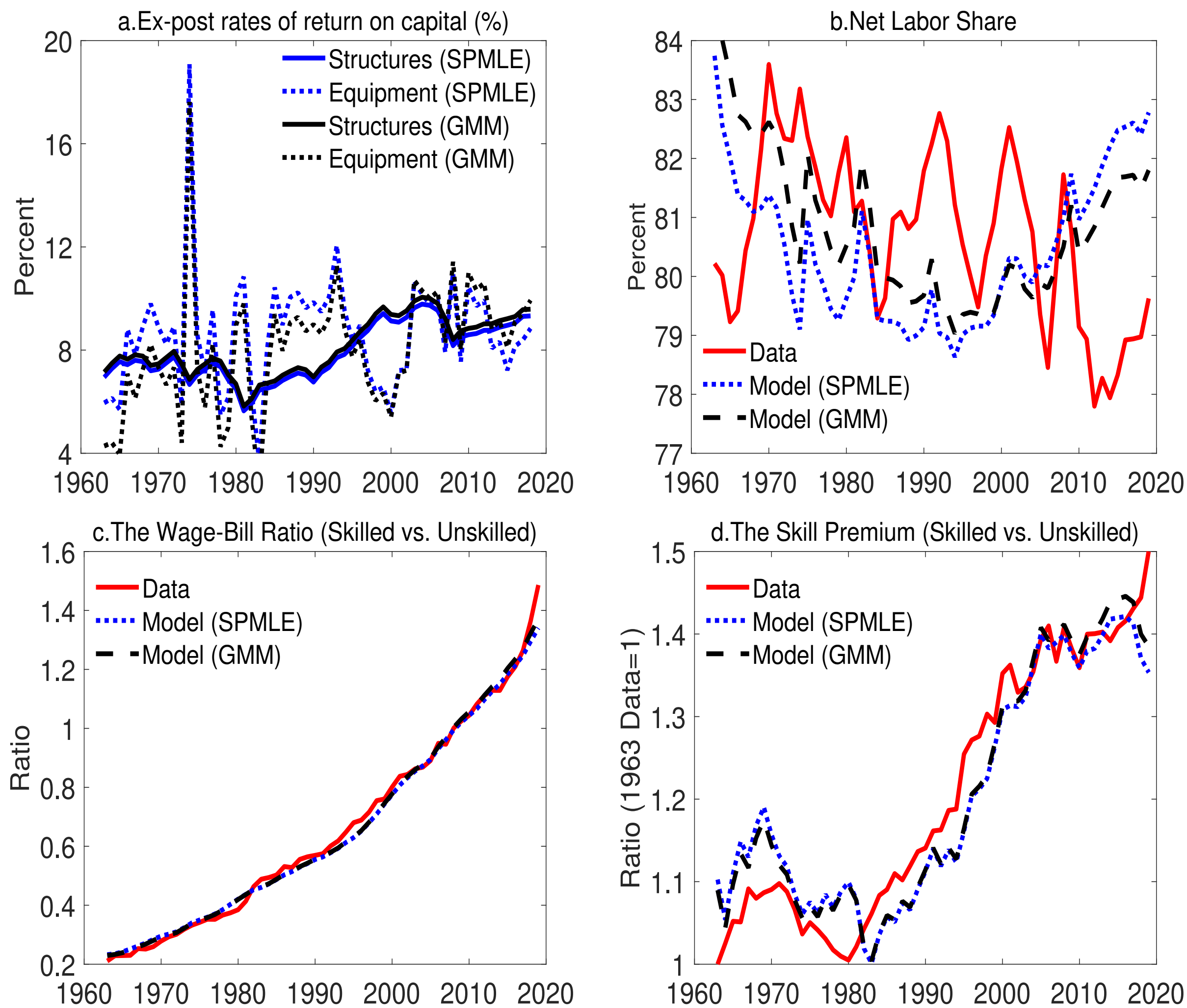
The lower block of Table 2.2 reports the parameter estimates with KORV's definition of net labor share for both periods we study. Similar to our baseline parameter estimates, capital-skill complementarity remains quantitatively important and significant with this definition of the labor share. The parameter $\rho$ is estimated to be slightly more complementary with equipment for the entire period of the study when the net labor share is targeted. With the SPMLE estimation for the 1963-2019 period, we obtain -0.381 compared to -0.309 (gross labor share) for $\rho$, which represents a lower elasticity of substitution between skilled labor and equipment capital, 0.72 compared to 0.76 estimated with gross labor share.

Figure 2.8 presents the model fit with net labor share for the 1963–2019 period with the two estimation methodologies we used. The model fit for the skill premium and wage-bill ratio improves slightly, especially for the latter period relative to our baseline case with gross labor share (see Table 2.5). More importantly, the model is consistent with the observed lower rate of return on equipment capital over the past two decades. Without the challenge of having to fit the persistent negative trend in gross labor's share, the model captures the relative stability of (net) labor's share, with an average net labor share of 80.5 percent, compared with 80.7 percent in data. That being said, using the net labor share as one of the targets instead of using the gross labor share worsens the fit of the gross labor share. This point is depicted in Figure 2.18 in Appendix 2.9, which plots the model fits for the gross labor share when either the net labor share or the gross labor share is used in the estimation. As seen in the figure, the model generates a counterfactual rise in the gross labor share when the net labor share is targeted, consistent with the fact that substitution parameters are closer to original KORV parameters in this case.

#### 2.5.2.3 Summary of elasticity estimates across alternative labor share definitions

The different cases studied in this paper confirmed that capital-skill complementarity remains important. While both the unskilled and skilled estimated substitution elasticities with equipment are somewhat higher than in KORV, the difference between the two elasticities is very similar to that in KORV. We averaged the elasticity estimates we obtained using the four different labor share types and two estimation methodologies for the two periods we studied. As seen in Table 2.4, on average, our elasticity estimates are almost unchanged from what Krusell et al. (2000) report for the 1963-1992 period: 1.67 for the elasticity between unskilled labor and the composite of equipment and skilled labor and 0.65 for the elasticity

Figure 2.8: Model Fit for the 1963-2019 Period (with KORV's Definition of Net Labor Share)




Notes: These charts are produced using the observable factor inputs and the parameters estimated employing data for the 1963-2019 period. KORV's definition of labor share, net of depreciation, is used in estimation. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963-2019 period.

between skilled labor and equipment. When the 1963-2019 period is considered, a modestly larger pair of elasticities–1.70 and 0.75, respectively–are estimated in order to address the secular decline in the labor share.

Table 2.4: Averaged Elasticity Estimates

| | **Period** | $\frac{1}{1-\sigma}$ | $\frac{1}{1-\rho}$ |
|---|---|---|---|
| Krusell et al. (2000) | 1963–1992 | 1.67 | 0.67 |
| Updated estimate | 1963–1992 | 1.67 | 0.65 |
| Updated estimate | 1963–2019 | 1.70 | 0.75 |

Notes: Updated estimates are average of eight different estimates obtained using four labor share types: KORV's definition of gross and net labor shares and NFBS gross and net labor shares; and two estimation methodologies: SPMLE and GMM.

#### 2.5.2.4 Comparison of model fits across alternative labor share definitions

To analyze the model fits for the skill premium and labor shares, we compare normalized root mean squared model errors (NRMSEs) for both gross and net income for both the KORV and nonfarm business income definitions. Table 2.5 reports NRMSEs for both the the 1963-1992 and the 1963-2019 periods using the SPMLE methodology in estimation.

Table 2.5: Normalized RMSEs for the Skill Premium and the Labor Share

| | **Skill Premium** | | **Labor Share** | |
|---|---|---|---|---|
| | 1963–1992 | 1963–2019 | 1963–1992 | 1963–2019 |
| KORV Gross | **0.033** | 0.051 | **0.015** | **0.028** |
| KORV Net | 0.035 | 0.044 | 0.018 | **0.028** |
| NFBS Gross | **0.033** | 0.044 | 0.016 | 0.032 |
| NFBS Net | 0.035 | **0.042** | 0.020 | 0.033 |

Notes: RMSE stands for root mean squared error. Normalized RMSE is the RMSE divided by the mean value of the variable during the relevant period. Bold entries represent the smallest value in each column.

Regarding the skill premium, all labor share definitions perform similarly well for the 1963-2019 period, while we see a non-negligible improvement in the fit when we use KORV's net labor share definition as well as two types of NFBS labor shares. Regarding the labor share, estimation with KORV's gross labor share (our baseline) slightly outperforms all other versions for both periods once the RMSEs are corrected (normalized) to account for different levels of labor share definitions.

Overall, the findings are quantitatively similar, with no specific case appearing to be superior to the others in all aspects.

#### 2.5.2.5 Estimation with information and communication technologies capital

This section analyzes the implications of changing the complementary capital stock from total equipment capital to ICT capital. We use the quality-adjusted ICT price deflator of DiCecio (2009) to construct real ICT investment, and we apply the perpetual inventory method to construct the annual ICT capital stock. Non-ICT capital equipment is added to the stock of capital structures.

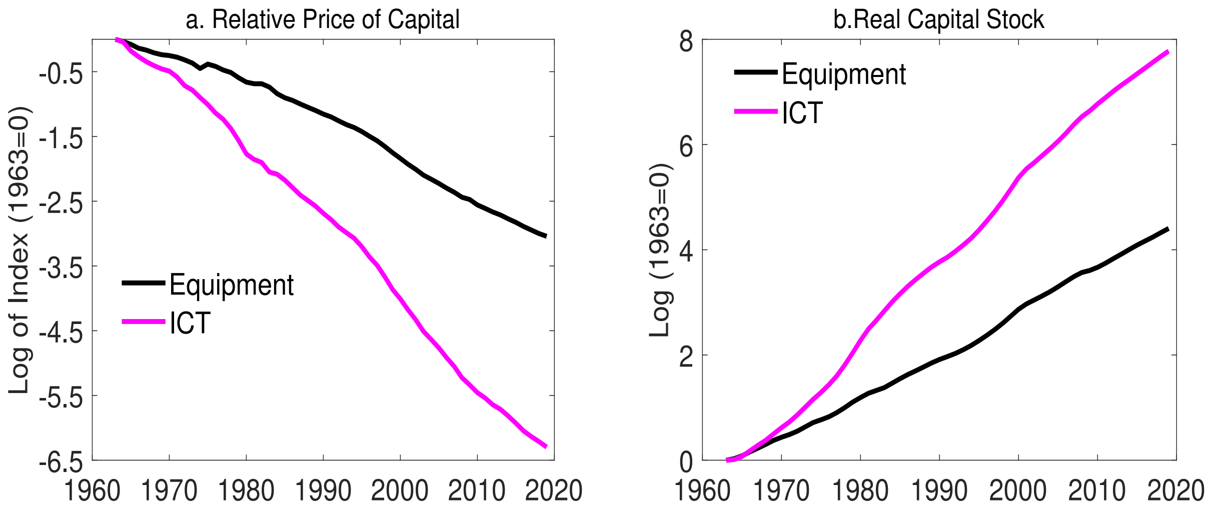
Extending the KORV framework along this dimension allows us not only to evaluate whether the model labor share will be closer to the data, but also focuses the analysis on a concept of a complementary capital stock that has exhibited the fastest technological change.

Panel a of Figure 2.9 shows the very rapid drop in the relative price of ICT capital compared to total capital equipment (including ICT capital), while panel b shows the corresponding very rapid rise in the real stock of ICT capital compared to the stock of total equipment. These differences reflect faster technological change in ICT capital compared to total equipment.

We estimate the model using both SPMLE as in Krusell et al. (2000) and using GMM. The results for the elasticity parameters, along with the results with equipment capital (our baseline) are presented in Table 2.6.

The parameter estimates show strong capital-skill complementarity using ICT capital as the complementary capital stock. We find an elasticity of substitution of 2.52 between unskilled labor and the composite output of ICT capital and skilled labor and an elasticity of substitution of 0.93 between ICT capital and skilled labor using the SPMLE methodology. Corresponding GMM elasticity estimates are 3.01 and 0.90, respectively.

Figure 2.9: Equipment vs ICT Capital




Source: Authors' calculations from the BEA's NIPA tables and ICT and equipment capital investment price deflators of DiCecio (2009).

Table 2.6: Comparison of Estimates of Parameters Governing Elasticities (with KORV's Definiton of Gross Labor Share)

| **Capital** | **Methodology** | $\sigma$ | $\rho$ |
|---|---|---|---|
| Equipment | SPMLE | 0.431 | −0.309 |
| | | (0.013) | (0.026) |
| | GMM | 0.461 | −0.298 |
| | | (0.007) | (0.013) |
| ICT | SPMLE | 0.603 | −0.077 |
| | | (0.050) | (0.008) |
| | GMM | 0.669 | −0.109 |
| | | (0.007) | (0.009) |

Notes: The values in parentheses are standard errors.

The larger unskilled labor elasticity compared to the baseline model with total equipment suggests that ICT capital is significantly more substitutable with unskilled labor than is the total stock of equipment.

This finding connects with the closely related literature on the skill content of jobs introduced in Autor et al. (2003), and further developed in Autor (2015), and related studies of automation replacing routine jobs, including Acemoglu and Restrepo (2018), Acemoglu and Restrepo (2019), Eden and Gaggl (2018), and Eden and Gaggl (2019). In particular, unskilled workers primarily occupy the routine jobs that are being replaced by automation, which in turn reflects rapid innovations in ICT capital.

Based on these related literatures, it is natural to expect a higher substitution elasticity. This analysis thus provides a quantitative estimate for how much higher this elasticity is.

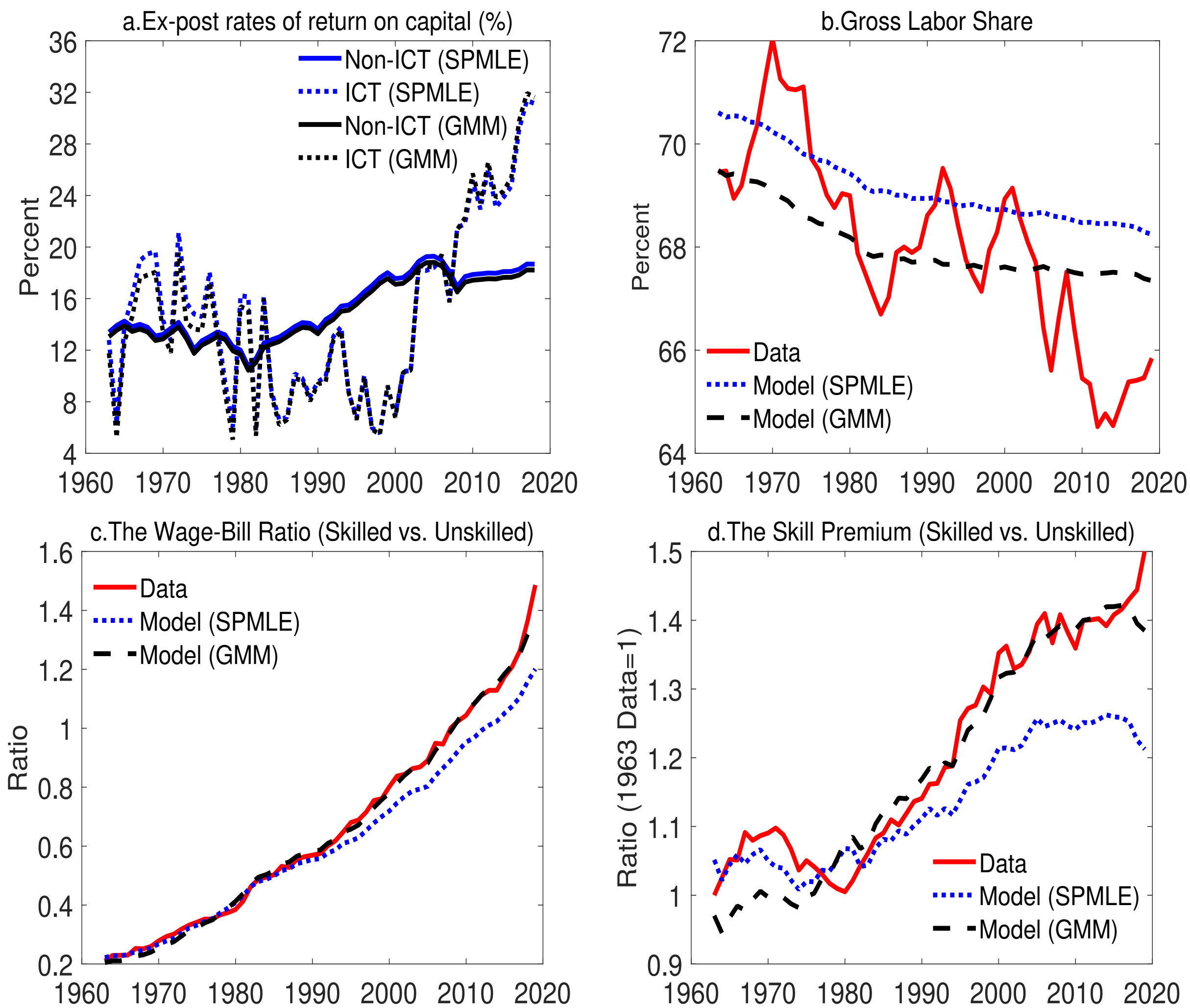
The differences in the substitution elasticities between SPMLE and GMM have implications for the overall model fit and the skill premium, which are presented in Figure 2.10.

Using ICT capital moderately improves the fit for the model labor share relative to using the total stock of equipment.[17] Both the SPMLE and GMM estimates (see panel b of Figure 2.10) show about a three percentage point decline in labor share over the full period of analysis (1963-2019) using ICT capital. However, neither estimation method produces a significant labor share decline after 1992, a period in which the actual labor share declines by about five percentage points. This finding is consistent with the results in Eden and Gaggl (2018), who use a different classification of labor input (routine and non-routine tasks), and a different price deflator and investment series for the construction of the ICT capital stock to study the welfare effects of automation (growth of ICT capital) and find also that the increasing use of ICT has been responsible for about half of the decline in the labor income share.[18] Our findings are also similar to Eden and Gaggl (2018) in that both analyses

[17] See Figure 2.19 in the Appendix for a comparison of model fits for KORV's gross labor share when either equipment or ICT capital is used as complementary capital.

[18] Eden and Gaggl (2018) construct nominal ICT investment series based on data from the World Information Technology and Services Alliance (WITSA) and the International Telecommunication Union (ITU)

Figure 2.10: Model Fit for the 1963-2019 Period (with ICT Capital and KORV's Definition of Gross Labor Share)




Notes: These charts are produced using the observable factor inputs and the parameters estimated employing data for the 1963-2019 period. KORV's definition of gross labor share and ICT capital are used in estimation. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963-2019 period.

suggest that capital-biased technological change in conjunction with empirically plausible substitution possibilities across factors do not fully account for the change in labor share.

The two estimation methods fit the no-arbitrage condition equally well, though neither captures the large increases in the marginal product of capital equipment that occur after 2010. The GMM parameter estimates fit the skilled-unskilled wage bill ratio very closely over the full period, while the SPMLE parameter estimates generate a wage bill ratio that grows more slowly than the actual wage bill ratio after around 2000.

In contrast to all the other cases considered, the model skill premium differs between the two estimation methods. Panel d of Figure 2.10 shows the model skill premium for both estimation methods. The model skill premium estimated with GMM tracks the actual skill premium quite closely over the entire period, including capturing the slowdown in the growth of the skill premium between 2002 and 2016, a period in which the normalized skill premium remains at around 40 percent above its 1963 level. The SPMLE model skill premium does not track the data as closely, particularly the increases that occur after 2005.

The more accurate model skill premium produced using GMM reflects a larger degree of capital skill complementarity compared to SPMLE. Both methods estimate a skilled labor - ICT capital elasticity of around 0.9, but the estimated elasticity between unskilled labor and the composite output of ICT capital and skilled labor is about 3 for GMM, and about 2.5 for SPMLE.

To understand the importance of these differences for the model skill premium, recall from equation 2.8 that the difference between the parameters governing elasticities ($\sigma$ and $\rho$) is a key factor driving the skill premium. With GMM, the difference between $\sigma$ and $\rho$ is about 0.78, whereas this difference with SPMLE is about 0.68, which correspond to a difference between $\frac{1}{1-\sigma}$ and $\frac{1}{1-\rho}$ of 2.1 and 1.6, respectively.

databases. They treat non-ICT investment as the residual between total capital investment and this ICT investment. Instead, we take "Information processing equipment investment" (line 10 from NIPA Table 5.3.5) as the nominal ICT investment and use the difference between total equipment investment (line 9 from NIPA Table 5.3.5) and "Information processing equipment investment" (line 10 from NIPA Table 5.3.5) as nominal non-ICT equipment investment. Additionally, Eden and Gaggl (2018) use an ICT price deflator estimated based on the BEA's fixed asset accounts, while we used the DiCecio (2009) deflator.

These results raise the question of why the SPMLE and GMM estimates are different in this case, whereas they are very similar in all other cases. The primary distinction between the two estimation methods is that GMM does not need to estimate the variances of the latent labor efficiency process, whereas these variances are needed to prevent singularity in the full-information SPMLE estimation method used originally in Krusell et al. (2000).

We found difficulty in achieving convergence with SPMLE without a very volatile labor efficiency shock process, one in which the standard deviation of labor efficiencies is nearly 30 percent per year. This strikes us as being implausibly large. We conjecture that this high volatility may be due to the fact that ICT capital grows so much more quickly than the stock of total equipment; real ICT capital rises by a factor of nearly 2,500 since 1963, while the total stock of equipment rises by about a factor of 83 over the same period.

Given that GMM provided reasonable results in all the other cases evaluated here, we focus on the GMM estimates for the case of ICT capital, which yield substitution elasticities of about 3 between unskilled labor and the composite output of ICT capital and skilled labor, and about 0.9 between skilled labor and capital, and leave the technical issues regarding SPMLE estimation for future research.

## 2.6 Conclusion

This paper analyzes the quantitative importance of capital-skill complementarity as a determinant of U.S. wage inequality, using data through 2019, compared with Krusell et al. (2000), which used data through 1992.

We first study the out-of-sample performance of the original Krusell et al. (2000) framework, which predicts a rise in the skill premium after 1992, but a counterfactual rise in labor's share of income. We then study how alternative measures of income, labor's share of income, depreciation, the conceptual definition of the complementary capital stock, and an alternative estimation method, affect the model's ability to jointly capture the skill premium and labor's share.

We find that capital-skill complementarity continues to be a quantitatively important determinant of U.S. wage inequality, despite the five percentage point decline in labor's share that has occurred after the Krusell et al. (2000) estimation period ended. In all of the estimation cases, the post-1992 decline in labor's share results in an estimated degree of complementarity between skilled labor and capital that is slightly lower than in Krusell et al. (2000), as well as a modestly higher estimated elasticity between unskilled labor and composite output of equipment and skilled labor. The reason for this is because slightly higher elasticities, ceteris paribus, reduce the marginal productivities compared to lower elasticities, and thus do not put as much upward pressure on the model's labor share. The largest departure we consider from Krusell et al. (2000) is replacing the total stock of equipment capital with information-communications-technology capital as the capital stock that is complementary with skilled labor. When estimated with GMM, the model captures the skill premium closely and produces a slowly declining labor share between 1963-1992. However, the model does not capture the drop in labor's share that occurs after 2000.

For economic models that posit the total stock of equipment as the complementary capital stock, we find that the KORV elasticity estimates change very little, from about 1.67 to about 1.70 for unskilled labor and equipment, and from about 0.67 to about 0.75 for skilled labor and equipment. For models that posit ICT capital as the complementary capital stock, we find the unskilled labor elasticity rises to about 3, which may reflect the process of routine jobs being replaced by automation, as analyzed in Autor (2015) and Acemoglu and Restrepo (2018). For the ICT case, the substitution elasticity between skilled labor and capital is about 0.9.

This study finds that capital-skill complementarity remains strong in U.S. data between 1963-2019, accounting for much of the change in wage inequality between highly skilled and less-skilled workers. However, the KORV framework does not yet capture the drop in labor's share that has occurred more recently. A task for future research is to focus on jointly accounting for the dynamics of wage inequality and the decline in labor's share within this framework.

## 2.7 Appendix: Data description

### 2.7.1 Construction of labor inputs and wage rates

Our labor and wage series construction follow earlier studies, including Katz and Murphy (1992), Krusell et al. (2000), Autor et al. (2008) and Domeij and Ljungqvist (2019). We use all of the person-level data, excluding the agents who are younger than 16 or older than 70, unpaid family workers, those working in the military, who were not in the labor force in the previous year, and who did not report their education level. We included the self-employed when constructing labor inputs, even though we excluded them from our wage sample. This gave us a better match of original KORV data, but excluding or including the self-employed from the labor input construction did not have any significant effect on our findings. In our wage sample, we also dropped the observations reporting working less than 40 weeks or 35 hours a week or both. Following Domeij and Ljungqvist (2019), we also dropped individuals with allocated income, those with hourly wages below half of the minimum federal wage rate, and those whose weekly pay was less than $62 in 1980 dollars from our wage sample.

For each person, we record their personal characteristics: age, sex, race; employment statistics: employment status (empstat), class of worker (classwly), weeks worked last year (wkswork1 and wkswork2), usual hours worked per week last year (uhrsworkly and shrsworky), income—total wage and salary income (incwage)—and CPS personal supplement weights: asecwt. Then, each person is assigned to one of 264 groups created by age, race, sex, and skill (education). Age is divided into 11 five-year groups: 16–20, 21–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, an 66–70. Race is divided into three: white, black, others; sex is divided into male and female, and education is divided into four groups: below high school, high school, some college and college graduates and beyond.

Following Krusell et al. (2000), we did not do any correction for topcodes. Alternatively, we also adjusted topcoded income variables using the “revised income top-codes files" published by the Census Bureau to swap top-coded values in 1976-2010 CPS files with these revised values. This procedure replaces the top-coded values with new values based on the Income Component Rank Proximity Swap method, which was introduced in 2011. With this method, we had the top-coding methodology consistent and comparable for most of the years in our sample. Because the main results remained unchanged in this alternative case, we reported only findings without corrections for topcodes, for the sake of comparability with Krusell et al. (2000).

For the CPS years after 1975, CPS has usual hours worked per week and weeks worked last year. Thus, calculating the annual hours for a person is straightforward: We simply multiply weeks worked last year by usual hours worked. Hence, for CPS years 1976 and after, total hours are:

$$hours_{i,t-1} = wkswork1_{i,t-1} \times uhrswork_{i,t-1}$$

where $i$ is individual observation and $t$ is the CPS year.

For earlier years, we need to do two adjustments. First, weeks worked are available only as intervals, and we need to approximate a scalar value for each interval. Fortunately, both the intervals and actual weeks are available for years after 1975. Therefore, we calculated the average weeks worked after year 1975 for each interval and replaced the earlier years with those values.

Second, we have to use the “hours worked last week" variable as a proxy to usual hours worked per week last year. However, there are many agents who were not employed the week before the survey or who were employed but not at work for some reason, despite reporting a positive income for the previous year. Rather than dropping those observations, we replaced the hours they worked per week with the average of the hours worked by the people in their group in that particular year. We also paid attention to whether the person was employed part time or full time when doing this replacement.

Hourly wage is calculated as

$$wage_{i,t-1} = \frac{incwage_{swapped_{i,t-1}}}{hours_{i,t-1}}$$

Later, observations with weeks worked less than 40 hours, weekly hours less than 35, hourly wage less than half the minimum wage, and weekly pay less than \$62 in 1980 dollars are dropped to smooth out the effect of outliers and misreporting. Following this, for each groups and year, we calculate group weights as $\mu_{g,t} = \sum_{i \in g} \mu_{i,t}$, where $i \in g$ is set of groups. Then, average hours and wage measures for each group and year are calculated as follows:

$$hours_{g,t-1} = \frac{\sum_{i \in g} \mu_{i,t} \times hours_{i,t-1}}{\mu_{g,t}}$$
$$wage_{g,t-1} = \frac{\sum_{i \in g} \mu_{i,t} \times wage_{i,t-1}}{\mu_{g,t}}.$$

To aggregate across groups into aggregate task groups, we follow Krusell et al. (2000) and use the group wages of 1980 as the weights. We have total hours

$$N_{t-1} = \sum_{g \in G} hours_{g,t-1} \times \mu_{g,t} \times wage_{g,80}$$

and the average hourly wage is

$$W_{t-1} = \frac{\sum_{g \in G} hours_{g,t-1} \times \mu_{g,t} \times wage_{g,80}}{N_{t-1}}.$$

### 2.7.2 Labor share

#### 2.7.2.1 Gross labor share

As the baseline, we followed KORV's definition of labor share, which we constructed from the BEA National Income and Product Accounts (NIPA) Tables 1.10 and 1.17.5. To calculate the gross labor share, we first constructed the capital's income share following Cooley and Prescott (1995) as the ratio of the sum of unambiguous capital income (net interest and miscellaneous payments (domestic industries), rental income of persons with capital consumption adjustment, corporate profits with inventory valuation and capital consumption adjustments (domestic industries) and depreciation (consumption of fixed capital) to the difference between gross domestic income and proprietors' income. We then subtracted this ratio from 1 to obtain the gross labor share.

An alternative measure uses the nonfarm business sector labor share, defined as "total employee compensation in the nonfarm business sector excluding self-employment income" divided by "gross value added in the nonfarm business sector excluding self-employment income." It is constructed using NIPA Tables 1.3.5, 1.12, 1.13 and 6.2. To construct the numerator, we take "compensation of employees, domestic industries (NIPA Table 6.2, line 2)" and subtract the following items from it: farm compensation (Table 6.2, line 5), federal general government compensation (Table 6.2, line 88), state and local general government compensation (Table 6.2, line 93), compensation of households (Table 1.13, line 43), and compensation of institutions (Table 1.13, line 50). To obtain the denominator, we take "gross value added in the nonfarm business sector (Table 1.3.5, line 3)" and subtract "sole proprietors income in the nonfarm business sector (Table 1.12, line 11)" from it. The labor share still demonstrates about a 3 percentage points decline, a little less than about 5 percentage points seen when the KORV's measure of labor share is considered.

In short, when calculating the alternative labor share, we subtract farm and government compensations from both the numerator and denominator from the KORV's measure. Economically, the difference between KORV's definition and the alternative measure of labor share is the sectors: the former takes the whole economy while the latter focuses on the nonfarm business sectors.

#### 2.7.2.2 Net labor share

Our first measure of net labor share is comparable to the first measure of gross labor share (the KORV version), only excluding consumption of fixed capital from the numerator and denominator when calculating the income share of capital. The alternative net labor share measure is the nonfarm business sector labor share net of depreciation. To build it, we only replace the "gross value added in the nonfarm business sector (Table 1.3.5, line 3)" with "net value added in the nonfarm business sector (Table 1.9.5, line 3)." We see a slight increase in this measure of labor share over years after taking into account the effect of depreciation. The surge around 2000 is largely attributed to an increase in employee compensations, and a stable series of value added.

## 2.8 Appendix: Estimation techniques

### 2.8.1 Simulated Pseudo-Maximum Likelihood Estimation

The estimation process entirely follows Krusell et al. (2000). The process is a simulated two-stage pseudo-maximum likelihood estimation (SPMLE) method developed by White (1996). Here, we are providing a brief description borrowed from KORV. Further details can be found in the original paper, particularly in the working paper version.

In the first stage, we treat the skilled and unskilled labor input as endogenous, and project them onto a constant and a trend; current, and lagged stocks of capital equipment and structures; the lagged relative price of equipment; and the lagged value of the U.S. Commerce Department's composite index of business cycle indicators. Then in the second stage, we use the fitted values of skilled and unskilled labor input from the regression in stage 1 to estimate the model. We define the vector $\tilde{X}_t$ as consisting of the stocks of equipment and structures and of the instrumented values of skilled and unskilled labor input: $\tilde{X}_t = \left\{k_{st,t}, k_{eq,t}, \hat{h}_{s,t}, \hat{h}_{u,t}, \delta_{eq,t}, \delta_{st,t}\right\}$, where $\hat{h}_{s,t}$ and $\hat{h}_{u,t}$ stand for the fitted values for skilled and unskilled labor.

In the second stage, we use the instruments and the instrumented values of the labor input series in SPMLE to estimate the parameters of the model. This proceeds as follows: Given the distributional assumptions on the error terms, for each date $t$ observation, we generate $S$ realizations of the dependent variables, each indexed by $i$, by following two steps:

$$\begin{aligned} &\text{Step 1:} \quad \varphi_t^i = \varphi_0 + \gamma t + \omega_t^i \\ &\text{Step 2:} \quad Z_t^i = f\left(\tilde{X}_t, \psi_t^i, \varepsilon_t^i; \phi\right). \end{aligned} \tag{2.15}$$

In Step 1, a realization of $\omega_t$ is drawn from its distribution and used to construct a year $t$ value for $\varphi_t$. In Step 2, this realization of $\varphi_t$, together with a draw of $\varepsilon_t$ allows us to generate a realization of $Z_t$. By simulating the model, we obtain the first and second moments of $Z_t$:

$$\begin{aligned} m_S\left(\tilde{X}_t;\phi\right) &= \tfrac{1}{S}\textstyle\sum_{i=1}^{S} f\left(\tilde{X}_t,\psi_t^i,\varepsilon_t^i;\phi\right) \\ V_S\left(\tilde{X}_t;\phi\right) &= \tfrac{1}{S-1}\textstyle\sum_{i=1}^{S}\left(Z_t^i - m_S\left(\tilde{X}_t;\phi\right)\right)\left(Z_t^i - m_S\left(\tilde{X}_t;\phi\right)\right)'. \end{aligned} \tag{2.16}$$

On the basis of these moments constructed for each $t = 1...T$, we can write the second stage objective function as:

$$\begin{aligned} \ell^2\left(Z^T;\tilde{X}_t,\phi\right) = -\frac{1}{2T}\sum_{t=1}^{T}\Big\{&\left[Z_t - m_S\left(\tilde{X}_t;\phi\right)\right]'\left(V_S\left(\tilde{X}_t;\phi\right)\right)^{-1}\left[Z_t - m_S\left(\tilde{X}_t;\phi\right)\right] \\ &- \log\det\left(V_S\left(\tilde{X}_t;\phi\right)\right)\Big\}. \end{aligned} \tag{2.17}$$

The SPML estimator $\hat{\phi}_{ST}$ is defined as the maximizer of equation (2.17). Following Krusell et al. (2000), we compute the standard errors using Theorem (6.11) in White (1996).

*Standard errors:*

The computations of the exact asymptotic standard errors take into account the first-stage parameter uncertainty in the instrumental variable estimation. Define the set of potentially endogenous variables as $X^T$ and the set of instruments as $W^T$ in the first stage. Clearly, the projection in the first stage can be regarded as a special case of maximum likelihood estimation, and we denote the first-stage likelihood function as $\ell^1\left(X^T;W^T,\theta\right)$, where $\theta$ is the parameters of this first-stage likelihood function. The second-stage likelihood function is $\ell^2\left(Z^T;\tilde{X}^T\left(W^T,\theta^*\right),\phi\right)$, where $\tilde{X}^T\left(W^T,\theta^*\right)$ is the linear projection of $X^T$ in the space of $W^T$, and the "*" parameters denote the pseudo-true values.

Let $\nabla_\theta$ and $\nabla_{\theta\theta}$ denote the first and second derivative with respect to $\theta$. The Hessian matrix and information matrix are as follows:

$$H^* = \begin{bmatrix} \nabla_{\theta\theta}\ell^1\left(\theta^*,\phi^*\right) & \nabla_{\theta\phi}\ell^1\left(\theta^*,\phi^*\right) \\ \nabla_{\phi\theta}\ell^2\left(\theta^*,\phi^*\right) & \nabla_{\phi\phi}\ell^2\left(\theta^*,\phi^*\right) \end{bmatrix} = \begin{bmatrix} \nabla_{\theta\theta}\ell^1\left(\theta^*,\phi^*\right) & 0 \\ \nabla_{\phi\theta}\ell^2\left(\theta^*,\phi^*\right) & \nabla_{\phi\phi}\ell^2\left(\theta^*,\phi^*\right) \end{bmatrix} \tag{2.18}$$

$$I^* = \begin{bmatrix} \nabla_\theta \ell^1(\theta^*) \cdot \nabla'_\theta \ell^1(\theta^*) & \nabla_\theta \ell^1(\theta^*) \cdot \nabla'_\phi \ell^2(\theta^*, \phi^*) \\ \nabla_\phi \ell^2(\theta^*, \phi^*) \cdot \nabla'_\theta \ell^1(\theta^*) & \nabla_\phi \ell^2(\theta^*, \phi^*) \cdot \nabla'_\phi \ell^2(\theta^*, \phi^*) \end{bmatrix} \tag{2.19}$$

Theorem 6.11 in White (1996) establishes that the asymptotic variance-covariance matrix of $\hat{\phi}_T$ is $\text{var}\left(\hat{\phi}_T\right) = H_{22}^{*-1}\left[I_{22}^* - H_{21}^{*\prime} H_{11}^{*-1} I_{12}^* - I_{21}^* H_{11}^{*-1} H_{21}^* + H_{21}^{*\prime} H_{11}^{*-1} I_{11}^* H_{11}^{*-1} H_{21}^*\right] H_{22}^{*-1}$. To compute the asymptotic variance of our simulation-based estimates of the parameters, we replace in the above expressions $\theta^*$ by $\hat{\theta}^T$ as well as $\phi^*$ and $\hat{\phi}_T$ by $\hat{\phi}_{ST}$.

### 2.8.2 Generalized Method of Moments

Denote $Y_t$ as the annual sample data used for estimation, and $\theta = (\alpha, \sigma, \mu, \rho, \lambda, \varphi_u)$ the parameters to estimate. To apply GMM, we use the same three moment conditions as in the SPMLE methodology: the wage-bill ratio, the labor share, and the no-arbitrage condition.

$$g(Y_t, \theta) = \begin{bmatrix} \frac{w_{s,t} h_{s,t}}{w_{u,t} h_{u,t}} - \text{wbr}_t \\ \frac{w_{s,t} h_{s,t} + w_{u,t} h_{u,t}}{Y_t} - \text{lshare}_t \\ q_t A_{t+1} G_{eq,t+1} + \left(1 - \delta_{eq,t+1}\right) E\left(\frac{q_t}{q_{t+1}}\right) - A_{t+1} G_{st,t+1} - \left(1 - \delta_{st,t+1}\right) \end{bmatrix} \tag{2.20}$$

The goal of estimation is to find a set of parameters $(\theta_0)$ that satisfy:

$$m(\theta_0) \equiv \mathrm{E}\left[g(Y_t, \theta_0)\right] = 0, \tag{2.21}$$

where $m(\theta_0)$ is the value of the function $g(Y_t, \theta)$ evaluated at $\theta_0$ and $E$ is expected value. Then, we replace the theoretical expected value $E[\cdot]$ with its sample average:

$$\hat{m}(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta). \tag{2.22}$$

We then minimize the norm of this expression with respect to $\theta$. Therefore, the GMM estimator can be written as:

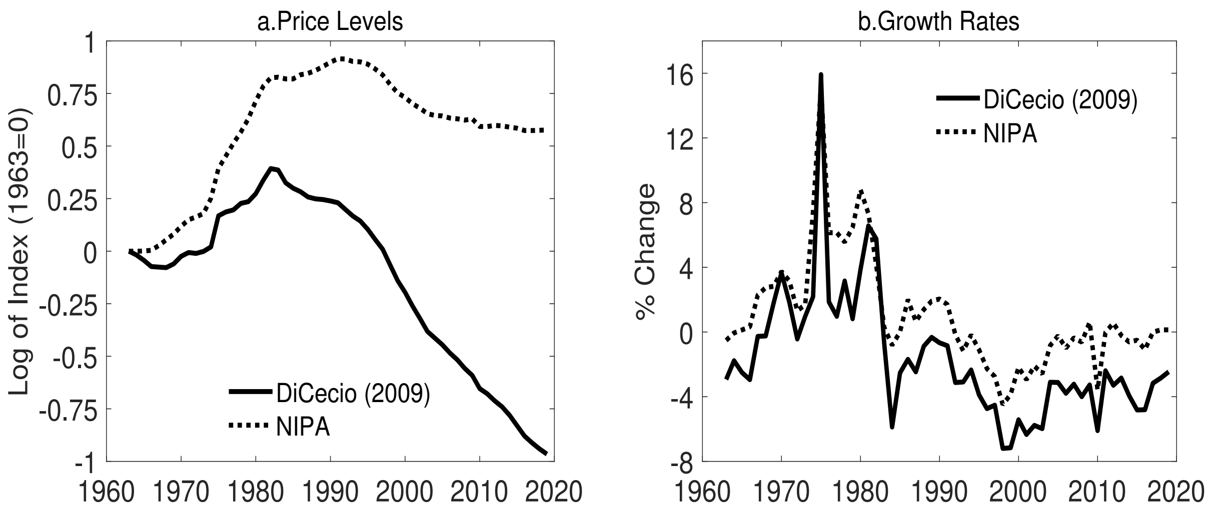
$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left( \frac{1}{T} \sum_{t=1}^{T} g\left(Y_t, \theta\right) \right)' \hat{W} \left( \frac{1}{T} \sum_{t=1}^{T} g\left(Y_t, \theta\right) \right), \tag{2.23}$$

where $\hat{W}$ is the weighting matrix computed based on the available data set.

In the estimation, we use STATA's default two-state estimator with modified Newton–Raphson algorithm. We also use "unadjusted" weight matrix that assumes the moment equations are independent and identically distributed, and that errors are homoskedastic. As instruments, we use current and lagged stocks of equipment and structures, lagged relative price of equipment capital, a time trend, and the lagged value of the U.S. Commerce Department's composite index of business cycle indicators without a constant.

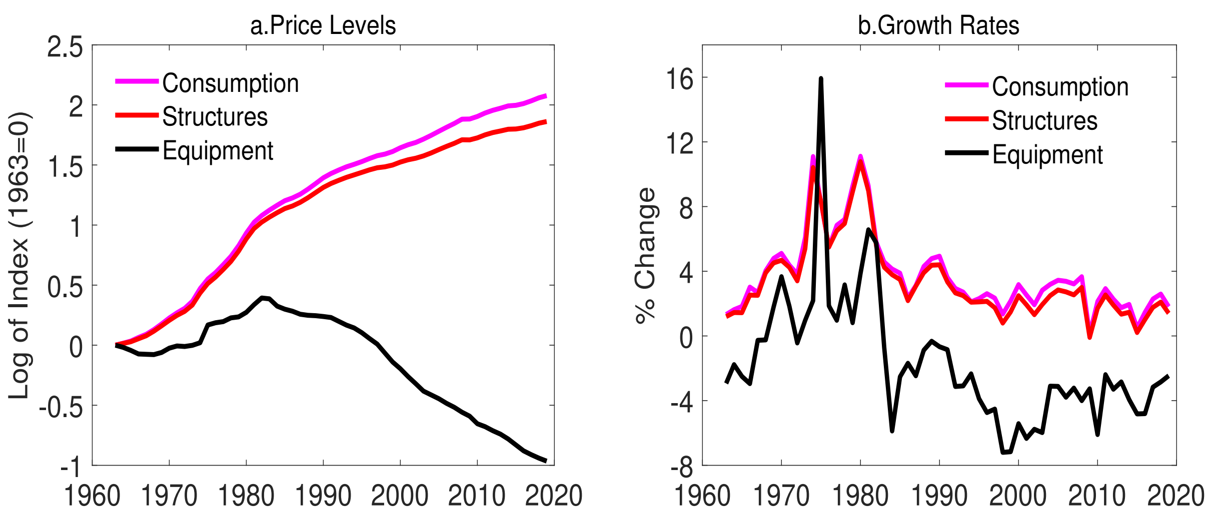
## 2.9 Appendix: Additional figures and tables

Figure 2.11: Price of Equipment Capital



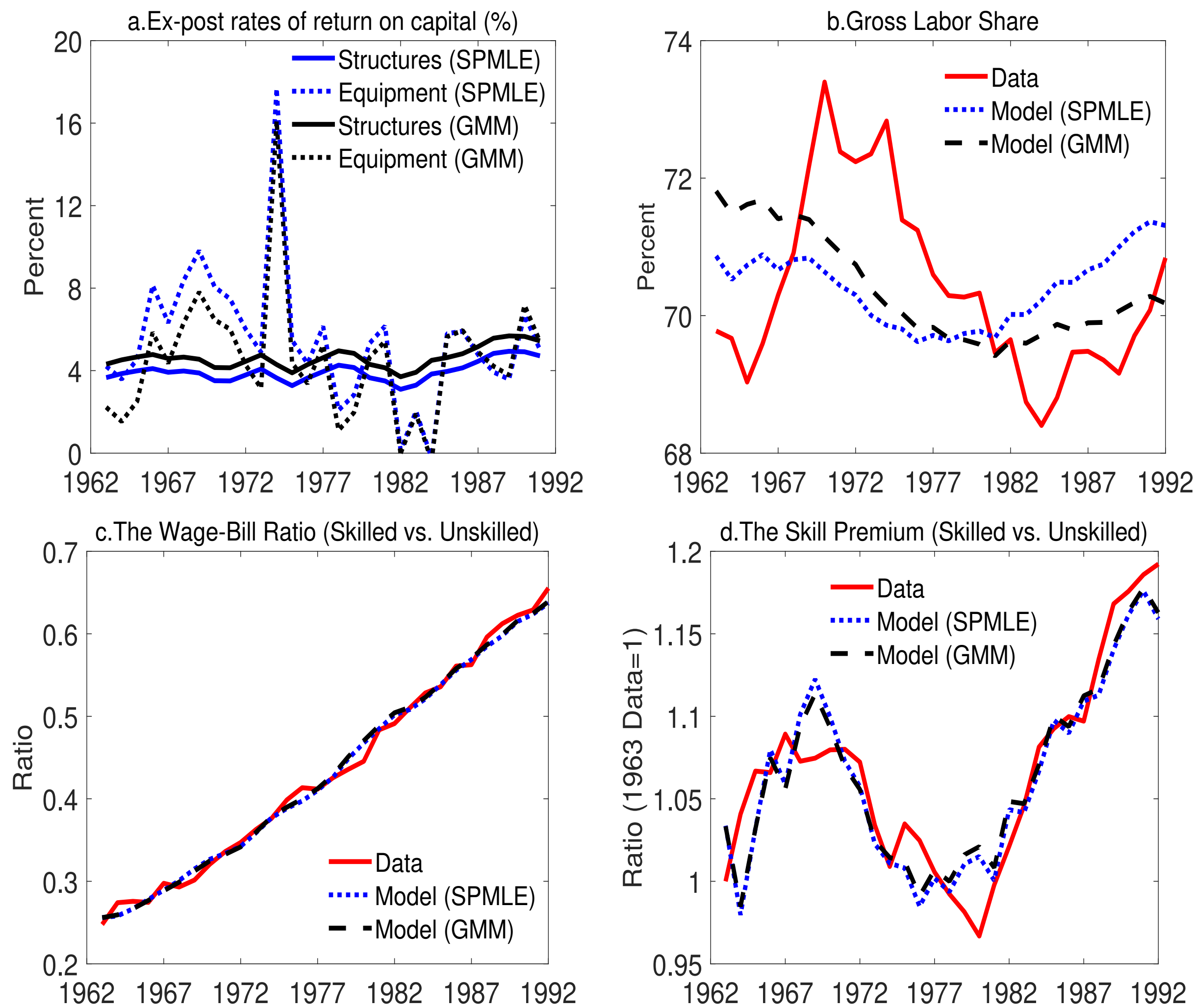

Source: Authors' calculations from the BEA's NIPA tables and from the investment prices series of DiCecio (2009).

Figure 2.12: Prices of Consumption, Structures, and Equipment Capital




Source: Authors' calculations from the BEA's NIPA tables and from the investment prices series of DiCecio (2009).

Figure 2.13: Replication with Original KORV Data




Notes: These charts are produced using the observable factor inputs and the parameters estimated employing the original KORV data, which covers the period between 1963 and 1992. While panel a runs through 1991, the rest of the panels plot the data and the model fit for the entire 1963–1992 period.

Table 2.7: Comparison of Parameter Estimates for the 1963-1992 Period for Alternative Labor Shares

| **Methodology** | **Labor Share** | $\sigma$ | $\rho$ | $\alpha$ | $\eta_\omega$ |
|---|---|---|---|---|---|
| I. SPMLE | KORV Gross | 0.438 | $-0.520$ | 0.105 | 0.083 |
| | | (0.020) | (0.043) | (0.002) | (0.007) |
| II. GMM | | 0.467 | $-0.478$ | 0.106 | – |
| | | (0.018) | (0.035) | (0.002) | – |
| III. SPMLE | KORV Net | 0.412 | $-0.606$ | 0.098 | 0.111 |
| | | (0.024) | (0.048) | (0.002) | (0.015) |
| IV. GMM | | 0.428 | $-0.592$ | 0.098 | – |
| | | (0.022) | (0.041) | (0.002) | – |
| V. SPMLE | NFBS Gross | 0.380 | $-0.494$ | 0.155 | 0.088 |
| | | (0.018) | (0.040) | (0.002) | (0.012) |
| VI. GMM | | 0.390 | $-0.484$ | 0.156 | – |
| | | (0.016) | (0.027) | (0.002) | – |
| VII. SPMLE | NFBS Net | 0.346 | $-0.565$ | 0.158 | 0.111 |
| | | (0.021) | (0.042) | (0.002) | (0.019) |
| VIII. GMM | | 0.343 | $-0.582$ | 0.158 | – |
| | | (0.020) | (0.033) | (0.002) | – |

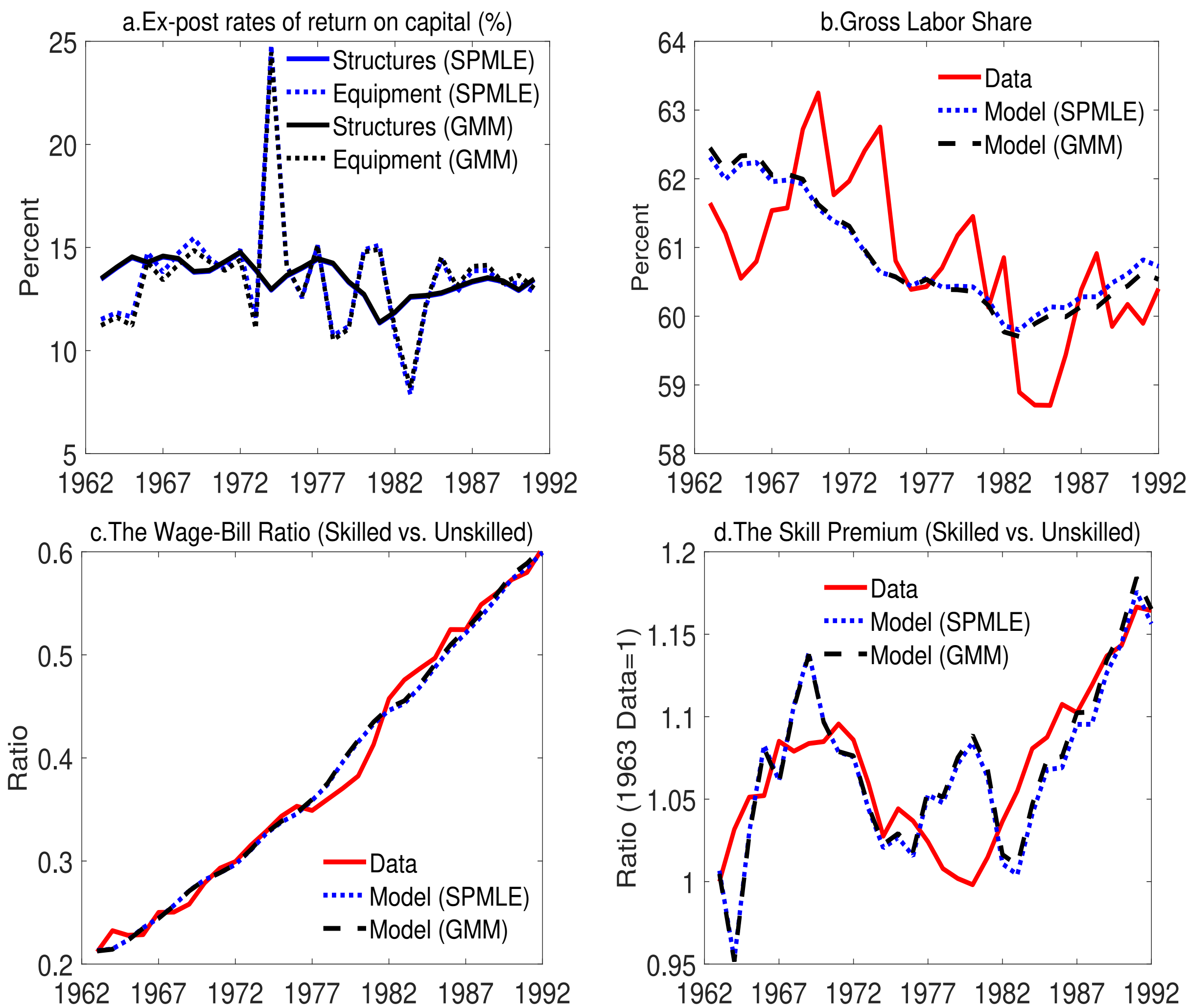
Notes: The values in parentheses are standard errors.

Table 2.8: Comparison of Parameter Estimates for the 1963-2019 Period for Alternative Labor Shares

| **Methodology** | **Labor Share** | $\sigma$ | $\rho$ | $\alpha$ | $\eta_\omega$ |
|---|---|---|---|---|---|
| I. SPMLE | KORV Gross | 0.431 | $-0.309$ | 0.109 | 0.085 |
| | | (0.013) | (0.026) | (0.002) | (0.005) |
| II. GMM | | 0.461 | $-0.298$ | 0.112 | – |
| | | (0.007) | (0.013) | (0.002) | – |
| III. SPMLE | KORV Net | 0.422 | $-0.381$ | 0.097 | 0.090 |
| | | (0.016) | (0.032) | (0.002) | (0.006) |
| IV. GMM | | 0.460 | $-0.339$ | 0.098 | – |
| | | (0.008) | (0.014) | (0.002) | – |
| V. SPMLE | NFBS Gross | 0.381 | $-0.325$ | 0.156 | 0.086 |
| | | (0.008) | (0.019) | (0.002) | (0.005) |
| VI. GMM | | 0.395 | $-0.296$ | 0.158 | – |
| | | (0.006) | (0.011) | (0.002) | – |
| VII. SPMLE | NFBS Net | 0.356 | $-0.386$ | 0.153 | 0.123 |
| | | (0.010) | (0.020) | (0.002) | (0.007) |
| VIII. GMM | | 0.376 | $-0.335$ | 0.153 | – |
| | | (0.006) | (0.012) | (0.002) | – |

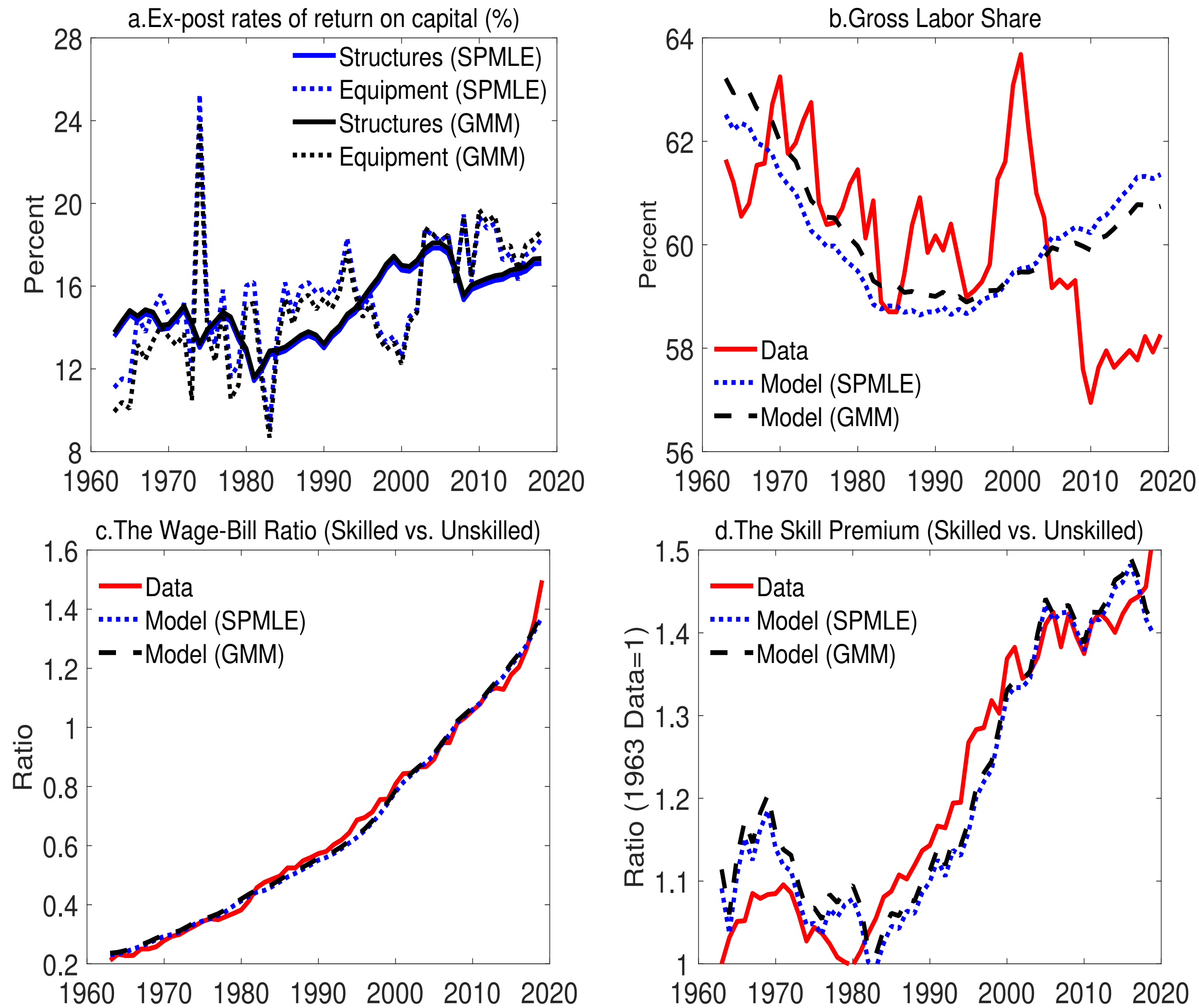
Notes: The values in parentheses are standard errors.

Figure 2.14: Model Fit for the 1963-1992 Period (with Nonfarm Business Sector Gross Labor Share)



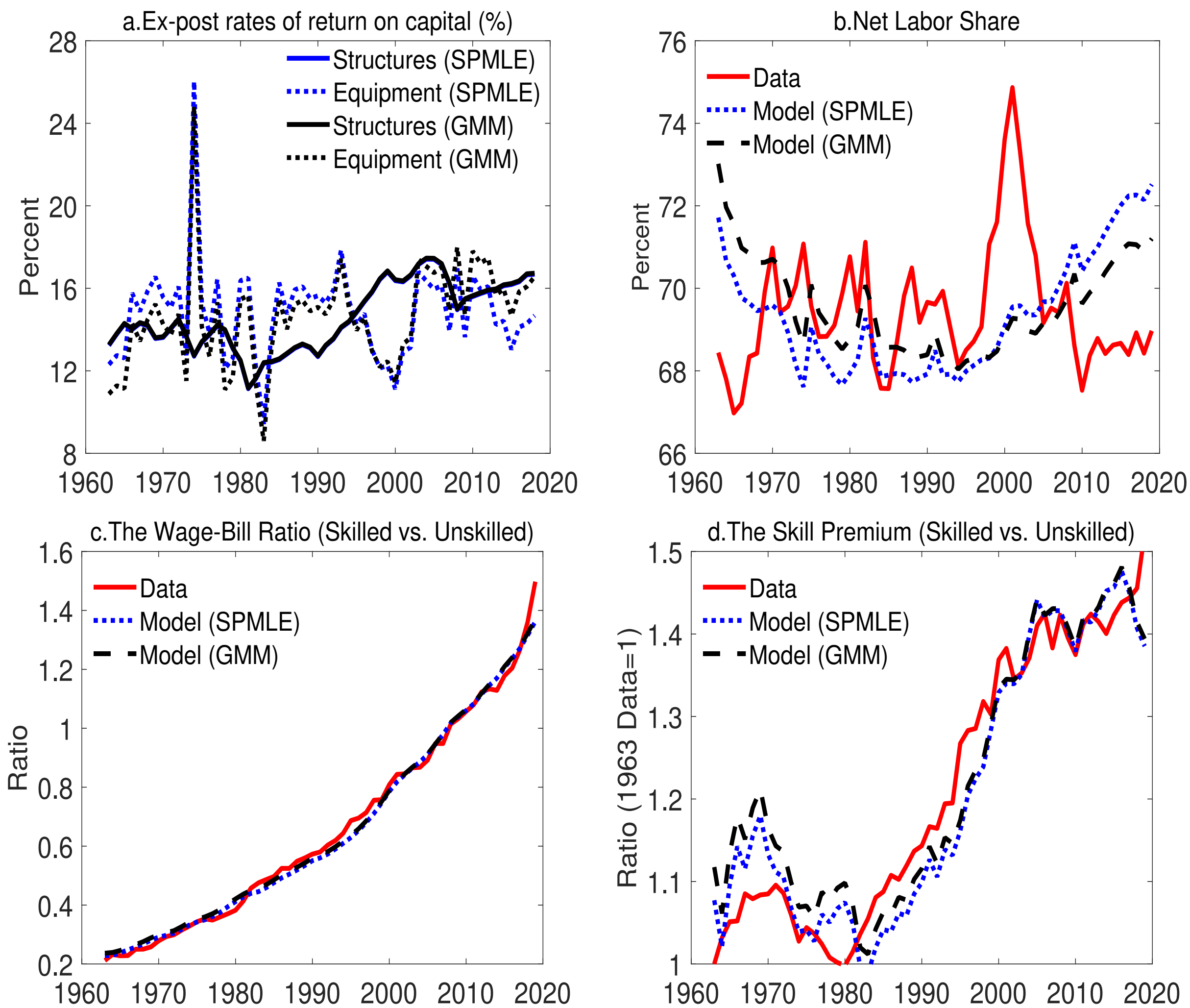

Notes: These charts are produced using the observable factor inputs and the parameters estimated employing data for the 1963–1992 period. Nonfarm business sector gross labor share is used in estimation. While panel a runs through 1991, the rest of the panels plot the data and the model fit for the entire 1963–1992 period.

Figure 2.15: Model Fit for the 1963-2019 Period (with Nonfarm Business Sector Gross Labor Share)




Notes: These charts are produced using the observed factor inputs and the parameters estimated employing data for the 1963–2019 period. Nonfarm business sector gross labor is used. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963–2019 period.

Figure 2.16: Model Fit for the 1963-2019 Period (with Nonfarm Business Sector Net Labor Share)



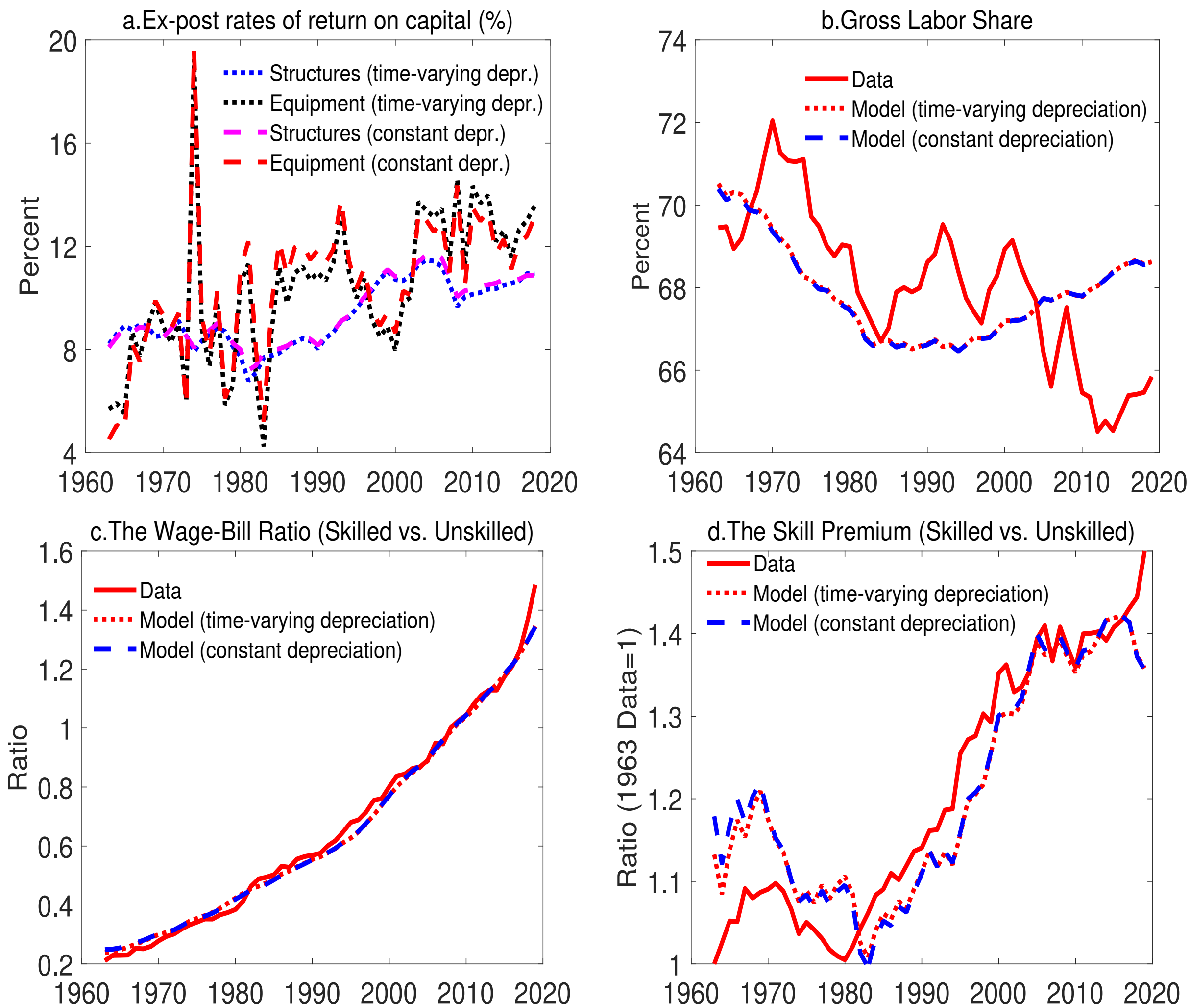

Notes: These charts are produced using the observable factor inputs and the parameters estimated employing data for the 1963–2019 period. Nonfarm business sector labor share net of depreciation is used in estimation. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963–2019 period.

Table 2.9: Comparison of Parameter Estimates with Constant Depreciation Rates (with KORV's Definition of Gross Labor Share)

| | **1963-1992** | | **1963-2019** | |
|---|---|---|---|---|
| | **Time-varying** | **Constant** | **Time-varying** | **Constant** |
| $\sigma$ | 0.438 | 0.435 | 0.431 | 0.418 |
| | (0.020) | (0.017) | (0.013) | (0.011) |
| $\rho$ | $-0.520$ | $-0.534$ | $-0.309$ | $-0.293$ |
| | (0.043) | (0.048) | (0.026) | (0.024) |
| $\alpha$ | 0.105 | 0.106 | 0.109 | 0.109 |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| $\eta_\omega$ | 0.083 | 0.082 | 0.085 | 0.090 |
| | (0.007) | (0.026) | (0.005) | (0.005) |

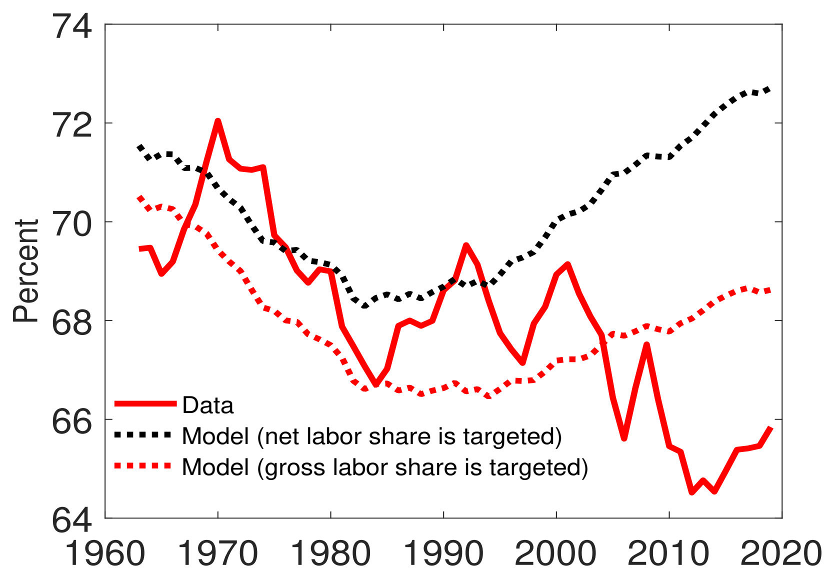
Notes: The values in parentheses are standard errors.

Figure 2.17: Comparison of Model Fit with Time-varying and Constant Depreciation Rates (with KORV's Definition of Gross Labor Share)



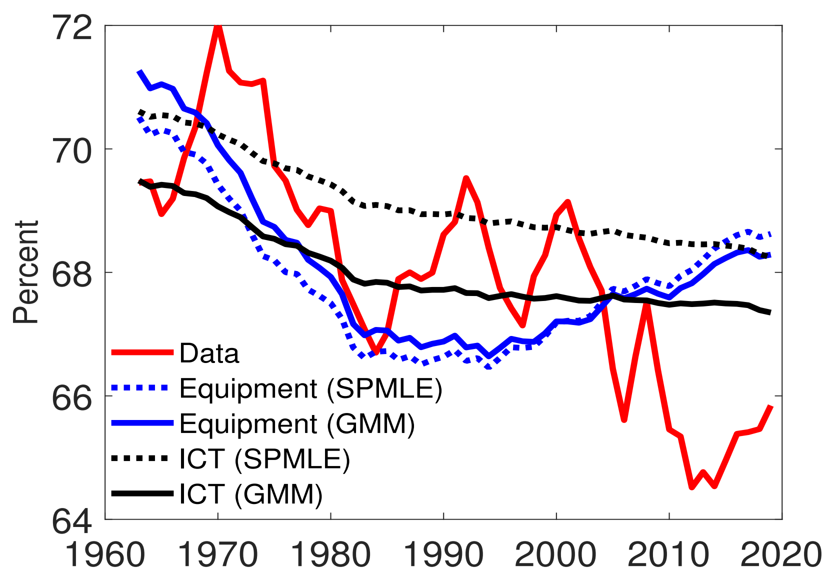

Notes: These charts are produced using the observable factor inputs and the parameters estimated with the SPMLE methodology employing the original KORV data, which covers the period between 1963 and 2019. While panel a runs through 2018, the rest of the panels plot the data and the model fit for the entire 1963–2019 period.

Figure 2.18: Model Fit for KORV's Gross Labor Share (Using Either Net or Gross Labor Shares as Targets)




Notes: This chart is produced using the observable factor inputs and the parameters estimated employing data for the 1963–2019 period. The black dotted line is implied the gross labor share when KORV's definition of the net labor share is used in estimation. The red dotted line is the gross labor share when KORV's definition of gross labor share is used in estimation.

Figure 2.19: Model Fit for KORV's Gross Labor Share (Using Either Equipment or ICT Capital as Complementary Capital)




Notes: This chart is produced using the observable factor inputs and the parameters estimated employing data for the 1963–2019 period. Blue lines are obtained when equipment capital is used as complementary capital, while black lines are obtained when ICT capital is used as complementary capital.

# Chapter 3

# Private Information, Adverse Selection and Small Business Financing

## 3.1 Introduction

It has been widely acknowledged that small businesses in general are under-served, especially in terms of finance. According to Robb and Robinson (2012), among all the debt and equity funding sources, banks are the most common providers of external credit for small firms, especially the new, innovative startups, and therefore a well-developed bank-lending market is important for allocating credit and boosting economic growth. Ever since Akerlof (1976), studies have shown that bank lending can be inefficient and cause "adverse selection" problems due to asymmetric information. However, banks are not doing nothing with it. In fact, to alleviate this problem, banks design differentiated lending contracts and rely heavily on collateral to search for and screen qualified borrowers, but this effort has largely been neglected in the literature. This paper studies the efficiency of bank's collateral-based screening practice in a competitive search framework with adverse selection.

To do this, I develop a competitive search model of bank-borrower relationship, where banks post loan contracts and borrowers direct their search to the contract that they like the most. Borrowers have projects that deliver the same expected profits but differ in terms of the probability of success, which is private information to themselves. Due to the search frictions, the borrowers potentially face a trade-off between interest rate and the probability of getting matched.

Since banks cannot distinguish borrowers with different levels of risk, they use the terms of loan contracts for screening. I show that information frictions and search frictions may lead to a lending strategy that hurts the borrowers with less risky investment projects[1], but the adverse selection can be different from the classical Akerlof (1976) paper. It depends on the enforceability of banks upon project failure. Enforeability is important because banks recover the principal and interest in a loan, and thus they care more about borrowers' likelihood of repaying debts rather than their profitability upon success.

In a credit market where banks can hardly enforce the repayments, banks tend to post too few contracts designed to attract the safer borrowers, with the purpose to deter risky borrowers from imitating the safe types. Since high repayment requirement is not an effective threat in this case, market tightness becomes the screening device. Nevertheless, when banks have higher enforcement, which means debt obligations are hard to escape, interest rate becomes a better screening device. To stand out from the risky ones, borrowers with safe projects accept the contract in which they are over-charged and over-offered the loans. I show that this can be Pareto improved by reducing safe types' interest rates and increasing risky types' chances of being funded, as long as the share of risky borrowers in the society is not very high.

In the past decades, the rapid development of digital technologies and the rise of online fintech lenders transforms lender's enforcement when granting loans. Bank and fintech lending fundamentally differ in their enforcement technologies. With their data-driven services, the fintech lenders can obtain more information on the borrower's future cash flows, and seize a fraction of it. Therefore, the new technologies increases the enforeability of lenders and thus affects credit allocations and market efficiency. This is consistent with the fact that compared to traditional banks, fintech lenders typically charge

[1] Adverse selection is more likely to take place when loan default is not costly, the less risky type are not "safe" enough or the average social productivity is quite low.

higher interest rates but are more likely to grant loans to small businesses.

The efficiency analysis shows that although there are bank screening in the economy, distortions may still occur under private information[2] as the safer types always experience a utility loss in order to stand out from their risky fellows. The negative externality here is rooted in three types of frictions. First, when the legal and institutional quality of bank-lending market is not good enough, the risky borrowers have strong incentives to imitate the safe ones. To avoid this, the safe types accept contract terms that make them less likely to get credits compared to the first-best case. Second, banks are subject to asymmetric information and thus have to resort to suboptimal allocations as screening devices. Even when banks have a high level of repayment enforceability, their lack of information can lead to distortions like a "rat race" in which the safe borrowers overwork. Third, there are searching costs in the bank-borrower matching process, creating misallocations on the extensive margin.

Credit misallocation in this paper is caused by asymmetric information, a big issue in bank lending, and even more so with the recent trend of consolidation in the US banking system, which gives rise to more large banks and fewer small banks. Usually, small banks are better at collecting soft information and more flexible in granting small business loans, while large banks, as Berger and Black (2011) point out, tend to lend to relatively large, information-transparent borrowers using "hard" information. In this way, large banks are more likely to facilitate a high-volume, economies-of-scale business model and perhaps screen for more established firms. In general, banks have limited knowledge of the riskness and profitability of their small business borrowers, which provides empirical motivations for this paper.

This paper also discusses the impact of the rise of equity-financing institutions such as venture capital and private equity. In the past decades, VC has emerged as another major source of financing for young, in-mature firms, greatly reduces firms reliance on collateral. This increases the funding probability of borrowers with little collateral but higher

[2] Even if there are fintech lenders, borrower type is still private information before lending.

probability of success, but market inefficiency still prevails in the form of overcharging as long as there is information and search frictions.

## 3.2 Literature Review

This paper contributes to the literature on small business lending by examining how adverse selection and search frictions affect credit allocation. As far as I know, this is the first paper to model bank-borrower relationship allowing for adverse selection among ex ante heterogeneous projects in a search environment. Although the search tools have been extensively used in labor, housing and asset markets, it is rarely used in credit markets, especially the bank-lending relationship. A lot of studies focus on how the venture capital markets influence the financing of startups in a search setting (Inderst and Müller (2004), Michelacci and Suarez (2004), Hellmann and Thiele (2015), Jovanovic and Szentes (2013), Winegar (2018)). My paper, on the other hand, studies debt financing which has a very different payoff structure from theirs: debt providers care more about risks while the capital suppliers value high returns and have much greater tolerance for failure. Consequently, the misallocation in my paper features over-exploiting the safe types while theirs emphasize the insufficient financing for risky innovative startups.

Another key source of friction is adverse selection caused by asymmetric information, which has been a classical topic in credit markets. Stiglitz and Weiss (1981) show that an increase in interest rate needs not necessarily increase profits, because the risky firms are willing to borrow at high interest rates as they perceive their probability of repaying the loan to be low. Adverse selection refers to the phenomenon that less risky firms drop out of the market. Their paper relies on an important assumption that the riskier firms have higher expected profit, and thus are more willing to borrow when the interest rates are high. This is similar to one of the cases in my paper where banks are very tolerant for non-performing loans, which means firms do not get punished for failing to pay back the loans.

Not surprisingly, the distortion generated under this situation is similar to that in Stiglitz and Weiss (1981). However, my paper also considers the case where banks have high level of enforceability on repayment and the risky firms cannot escape from their debt obligations even in bad states. Distortions in this scenario, however, features overworking of the safe types. This is similar in spirit to the rat race problem in Akerlof (1976), where the good workers are less sick of working long hours and hence willing to be over-employed. Likewise, in this paper, borrowers with safe projects have a higher probability to succeed and thus are more tolerant of high repayment requirements, which explains why they are offered more contracts and higher credit lines in this type of credit market.

My paper studies adverse selection in the bank-lending markets by embedding it into a search environment. Two pioneer studies have combined adverse selection and search frictions, the screening-type game proposed by Guerrieri et al. (2010), and the signaling-type game in Delacroix and Shi (2013). My paper follows the baseline model in Guerrieri et al. (2010) and sheds light on the role of financial intermediaries in nurturing borrowerial firms. Their paper proposes a general framework of a screening-type game in a competitive search environment, where uninformed principals post terms of trade, and agents who have private information on their types choose where to direct their search, and the two parties match bilaterally. Under three key assumptions, there always exists a separating equilibrium where each type applies to a different contract and it is not generally efficient.

There is a large literature that embeds search into financial markets, most of which are on equity financing(Hellmann (2002); Sørensen (2007); Lerner et al. (2011); Tian and Wang (2011); Manso (2011)). Nanda and Rhodes-Kropf (2013) argue that investors respond to financing risk by modifying their focus to financing less innovative firms. I find similar results in debt markets and contribute to this line of research by emphasizing how adverse selection and search frictions lead to the phenomenon. A closely related paper is Winegar (2018), which also highlight these two types of frictions, but he adopts a signaling-type game where the informed borrowers post contracts, and the paper focuses exclusively on the venture capital markets which is greatly impacted by the supply of capital. My paper, on the other hand, considers bank financing.

My paper also contributes by broadening the applications of search models, which have been shown a powerful tool in labor markets and OTC asset markets. However, it has not been widely used in studying the relationships between lenders and firm borrowers. Two recent studies apply the search models to investigate how the bank lending relationships account for the sluggish credit recovery from the crisis. Boualam (2015) argues that two types of frictions make the bank-firm relationship hard to recover once broken, the search frictions including the bank sunk cost at origination and the costly and time-consuming relationship building, and the frictions in long-term contract including limited enforceability and institutional environment of credit markets. Payne (2018) also studies the disruption of bank business credit during crisis in a directed search model, but introduces heterogeneous investment projects and heterogeneous shocks to banks and tries to match four stylized facts in US credit markets. My study complements this strand of literature by introducing asymmetric information between lenders and borrowers, arguing that adverse selection in addition to traditional search frictions also lead to misallocation, and my model is able to discuss whether the distortion happens on the intensive or extensive margin.

The paper is organized as follows. In Section 3.3, I will lay out the model and define equilibrium. Next, in Section 3.4, I will discuss under which circumstances will credit misallocation arise, and I will also solve the model under asymmetric information and make a comparison between the model equilibrium and the first-best contract terms. In Section 3.5, I show that efficiency can be improved for a specific type of distortion. I discuss the potential effect of the rise of venture capital in Section 3.6 Finally, conclusions will be drawn in Section 3.7.

## 3.3 Model

### 3.3.1 Setup

Consider an economy with two types of agents: banks and borrowers. The risk-neutral banks offer loans to borrowers who are heterogeneous in their probability of success. The borrowers should be risk-averse, but for simplicity, I assume they are risk-neutral and therefore maximize their profits, but this can be extended to include risk-aversion. In this bank-borrower relationship, whether a firm will be productive or not is initially unknown by both the bank and the borrower, but the probability of success is private information of the latter. Some of them are more likely than others to succeed in investing in a project, and the borrowers themselves know their likelihood of success while banks do not.borrower

Although bank loans usually run long-term, I keep things simple here and model the contracting problem in a static setting with only one period, but the essential features would carry over to an intertemporal model. At the beginning, borrowers seek an external finance $I$ for their investment, and they are required by the bank to repay $R$. There are two types of borrowers in the economy: with probability $p_i$, type-$i$ invest successfully and generate a revenue of $z_i I$ at the end of the period ($z_i > 1$); with probability $1 - p_i$, the project fails and generates 0, but the borrower is still subject to loan repayment $R$, although she may not be able to return all of it. Denote $\xi$ as the fraction the borrower repay upon project failure. $\xi$ reflects banks' tolerance of non-performing loans, or their ability to enforce loan repayments. A small $\xi$ indicates that loan defaulting is not costly, so this parameter describes the quality of the legal and institutional environment of the credit system. Assume $p_1 < p_2$, so type-1 borrower is more risky and less likely to succeed. I also assume $p_1 z_1 = p_2 z_2$, although the key results in this paper hold under looser parameter restrictions. I set it this way because it simplifies the calculation a lot, and at the mean time is not restrictive as it allows the highly risky firms to earn extremely high revenues but still satisfies the condition.borrower

The revenues above are derived without liquidating any assets, but the firm can also obtain additional cash by liquidating assets at the expense of deteriorating the future value. Denote the future value of the firm's assets by $A(l)$, where $l \in [0, L]$ is the additional cash generated by liquidating assets. $A'(l) < 0$ because liquidating hurts the firm value. Following Berlin and Mester (1999), by liquidation I mean any activity that reduces the value of future production activities to generate revenue today, and I assume that liquidating assets to produce revenue reduces the total value of the firm, ex post, that is

$$\frac{d(l + A(l))}{dl} = 1 + A'(l) < 0 \tag{3.1}$$

I also assume that the cost function of liquidation is convex in the amount of cash generated today: $A''(l) < 0$. This assumption is based on the micro foundation that one additional dollar a firm liquidates increasingly adds to the likelihood of the firm getting into trouble. Any function that satisfies the conditions above gives rise to similar key results, but in order to derive clear analytical solutions for first-best contract, I set $A(l) = V - l - l^2$, where $V$ is the firm's present value without liquidation. Since we do not care about absolute value of the firm's payoff, I normalize $V$ to be 0. Given the contract term and the economic states, borrowers choose $l$ to maximize profits. borrower

There is a set of ex ante homogeneous banks, each of whom may or may not participate in the market. If a bank enters, it posts a contract $y$ at the cost of $c$. Upon bilateral matching, the uninformed bank makes a decision on the loan size $I$ and the repayment $R$. Therefore, the contract that banks post is $y = \{I, R\}$. The utility of a type-$i$ borrower who applies to contract $y$ and is matched is

$$u_i(I, R) = \max_{l_u, l_d} p_i \left[z_i I - R + l_u + A\left(l_u\right)\right] + (1 - p_i)\left[-\xi R + l_d + A\left(l_d\right)\right] \tag{3.2}$$

where $z_i > 1$ is the payoff of a successful project with unit investment, and hence $p_i z_i$ is the expected value of a project. I assume $p_i z_i \geq 1$ otherwise the expected profits are negative and borrowers would never choose to start a business.borrower

Take the first order condition of $u_i$ with respect to $l_u$ and $l_d$. Since $1 + A'(l) < 0$, we want $l_u$ and $l_d$ to be as small as possible, which means they should just offset the arrears if profits are negative, and 0 if profits are positive. Given the contract $y = \{I, R\}$, an borrower earns profit $z_i I - R$ if succeeds, and $-\xi R$ if fails. I assume $z_i$ is large enough to guarantee $z_i I > R$, so there is no need to liquidate once the project is successful, thus $l_u = 0$. Also, $l_d \geq \xi R$, which boils down to $l_d = \xi R$. Based on the analysis, we can rewrite the firm's payoff (profit) function as

$$\begin{aligned} u_i(I, R) &= p_i \left[ z_i I - R \right] + (1 - p_i) A \left( \xi R \right) \\ &= p_i z_i I - \left\{ \left[ p_i + (1 - p_i)\xi \right] R + (1 - p_i)\xi^2 R^2 \right\} \\ &\equiv p_i z_i I - \phi_i(R) \end{aligned} \tag{3.3}$$

It is clear that $\phi_i(R) = \left[ p_i + (1 - p_i)\xi \right] R + (1 - p_i)\xi^2 R^2$ is a convex function of $R$.borrower

The payoff of a bank posting $y$ and is matched with a type-$i$ firm is

$$v_i(I, R) = \left[ p_i + (1 - p_i)\xi \right] R - I \equiv f_i(R) - I \tag{3.4}$$

As mentioned, the borrower only returns $\xi$ share of the required repayment to the bank and defaults for the rest of their debt obligation. In the Appendix, I verify that the model satisfies the three key assumptions in Guerrieri et al. (2010): monotonicity, local non-satiation and sorting. The last assumption, which is a modified version of the single crossing property, is particularly important to derive the separating equilibrium. borrower

The matching process is characterized by the banks posting contracts, and the borrowers directing their search to any one they prefer. Matching is bilateral, so at most one borrower ever contacts a bank. Let $\theta(y)$ denote the market tightness, or bank-borrower ratio in the submarket with contract $y$, and let $s_i(y)$ denote the share of agents applying to $y$ that are type $i$. These two variables are determined endogenously in equilibrium. Assume the matching function is $\mu(\theta) = \theta^\alpha$, where $\alpha \in (0,1)$. As can be seen, the matching function for borrower is concave. Thus, a bank offering $y$ matches with a firm with probability $\eta(\theta) = \mu(\theta)/\theta = \theta^{\alpha-1}$, Given this, the expected payoff of a bank who posts contract $y$ is:

$$\frac{\mu(\theta(y))}{\theta(y)} \sum_i s_i(y) v_i(y) - c \tag{3.5}$$

and the expected utility of a type-$i$ borrower who applies to contract $y$ is

$$\mu(\theta(y)) u_i(y) \tag{3.6}$$

### 3.3.2 Competitive Equilibrium

A competitive search equilibrium is defined as a vector $\overline{U} = \{\overline{U}_i,\text{i} \in \{1,2\}\} \in \mathbb{R}^2_+$, a measure $\lambda$ on $Y$ with support $Y^P$, a function $\theta : Y \mapsto [0,\infty)$, and a function $S : Y \mapsto \Delta^2$ that satisfy the following conditions:

(i) Banks' *Profit Maximization* and *Free Entry*: For any $y \in Y$,

$$\eta(\theta(y)) \sum_i s_i(y) v_i(y) \leq c \tag{3.7}$$

with equality if $y \in Y^P$.

(ii) borrower's *Optimal Search*: Let $\overline{U}_i = \max \left\{0, \max_{y' \in Y^P} \mu\left(\theta\left(y'\right)\right) u_i\left(y'\right)\right\}$, and $\overline{U}_i = 0$ if $Y^P = \emptyset$. Then for any $y \in Y$ and $i \in \{1,2\}$,

$$\overline{U}_i \geq \mu(\theta(y)) u_i(y) \tag{3.8}$$

with equality if $\theta(y) < \infty$ and $s_i(y) > 0$.

(iii) *Market Clearing*: Let $\pi_i$ be the fraction of borrowers who are of type $i$. Then

$$\int_{Y^P} \frac{s_i(y)}{\theta(y)} d\lambda(\{y\}) \leq \pi_i \quad \text{for } i \in \{1,2\} \tag{3.9}$$

## 3.4 Economic Analysis

### 3.4.1 First-best Contract

Under asymmetric information, for each type $i = \{1, 2\}$, we derive equilibrium allocations by solving the following maximization problem (P-$i$):

$$\begin{aligned}\overline{U}_i = \max_{\theta\in[0,\infty), y\in Y} & \mu(\theta)\{p_i z_i I - \phi_i(R)\} \\ \text{s.t. } & \mu(\theta)(f_i(R) - I) \geq \theta c \\ & \mu(\theta)\{p_j z_j I - \phi_j(R)\} \leq \overline{U}_j \quad \text{for all } j < i\end{aligned} \tag{3.10}$$

Before characterizing equilibrium, I will describe the efficient allocations for both types under full information. For type $i$, this is given by the solution to problem (P-$i$) but eliminating the second constraint. The results are summarized in the following Proposition.borrower

Proposition 1: *The first-best allocations for type-i borrowers under full information are given by*

$$\begin{aligned}R_i^* &= \frac{(p_i z_i - 1)\left[p_i + (1-p_i)\xi\right]}{2\xi^2(1-p_i)}. \\ \theta_i^* &= \left[\frac{\alpha}{p_i z_i c}\frac{(p_i z_i - 1)^2\left[p_i + (1-p_i)\xi\right]^2}{4\xi^2(1-p_i)}\right]^{\frac{1}{1-\alpha}}. \\ I_i^* &= \frac{(p_i z_i - 1)\left[p_i + (1-p_i)\xi\right]^2}{2\xi^2(1-p_i)} - \frac{\alpha}{p_i z_i}\frac{(p_i z_i - 1)^2\left[p_i + (1-p_i)\xi\right]^2}{4\xi^2(1-p_i)}.\end{aligned} \tag{3.11}$$

Clearly, $R_1^* < R_2^*$, $\theta_1^* < \theta_2^*$ and $I_1^* < I_2^*$. So, type-2 borrowers have a higher probability of getting loans, and upon being financed, they borrow more money and also make more payments. But if we calculate and compare the gross interest rates, we find $R_2^*/I_2^* < R_1^*/I_1^*$. To sum up, under symmetric information, type-2 enjoys more favorable loan terms than type-1 thanks to their lower riskiness.

With asymmetric information, the problem for type-1 is still (P-1), and consequently the equilibrium loan size, repayment and market tightness are the same as first-best allocations for type-1 borrowers: $I_1 = I_1^*$, $R_1 = R_1^*$, and $\theta_1 = \theta_1^*$. $\overline{U}_1 = \theta_1^{*\alpha}[p_1 z_1 I_1^* - \phi_1(R_1^*)]$.

Next, consider (P-2) where the second constraint $\mu(\theta)\{p_1 z_1 I - \phi_1(R)\} \leq \overline{U}_1$ is additionally imposed to the full information problem. If this constraint is satisfied when plugging in the efficiency allocations given by (3.11),

$$\theta_2^{*\alpha}\{p_1 z_1 I_2^* - \phi_1(R_2^*)\} \leq \theta_1^{*\alpha}\{p_1 z_1 I_1^* - \phi_1(R_1^*)\} \tag{3.12}$$

it indicates type-1 borrowers prefer their own first-best contract terms ($I_1^*$,$R_1^*$) and trading probabilities ($\theta_1^*$) to those of type-2 borrowers, and thus there is no distortion and each type obtains their first-best allocations in equilibrium.borrower

In next section I will discuss under what set of parameters the inequality (3.12) is violated and distortions arise, and upon this how loan size, repayments and market tightness change relative to their first-best levels.

### 3.4.2 Equilibrium Contract under Asymmetric Information

Whether condition (3.12) holds depend on the key exogenous parameters $p_i$, $z_i$, $\xi$, $\alpha$, $c$. If the inequality is rarely violated, the lending market operates at social optimum most of the time, and there is no need to worry about credit misallocation generated by asymmetric information between lenders and borrowers. Thus, it is crucial to identify under what circumstances will adverse selection arise. However, the relationship between the left-hand side and right-hand side of (3.12) is not monotone in many parameters, and it is impossible to analytically derive a neat range where distortions take place. To deal with this, I fix some parameters and use the plots in Panel 1 of Figure 1-3 to identify the effects of other parameters.borrower

Of course we can investigate the effect of each parameter on distortions, but to save time and space, I fix $p_1 = 0.4$, $\alpha = 0.7$ and $c = 5$, and discuss how $\xi$, $p_2$ and $z_2$ (equivalently

Figure 3.1: Equilibrium vs First Best for Type-2. $\xi = 0.1$.

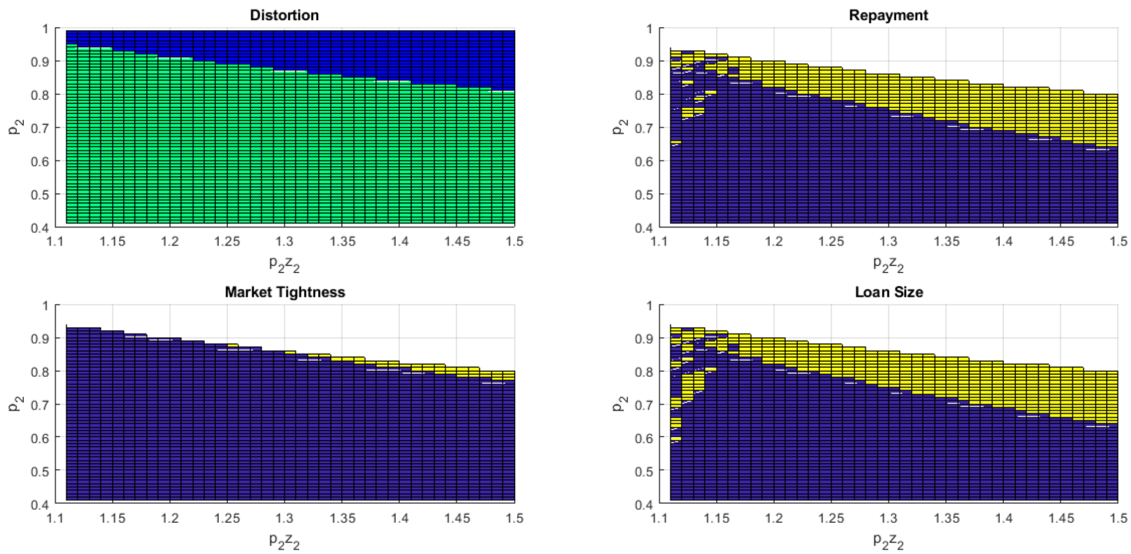



Figure 3.2: Equilibrium vs First Best for Type-2. $\xi = 0.5$.

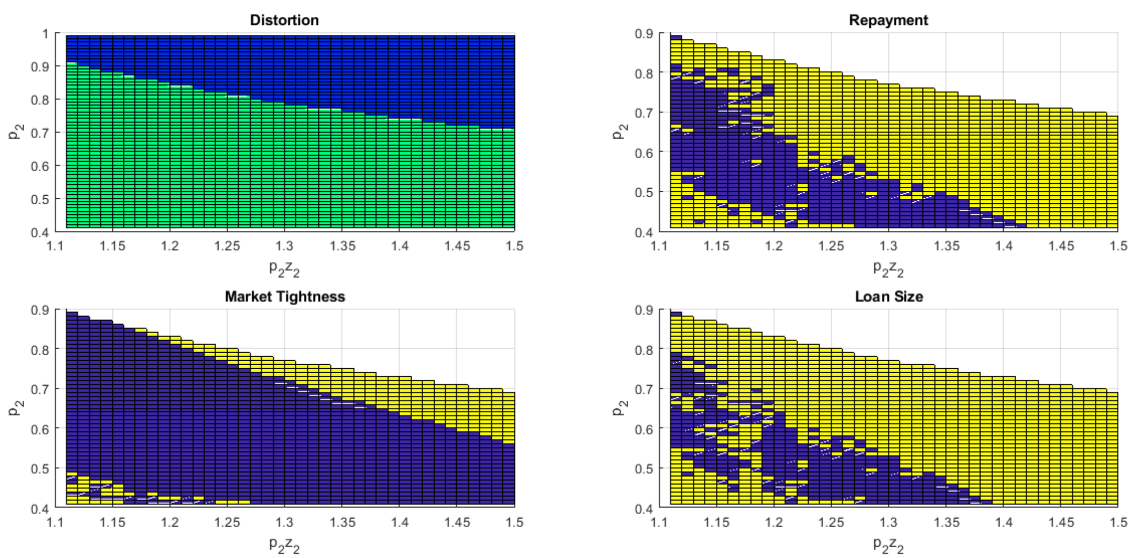
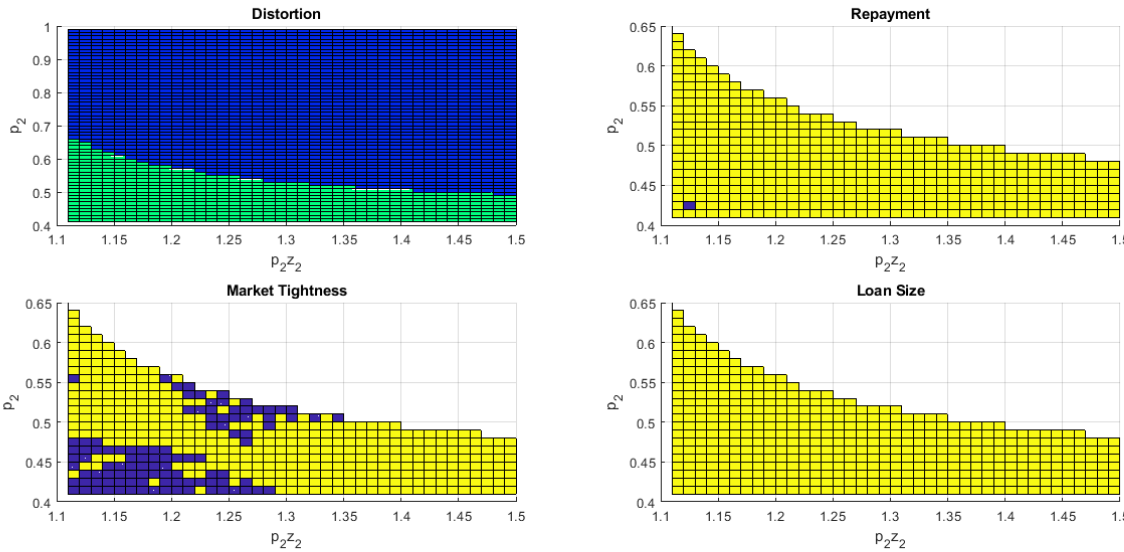


Figure 3.3: Equilibrium vs First Best for Type-2. $\xi = 0.9$.




$p_2z_2$) affect the results. We can study the effects $p_1$, $\alpha$ and $c$ using the same method. First, consider when $\xi$ is very small, say, $\xi = 0.1$, which means the borrowers only need to pay back a small proportion (10%) of their obligations in bad states. A small $\xi$ refers to the case where firms can default without severe punishment. The presence of distortions ((3.12) violated) is represented by the green area of Panel 1, while the absence is the blue area. borrower

For small values of $\xi$ (Figure 1), distortions take place for a wide range of parameter combinations. Given expected revenue ($p_2z_2$), the condition is violated when $p_2$ is not very close to 1. In other words, if type-2 is highly likely to succeed, their first-best contract is not attractive to type-1. Intuitively, the efficiency repayment $R_2^*$ for extremely safe type might be too high for type-1 to endure, so the latter have no incentives to mimic type-2 and would stick to their own first-best allocations. However, when type-2 is not very safe, type-1 prefer type-2's first-best contract to their own, because the benefits of large market tightness exceeds the cost of high repayment. To screen out type-1, the bank posts a contract for type-2 that makes them worse-off compared to the first-best level, but type-2 accepts this in order to stand out from type-1. As we increase the average productivity of the whole society,

$p_2z_2$ (also $p_1z_1$), distortions become less likely to happen. With a larger $p_2z_2$, the influence of $p_2$ is still the same as mentioned above although the range for type-2 to be "safe" enough to deter the imitation of type-1 is larger. This implies that a less productive economy is more likely to suffer from credit distortion. Panel 1 in Figure 2 and 3 depict the cases for $\xi = 0.5$ and $\xi = 0.9$. Still, the way that $p_2$ and $p_2z_2$ generate distortions are similar, but less so as $\xi$ increases, which means a banking system with higher enforceability of loans are less likely to observe credit misallocation.borrower

Next, I will show the relationship between type-2's equilibrium and first-best allocations. It turns out there are different types of distortions for different values of $\xi$. The results are shown in Panel 2-4 of Figure 1-3. A grid is purple-colored if the equilibrium allocation for type-2 ($R_2$, $\theta_2$ or $I_2$) is smaller than the first-best level ($R_2^*$, $\theta_2^*$ or $I_2^*$) at that specific combination of $p_2$ and $p_2z_2$, and is yellow if the opposite is true.borrower

In Figure 1 where $\xi = 0.1$, for most parameter combinations type-2 borrowers receive less funding on both the intensive margin ($I_2 < I_2^*$) and extensive margin ($\theta_2 < \theta_2^*$), and they also pay back less ($R_2 < R_2^*$) compared to first-best case[3]. As $p_2$ and $p_2z_2$ increase but are not high enough to avoid distortions, the equilibrium loan size $I_2$ and repayments $R_2$ exceed $I_2^*$ and $R_2^*$, while market tightness $\theta_2$ is almost always below $\theta_2^*$. This could be explained by the behavior of type-1. Compared to their own first-best contract, type-1 prefers the higher market tightness ($\theta_2^* > \theta_1^*$) of type-2's first-best contract but dislikes the higher repayment ($R_2^* > R_1^*$), which causes more utility loss for type-1 than type-2 due to riskiness. However, when $\xi$ is small, the burden of investment failure is minor, and type-2's contract becomes more appealing to type-1 because of the higher $\theta$. If banks post $y_2^* = (I_2^*, R_2^*)$, they would attract a lot of type-1, which is no good for banks because what they care about is the borrower's probability of success ($p_i$) rather than their profitability ($z_i$). In order to prevent type-1 from mimicking type-2, banks instead post a contract $y_2 = (I_2, R_2)$ that generates

[3] The equilibrium is solved using Matlab function "fmincon", but the algorithm and the shape of the function make it hard to converge under a few paramter settings. This explains why there are some "noisy" blocks where there are a few yellow inside the purple area or vice versa.

lower market tightness ($\theta_2 < \theta_2^*$). Although this is not optimal for type-2, they compromise so as to get distinguished from type-1. When $p_2$ and $z_2$ are relatively small, type-2 receive less credit on both the intensive and extensive margin. As these two parameters increase, type-2 are safer and more productive, and are therefore loaded with more loans and repayments than optimal, in addition to still being rationed on the extensive margin, for the sake of preventing type-1's imitation.borrower

Similar patterns hold for $\xi = 0.5$ and $\xi = 0.9$. As $\xi$ increases, the repayment requirement becomes more burdensome for type-1, so the least costly way to screen becomes increasing $R_2$ rather than decreasing $\theta_2$. As is shown in Figure 2, when $\xi = 0.5$, although there is still distortion that relies on a lower probability of getting credits ($\theta_2$), most distortions emphasizes a larger repayment ($R_2$). Panel 2 and 4 indicate that for most combinations of $p_2$ and $p_2 z_2$, $R_2 > R_2^*$ and $I_2 > I_2^*$, borrowers with safe projects are excessively financed due to the fact that banks want to avoid risky borrowers but do not have sufficient information on borrowers' types. When $\xi = 0.9$, under almost all parameter settings type-2 receive more funding on both the intensive and extensive margin. This is because given a high $\xi$, a higher repayment substantially harms the risky borrowers. As for type-2, they now are willing to undertake more repayments in order to be distinguished from type-1. And to compensate for the high payments, they obtain more credits on both the extensive and intensive margin (higher $\theta_2$ and $I_2$).borrower

To sum up, banks that are subject to asymmetric information screen their borrowers by designing different contracts. The contract terms that are used as screening devices, however, vary with banks' tolerance of bad loans. When banks tend to let go of the non-performing loans, it is better to screen via the market tightness. To distinguish themselves from type-1, type-2 borrowers are willing to accept a lower borrowing probability. As $p_2$ increases, the safe types are better able to endure high payments, and at the same time banks are more eager to get rid of the risky type, so the contract for type-2 features repayments higher than optimal while still keep market tightness below first-best.borrower

When $\xi$ is high, a higher repayment significantly drags down type-1's utility, and thus replaces market tightness to serve as the screening device, which captures the idea of rat race in Akerlof (1976). To distinguish themselves from other risky fellows, type-2 borrowers make do with the inordinate payments and produce excessively, as if in a rat race. Since type-2 are more likely to succeed and stand the high repayments than type-1, banks give them a lot of money and make them "overwork". This hurts the safe type a lot. If, nonetheless, the capital resources are allocated to the risky but innovative startups, the efficiency of the whole society will get increased, as will be shown immediately in next section. Competitive search is essential for the results.

## 3.5 Efficiency

In this section, I show that when $\xi$ is large - the rat rate case, efficiency can be enhanced by granting more credits to the risky types. Consider an allocation that treats the two types of borrowers identically. All banks post the same contract $y = (I, R) = (I^p, R^p)$ and the bank-borrower ratio is $\theta^p$, where $R^p = R_2^*$ is the first-best level repayment for type-2 borrowers, $\theta^p = \theta_2^*$ is the first-best market tightness for type-2, and $I^p$ ensures banks earn zero expected profits:

$$c^p = \pi_1 f_1(R^p) + \pi_2 f_2(R^p) - \frac{\theta^p c}{\mu(\theta^p)} \tag{3.13}$$

The share of type-$i$ borrowers searching for this contract satisfies $s_i(I^p, R^p) = \pi_i$ and there are enough of these contracts to be consistent with market clearing. It is straightforward to verify that this allocation is feasible, and I claim that if there are sufficiently few type-1 borrowers - which is the case in reality as risky and innovative small firms are relatively scarce compared to the safe ones, it is also a Pareto improvement over the equilibrium:borrower

Proposition 2: Assume condition (3.12) is not satisfied. When $\xi$ is large, for fixed values of the other parameters, there exists a $\overline{\pi} > 0$ such that if $\pi_1 < \overline{\pi}$, the equilibrium is Pareto dominated by the pooling allocation where all banks post $(I^p, R^p)$ with associated market tightness $\theta^p$.borrower

Please see Appendix for proof. In equilibrium, banks that want to attract type-2 borrowers need to screen out type-1. The cost of screening, which is reflected in the difference between type-2's equilibrium contract and their first-best allocations ($\overline{U}_2^p \geq \overline{U}_2$), is independent of the share of type-1 borrowers, while the collective benefit of screening depends on the share of type-1 borrowers. When the share of type-1 is small enough, type-2 would prefer to cross-subsidize type-1 to avoid costly screening. In this case, the cost of cross-subsidization is worth the increased efficiency of trade for type-2. However, this is inconsistent with equilibrium, since any individual borrower would prefer a contract that screens out the risky types. For example, both types would prefer their first-best allocations compared to the pooling allocations.borrower

Why does the inefficiency arise in this environment? Like the rat race, the return for higher probability of success exceeds the additional total revenue. The safe types attract more capital flows, not only from the revenue generated from production, but also because of the greater estimate of their quality by the bank. Efficiency can be improved if more funding resources is reallocated to the risky type.

## 3.6 Discussion of Venture Capital

### 3.6.1 Model Setup

This section describes the extended model framework with two types of lenders.

#### 3.6.1.1 Borrowers

The economy is populated with $N_B$ borrowers. At the beginning of the period, each borrower has one project that requires an initial investment $\omega$ to install. Assume $\omega$ must be externally financed and is the same for all borrowers. Borrowers are heterogenous across two dimensions, the amount of physical collateral $c$, which is observable, and the probability of having a successful project $\alpha$, which is private information to borrowers. Let the cumulative density function (CDF) $F(c,\alpha) = G(c)H(\alpha)$ summarize the prior distribution of collateral and probability of success. A project generates a constant cash flow $y^s$ if it is successful and $y^f$ if it fails[4].

#### 3.6.1.2 Lenders

Lenders are ex ante homogenous[5] and endowed with capital $\omega$. Lenders are risk-neutral and can enter at cost $k$. We fist consider traditional lenders like banks, then introduce venture capital that leverage data-driven technologies to grant funding, and see how this change affects the allocation of credit and social welfare.

**Banks.** Even when a project starts, banks cannot discover the borrower types and thus heavily rely on physical collateral to assign loans. Banks post contracts $s = (e, r) \in \mathcal{S} = [0, 1] \times \mathbb{R}_+$, which specifies a lending probability $e$ and an interest rate $r$. If the project succeeds, the bank gets repaid $(1 + r)\omega$; if it fails, the bank collects the collateral $c$.

**Venture capital.** Equity financing takes the form of funding given in exchange for partial ownership and future profits. Therefore, for the equity financiers like VC, instead of investigating the physical collateral, they obtain a fraction of the project cash flow $y^j, j = s, f$. Venture capital financiers post contracts $s = (e, l) \in \mathcal{S} = [0, 1] \times [0, 1]$, which specifies a lending probability $e$ and the equity share $l$. Therefore, although VC cannot tell the borrower types before giving funds, their profits are correlated with the true type's project benefits.

### 3.6.2 Financial Market

Suppose agents meet pairwise. The matching process is characterized by the lenders posting contract$s$ and borrowers direct their search to the one they prefer after observing all posted contracts. Matching is bilateral, so at most one borrower ever contacts one lender. Any

[4] Assume the project benefits are sufficiently large so that borrowers would always want to borrow and invest: $\alpha\left[y^s - (1+r)\omega + c\right] + (1-\alpha)\, y^f > c$.

[5] In an extended version of the model, we consider co-existence of two types of lenders.

contract $s$ is associated with a submarket with tightness $\theta(s)$, and a density $\phi(c,\alpha;s)$ of type-$(c,\alpha)$ borrowers. Denote the probability a borrower matches is $q\left[\theta(s)\right]=\theta(s)^{\gamma}$, where $0<\gamma<1$ [6]. Therefore, the probability a financier matches with a borrower $(c,\alpha)$ is $\frac{q\left[\theta(s)\right]\phi(c,\alpha;s)}{\theta(s)}$.

Conditional on meetings, the expected value of a bank granting loans to a borrower with $(c,\alpha)$ is

$$v_L^{bank}(s;c,\alpha)=e\left[\alpha r\omega+(1-\alpha)(c-\omega)\right]. \tag{3.14}$$

When the lender is a venture capital, the value is

$$v_L^{vc}(s;c,\alpha)=el\left[\alpha y^s+(1-\alpha)y^f\right]. \tag{3.15}$$

The expected payoff of a borrower $(c,\alpha)$ borrowing from a bank posting $s$ is given by

$$\begin{aligned} v_B^{bank}(s;c,\alpha)&=e\left\{\alpha\left[y^s-(1+r)\omega+c\right]+(1-\alpha)y^f\right\}+(1-e)c,\\ v_B^{vc}(s;c,\alpha)&=e(1-l)\left[\alpha y^s+(1-\alpha)y^f\right]+c. \end{aligned} \tag{3.16}$$

### 3.6.3 Competitive Equilibrium

**Definition.** A *competitive search equilibrium* is defined by a vector of borrowers' market utilities $\overline{U}(c,\alpha)$, market tightness $\theta(s)$, market composition $\phi(c,\alpha;s)$, defined over $\mathcal{S}$, a CDF $Z(s)$, and a set of posted contracts $\mathcal{S}^P\in\mathcal{S}$ that satisfy the following conditions:

(i) Financiers' *Profit Maximization* and *Free Entry*: For any $s\in\mathcal{S}$,

$$\frac{q\left[\theta(s)\right]}{\theta(s)}\int\phi(c,\alpha;s)v_L(s;c,\alpha)-k\leq 0 \tag{3.17}$$

[6] This implies that the number of bilateral meetings takes a Cobb-Douglas form.

with equality if $s \in \mathcal{S}^P$.

(ii) Borrowers' *Optimal Search*: For any $s \in \mathcal{S}$ and $(c, \alpha)$,

$$q\left[\theta(s)\right] v_B(s; c, \alpha) + \Big\{1 - q\left[\theta(s)\right]\Big\} c \leq \overline{U}(c, \alpha) \tag{3.18}$$

with equality if $\theta(s) < \infty$ and $\phi(c, \alpha; s) > 0$, where

$$\overline{U}(c, \alpha) = \max_{s \in \mathcal{S}} q\left[\theta(s)\right] v_B(s; c, \alpha) + \Big\{1 - q\left[\theta(s)\right]\Big\} c \tag{3.19}$$

(iii) *Market Clearing*:

$$\int_{\mathcal{S}^P} \phi(c, \alpha; s)\theta(s) dZ(s) \leq F(c, \alpha) \tag{3.20}$$

### 3.6.4 Discussion

Since venture capital does not require collateral, they are more likely to grant funding to borrowers with low $c$ and high $\alpha$ compared to the banks. However, since there is still information friction, there is credit rationing towards the safe types to deter the risky types from imitating.

## 3.7 Conclusion

In this paper, I develop a competitive search model of the bank-borrower relationship to explain how adverse selection and search frictions lead to a lending strategy that is detrimental to the less risky firms. I show that when banks do not enforce loan repayments firmly, and the average productivity is not very high, the safe type borrowers get credit-rationed on both the intensive and extensive margin. As firm productivity increases, they become over-charged but still less likely to obtain loans compared to the first-best level.borrower

When banks are relatively aggressive in collecting loan payments, fewer distortions occur, but once they are present, they take the form of overworking borrowers who are more likely to be successful in project investment. Similar in spirit to Akerlof (1976), the safer types have more tolerance for the high repayment requirements, which explains why they are offered more contracts and large loan size. However, I show that this is inefficient relative both to the first-best case and a pooling allocation where two types of borrowers are treated identically. This sheds light on how the innovative policies can guide the capital flows to achieve better efficiency by financing small businesses at their early age.

## 3.8 Appendix: Verification of Three Assumptions

The three key assumptions can be summarized as follows (Let $B_{\varepsilon}(y) \equiv \{y' \in Y | d(y, y') < \varepsilon\}$ be a ball of radius $\varepsilon$ around $y$):

Assumption A1 - Monotonicity: For all $y \in \overline{Y}$, $v_1(y) \leq v_2(y) \leq \cdots \leq v_I(y)$.

Assumption A2 - Local Nonsatiation: For all $i \in \mathbb{I}, y \in \overline{Y}_i$, and $\varepsilon > 0$, there exists a $y' \in B_{\varepsilon}(y)$ such that $v_i(y') > v_i(y)$ and $u_j(y') \leq u_j(y)$ for all $j < i$.

Assumption A3 - Sorting: For all $i \in \mathbb{I}, y \in \overline{Y}_i$, and $\varepsilon > 0$, there exists a $y' \in B_{\varepsilon}(y)$ such that $u_j(y') > u_j(y)$ for all $j \geq i$ and $u_j(y') < u_j(y)$ for all $j < i$.

The payoff function for bank $v_i(I, R) = [p_i + (1 - p_i)\xi] R - I$ is strictly increasing in $p_i$, which verifies A1. For A2, for any contract $y$, if we increase $R$, $v_i$ strictly increases and $u_j$ strictly decreases, for any $i$ and $j$. To verify A3, note that $\phi_i'(R) = [p_i + (1 - p_i)\xi] + 2(1 - p_i)\xi^2 R$, and $\frac{\partial \phi_i'(R)}{\partial p_i} = 1 - \xi - 2\xi^2 R$. For any given contract $y = (I, R)$, if $1 - \xi - 2\xi^2 R > 0$, then $\phi_1'(R) < \phi_2'(R)$. Consider an incremental decrease in $R$ to $R' = R - dR$, and an incremental decrease in $I$ to $I' = I - dI$. For a type-$i$ borrower, this raises her payoff by approximately $-p_i z_i dI + \phi_i'(R) dR$, which is positive if and only if $\frac{dI}{dR} < \frac{\phi_i'(R)}{p_i z_i}$. Since $p_1 z_1 = p_2 z_2$, and $\phi_1'(R) < \phi_2'(R)$, An appropriate choice of $\frac{dI}{dR}$ yields an increase in utility of type-2 and a decrease in utility of type-1, which verifies A3. Similarly, if $1 - \xi - 2\xi^2 R < 0$, then

$\phi_1'(R) > \phi_2'(R)$. Consider an incremental increase in $R$ to $R' = R + dR$, and an incremental increase in $I$ to $I' = I + dI$. For a type-$i$ borrower, this raises her payoff by approximately $p_i z_i dI - \phi_i'(R)dR$, which is positive if and only if $\frac{dI}{dR} > \frac{\phi_i'(R)}{p_i z_i}$. We can also find a proper $\frac{dI}{dR}$ that yields an increase in utility of type-2 and a decrease in utility of type-1.

## 3.9 Appendix

### 3.9.1 Proof of Proposition 1

It is easy to verify that the zero-profit constraint is binding: $\frac{\mu(\theta)}{\theta}(f_i(R) - I) = c$. Eliminate $I$ and reduce the problem to

$$\overline{U}_i = \max_{\theta,R,I} \mu(\theta)\left[p_i z_i I - \phi_i(R)\right] = \max_{\theta,R} \mu(\theta)\left[p_i z_i f_i(R) - \phi_i(R)\right] - p_i z_i \cdot \theta c \tag{3.21}$$

At the solution, $R_i^*$ solves $p_i z_i f_i'(R) = \phi_i'(R)$ and $\theta_i^*$ solves $\mu'(\theta)\left[p_i z_i f_i(R_i^*) - \phi_i(R_i^*)\right] = p_i z_i \cdot c$. Solving these two first order conditions gives us the analytical solutions in (3.11). Substituting this back into the constraint delivers $I_i^*$, and the objective function delivers $\overline{U}_i$. borrower

### 3.9.2 Proof of Proposition 2

I first prove that if the pooling contract raises the utility of type-2 borrowers relative to the equilibrium level, it raises the utility of type-1 as well. Since the constraint to exclude type-1 borrowers from the type-2 contract binds in equilibrium, the following equality holds:

$$p_1 z_1 \overline{U}_2 - p_2 z_2 \overline{U}_1 = \mu(\theta_2)\left(p_2 z_2 \phi_1(R_2) - p_1 z_1 \phi_2(R_2)\right) \tag{3.22}$$

In the pooling contract, both types pay the same amount back to the banks and have the same market tightness. Therefore,

$$p_1 z_1 \overline{U}_2^p - p_2 z_2 \overline{U}_1^p = \mu(\theta_2)\left(p_2 z_2 \phi_1(R^p) - p_1 z_1 \phi_2(R^p)\right) \tag{3.23}$$

Take differences, we obtain

$$p_1 z_1(\overline{U}_2 - \overline{U}_2^p) - p_2 z_2(\overline{U}_1 - \overline{U}_1^p) = \mu(\theta_2)\left[p_2 z_2(\phi_1(R_2) - \phi_1(R^p)) - p_1 z_1(\phi_2(R_2) - \phi_2(R^p))\right] \tag{3.24}$$

Rearranging it,

$$p_1 z_1(\overline{U}_2^p - \overline{U}_2) = p_2 z_2(\overline{U}_1^p - \overline{U}_1) - \mu(\theta_2)\left[p_2 z_2(\phi_1(R_2) - \phi_1(R^p)) - p_1 z_1(\phi_2(R_2) - \phi_2(R^p))\right] \tag{3.25}$$

From $\phi_i'(R) = [p_i + (1-p_i)\xi] + 2(1-p_i)\xi^2 R$, we have $\phi_1'(R) - \phi_2'(R) = (p_1 - p_2)(1 - \xi - 2\xi^2 R)$. When $\xi$ is close to 1, $\phi_2'(R) < \phi_1'(R)$. Integrating from $R^p$ to $R_2$ on both sides, we have

$$\mu(\theta_2)\left[p_2 z_2(\phi_1(R_2) - \phi_1(R^p)) - p_1 z_1(\phi_2(R_2) - \phi_2(R^p))\right] > 0 \tag{3.26}$$

It immediately follows that if $\overline{U}_2^p \geq \overline{U}_2$, then $\overline{U}_1^p > \overline{U}_1$.borrower

Next, because condition (3.12) is not satisfied, the equilibrium expected utility of type-2 borrowers, $\overline{U}_2$, is strictly less than the first-best level, $\overline{U}_2^*$. In addition, by construction $R^p$ and $\theta^p$ are equal to the first-best level for type-2, $(R_2^*, \theta_2^*)$, and $I^p$ is continuous in $\pi_1$ and converges to the first-best level $I_2^*$ as $\pi_1$ converges to 0. It follows that $\overline{U}_2^p$ is continuous in $\pi_1$ and converges to $\overline{U}_2^* > \overline{U}_2$ as $\pi_1$ converges to 0. This proves there is a $\overline{\pi}$ such that for all $\pi_1 < \overline{\pi}$, $\overline{U}_2^p > \overline{U}_2$.

# Bibliography


Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, 40(1):7–72.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91.

Acemoglu, D. and Autor, D. (2011). *Skills, Tasks and Technologies: Implications for Employment and Earnings*, volume 4 of *Handbook of Labor Economics*, chapter 12, pages 1043–1171. Elsevier.

Acemoglu, D. and Autor, D. (2012). What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology. *Journal of Economic Literature*, 50(2):426–63.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Afrouzi, H., Drenik, A., and Kim, R. (2020). Growing by the masses-revisiting the link between firm size and market power.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2019). A theory of falling growth and rising rents. Technical report, National Bureau of Economic Research.

Aiyagari, S. R. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, 109(3):659–684.

Akcigit, U. and Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443.

Akerlof, G. (1976). The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics*, pages 599–617.

Albanesi, S., Gregory, V., Patterson, C., and Sahin, A. (2013). Is job polarization holding back the labor market?

Alvarez-Cuadrado, F., Long, N., and Poschke, M. (2015). Capital-labor substitution, structural change and the labor income share. IZA Discussion Papers 8941, Institute for the Study of Labor (IZA).

Antonczyk, D., DeLeire, T. C., and Fitzenberger, B. (2010). Polarization and Rising Wage Inequality: Comparing the U.S. and Germany. IZA Discussion Papers 4842, Institute for the Study of Labor (IZA).

Argente, D., Fitzgerald, D., Moreira, S., and Priolo, A. (2021a). How do entrants build market share? the role of demand frictions. Technical report, Mimeo, Penn State University.

Argente, D., Fitzgerald, D., Moreira, S., and Priolo, A. (2021b). How do firms build market share? *Available at SSRN 3831706*.

Arkolakis, C. (2016). A unified theory of firm selection and growth. *The Quarterly Journal of Economics*, 131(1):89–155.

Armenter, R. (2015). A bit of a miracle no more: the decline of the labor share. *Business Review*, (Q3):1–9.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Reenen, J. V. (2017). Concentrating on the fall of the labor share. Working Paper 23108, National Bureau of Economic Research.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*.

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–97.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2006). The Polarization of the U.S. Labor Market. *American Economic Review*, 96(2):189–194.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics*, 90(2):300–323.

Autor, D. H., Katz, L. F., and Krueger, A. B. (1998). Computing Inequality: Have Computers Changed the Labor Market? *The Quarterly Journal of Economics*, 113(4):1169–1213.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content Of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

Bai, J., Chen, M., Liu, J., Mu, X., and Xu, D. Y. (2020). Search and information frictions on global e-commerce platforms: Evidence from aliexpress. Technical report, National Bureau of Economic Research.

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

Beaudry, P., Green, D. A., and Sand, B. M. (2013). The Great Reversal in the Demand for Skill and Cognitive Tasks. NBER Working Papers 18901, National Bureau of Economic Research, Inc.

Beck, T. and Levine, R. (2002). Industry growth and capital allocation:: does having a market-or bank-based system matter? *Journal of financial economics*, 64(2):147–180.

Becker, G. S. and Murphy, K. M. (1993). A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics*, 108(4):941–964.

Benhabib, J. and Bisin, A. (2002). Advertising, mass consumption and capitalism. *manuscript, Department of Economics NYU.*

Benhabib, J. and Bisin, A. (2011). Social construction of preferences: advertising. In *Handbook of Social Economics*, volume 1, pages 201–220. Elsevier.

Benkard, C. L. (2000). Learning and forgetting: The dynamics of aircraft production. *American Economic Review*, 90(4):1034–1054.

Berger, A. N. and Black, L. K. (2011). Bank size, lending technologies, and small business finance. *Journal of Banking & Finance*, 35(3):724–735.

Berlin, M. and Mester, L. J. (1999). Deposits and relationship lending. *The Review of Financial Studies*, 12(3):579–607.

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K., and Moxnes, A. (2022). The origins of firm heterogeneity: A production network approach. *Journal of Political Economy*, 130(7):1765–1804.

Bewley, T. (1977). The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292.

Bhandari, A., Martellini, P., and McGrattan, E. (2021). A theory of business transfers. *WP, UMN.*

Bisello, M. (2013). Job polarization in Britain from a task-based perspective.Evidence from the UK Skills Surveys. Discussion Papers 2013/160, Dipartimento di Economia e Management (DEM), University of Pisa, Pisa, Italy.

Blanchard, O. J. (1985). Debt, Deficits, and Finite Horizons. *Journal of Political Economy*, 93(2):223–47.

Blinder, A. S. (2009). How Many US Jobs Might be Offshorable? *World Economics*, 10(2):41–78.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44.

Böckerman, P. and Maliranta, M. (2011). Globalization, creative destruction, and labour share change: evidence on the determinants and mechanisms from longitudinal plant-level data. *Oxford Economic Papers*, 64(2):259.

Bornstein, G. et al. (2018). Entry and profits in an aging economy: The role of consumer inertia.

Boualam, Y. (2015). Bank lending and relationship capital. *Job Market Paper, University of Pennsylvania.*

Bridgman, B. (2018). Is Labor's Loss Capital's Gain? Gross Versus Net Labor Shares. *Macroeconomic Dynamics*, 22(8):2070–2087.

Butters, G. R. (1977). Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, 44(3):465–491.

Caillaud, B. and Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics*, pages 309–328.

Caselli, F. and Coleman, W. J. (2006). The World Technology Frontier. *American Economic Review*, 96(3):499–522.

Cavenaile, L., Celik, M. A., Roldan-Blanco, P., and Tian, X. (2021). Style over substance? advertising, innovation, and endogenous market structure. *Advertising, Innovation, and Endogenous Market Structure (March 13, 2021).*

Cavenaile, L. and Roldan-Blanco, P. (2021). Advertising, innovation, and economic growth. *American Economic Journal: Macroeconomics*, 13(3):251–303.

Chatterjee, S. and Eyigungor, B. (2020). Policy inertia, election uncertainty, and incumbency disadvantage of political parties. *The Review of Economic Studies*, 87(6):2600–2638.

Chiavari, A. (2022). The macroeconomics of rising returns to scale: Customers acquisition, markups, and dynamism.

Chirinko, R. S. (2008). [sigma]: The long and short of it. *Journal of Macroeconomics*, 30(2):671–686.

Cooley, T. F. and Prescott, E. C. (1995). *Frontiers of Business Cycle Research*, pages 1–38. Princeton University Press.

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. *Journal of Political Economy*, 118(4):747–786.

Cummins, J. G. and Violante, G. L. (2002). Investment-Specific Technical Change in the US (1947-2000): Measurement and Macroeconomic Consequences. *Review of Economic Dynamics*, 5(2):243–284.

Dabla-Norris, E., Kochhar, K., Suphaphiphat, N., Ricka, F., and Tsounta, E. (2015). Causes and Consequences of Income Inequality; A Global Perspective. IMF Staff Discussion Notes 15/13, International Monetary Fund.

Das, S. P. (2012). International trade and polarization in the labor market. *Economics - The Open-Access, Open-Assessment E-Journal*, 6:1–44.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

De Ridder, M. (2019). Market power and innovation in the intangible economy.

Delacroix, A. and Shi, S. (2013). Pricing and signaling with frictions. *Journal of Economic Theory*, 148(4):1301–1332.

DiCecio, R. (2009). Sticky wages and sectoral labor comovement. *Journal of Economic Dynamics and Control*, 33(3):538–553.

Dinlersoz, E. M. and Yorukoglu, M. (2012). Information and industry dynamics. *American Economic Review*, 102(2):884–913.

Dixit, A. and Norman, V. (1978). Advertising and welfare. *The Bell Journal of Economics*, pages 1–17.

Domeij, D. and Ljungqvist, L. (2006). Wage Structure and Public Sector Employment: Sweden versus the United States 1970-2002. CEPR Discussion Papers 5921, C.E.P.R. Discussion Papers.

Domeij, D. and Ljungqvist, L. (2019). Public Sector Employment and the Skill Premium: Sweden versus the United States 1970–2012. *Scandinavian Journal of Economics*, 121(1):3–31.

Dorn, D., Katz, L. F., Patterson, C., Van Reenen, J., et al. (2017). Concentrating on the fall of the labor share. *American Economic Review*, 107(5):180–85.

Drozd, L. A. and Nosal, J. B. (2012). Understanding international prices: Customers as capital. *American Economic Review*, 102(1):364–95.

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisiting the German Wage Structure. *The Quarterly Journal of Economics*, 124(2):843–881.

Eden, M. and Gaggl, P. (2018). On the Welfare Implications of Automation. *Review of Economic Dynamics*, 29:15–43.

Eden, M. and Gaggl, P. (2019). Capital Composition and the Declining Labor Share. Technical report.

Eggertsson, G. B., Mehrotra, N. R., and Robbins, J. A. (2019). A model of secular stagnation: Theory and quantitative evaluation. *American Economic Journal: Macroeconomics*, 11(1):1–48.

Einav, L., Klenow, P. J., Levin, J. D., and Murciano-Goroff, R. (2021). Customers and retail growth. Technical report, National Bureau of Economic Research.

Eisfeldt, A. L., Falato, A., and Xiaolan, M. Z. (2021). Human capitalists. Technical report, National Bureau of Economic Research.

Eisfeldt, A. L., Falato, A., and Xiaolan, M. Z. (2022). Human Capitalists. In *NBER Macroeconomics Annual 2022, volume 37*, NBER Chapters. National Bureau of Economic Research, Inc.

Elsby, M., Hobijn, B., and Sahin, A. (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity*, 47(2 (Fall)):1–63.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity. Federal Reserve Bank of San Francisco.

Fernandez-Macias, E. (2012). Job polarization in europe? changes in the employment structure and job quality, 1995-2007. *Work and Occupations*, 39(2):157–182.

Firpo, S., Fortin, N. M., and Lemieux, T. (2011). Occupational Tasks and Changes in the Wage Structure. IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

Fishman, A. and Rob, R. (2003). Consumer inertia, firm growth and industry dynamics. *Journal of Economic theory*, 109(1):24–38.

Fitzgerald, D., Haller, S., and Yedid-Levi, Y. (2017). Appendix for how exporters grow. Technical report, Federal Reserve Bank of Minneapolis.

Flood, S., King, M., Ruggles, S., and Warren, J. R. (2015). Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [machine-readable database]. `https://cps.ipums.org/cps/`.

Foote, C. L. and Ryan, R. W. (2014). Labor market polarization over the business cycle. Working Papers 14-16, Federal Reserve Bank of Boston.

Foster, L., Grim, C., Haltiwanger, J., and Wolf, Z. (2016). Firm-level dispersion in productivity: is the devil in the details? *American Economic Review*, 106(5):95–98.

Garcia-Macia, D., Hsieh, C.-T., and Klenow, P. J. (2019). How destructive is innovation? *Econometrica*, 87(5):1507–1541.

Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization*, 44(2):115–129.

Goos, M. and Manning, A. (2007). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *The Review of Economics and Statistics*, 89(1):118–133.

Goos, M., Manning, A., and Salomons, A. (2011). Explaining job polarization: the roles of technology, offshoring and institutions. Center for Economic Studies - Discussion papers ces11.34, Katholieke Universiteit Leuven, Centrum voor Economische Studiën.

Gordon, R. J. (1990). *The Measurement of Durable Goods Prices*. Number gord90-1 in NBER Books. National Bureau of Economic Research, Inc.

Gort, M., Greenwood, J., and Rupert, P. (1999). Measuring the Rate of Technological Progress in Structures. *Review of Economic Dynamics*, 2(1):207–230.

Gourio, F. and Rudanko, L. (2014). Customer capital. *Review of Economic Studies*, 81(3):1102–1136.

Greenwood, J., Hercowitz, Z., and Krusell, P. (1997). Long-Run Implications of Investment-Specific Technological Change. *American Economic Review*, 87(3):342–62.

Greenwood, J., Ma, Y., and Yorukoglu, M. (2021). 'you will:'a macroeconomic analysis of digital advertising. Technical report, National Bureau of Economic Research.

Griliches, Z. (1969). Capital-skill complementarity. *The review of Economics and Statistics*, pages 465–468.

Grossman, G., Helpman, E., Oberfield, E., and Sampson, T. (2018). The Productivity Slowdown and the Declining Labor Share. Technical report.

Grossman, G. M. and Rossi-Hansberg, E. (2008). Trading Tasks: A Simple Theory of Offshoring. *American Economic Review*, 98(5):1978–97.

Grossman, G. M. and Shapiro, C. (1984). Informative advertising with differentiated products. *The Review of Economic Studies*, 51(1):63–81.

Guerrieri, V., Shimer, R., and Wright, R. (2010). Adverse selection in competitive search equilibrium. *Econometrica*, 78(6):1823–1862.

Guerriero, M. (2012). The labour share of income around the world: Evidence from a panel dataset. Development Economics and Public Policy Working Paper 32/12, University of Manchester.

Guerriero, M. and Sen, K. (2012). What Determines the Share of Labour in National Income? A Cross-Country Analysis. IZA Discussion Papers 6643, Institute for the Study of Labor (IZA).

Hansen, G. D. and Ohanian, L. E. (2016). Neoclassical models in macroeconomics. Working Paper 22122, National Bureau of Economic Research.

Haskel, J. and Westlake, S. (2017). Capitalism without capital. In *Capitalism without Capital*. Princeton University Press.

Hellmann, T. (2002). A theory of strategic venture investing. *Journal of financial economics*, 64(2):285–314.

Hellmann, T. and Thiele, V. (2015). Friends or foes? the interrelationship between angel and venture capital markets. *Journal of Financial Economics*, 115(3):639–653.

Holmes, C. and Mayhew, K. (2010). Are UK Labour Markets Polariz-ing. Working Paper 97, SKOPE.

Hopenhayn, H., Neira, J., and Singhania, R. (2018). The rise and fall of labor force growth: Implications for firm demographics and aggregate trends. *NBER Working Paper*, pages 1–28.

Hsieh, C.-T. and Rossi-Hansberg, E. (2019). The industrial revolution in services. Technical report, National Bureau of Economic Research.

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

Ikenaga, T. and Kambayashi, R. (2010). Long-term Trends in the Polarization of the Japanese Labor Market: The Increase of Non-routine Task Input and Its Valuation in the Labor Market.

Imrohoroglu, A. (1989). Cost of Business Cycles with Indivisibilities and Liquidity Constraints. *Journal of Political Economy*, 97(6):1364–83.

Inderst, R. and Müller, H. M. (2004). The effect of capital market characteristics on the value of start-up firms. *Journal of Financial Economics*, 72(2):319–356.

Jaimovich, N. and Siu, H. E. (2012). The Trend is the Cycle: Job Polarization and Jobless Recoveries. NBER Working Papers 18334, National Bureau of Economic Research, Inc.

Jones, C. I. (1995). R & d-based models of economic growth. *Journal of political Economy*, 103(4):759–784.

Jones, C. I. (2020). The end of economic growth? unintended consequences of a declining population. Technical report, National Bureau of Economic Research.

Jorgenson, D. W. (2012). The World KLEMS Initiative. *International Productivity Monitor*, 24:5–19.

Jovanovic, B. and Szentes, B. (2013). On the market for venture capital. *Journal of Political Economy*, 121(3):493–527.

Jung, J. and Mercenier, J. (2014). Routinization-Biased Technical Change And Globalization: Understanding Labor Market Polarization. *Economic Inquiry*, 52(4):1446–1465.

Kaldor, N. (1957). A model of economic growth. *The Economic Journal*, 67:591–624.

Kampelmann, S. and Rycx, F. (2011). Task-Biased Changes of Employment and Remuneration: The Case of Occupations. IZA Discussion Papers 5470, Institute for the Study of Labor (IZA).

Karabarbounis, L. and Neiman, B. (2014a). Capital Depreciation and Labor Shares Around the World: Measurement and Implications. NBER Working Papers 20606, National Bureau of Economic Research, Inc.

Karabarbounis, L. and Neiman, B. (2014b). The Global Decline of the Labor Share. *The Quarterly Journal of Economics*, 129(1):61–103.

Karahan, F., Pugsley, B., and Şahin, A. (2019). Demographic origins of the startup deficit. Technical report, National Bureau of Economic Research.

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of political economy*, 112(5):986–1018.

Koh, D., Santaeulalia-Llopis, R., and Zheng, Y. (2015). Labor Share Decline and the Capitalization of Intellectual Property Products.

Kroen, T., Liu, E., Mian, A. R., and Sufi, A. (2021). Falling rates and rising superstars. Technical report, National Bureau of Economic Research.

Krusell, P., Ohanian, L. E., Rios-Rull, J.-V., and Violante, G. L. (1997). Capital-skill complementarity and inequality: a macroeconomic analysis. Technical report.

Krusell, P., Ohanian, L. E., Rios-Rull, J.-V., and Violante, G. L. (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica*, 68(5):1029–1054.

Lawrence, R. Z. (2015). Recent declines in labor's share in us income: A preliminary neoclassical account. Working Paper 21296, National Bureau of Economic Research.

Lentz, R. and Mortensen, D. T. (2008). An empirical model of growth through product innovation. *Econometrica*, 76(6):1317–1373.

Leon-Ledesma, M. A., McAdam, P., and Willman, A. (2010). Identifying the Elasticity of Substitution with Biased Technical Change. *American Economic Review*, 100(4):1330–57.

Leon-Ledesma, M. A., McAdam, P., and Willman, A. (2015). Production Technology Estimates and Balanced Growth. *Oxford Bulletin of Economics and Statistics*, 77(1):40–65.

Lerner, J., Sorensen, M., and Strömberg, P. (2011). Private equity and long-run investment: The case of innovation. *The Journal of Finance*, 66(2):445–477.

Maliar, L., Maliar, S., and Tsener, I. (2020). Capital-Skill Complementarity and Inequality: Twenty Years After. CEPR Discussion Papers 15228, C.E.P.R. Discussion Papers.

Maliar, L., Maliar, S., and Tsener, I. (2022). Capital-skill complementarity and inequality: Twenty years after. *Economics Letters*, 220:110844.

Manso, G. (2011). Motivating innovation. *The Journal of Finance*, 66(5):1823–1860.

Marx, M., Mojon, B., and Velde, F. R. (2019). Why have interest rates fallen far below the return on capital. BIS Working Papers 794, Bank for International Settlements.

McAdam, P., Muck, J., and Growiec, J. (2015). Endogenous Labor Share Cycles: Theory and Evidence. Technical report.

Mehra, R. and Prescott, E. C. (1985). The Equity Premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161.

Michelacci, C. and Suarez, J. (2004). Business creation and the stock market. *The Review of Economic Studies*, 71(2):459–481.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of political economy*, 94(4):796–821.

Molinari, B. and Turino, F. (2009). Advertising and business cycle fluctuations. *Instituto Valenciano de Investigaciones Económicas Working Paper No. AD*, 9.

Monte, F. (2011). Skill bias, trade, and wage dispersion. *Journal of International Economics*, 83(2):202–218.

Morlacco, M. and Zeke, D. (2021). Monetary policy, customer capital, and market power. *Journal of Monetary Economics*, 121:116–134.

Nanda, R. and Rhodes-Kropf, M. (2013). Investment cycles and startup innovation. *Journal of Financial Economics*, 110(2):403–418.

Naticchioni, P., Ragusa, G., and Massari, R. (2014). Unconditional and Conditional Wage Polarization in Europe. IZA Discussion Papers 8465, Institute for the Study of Labor (IZA).

Nellas, V. and Olivieri, E. (2012). Job Polarization and Labor Market Institution.

Nelson, P. (1974). Advertising as information. *Journal of political economy*, 82(4):729–754.

Oberfield, E. and Raval, D. (2014). Micro Data and Macro Technology. NBER Working Papers 20452, National Bureau of Economic Research, Inc.

Ohanian, L. E. and Orak, M. (2016a). Capital-Skill Complementarity, Inequality, and Labor's Share of Income, 1963–2013. Technical report.

Ohanian, L. E. and Orak, M. (2016b). Capital-Skill Complementarity, Inequality, and Labor's Share of Income, 1963-2013. Discussion Paper.

Olmstead-Rumsey, J. (2019). Market concentration and the productivity slowdown.

O'Mahony, M. and Timmer, M. P. (2009). Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database. *Economic Journal*, 119(538):F374–F403.

Orak, M. (2013). From Job Polarization to Wealth Inequality: A Quantitative Test of Routinization Hypothesis.

Orak, M. (2015). Is Information Technology Hurting Labor?

Orak, M. (2017). Capital-Task Complementarity and the Decline of the U.S. Labor Share of Income. International Finance Discussion Papers 1200, Board of Governors of the Federal Reserve System (U.S.).

Paciello, L., Pozzi, A., and Trachter, N. (2019). Price dynamics with customer markets. *International Economic Review*, 60(1):413–446.

Perla, J. (2017). Product awareness, industry life cycles, and aggregate profits. *Mimeograph, University of British Columbia.*

Perla, J. (2019). A model of product awareness and industry life cycles. *Unpublished Paper, University of British Columbia.*

Peters, M. and Walsh, C. (2021). Population growth and firm dynamics. Technical report, National Bureau of Economic Research.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Handbook of Labor Economics. Belknap Press.

Piketty, T. and Zucman, G. (2014). Capital is Back: Wealth-Income Ratios in Rich Countries 1700?2010. *The Quarterly Journal of Economics*, 129(3):1255–1310.

Polgreen, L. and Silos, P. (2005). Capital-skill complementarity and inequality: a sensitivity analysis. Technical report.

Polgreen, L. and Silos, P. (2008). Capital-Skill Complementarity and Inequality: A Sensitivity Analysis. *Review of Economic Dynamics*, 11(2):302–313.

Polgreen, L. and Silos, P. (2009). Crude substitution: The cyclical dynamics of oil prices and the skill premium. *Journal of Monetary Economics*, 56(3):409–418.

Rachel, L. (2021). Leisure-enhancing technological change. *Job market paper, London School of Economics*.

Reichheld, F. F., Teal, T., and Smith, D. K. (1996). *The loyalty effect*, volume 1. Harvard business school press Boston, MA.

Robb, A. and Robinson, D. (2012). The capital structure decisions of startup firms. *Review of Financial Studies*, 1(1):1–27.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029.

Rognlie, M. (2015). Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity? *Brookings Papers on Economic Activity*, 46(1 (Spring):1–69.

Roldan-Blanco, P. and Gilbukh, S. (2021). Firm dynamics and pricing under customer capital accumulation. *Journal of Monetary Economics*, 118:99–119.

Shao, E. and Silos, P. (2014). Accounting For The Cyclical Dynamics Of Income Shares. *Economic Inquiry*, 52(2):778–795.

Sherk, J. (2016). Labor's Share of Income Little Changed Since 1948. *The Heritage Foundation Report*, (May 31, 2016).

Smith, C. L. (2013). The dynamics of labor market polarization. Technical report.

Sørensen, M. (2007). How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance*, 62(6):2725–2762.

Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, 24(2):235–270.

Stewart, M. B. (2012). Wage inequality, minimum wage effects, and spillovers. *Oxford Economic Papers*, 64(4):616–634.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, 71(3):393–410.

Stokey, N. and Lucas, R. (1996). *Recursive methods in economic dynamics*. Harvard Univ. Press, Cambridge, Mass. [u.a.], 4. print edition.

Thomas, J. S. (2001). A methodology for linking customer acquisition to customer retention. *Journal of marketing research*, 38(2):262–268.

Tian, X. and Wang, T. Y. (2011). Tolerance for failure and corporate innovation. *The Review of Financial Studies*, 27(1):211–255.

Unel, B. (2010). Analyzing skilled and unskilled labor efficiencies in the US. *Journal of Macroeconomics*, 32(4):957–967.

vom Lehn, C. (2018). Understanding the decline in the u.s. labor share: Evidence from occupational tasks. *European Economic Review*, 108:191 – 220.

White, H. (1996). *Estimation, inference and specification analysis.* Number 22. Cambridge university press.

Winegar, A. (2018). Adverse selection, capital supply, and venture capital allocation. *Capital Supply, and Venture Capital Allocation (April 15, 2018).*

Zambrano, A. (2013). Endogenous Employment and Incomplete Markets.

Òscar Jordà, Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). The Rate of Return on Everything, 1870–2015. *The Quarterly Journal of Economics*, 134(3):1225–1298.