The London School of Economics and Political Science

Essays on Firms, Technology, and Macroeconomics

Xijie Gao

A thesis submitted to the Department of Economics for the degree of Doctor of Philosophy, London, September 2020

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of approximately 26,000 words.

Statement of co-authored work

I confirm that Chapter 3 was jointly co-authored with Jiajia Gu and I contributed 50% of this work.

Abstract

This thesis consists of three chapters. Chapter 1 studies the role of marketing in the economy. Using aggregate and firm-level data, I find that aggregate marketing intensity in the US increased sharply around the mid-1990s, which coincides with a rapid rise of elasticity between firm-level Marketing Production Cost Ratio and markup. To explain these facts, I develop a model with heterogeneous firms and endogenous markups in which firms engage in marketing to signal their quality. I use a calibrated version of the model to quantify the impact of information frictions and marketing on aggregate productivity. I find that quality information revealed by marketing is valuable and access to marketing cannot undo the information frictions completely.

Chapter 2 examines the impact of zombie firms on resource allocation. Using firm-level data in China, I show that zombie firms are larger, less productive, and receive a higher subsidy rate on average. The difference in average subsidy rate between zombies and non-zombies reflects both the selection criterion of zombies and the underlying joint distribution of subsidy rate and productivity. I develop a model with heterogeneous firms to quantify the impact of zombies on aggregate productivity. Quantitative exercise shows that reducing the dispersion in subsidy rate across firms can lead to significant productivity gains, while policies that increase the exit rate of zombies have limited productivity effects.

Chapter 3 establishes two facts along with the rise of information technology: (i) the output from the information sector is more intensively used as an intermediate input; (ii) the wage of information workers and their total employment increase relative to those of non-information workers. To understand the causes, we develop a two-sector accounting framework with sector-factor specific technical changes. We find that labor-augmenting technical change is important in explaining the observed change in wage premium and intermediate shares.

Acknowledgements

I am deeply indebted to my supervisor Rachel Ngai, for her careful guidance and continued support throughout my PhD journey. The completion of this thesis would not have been possible without her encouragement. I am also grateful to Wouter Den Haan as my supervisor during the later stage of my PhD studies. His insights and suggestions help me immensely in improving my work. I thank Christopher Pissarides, John Van Reenen and Ricardo Reis for all the help they gave me during the job market. I thank Shengxing Zhang for the reading group he organized, which helps me learn how to read papers and explain intuitions. I thank Carol Corrado and Janet Hao for their help on estimation of marketing spending in Chapter 1. Finally, I would like to thank all participants at LSE Money/Macro Student WiP seminars for useful feedback.

Contents

Abstract **3**

Acknowledgements **4**

1 Marketing, Market Power, and Aggregate Productivity **15**

1.1 Introduction 15

1.2 Data and Measurement 20

1.2.1 Aggregate data on marketing spending 20

1.2.2 Firm level data 21

1.3 Empirical Findings 23

1.3.1 Aggregate trend of marketing intensity 23

1.3.2 Marketing-production cost ratio and markup 26

1.4 Model 26

1.4.1 Setup 28

1.4.2 Complete information 31

1.4.3 Incomplete information 32

1.4.4 Resource allocation and aggregation productivity 37

1.5 Calibration 40

1.5.1 Assigned parameters 41

1.5.2 Calibrated parameters 41

1.6 Results 45

1.6.1 MPCR-markup elasticity 46

1.6.2 Can access to marketing undo information frictions? . . . 47

1.6.3 The impact of marketing on aggregate productivity . . . 49

1.7 Implication and Discussion 51

1.7.1 Trend of market power in the data 52

1.7.2 Marketing and economic development 52

1.7.3 Marketing vs Advertising 54

1.8 Conclusion 55

1.9 Appendix 57

1.9.1 Measurement of marketing spending 57

1.9.2 Firm-level data and markup estimation 61

1.9.3 Marketing as a fixed cost 63

1.9.4 MPCR-markup elasticity: alternative regression 64

1.9.5 MPCR-markup elasticity: Translog production function . 66

1.9.6 SGA-production cost ratio vs markup 66

1.9.7 Proof of Propositions and Lemmas 68

1.9.7.1 Proof of Lemma 1 68

1.9.7.2 Proof of Lemma 2 69

1.9.7.3 Solving the LCSE 70

1.9.7.4 Proof of Proposition 2 71

1.9.8 Robustness check 73

1.9.8.1 Importance of u and ξ 73

1.9.8.2 Step size of quality ladder Δ 74

1.9.8.3 Oligopolistic competition 74

2 Zombie Firms, State Subsidies, and Resource Misallocation 81

2.1 Introduction 81

2.2 Empirical findings 84

2.2.1 Data and measurement 84

2.2.2 Characteristics of zombie firms 86

2.2.3 Zombies and state subsidies 90

2.2.4 Alternative measures of zombies 93

2.3 Model 95

2.3.1 Model setup 96

2.3.2 Zombie firms in the model 98

2.4 Calibration 101

2.5 Results 104

2.5.1 Productivity gain from removing distortions 104

2.5.2 Productivity gain from rising private firms 106

2.5.3 How effective is the policies that increase the exit rate of zombies? 107

2.6 Conclusion 109

2.7 Appendix 110

2.7.1 Firm-level data from China 110

2.7.2 Alternative calibration 111

2.7.3 Supplementary figures 113

3 The Rise of Information Inputs in Production 121

3.1 Introduction 121

3.2 Empirical facts 124

3.2.1 Data and measurement 124

3.2.2 Intermediate input from information sector 124
3.2.3 Rise of info labor 126
3.3 Theoretical framework 131
3.3.1 Sectoral production 132
3.3.2 Inferring technologies 133
3.3.3 Data and implementation 134
3.3.4 Elasticities 137
3.4 Quantitative results 138
3.4.1 Intermediate shares and technical changes 139
3.4.2 Wage premium and labor-augmenting technical change . 141
3.4.3 Robustness check 142
3.4.3.1 Alternative elasticity of substitution 142
3.4.3.2 Sector-specific price for intermediate inputs . . 143
3.5 Conclusion 144
3.6 Appendix 145
3.6.1 Data appendix 145
3.6.2 Supplementary tables and figures 149

List of Tables

1.1 Parameterization . . . 42

1.2 Moments . . . 44

1.3 MPCR-markup elasticity . . . 46

1.4 Co-movement between MPCR-markup elasticity and aggregate marekting intensity . . . 46

1.5 Percentage losses of aggregate productivity . . . 48

1.6 Decomposition of *between-firm* component . . . 48

1.7 Productivity losses and rising returns of marketing . . . 49

1.8 Productivity gains from marketing . . . 50

1.9 Aggregate markup and markup distribution . . . 53

1.10 Marketing occupations . . . 59

1.11 MPCR-markup elasticity: alternative regression specification . . . 65

1.12 Importance of u and ξ . . . 74

1.13 Results with calibrated Δ . . . 75

1.14 Oligopolistic competition . . . 77

2.1 Parameterization . . . 102

2.2 Moments . . . 102

2.3 Productivity gain from removing distortions . . . 105

2.4 Alternative calibration: parameterization . . . 112

2.5 Alternative calibration: moments 112

2.6 Decomposition of productivity gain from removing distortions: alternative calibration 112

3.1 Regression of income on info score 127

3.2 Annual growth rate of technologies 139

3.3 Mean and median information score by sector: 1980 149

3.4 Annual growth rate of technologies with alternative σ_I and σ_N . 150

3.5 Annual growth rate of technologies with alternative σ_C 151

3.6 Annual growth rate of technologies with alternative σ_k 152

3.7 Annual growth rate of sector-factor augmenting technologies with sector-factor specific intermediate prices 152

3.8 Distribution of occupations 153

List of Figures

1.1 Aggregate marketing trend in the US 23

1.2 Marketing expenditure decomposition 24

1.3 Employment shares of marketing-related activities in the US . . 25

1.4 Binscatter plot between log(MPCR) and log(markup) 27

1.5 Binscatter plot between log(MPCR) and log(markup): before and after mid 1990s 27

1.6 MPCR-markup elasticity 28

1.7 Quality, Firm Size and Markup 32

1.8 MPCR and markup: different returns to marketing 36

1.9 Illustration: u_k, ξ_k and marketing intensity 44

1.10 Aggregate markup trend in Compustat data 53

1.11 Markup dispersion in Compustat data 53

1.12 Advertising and economic development 54

1.13 Marketing expenditure decomposition since 1987 60

1.14 Aggregate marketing intensity: IT intensive vs non IT-intensive 61

1.15 MPCR-markup elasticity: alternative specification 65

1.16 Binscatter plot between log(MPCR) and log(markup) with Translog production function 66

1.17 Binscatter plot between log(MPCR) and log(markup) with Translog production function: before and after mid 1990s 67

1.18 MPCR-markup elasticity with Translog production function . . 67

1.19 MPCR-markup elasticity: SGA as a measure of marketing spending 68

2.1 Zombie rate in China 86

2.2 Zombie rate: SOEs vs Non-SOEs 87

2.3 Zombie rate by industry groups in 1999 87

2.4 Zombie rate by provinces in 1999 88

2.5 Size distribution: zombies vs non-zombies 89

2.6 Output and productivity: zombies vs non-zombies 89

2.7 Subsidy: zombies vs non-zombies 91

2.8 Regression-based subsidy gap for zombie firms 91

2.9 Subsidy rate and productivity 92

2.10 Alternative measures of zombie firms 94

2.11 Subsidy rate and zombie firms: alternative zombie measures . . 95

2.12 Zombie classification 99

2.13 Productivity distribution: zombies vs non-zombies 103

2.14 Productivity and subsidy rate: zombies vs non-zombies 103

2.15 Subsidy dispersion in the data 107

2.16 Aggregate productivity, zombie rate and dispersion of subsidy . 107

2.17 Aggregate productivity and exit rate of zombie firms 108

2.18 Aggregate productivity, zombie rate and dispersion of subsidy: alternative calibration 112

2.19 Aggregate productivity and exit rate of zombie firms: alternative calibration 113

2.20 Rising share of non-SOE firms 114

2.21 Decomposing the decline of zombie rate by industries and provinces 114

2.22 Regression-based size gap between zombies and non-zombies . . 115

2.23 Regression-based productivity gap between zombies and non-zombies 115

2.24 Zombie rate and age restriction 116

2.25 Size and productivity gap between zombies and non-zombies: alternative zombie measures 116

2.26 Measured exit rate 117

3.1 Aggregate intermediate share 125

3.2 Aggregate intermediate share decomposition 126

3.3 Intermediate share by sector 127

3.4 Returns to info occupations 128

3.5 Wage premium of info worker 129

3.6 Employment and hours worked: info worker and non-info worker 130

3.7 Partial R^2 net of experience, education, race and gender 131

3.8 Labor share 135

3.9 Inferred wage premium of info labor by sector 136

3.10 Quantity and price indices by sector 136

3.11 Technical changes and intermediate shares 140

3.12 Wage premium and labor-augmenting technical change 142

3.13 Technical changes and intermediate shares: $\sigma_I = \sigma_N = 1.8$ . . . 153

3.14 Technical changes and intermediate shares: $\sigma_I = \sigma_N = 2.8$ . . . 154

3.15 Technical changes and intermediate shares: $\sigma_I = 2.2$ and $\sigma_N = 2$ 155

3.16 Technical changes and intermediate shares: $\sigma_C = 0.4$ 156

3.17 Technical changes and intermediate shares: $\sigma_C = 0.8$ 157

3.18 Technical changes and intermediate shares: $\sigma_k = 0.4$ 158

3.19 Technical changes and intermediate shares: $\sigma_k = 0.8$ 159

3.20 Technical changes and intermediate shares: sector-factor specific intermediate prices 160

3.21 Wage premium and labor-augmenting technical change with alternative elasticities 161

3.22 Wage premium and labor-augmenting technical change with alternative elasticities (continued) 162

Chapter 1

Marketing, Market Power, and Aggregate Productivity

1.1 Introduction

Firms spend substantial resources on marketing. In 2012, around \$140 billion was spent on media advertising in the US. This represents 0.9% of US GDP and \$444 per capita.[1] As Bagwell (2007) noted, marketing is a "prominent feature of economic life".[2] The massive spending creates information for consumers, provides a channel for firms to differentiate from their competitors and gain market power. Despite the prominence of marketing, at the aggregate level, there is little empirical evidence on the magnitude and trend of this spending, and how the marketing spending correlates with market power. Quantitatively, few attempts have been made to quantify the impact of marketing on aggregate productivity by accounting for both the positive effect of marketing through information revelation and the negative effect of marketing via markup dispersion.

In this chapter, I estimate aggregate marketing intensity in the US, which is defined as the ratio of aggregate marketing spending to GDP. Using firm-level data, a positive correlation between Marketing Production Cost Ratio (MPCR) and market power is documented. Moreover, the *cross-sectional* MPCR-markup elasticity *co-moves* closely with aggregate marketing intensity. To explain these facts, I develop a model with heterogeneous firms and

[1]Belleflamme and Peitz (2015).

[2]The original phrase refers to advertising. Since advertising is a major component of marketing, I adopt the quote here.

endogenous markups where firms engage in marketing to signal their quality. The existence of information frictions could explain both the level of elasticity and its co-movement with aggregate marketing intensity. I then use the model to evaluate the impact of information frictions and marketing on aggregate productivity. Although access to marketing cannot restore the complete information allocation, marketing generates significant productivity gains by revealing valuable information to consumers. The gain from quality information revealed by marketing tends to dominate the loss due to markup dispersion.

I begin by estimating the aggregate marketing intensity in the US. The aggregate marketing expenditure contains three components: advertising expenses, purchased marketing services, and own-account marketing labor compensation. It is observed that since the mid-1990s — the beginning of the Internet era — aggregate marketing intensity in the US increased by about 20% until 2000. From 2000 onwards, the aggregate marketing intensity flattened out. The further decomposition shows that the rise of aggregate marketing intensity is a within-industry phenomenon. The within-industry component continues to rise until the financial crisis. Therefore, it is the reallocation of output between industries that results in a flattened aggregate marketing intensity since the 2000s. One possible explanation of rising marketing intensity is that the rise of information and communications technology expands the marketing opportunities and provides more channels for the interaction between firms and consumers.

I then investigate whether marketing spending is related to the market power of the firm. Using Compustat data, I document that marketing is positively correlated with the markup. Firms with high markups tend to spend more on marketing relative to production, and thus have high MPCR. In addition, the *cross-sectional* MPCR-markup elasticity increases sharply around mid-1990s which *co-moves* with aggregate marketing intensity closely .

To explain the findings in the data, I develop a model with heterogeneous firms and endogenous variable markups where firms engage in marketing to signal their quality. The demand of a firm depends on its taste shifter, which is assumed to be a geometric average of the exogenous quality and endogenous appeal of the firm. Each firm has access to a marketing technology, where they can increase appeal using labor inputs. I assume it is less costly for high-quality firms to increase the appeal. Thus, when quality of the firm is unobservable, firms can credibly convey quality information via appeal. In the presence of information frictions, I use the notion of Perfect Bayesian Equilibrium (PBE) to solve the model. One drawback of PBE is the multiplicity of equilibria due to the flexibility of off-equilibrium beliefs. I focus on the *least-cost separating equilibrium* (LCSE) in this chapter.

In the model, quality is the only source of heterogeneity between firms. Markup is positively correlated with firm size and quality. When there are no information frictions, firms would still spend on marketing to increase appeal as it is complementary to consumption. The MPCR is constant across firms within the same industry and independent of the markup. However, with information frictions, the signaling competition between firms generates a positive correlation between MPCR and markup. Technical changes that increase the returns to marketing fuel the signaling competition, which would increase the MPCR-markup elasticity and aggregate marketing intensity simultaneously. The intuition is that higher returns to marketing lower the cost for low-quality firms to mimic high-quality firms. Thus, more intense signaling competition between firms generates a larger dispersion of MPCR for given amount of markup dispersion.

I calibrate the model to match the industry-level distribution on marketing intensity and firm size. There are two sources of heterogeneity across industries. The first difference is the weight of appeal in consumers' utility function. For instance, the weight could be higher for some goods such as cosmetics and clothing. The second source of heterogeneity is the level of quality differentiation within an industry, which is measured by the tail parameter of quality distribution. The tail parameter captures the extent of signaling competition. These two sources of heterogeneity determine the marketing intensity and firm size distribution within an industry. The calibrated model implies an MPCR-markup elasticity similar to the data, and a rise in the returns to marketing generates a co-movement between MPCR-markup elasticity and aggregate marketing intensity.

The calibrated model allows us to quantify the effects of marketing and information frictions on aggregate productivity. I begin by asking to what extent access to marketing can undo the information frictions. Compared to the complete information allocation, the productivity loss from information frictions is around 3 percent. I decompose the loss into two channels: (i) a *between-firm* component that summarizes the misallocation between firms due to markup and information distortion, and (ii) a *between-activity* component that summarizes the misallocation of labor between production and marketing in the aggregate. I find that *between-activity* channel accounts for almost all the losses from information frictions. Moreover, the recent technical changes that induce higher returns to marketing, such as the Internet and search engines, would increase the loss from information frictions.

Next, I move on to study the aggregate impact of marketing, which has been debated at length in the theoretical economic literature. Using the calibrated model, I attempt to study this question quantitatively. On the one hand, marketing is *informative* as it conveys valuable quality information to consumers. On the other hand, the information revealed by marketing would result in markup dispersion and hence resource misallocation between firms. I restrict a firm's ability to signal its quality via appeal and compare the LCSE with a *no-information* allocation. I find the productivity gains from information revealed by marketing can be substantial. However, the result comes with caveats. Firstly, LCSE implies a minimum loss in the process of information revelation. The deviation from LCSE would lower the gains from marketing. Secondly, as in Edmond et al. (2018), markup is related only to firm size in the model. If marketing generates dispersion in the markup which is not related to size, then it would involve a larger loss from misallocation. Furthermore, the model presumes each sector faces the same information frictions, whereas in reality quality information could be observable for some industries even without marketing.[3]

Related literature

This chapter is related to several strands of literature. Firstly, the measurement of marketing spending is related to those of intangible capital. Corrado et al. (2005) and Corrado and Hao (2014a) measure brand equity as a component of intangible capital by capitalizing several items of marketing spending. In this chapter, I estimate marketing spending directly and show its trend since 1987 in the US. The technological interpretation of the rising marketing intensity is consistent with Haskel and Westlake (2018).

In the model, firms spend on marketing to signal their quality, increase the appeal, and gain market power. The variation in firm size is driven by the quality information revealed by marketing, which is related to the recent literature on

[3]Beyond the mechanism in the model, there are many other ways that marketing could affect aggregate productivity. Some of them focus on the positive role of marketing, and others focus on the detrimental role of marketing. It is not possible to account for all those channels into a single framework. Nevertheless, the result remains as one of the first attempts to quantify the aggregate impact of marketing.

the determinants of firm performance (Hottman et al., 2016). Besides, the revealed information induces dispersion in markup and lowers the aggregate productivity as in Edmond et al. (2018) and Baqaee and Farhi (2018). Other theoretical papers discuss different ways of how marketing affects demand, such as reaching more customers (Arkolakis, 2010) and forming new customer-relationships (Gourio and Rudanko, 2014). Nevo (2001) studies the pricing behavior in the ready-to-eat cereal industry and discovers that the observed high price-cost margins can be partly attributed to the fact firms spend a large number of resources on marketing to influence the perceived quality.

The positive correlation between MPCR and markup in this chapter is related to Traina (2018) and De Loecker et al. (2018). They document that markup is positively correlated with "Selling, General and Administrative Expenses" (SGA) expenses of firms. However, as noted by Ptok et al. (2018), SGA is not a measure of marketing expenditure. SGA-based marketing spending measurement fails to generate the co-movement pattern with aggregate marketing intensity.

Lastly, this chapter is also related to the discussion of marketing and welfare. The idea of using marketing spending to signal quality is first proposed by Nelson (1974), which highlights the *informative* role of marketing. The model could entertain the other two views on marketing as documented by Bagwell (2007) — *complementary* view and *persuasive* view. Results show that the gains from information tend to dominate the loss due to markup dispersion, and the aggregate effects of marketing are welfare-enhancing. Rauch (2013) investigates the effect of a change in the marginal costs of advertising on consumer prices. He posits the aggregate effect of advertising to be informative since advertising tends to decease consumer price.

The chapter is organized as follows. Section 1.2 introduces the data and measurement of marketing spending and markup. Section 1.3 presents the empirical facts regarding marketing intensity and MPCR-markup elasticity. Section 1.4 illustrates the model. Section 1.5 explains how we calibrate the model to match cross-industry marketing intensity distribution and concentration facts. Section 1.6 presents the quantitative results. Section 1.7 discusses the implications of the model. Section 1.8 concludes.

1.2 Data and Measurement

1.2.1 Aggregate data on marketing spending

Firms spend substantial resources on marketing to attract consumers and generate revenues. As noted by Corrado and Hao (2014a), marketing spending is "outlays designed to augment the demand for a firm's products and services — that is, to shift its price-quantity demand schedules upward, so that more will be sold at a given price." Following the literature, I estimate the marketing spending in the US using data from different sources.[4]

The marketing expenditure consists of three components. The first component is the advertising expense. Advertising is one important part of marketing which is used to raise awareness and convey information concerning the products to the consumer. Beyond advertising, firms also purchase marketing research and marketing consulting services. Those services help them to create and develop a successful marketing strategy. Lastly, there is an additional in-house component where firms hire employees, such as marketing managers and public relation specialists, to conduct marketing activities. I estimate these three components separately and add them up to get the total marketing spending in the US.

Advertising expenses. I extract advertising expense data from IRS tax data, both at the industry-level and aggregate level. Firms report advertising expenses to the IRS for the tax deduction purpose. The data are reported by legal forms of organization, major industry, and ownership.

The advertising expense includes all the direct costs of advertising. For example, it would cover the advertising expenses via different media channels (including newspapers, TV, Internet, etc) and public relations expenses (e.g. sponsorship of sports teams, publicity campaign or events), and cost of producing promotional items like t-shirts and mugs.

[4] The measurement of marketing spending is related to the literature on the measurement of intangible capital. Corrado et al. (2005) summarize the intangible capital into three categories: computerized information, innovative property, and economic competencies. Within the category of economic competencies, they capitalize several spending on marketing and label it as *brand equity*. Due to the process of capitalization, only a fraction of marketing spending is accounted as an investment in the brand and the level of the fraction varies across different items of marketing. Thus, the resulting trend of *brand equity* may be different compared to the total marketing spending itself. In this chapter, I estimate directly the magnitude and trend of marketing spending in the US. For details on the measurement of *brand equity*, see Corrado and Hao (2014a).

Marketing research and consulting. The second component of marketing spending is purchased marketing research and marketing consulting services. I estimate the total spending on those activities using the revenue of the corresponding industries, specifically NAICS 541613 for marketing consulting services and NAICS 541910 for marketing research.[5] For the industry-level purchase of the output from these two industries, I use the 2002 benchmark input-output table.

Marketing labor compensation. The last component of marketing spending is the own-account component, where the own-account component is measured by the wage compensation to marketing workers. Specifically, I first identify a list of marketing occupations and then estimate the labor compensation associated with those marketing occupations for each industry.[6]

Notice that the marketing labor compensation in the media and marketing related industries may already be part of purchased advertising expenses for other industries. To avoid this double counting problem, I only estimate the marketing labor compensation for the private, non-agricultural, non-media, non-marketing related industries. Besides, I do not estimate the capital spending in the marketing department for two reasons. Firstly, to my knowledge, there is no data available that separately report the capital investment for different function departments within the firm. Secondly, the measurement of marketing spending is recorded as operating expenses in the income statement, which would be consistent with the firm-level data.

1.2.2 Firm level data

For the firm-level data, I use the Compustat Fundamental Annual file, where I observe sales, capital stock information, operating expenses, and industry classification. The choice of data is driven by two reasons. Firstly, Compustat covers a wide range of industries over a substantial period of time. Additionally, it reports detailed components of a firm's operating expenses, which allows for the estimation of markup and marketing spending.

Following De Loecker and Warzynski (2012), I apply the production-based estimate of markup to the Compustat data. Markup is measured as a ratio

[5] The survey-based measures are available from the Service Annual Survey (SAS).

[6] See section 1.9.1 in Appendix for the list of occupation that is marketing related and the method to estimate the marketing labor compensation.

between output elasticity of variable input and revenue share of variable input

$$\mu_{it} = \theta_V \frac{P_{it}Q_{it}}{P_{it}^V V_{it}}$$

In Compustat data, I use "Cost of Goods Sold" (COGS) as a measure of variable input and assume an industry-specific production function to estimate θ_V.[7]

Marketing spending is measured using advertising expenses (XAD). Marketing-Production Cost Ratio (MPCR) is measured as the ratio of marketing spending to production cost (COGS).[8] For each year, I define the *cross-sectional* MPCR-markup elasticity as coefficient ρ from the following regression

$$log\,(MPCR_{is}) = \rho\ log(\mu_{is}) + \delta_s + \epsilon_{is}$$

where δ_s is the industry fixed effect.[9]

It is worth noting that "Selling, General and Administrative Expenses" (SGA) is not a valid measure of firms' marketing spending. Among 29 items that constitute SGA, only 2 items — advertising expenses and marketing expenses — relate directly to marketing spending. Ptok et al. (2018) show that advertising expenses reported by Compustat is highly correlated with total marketing spending of the firm.[10,11] To this end, I view the advertising expenses as a valid proxy of the marketing spending of the firm.

[7] The lack of firm-specific deflator for input and output may lead to a biased estimate of θ_v. But it will not affect the variation of markup across firms in a given year or over time. See Brandt et al. (2017), De Loecker et al. (2018) for details.

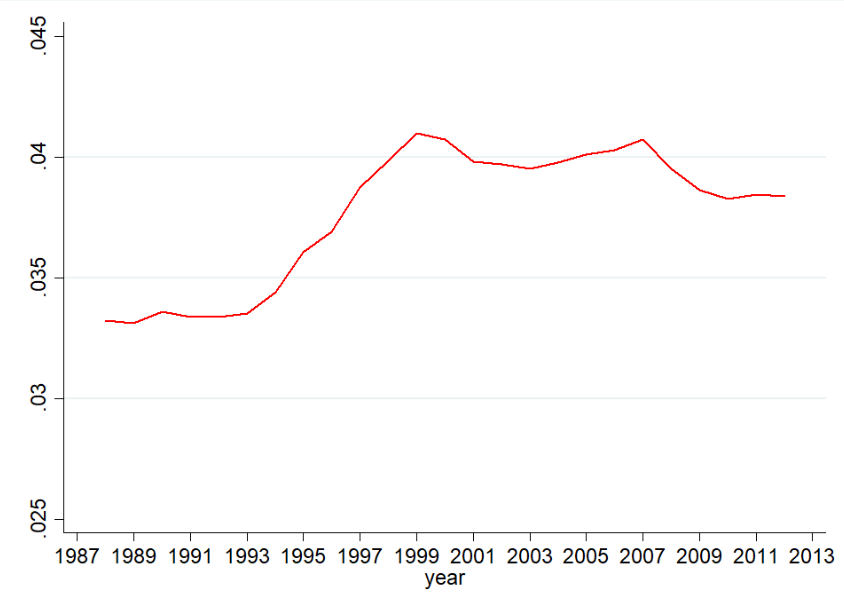
[8] Both components are part of the operating expenses of the firm. COGS could be viewed as the total variable cost of production.

[9] The COGS appears in the denominator of both the dependent variable and regressor. In section 1.9.4 of Appendix, I implement an alternative specification $log\,(XAD_{is}) = \rho_0\ log(\mu_{is}) + \rho_1 log(COGS_{is}) + \delta_s + \epsilon_{is}$ and find the estimated ρ_0 is similar to the estimated ρ in the main regression.

[10] Ptok et al. (2018) collect marketing data from an alternative source *Advertising Age* and compare it to SGA based marketing estimates. As they mention in the paper, "ADV offers a good measure of advertising spending and a partial measure of total marketing spending, but SGA fails to capture marketing spending or any of its subconstructs."

[11] In section 1.9.6 of Appendix, I estimate the marketing spending using SGA expenses. The SGA-based marketing measure fails to generate the co-movement pattern in section 1.3.

Figure 1.1: Aggregate marketing trend in the US



Notes: This figure plots the ratio of marketing expenditure to GDP in the US since 1987.

1.3 Empirical Findings

1.3.1 Aggregate trend of marketing intensity

Figure 1.1 shows the aggregate marketing intensity in the US. Since the mid-1990s, there is an upward trend in marketing spending in the US from about 3.3 percent of GDP to 4 percent. The rising trend vanishes after the 2000s.[12]

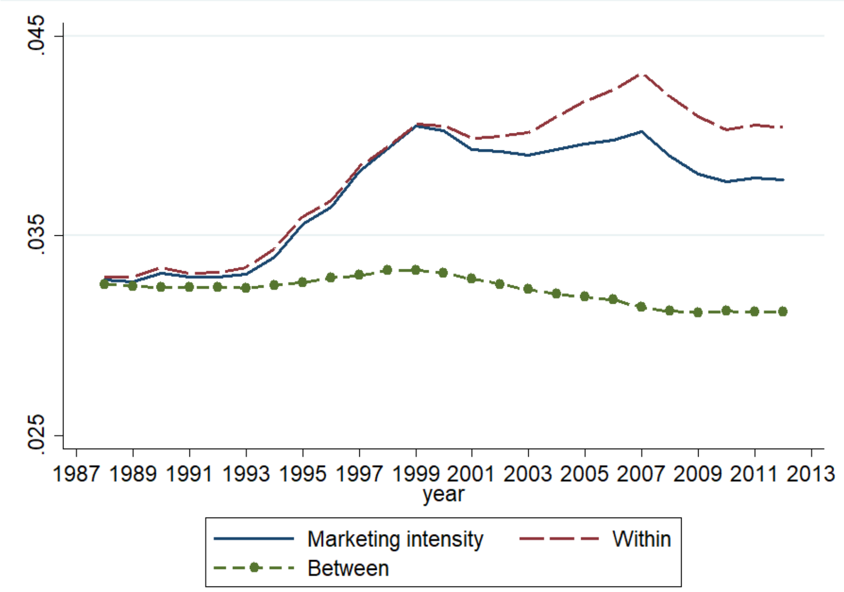
Aggregate marketing intensity is output weighted industry-level marketing intensity,

$$\frac{M_t}{Y_t} = \sum_s \frac{M_{st}}{Y_t} = \sum_s \frac{Y_{st}}{Y_t}\frac{M_{st}}{Y_{st}}$$

where M_{st} and Y_{st} are marketing spending and output for industry s at year t. The change in aggregate marketing intensity could be driven by either within-industry change in marketing intensity or between-industry reallocation of output. To understand the source of the change, I decompose the aggregate marketing intensity into a *between* and *within* component. For the *between* component, I fix the marketing intensity of each industry as in 1987 and allow

[12] In section 1.9.1 of Appendix, I plot the decomposition of marketing intensity into three components— advertising, marketing service, and own-account marketing labor compensation. In the mid-1990s, all three components of marketing spending were rising. However, since 2000, advertising intensity was declining while marketing service and own-account components were rising. This reflects the shift from *outbound* marketing towards *inbound* marketing.

Figure 1.2: Marketing expenditure decomposition



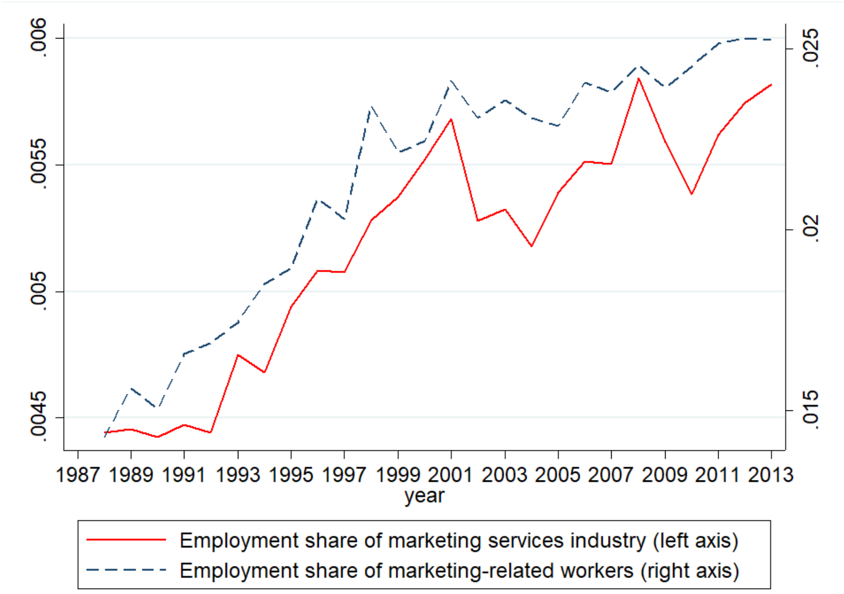
Notes: This figure decomposes the aggregate marketing intensity into two components: between-industry component and within-industry component. The solid line shows the aggregate marketing trend in the US. The *within* component shows the aggregate marketing intensity when fixing the industry value added share as in 1987. The *between* component shows the aggregate marketing intensity if I fix industry marketing intensity as in 1987.

for reallocation of output. On the other hand, for the *within* component, I keep the output share as in 1987 and allow for change of within-industry marketing intensity.

Figure 1.2 shows the result of this decomposition. The rise in the aggregate marketing intensity is driven by the *within* component. The *within* component follows closely the aggregate marketing intensity before 2000. After 2000, there is a growing discrepancy between the two. The *within* component keeps rising until the burst of financial crisis whereas the aggregate marketing intensity flattens out. The discrepancy can be explained by the change of *between* component. Before 2000, the change of this component is negligible. However, it is on a decreasing trend after 2000. As a result, the stagnation of aggregate marketing intensity since 2000 is due to between-industry reallocation.

Returns to marketing in the Internet era. Thus far, I show that since the mid-1990s, there is an increasing trend in marketing intensity in the US. The reason behind this change could be due to technology. According to the Wikipedia page on the history of marketing, the key innovation in the mid-1990s that influenced marketing practice is the rise of customer relationship management (CRM) technology and search engines. These innovations expand the marketing opportunities for the firms and increase the returns to market-

Figure 1.3: Employment shares of marketing-related activities in the US



Notes: This figure plots the employment shares of marketing services industry (left axis) and marketing-related workers (right axis). The marketing-related workers contain marketing labor in own-account component and workers in the marketing services industry.

ing spending. For instance, CRM software digitizes processes and automates tasks to improve the effectiveness of customer relationship management. A recent survey by CMO shows that in the year 2018, more than 40% of the marketing spending is on digital marketing. Besides, the availability of a vast stream of digital data presents opportunities for firms to identify customers who would be potentially interested in the product and hence increases the effectiveness of marketing. Goldfarb and Tucker (2011) find that the effectiveness of marketing depends on the advertiser's ability to collect data on web users.[13] Consistent with the technological interpretation of the rise in marketing intensity, Haskel and Westlake (2018) also argue that the current wave of digital technologies has made production more scalable.[14] Higher returns to marketing should reallocate labor towards marketing activities. In Figure 1.3, I plot the employment shares of marketing services industry and marketing-related occupations. The result implies there is a significant reallocation of labor towards marketing-related activities.

[13] See Goldfarb (2014) for a review of online marketing literature.

[14] In Table 1.14 of Appendix, I classify the industries into two groups: IT-intensive and non IT-intensive. I show that there is a larger increase in the marketing intensity for IT-intensive sector.

1.3.2 Marketing-production cost ratio and markup

In this section, I show that marketing spending at the firm level is related to market power.[15] Figure 1.4 is a binscatter plot between the logarithm of MPCR and the logarithm of markup. The figure implies a positive relationship between the markup and the MPCR of a firm. Firms with higher markups tend to spend more on marketing relative to production. This is consistent with the view that firms spend on marketing to differentiate from competitors and gain market power.

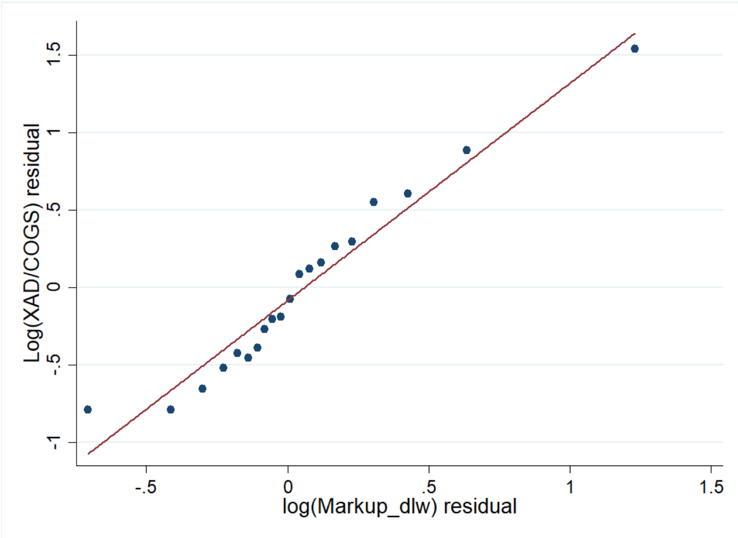
What is striking is that the MPCR-markup elasticity is co-moving with aggregate marketing intensity. Figure 1.5 presents the same binscatter plot but for two periods: before 1995 and after 1995. Comparing two sub-periods, the elasticity is higher after 1995. Indeed the elasticity between MPCR and markup increases by about 25 percent after 1995 as shown in the left panel of Figure 1.6. In the right panel of Figure 1.6, I plot the regression coefficient from year-on-year regressions. The level of *cross-sectional* elasticity between MPCR and markup co-moves almost perfectly with aggregate marketing intensity as in Figure 1.1. In the section below, I propose a model that could explain not only the correlation between MPCR and markup but more importantly the co-movement pattern between the MPCR-markup elasticity and aggregate marketing intensity.

1.4 Model

In this section, I present the model with heterogeneous firms and variable markups where firms engage in marketing to signal their quality. Section 1.4.1 and 1.4.2 discuss the model setup and resource allocation with complete information. In section 1.4.3, I show that, in the presence of information frictions, signaling competition between firms generates the positive correlation between MPCR and markup. More importantly, technical changes that induce higher returns to marketing generate co-movement between aggregate marketing intensity and MPCR-markup elasticity. Finally, section 1.4.4 characterizes how does the information frictions and markup dispersion affect aggregate productivity in the model.

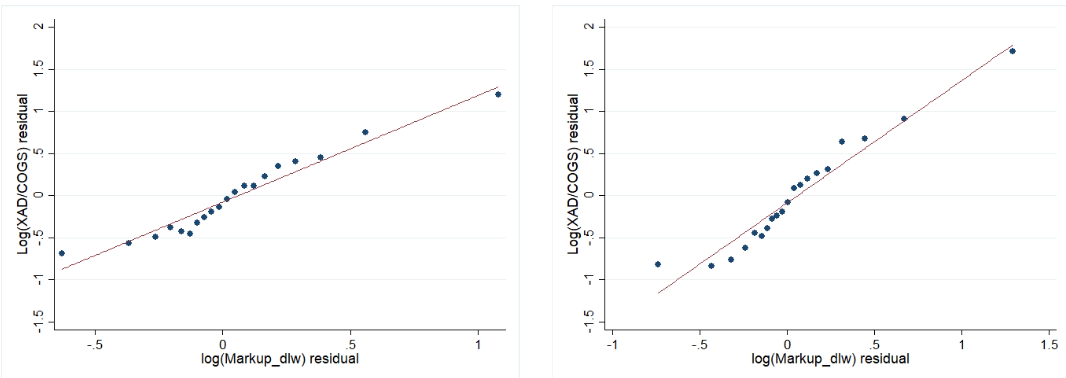
[15] In the main text, I estimate the markup by assuming an industry-specific Cobb-Douglas production function. In section 1.9.5 of Appendix, I show the results using Translog production function. The choice of production function does not change the empirical patterns documented in this section.

Figure 1.4: Binscatter plot between log(MPCR) and log(markup)



Notes: This binscatter plot shows the correlation between markup and marketing-production cost ratio. Both variables are residualized using industry-year fixed effects. Marketing cost is measured by advertising expense and production cost is measured by $COGS$.

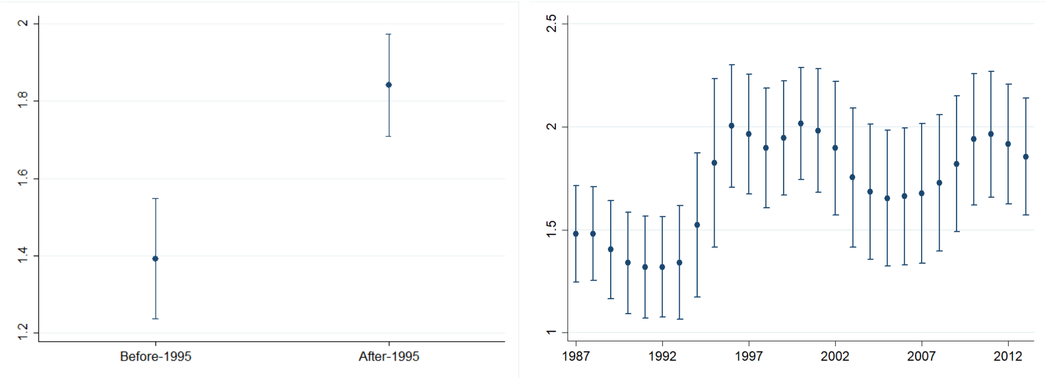
Figure 1.5: Binscatter plot between log(MPCR) and log(markup): before and after mid 1990s



(a) Before 1995 (b) After 1995

Notes: The left (right) panel shows a binscatter plot between markup and marketing-production cost ratio before (after) 1995. Both variables are residualized using industry-year fixed effects. Marketing cost is measured by advertising expenditure and production cost is measured by $COGS$.

Figure 1.6: MPCR-markup elasticity



Notes: This figure shows the elasticity between log(MPCR) and log(markup). The elasticity is obtained from a regression where I regress log(MPCR) on log(markup) and industry-year fixed effects. The left panel shows the result for two separate regressions: before 1995 and after 1995. The right panel shows the result for year-on-year regressions. For each year, I conduct the regression by combining data within a three-year moving window.

1.4.1 Setup

The economy has K industries. A representative consumer chooses consumption goods from all industries and all intermediate producers and supplies L units of labor inelastically. Intermediate firms are selling differentiated products with heterogeneous quality. The representative agent chooses consumption across all intermediate goods $\{C_k(\omega)\}_{k,\omega}$ to maximize the aggregate consumption

$$C = \prod_{k=1}^{K} (C_k)^{\frac{1}{K}} \tag{1.1}$$

subject to

$$\int_{\omega} \Upsilon\left(\frac{A_k(\omega)C_k(\omega)}{C_k}\right) d\omega = 1 \tag{1.2}$$

$$\sum_k \int_{\omega} P_k(\omega)C_k(\omega)d\omega = WL + \Pi. \tag{1.3}$$

The aggregate consumption C depends on the industry-level consumption good C_k, which is created by combining the varieties within each industry using the *Kimball aggregator* as in equation 1.2. $A_k(\omega)$ is the taste shifter of variety ω. Aggregator Υ is strictly increasing ($\Upsilon' > 0$) and concave ($\Upsilon'' < 0$) and satisfies $\Upsilon(1) = 1$.[16] The budget constraint, equation 1.3, shows that total spending on consumption should be equal to aggregate labor income WL plus aggregate profit Π from producers. W is the nominal wage rate and C is normalized as

[16]The aggregator nests the standard CES aggregator, i.e., $\Upsilon(q) = q^{\frac{\sigma-1}{\sigma}}$.

the numeraire good.

The taste shifter $A_k(\omega)$ determines the attractiveness of each variety. I assume it is a weighted-average of quality $Q_k(\omega)$ and appeal $\Phi_k(\omega)$ of variety ω,

$$A_k(\omega) = Q_k(\omega)^{u_k} \Phi_k(\omega)^{1-u_k}. \tag{1.4}$$

For each variety ω, $Q_k(\omega)$ is exogenously drawn from some quality distribution, while the appeal $\Phi_k(\omega)$ is endogenous where intermediate firms can spend on marketing to increase the appeal. I allow the weight $u_k \in [0, 1]$ to be industry-specific. For some industries like cosmetics and clothing, u_k would be lower, i.e., the agent cares more about the appeal component in the taste shifter, which in turns provides incentives for firms to spend more on marketing. In addition to the endogenous nature of the appeal, I assume appeal has an information advantage. Consumers cannot observe quality $Q_k(\omega)$ of a firm, but they can observe appeal $\Phi_k(\omega)$ instead. Thus, it is possible for firms to signal their quality using appeal.[17]

Following Klenow and Willis (2016), Gopinath and Itskhoki (2010), Edmond et al. (2018), I consider the following specification of function $\Upsilon(q)$

$$\Upsilon(q) = 1 + (\sigma - 1)exp\left(\frac{1}{\epsilon}\right)\epsilon^{\frac{\sigma}{\epsilon}-1}\left[\Gamma\left(\frac{\sigma}{\epsilon}, \frac{1}{\epsilon}\right) - \Gamma\left(\frac{\sigma}{\epsilon}, \frac{q^{\frac{\epsilon}{\sigma}}}{\epsilon}\right)\right] \tag{1.5}$$

where $\sigma > 1$, $\epsilon > 0$ and Γ is upper incomplete gamma function

$$\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt. \tag{1.6}$$

This specification permits a simple characterization of the dependence between demand elasticity and firm size. Suppose there is no information friction, i.e., both quality and appeal are observable. The FOC of the consumer implies the following demand function for intermediate good producers

$$P_k(\omega) = \Upsilon'(q_k(\omega)) A_k(\omega) D_k^{-1} P_k \tag{1.7}$$

where $q_k(\omega) = \frac{A_k(\omega)C_k(\omega)}{C_k}$ is related to the size of a firm with variety ω and $D_k = \int \Upsilon'(q_k(\omega)) q_k(\omega) d\omega$ is a demand shifter for varieties within industry k.[18] Firms spend on marketing to increase the appeal, which in turns shifts up the

[17] In the absence of information frictions, marketing can be interpreted as a generic type of quality improvement using labor inputs.

[18] The relative size of a firm with variety ω is $\frac{P_k(\omega)C_k(\omega)}{P_k C_k} = \Upsilon'(q_k(\omega)) q_k(\omega) D_k^{-1}$.

demand schedule for the firm. The demand elasticity for variety ω is

$$\varepsilon_k(\omega) = -\frac{\partial log(C_k(\omega))}{\partial log(P_k(\omega))} = -\frac{\Upsilon'(q_k(\omega))}{\Upsilon''(q_k(\omega))q_k(\omega)} = \sigma q_k(\omega)^{-\frac{\epsilon}{\sigma}}. \tag{1.8}$$

The ratio $\frac{\epsilon}{\sigma}$ is referred as *superelasticity* in the literature. It implies that the demand elasticity is decreasing with firm size. When $\epsilon = 0$, Kimball aggregator corresponds to CES production function with $\Upsilon(q) = q^{\frac{\sigma-1}{\sigma}}$. Atkeson and Burstein (2008) proposes an alternative way to generate variable demand elasticity. Their model features oligopoly competition with nested CES demand. In the robustness section of Appendix, I show that the results could be extended to the model of oligopolistic competition.

For intermediate firms, each firm produces one variety and draws quality $Q_k(\omega)$ from a distribution G_k. Labor is the only input of production. A firm in industry k with variety ω uses production labor $L_{kp}(\omega)$ produce output according to $Y_k(\omega) = L_{kp}(\omega)^{\alpha}$.[19] All the demand of a firm's output comes from the representative consumer, implying that $Y_k(\omega) = C_k(\omega)$. The firm also hires marketing labor $L_{ka}(\omega)$ to increase the appeal according to

$$\Phi_k(\omega) = Q_k(\omega)L_{ka}(\omega)^{\beta} \tag{1.9}$$

where β measures the returns to scale of marketing.

There are two determinants of firm appeal $\Phi_k(\omega)$, which is product quality $Q_k(\omega)$ and marketing labor input $L_{ka}(\omega)$. As documented by Bagwell (2007), marketing helps the firms to attract the consumers and increase the firm appeal, i.e., $\beta > 0$, but its effectiveness is diminishing, i.e., $\beta < 1$. With the rise of information and communication technology, there are more channels for firms to reach consumers. As I will discuss below, an increase in β would generate co-movement between MPCR-markup elasticity and aggregate marketing intensity.

I assume the marginal product of marketing labor is higher for high $Q_k(\omega)$ firms. For a given level of appeal Φ, it is less costly for high-quality firms to achieve. This assumption is crucial as firms can credibly signal their unobserved quality via appeal.

In the model, firms engage in marketing for three reasons. Firstly, as consumers cannot observe the quality of firms, firms engage in marketing to signal their

[19] In the model, the heterogeneity of firms comes from the quality variation instead of productivity. This is consistent with recent literature that studies the sources of firm heterogeneity, see Hottman et al. (2016) for example.

quality $Q_k(\omega)$. Marketing provides information about the quality of the firms which is consistent with the *informative view* of marketing.[20] Secondly, even without information frictions, marketing would increase the firm appeal $\Phi_k(\omega)$ as consumers prefer to consume a more appealing product. This corresponds to the *complementary view* of marketing, where marketing enters utility directly in a fashion that is complementary with the consumption of advertised product.[21] Lastly, spending on marketing helps the firm gain market power. Since marketing conveys direct information about firm quality, it will determine the firm size and hence markup distribution in the model.[22]

1.4.2 Complete information

Before analyzing the model with information frictions, I first solve the complete information allocation. The intermediate firm in industry k with variety ω maximizes the profit by choosing $\Phi_k(\omega)$ and $P_k(\omega)$

$$\Pi_k(\omega) = P_k(\omega)C_k(\omega) - WC_k(\omega)^{\frac{1}{\alpha}} - W\left(\frac{\Phi_k(\omega)}{Q_K(\omega)}\right)^{\frac{1}{\beta}} \tag{1.10}$$

subject to equation 1.7. The FOC implies

$$\frac{WC_k(\omega)^{\frac{1}{\alpha}}}{P_k(\omega)C_k(\omega)} = \frac{WL_{kp}(\omega)}{P_k(\omega)C_k(\omega)} = \frac{\alpha}{\mu_k(\omega)} \tag{1.11}$$

$$\frac{W\left(\frac{\Phi_k(\omega)}{Q_k(\omega)}\right)^{\frac{1}{\beta}}}{P_k(\omega)C_k(\omega)} = \frac{WL_{ka}(\omega)}{P_k(\omega)C_k(\omega)} = \frac{\beta(1-u_k)}{\mu_k(\omega)} \tag{1.12}$$

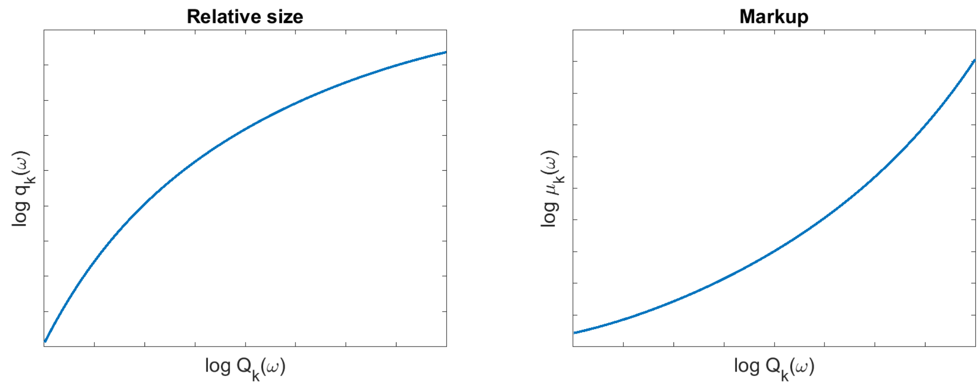
where $\mu_k(\omega) = \frac{\varepsilon_k(\omega)}{\varepsilon_k(\omega)-1} = \frac{\sigma}{\sigma - q_k(\omega)^{\frac{\epsilon}{\sigma}}}$ is markup for variety ω. The revenue share of production labor for a firm in industry k with variety ω is the ratio between output elasticity of production labor α and markup of the firm. On the other hand, the revenue share of marketing labor also depends on the industry-specific taste parameter u_k. In industries where consumers value more of the appeal of a product, the revenue share of marketing labor is higher. Combining

[20]The idea is first presented by Nelson (1974).

[21]As Stigler and Becker (1977) put it, when a firm advertises more, its product becomes more attractive to the consumer, since "the household is made to believe - correctly or incorrectly - that it gets a greater output of the commodity from a given input of the advertised product."

[22]Another view of marketing is the *persuasive view*, where marketing alters consumers' tastes and creates spurious product differentiation. Bertrand et al. (2010) implement a field experiment where a consumer lender sends direct-mail with randomized advertising content. They find that the advertising content which contains no information, including a photo of an attractive woman for instance, can significantly increase demand.

Figure 1.7: Quality, Firm Size and Markup



the two conditions, we can find that MPCR is not correlated with markup within each industry

$$\frac{WL_{ka}(\omega)}{WL_{kp}(\omega)} = \frac{\beta(1-u_k)}{\alpha}. \tag{1.13}$$

I summarize the results in the following proposition.

Proposition 1. *When there is no information friction, within each industry, the MPCR is a constant across firms and the MPCR-markup correlation is zero.*

Figure 1.7 illustrates the relationship between quality, firm size and markup in the model. Within each industry, a higher quality firm would have a larger relative size and higher markup. There is no misallocation of labor across activities (between production and marketing). The only source of misallocation comes from the markup dispersion as in the static model of Edmond et al. (2018), where they use the same Kimball aggregator to study the welfare cost of markup.[23]

1.4.3 Incomplete information

In an environment with incomplete information, higher quality firms are willing to increase appeal in order to differentiate from low quality firms. The equilibrium concept would be a perfect Bayesian equilibrium (PBE). Intermediate producers try to signal the quality of their products by using the joint signals of price and appeal $\{P_k(\omega), \Phi_k(\omega)\}$.[24] Consumers cannot observe the quality

[23] In their setting, there is no information frictions in the product market and no role for marketing. However, this chapter studies the role of marketing in a world with information frictions.

[24] In addition to the appeal, I assume consumers observe firms' prices, and the prices will be used as an additional signal. Equivalently, firms can choose signals of quantity and

of a variety, but could only observe the joint signals $\{P_k(\omega), \Phi_k(\omega)\}$. They form beliefs $b\left(P_k(\omega), \Phi_k(\omega)\right)$ and demand $C\left(P_k(\omega), \Phi_k(\omega)\right)$ units of output.[25] Formally,

Definition 1. *A PBE is a set of strategies for the consumers $\{C\left(P_k(\omega), \Phi_k(\omega)\right)\}_{k,\omega}$ and firms $\{P_k(\omega), \Phi_k(\omega)\}_{k,\omega}$, and posterior beliefs $\{b\left(P_k(\omega), \Phi_k(\omega)\right)\}_{k,\omega}$ such that:*

1. *Intermediate firms choose joint signals $\{P_k(\omega), \Phi_k(\omega)\}_{k,\omega}$ to maximize their profit,*
2. *Consumers choose demand $\{C\left(P_k(\omega), \Phi_k(\omega)\right)\}_{k,\omega}$ to maximize aggregate consumption,*
3. *$b(P_k(\omega), \Phi_k(\omega))$ is derived from the equilibrium strategies using Bayes' rule whenever possible.*[26]

In the model, firms can credibly signal quality information since the single-crossing property of profit function holds.[27] Higher quality firms are willing to increase marketing spending to signal their quality.

Lemma 1. *With information frictions, the marginal rate of substitution of signal $\Phi_k(\omega)$ for demand $C_k(\omega)$ is strictly decreasing with quality of firms when $P_k(\omega) > \frac{1}{\alpha} W C_k(\omega)^{\frac{1}{\alpha}-1}$.*

Lemma 1 shows the slope of iso-profit curve is lower for high-quality firms in the space of signal $\Phi_k(\omega)$ and demand $C_k(\omega)$.[28] Thus, it is possible for higher quality firms to differentiate from lower quality firms by increasing appeal Φ in return for an expansion of demand.

One drawback of PBE is flexibility in choosing the off-equilibrium beliefs, which would result in multiplicity of equilibria. Thus, I focus on the *least-cost separating equilibrium* (LCSE). The LCSE is of particular interest because it is

appeal $\{Y_k(\omega), \Phi_k(\omega)\}$. Consumers form belief $b\left(Y_k(\omega), \Phi_k(\omega)\right)$ after observing quantity and appeal, and generate inverse demand function $P\left(Y_k(\omega), \Phi_k(\omega)\right)$. Following the argument in this section, the same LCSE can be established. Therefore, the allocation of resources across firms and activities remain unchanged.

[25] The approach is drawn from Bagwell and Ramey (1988) and Bagwell (2007), where they study a similar signaling game with multiple signals. However, their models are analytical as there are only two types of firms, and they do not specify the demand side of the model.

[26] Beliefs are updated for signals on the equilibrium path.

[27] With multiple signals, the single-crossing condition will hold as long as the MRS of one signal for response is strictly decreasing with quality. See Ramey (1996) for details.

[28] Since each firm produces one differentiated variety and exhibits some market power, a profit-maximizing firm will always set price above marginal cost.

unique and also the one selected when we use standard refinement tool.[29,30] Suppose each firm draws their quality $Q_k(\omega)$ from a quality grid $Q_n = exp(n\Delta)$ for integers $n \in \{1, 2, ..., N\}$. In the LCSE, the lowest quality Q_1 type firm in each industry is "found out". They would select signal $\{P_k(Q_1), \Phi_k(Q_1)\}$ to maximize profit

$$\Pi_k(P_k(Q_1), \Phi_k(Q_1), b = Q_1, Q_1) \tag{1.14}$$

The FOC implies same labor allocation as in equation 1.11 and 1.12.

Suppose $\{P_k(Q_i), \Phi_k(Q_i)\}$ have been specified for type Q_i firms for $i = 2, ..., n-1$. For type Q_n firms, they will choose $\{P_k(Q_n), \Phi_k(Q_n)\}$ to maximize profit

$$\Pi_k(P_k(Q_n), \Phi_k(Q_n), b = Q_n, Q_n) \tag{1.15}$$

subject to the *incentive compatibility* (IC) condition

$$\Pi_k(P_k(Q_i), \Phi_k(Q_i), b = Q_i, Q_i) \geq \Pi_k(P_k(Q_n), \Phi_k(Q_n), b = Q_n, Q_i) \tag{1.16}$$

where $i \in \{2, ...n-1\}$. Given the cost structure of the model, the only IC condition could potentially bind is the one for quality Q_{n-1}. The intuition is that if mimicking Q_n firm is too costly for Q_{n-1} type, then it would be even more costly to do so for lower quality firms.

The solution of LCSE gives rise to the following allocation

$$\frac{WC_k(Q_n)^{\frac{1}{\alpha}}}{P_k(Q_n)C_k(Q_n)} = \frac{WL_{kp}(Q_n)}{P_k(Q_n)C_k(Q_n)} = \frac{\alpha}{\mu_k(Q_n)} \tag{1.17}$$

$$\frac{W\left(\frac{\Phi_k(Q_n)}{Q_n}\right)^{\frac{1}{\beta}}}{P_k(Q_n)C_k(Q_n)} = \frac{WL_{ka}(Q_n)}{P_k(Q_n)C_k(Q_n)} = \frac{\beta(1-u_k)}{\mu_k(Q_n)}X_k(Q_n) \tag{1.18}$$

where information distortion $X_k(Q_n) = \left(\frac{Q^\star_{kn}}{Q_n}\right)^{\frac{1}{\beta}} \geq 1$ and $Q^\star_{kn}$ is the virtual type of Q_n in industry k. When the IC condition is binding for quality Q_n firms in industry k, they would choose an allocation that exactly as if there were no information frictions, but her quality is higher. To see this point, we rearrange equation 1.18 and find

$$\frac{W\left(\frac{\Phi_k(Q_n)}{Q^\star_{kn}}\right)^{\frac{1}{\beta}}}{P_k(Q_n)C_k(Q_n)} = \frac{\beta(1-u_k)}{\mu_k(Q_n)} \tag{1.19}$$

[29]See Cho and Kreps (1987), Cho and Sobel (1990), and Ramey (1996) for example.

[30]LCSE also has a constrained efficiency property: among all the separating equilibria, it involves the least separation cost due to signaling. Given the quality distribution of firms, LCSE provides *lower* bound for the productivity losses due to information frictions.

where $Q^{\star}_{kn} > Q_n$. Thus, the allocation is also refered as *complete-information distortion* allocation in the literature.[31] Comparing equation 1.18 and equation 1.12, the presence of information frictions distorts the labor allocation and reallocate labor towards marketing. On the other hand, when the IC condition is not binding for Q_n firms, the virtual type would be the same as the quality of the firm, i.e., the Q_n firms allocate labor efficiently between activities.

Whether information frictions change the labor allocation depends on the value of quality information for the consumers. Consider the case when u_k is small. Recall u_k is the weight on the firm's quality in taste shifter. In this case, conditional on knowing the appeal Φ of each firm, quality information is of little value to the consumer. Thus, even with information frictions, complete information allocation is also LCSE. On the other hand, suppose u_k is relatively large, and quality information becomes valuable. Complete information allocation would no longer be incentive-compatible, thus information frictions lead to misallocation. Following lemma summarizes the results.[32,33]

Lemma 2. *When quality gap Δ is small, within each industry k, the complete information allocation would (not) be incentive-compatible when the weight of quality information u_k is small (large).*

Combining equation 1.17 and 1.18, the MPCR depends on the information distortion

$$\frac{WL_{ka}(Q_n)}{WL_{kp}(Q_n)} = \frac{\beta(1-u_k)}{\alpha} X_k(Q_n).$$

As demand becomes more and more inelastic for higher quality firms, the incentive to mimic higher quality firms would be higher. Thus, for a given industry, if the IC conditions are binding for lower quality firms, then they will also be binding for higher quality firms, and information distortion X increases with quality. Therefore, the MPCR would be positively correlated with markup.

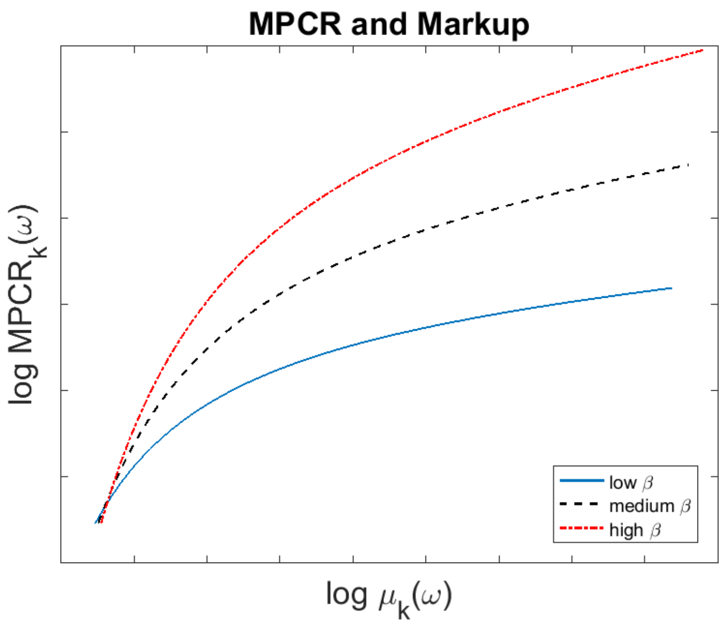
Proposition 2. *Within each industry k, demand is less elastic for higher quality firms. If $X_k(Q_i) > 1$ and $Q_j > Q_i$, then $X_k(Q_j) > X_k(Q_i) > 1$. Thus,*

[31] In Appendix, I show the derivation that LCSE has the feature of a *complete-information distortion* allocation. Also see Ramey (1996) for *complete-information distortion equilibrium.*

[32] The proof of the lemma is established in Appendix.

[33] Size of Δ is another determinant of whether the complete information allocation is incentive-compatible. As Δ becomes larger, the quality gap between the two types of firms becomes larger. Mimicking a firm with much higher quality is too costly for the low-quality firms. Thus, complete information allocation would be incentive-compatible.

Figure 1.8: MPCR and markup: different returns to marketing



MPCR is positively correlated with markup

$$MPCR_k(Q_n) = \frac{\beta(1-u_k)}{\alpha} X_k(Q_n) \propto \mu_k(Q_n).$$

The positive MPCR-markup correlation established in Proposition 2 only depends on fact that demand becomes less elastic for higher quality firms. Thus, it is robust to alternative specifications of Kimball aggregator and other demand systems such as the model with oligopolistic competition in Atkeson and Burstein (2008).[34]

Figure 1.8 plots the MPCR and markup across firms within an industry conditional on binding IC conditions. Firms with higher markup tend to spend more on marketing, and the MPCR-markup elasticity is increasing with the returns to marketing. The intuition is that when returns to marketing is higher, mimicking higher quality firms becomes less costly. More intense signaling competition among firms generates a larger dispersion of MPCR across firms and thus a larger MPCR-markup elasticity.[35]

To sum up, when there are information frictions, firms spend on marketing to signal their quality, which generates the positive correlation between MPCR and markup. Technical changes that increase the returns to marketing would lead to a rise of MPCR-markup elasticity and aggregate marketing intensity.

[34] The proof of Proposition 2 is established in Appendix.

[35] Since marketing spending is small compared to production cost, higher returns to marketing only generate a small variation in markup dispersion.

Besides, the existence of information frictions results in misallocation of labor between activities compared to complete information allocation. Both the MPCR-markup elasticity and the extent of misallocation depend on whether the quality information is valuable for the consumer or not. If the consumer only cares about the appeal of a product, then spending on marketing would be efficient. However, if consumer cares about quality information, signaling would result in misallocation of resources and generate co-movement between MPCR-markup elasticity and returns to marketing technology. In the calibration section below, I use the distribution of marketing intensity across industries to determine the size and variation of the importance of quality information u_k.

1.4.4 Resource allocation and aggregation productivity

So far I characterize the optimality conditions for each intermediate producer in the LCSE. I can now derive an aggregate production function for this economy to show how the markup and information distortion determine aggregate productivity. I will proceed in two steps: (i) within each industry k, combining cross-sectional distribution of firm size $q_k(Q_n)$ and the extent of information frictions $X_k(Q_n)$ yields an industry-level production function, (ii) then I aggregate industry-level production function to derive an aggregate production function.

Suppose the joint distribution of $\{q_k(Q_n), X_k(Q_n)\}$ for each industry has been solved. Let $M_k(Q_n)$ be the mass of firms with quality Q_n in industry k. Combining each intermediate firm's choices yields an industry-level production function which transforms industry-level labor input into industry-level consumption C_k

$$C_k = Z_k L_{kp}^{\alpha} L_{ka}^{\beta(1-u_k)} \tag{1.20}$$

where Z_k is industry-level efficiency, $L_{kp} = \sum_n L_{kp}(Q_n)M_k(Q_n)$ is the industry-level production labor input and $L_{ka} = \sum_n L_{ka}(Q_n)M_k(Q_n)$ is the industry-level marketing labor input. Let μ_k be industry-level markup and X_k is the industry-level information distortion. Both of them are implicitly defined by

$$\frac{WL_{kp}}{P_kC_k} = \frac{\alpha}{\mu_k} \tag{1.21}$$

$$\frac{WL_{ka}}{P_kC_k} = \frac{\beta(1-u_k)X_k}{\mu_k} \tag{1.22}$$

Some algebra shows industry-level markup μ_k and information distortion X_k

are the production cost-weighted average of of firm level markup and information distortion in industry k.[36]

$$\mu_k = \sum_n \mu_k(Q_n) \frac{WL_{kp}(Q_n)}{WL_{kp}} M_k(Q_n) \tag{1.23}$$

$$X_k = \sum_n X_k(Q_n) \frac{WL_{kp}(Q_n)}{WL_{kp}} M_k(Q_n) \tag{1.24}$$

Recall that $q_k(Q_n) = \frac{Q_n L_{ka}(Q_n)^{\beta(1-u_k)} L_{kp}(Q_n)^{\alpha}}{C_k}$. Together with the optimality condition of intermediate firms 1.17 and 1.18, industry-level efficiency is a weighted firm level quality

$$Z_k = \left[\sum_n \left(\frac{q_k(Q_n)}{Q_n} \right)^{\frac{1}{\alpha+\beta(1-u_k)}} \left(\frac{X_k}{X_k(Q_n)} \right)^{\frac{\beta(1-u_k)}{\alpha+\beta(1-u_k)}} M_k(Q_n) \right]^{-(\alpha+\beta(1-u_k))}. \tag{1.25}$$

The weight on quailty Q_n firms depends on relative size $q_k(Q_n)$ and the dispersion of information distortion $X_k(Q_n)$. Given the industry-level labor input L_{kp} and L_{ka}, Z_k summarizes how these inputs are allocated across producers. In a world with complete information, $\frac{X_k}{X_k(Q_n)} = 1$. Then industry-level efficiency Z_k only depends on the distribution of $q_k(Q_n)$.

Then we move on to study the cross-industry allocation of labor. Given L units of aggregate labor supply, together with equation 1.21 and 1.22, the labor allocation follows

$$L_{kp} = \frac{\alpha \mu_k^{-1}}{\sum_k \mu_k^{-1} (\alpha + \beta(1-u_k)X_k)} L = \psi_{kp} L \tag{1.26}$$

$$L_{ka} = \frac{\beta(1-u_k) X_k \mu_k^{-1}}{\sum_k \mu_k^{-1} (\alpha + \beta(1-u_k)X_k)} L = \psi_{ka} L \tag{1.27}$$

where industry-level labor allocation $\{\psi_{kp}, \psi_{ka}\}$ depends on markup μ_k, information distortion X_k, and preference parameter u_k. Compared to complete information allocation, the existence of information frictions X_k results in misallocation of labor across activities.

Combining industry-level production function and labor allocation rule, the aggregate consumption is given as

$$C = ZL^{\alpha+\beta(1-u)} \tag{1.28}$$

[36] Equivalently, the sales-weighted harmonic average of firm level markup and information distortion.

where u is the average weight of quality information for consumer, and $1-u$ is the average weight of appeal. Aggregate returns to labor input, $\alpha+\beta(1-u)$, is the sum of returns to production labor (α) and returns to marketing labor ($\beta(1-u)$). Z is *aggregate productivity* in the model which translates the aggregate labor input L into aggregate consumption C

$$Z = \prod_{k=1}^{K} \left(Z_k(\psi_{kp})^{\alpha}(\psi_{ka})^{\beta(1-u_k)}\right)^{\frac{1}{K}} \tag{1.29}$$

$$= \left\{\prod_{k=1}^{K}(Z_k)^{\frac{1}{K}}\right\}\left\{\prod_{k=1}^{K}\left((\psi_{kp})^{\alpha}(\psi_{ka})^{\beta(1-u_k)}\right)^{\frac{1}{K}}\right\}. \tag{1.30}$$

The aggregate productivity Z depends on two components. The first term $\{\prod_{k=1}^{K}(Z_k)^{\frac{1}{K}}\}$ is a geometric average of industry-level efficiency Z_k. It summarizes the *between-firm* allocative efficiency in the model. The second component $\{\prod_{k=1}^{K}((\psi_{kp})^{\alpha}(\psi_{ka})^{\beta(1-u_k)})^{\frac{1}{K}}\}$ is also a geometirc average of industry-level *between-activity* allocative efficiency. In the numerical section, I normalize $L=1$, then aggregate consumption would be identical to aggregate productivity.

Solution algorithm. From the aggregation results, the main challenge is to find the joint distribution $\{q_k(Q_n), X_k(Q_n)\}$. Once the within-industry allocation is solved, the cross-industry aggregation is straightforward.

Within each industry k, I first compute the complete information allocation $q_k^{CI}(Q_n)$. From equation 1.11 and 1.12, together with the definition of $q_k^{CI}(Q_n)$, we can find

$$\frac{\Upsilon'(q_k^{CI}(Q_n))q_k^{CI}(Q_n)}{\mu(q_k^{CI}(Q_n))} = \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta(1-u_k)}}\left(\frac{1}{\beta(1-u_k)}\right)^{\frac{\beta(1-u_k)}{\alpha+\beta(1-u_k)}} B_k\left(\frac{q_k^{CI}(Q_n)}{Q_n}\right)^{\frac{1}{\alpha+\beta(1-u_k)}} \tag{1.31}$$

where B_k is an industry statistic

$$B_k = \frac{W}{P_k D_k^{-1}} C_k^{\frac{1}{\alpha+\beta(1-u_k)}-1}. \tag{1.32}$$

Note in equation 1.31, B_k is the only unknown variable. Together with the following Kimball aggregator, we can solve for B_k and hence $q_k^{CI}(Q_n)$

$$\sum_n \Upsilon(q_k(Q_n))M_k(Q_n) = 1. \tag{1.33}$$

With the complete information allocation, we move on to check whether this allocation satisfies all the IC conditions. We start from the lowest quality Q_1 with $q_k^{CI}(Q_1)$ and $X_k(Q_1) = 1$. For $n > 1$, if Q_{n-1} has no incentives to mimic Q_n, then I let $q_k(Q_n) = q_k^{CI}(Q_n)$ and $X_k(Q_n) = 1$. However, if instead Q_{n-1} has incentives to mimic Q_n, then I compute $q_k(Q_n)$ and $X_k(Q_n)$ using the following conditions

$$
\begin{aligned}
\frac{\Upsilon'(q_k(Q_n))q_k(Q_n)}{\mu(q_k(Q_n))} = & \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta(1-u_k)}} \left(\frac{1}{\beta(1-u_k)}\right)^{\frac{\beta(1-u_k)}{\alpha+\beta(1-u_k)}} B_k \\
& \cdot \left(\frac{1}{X_k(Q_n)}\right)^{\frac{\beta(1-u_k)}{\alpha+\beta(1-u_k)}} \left(\frac{q_k(Q_n)}{Q_n}\right)^{\frac{1}{\alpha+\beta(1-u_k)}} \qquad (1.34)
\end{aligned}
$$

$$
\begin{aligned}
1 = & \frac{\Upsilon'(q_k(Q_n))q_k(Q_n)}{\Upsilon'(q_k(Q_{n-1}))q_k(Q_{n-1})} \\
& \cdot \frac{1-\alpha\mu(q_k(Q_n))^{-1} - \beta(1-u_k)X_k(Q_n)exp\left(\frac{\Delta}{\beta}\right)\mu(q_k(Q_n))^{-1}}{1-\alpha\mu(q_k(Q_{n-1}))^{-1} - \beta(1-u_k)X_k(Q_{n-1})\mu(q_k(Q_{n-1}))^{-1}} \qquad (1.35)
\end{aligned}
$$

Equation 1.34 is the counterpart of 1.31 when there are information frictions. Equation 1.35 indicates a binding IC condition for two consecutive types. Notice the solution to this procedure will result in an incentive-compatible allocation. We would repeat this procedure until we find an B_k that satisfies 1.33. With the joint distribution of $\{q_k(Q_n), X_k(Q_n)\}$, we could compute all the industry-level statistic μ_k, X_k, ψ_{ka} and ψ_{kp} and hence the aggregate productivity Z.

1.5 Calibration

In the model, the extent of information distortion depends on three sets of parameters: (i) returns to scale of marketing technology β, (ii) the relative importance of quality u_k and (iii) the amount of quality dispersion in each industry k. Following the literature, I assume the quality distribution G_k follows a power-law distribution with tail parameter ξ_k.[37]

[37] Using the quality grid $Q_n = exp(n\Delta)$, I assume the probability density function of G_k is given by $g_k(Q_n) = \frac{Q_n^{-(1+\xi_K)}}{\sum_n Q_n^{-(1+\xi_K)}}$.

In the data, I observe industry-level marketing intensity. Thus, if returns to marketing β varies with u_k across industries, we cannot separately estimate the two. To this end, I impose one identification assumption: different industries have access to a marketing technology with the same returns to scale, i.e., β is constant across industries. Intuitively, in reality, firms in different industries can use similar tools to do marketing. For instance, they could advertise their products via TV or on the Internet.[38] Then industry-level marketing intensity depends on the nature of industry u_k and the competitive environment within the industry ξ_k.

1.5.1 Assigned parameters

Panel A of Table 1.1 reports the choices of assigned parameters. I first set the returns to scale of production labor $\alpha = 0.9$.[39] For the *superelasticity* parameter $\frac{\epsilon}{\sigma}$, I use the benchmark estimate in Edmond et al. (2018), i.e., $\frac{\epsilon}{\sigma} = 0.14$. To be consistent with the data on marketing, I assume there are 47 industries in the economy. Finally, I assume there is unit mass of firms within each industry $M_k = \sum_n M_k(Q_n) = 1$. Then the only source of heterogeneity across industries is the dispersion of $\{u_k, \xi_k\}$.

1.5.2 Calibrated parameters

The remaining parameters are calibrated to match the data on marketing intensity and firm size distribution in the mid-1990s.[40] Specifically, I assume the weight of quality information u_k in each industry k is drawn from a Beta distribution with shape parameter $\{u_a, u_b\}$, and the tail parameter of quality distribution ξ_k is drawn from a Log-normal distribution with parameter $\{\xi_\mu, \xi_\sigma\}$. Then I jointly estimate the parameters $\{\beta, u_a, u_b, \xi_\mu, \xi_\sigma, \sigma\}$ to match a set of moment conditions documented in the data.[41]

[38] There are exceptions. For instance, legal restrictions are imposed on the marketing practice of tobacco products. See `https://truthinitiative.org/research-resources/tobacco-industry-marketing/what-do-tobacco-advertising-restrictions-look-today`. This concern is mitigated as our industry classification is broad. For example, tobacco belongs to industry *Food, beverage, and tobacco products*.

[39] The model generates a profit rate of 19% for the benchmark calibration. In the mid-1990s, the profit share measured by the ratio of net operating surplus to net value added is around 18% for the US corporate sector.

[40] For the data on firm size distribution, I use earliest available data in 1998. See `https://www.sba.gov/advocacy/firm-size-data`.

[41] I set the step size of quality ladder $\Delta = 0.02$ in the calibration. In Appendix, I show that a joint estimation of Δ yields similar results.

Table 1.1: Parameterization

Panel A: Assigned parameters

α	returns to scale of production labor	0.9
$\frac{\varepsilon}{\sigma}$	superelasticity of Kimball aggregator	0.14

Panel B: Calibrated parameters

		benchmark	high μ	high β
σ	average elasticity	13.52	7.48	13.83
β	returns to scale of marketing	0.054	0.056	0.071
u_a	shape parameter of Beta distr. u_k	0.047	0.060	0.042
u_b	shape parameter of Beta distr. u_k	0.009	0.012	0.010
ξ_μ	location parameter of Lognormal distr. ξ_k	0.28	0.19	0.31
ξ_σ	shape parameter of Lognormal distr. ξ_k	2.21	2.09	2.09
selected moments				
	marketing spending/GDP	0.03	0.03	0.04
	aggregate markup, μ	1.15	1.25	1.15

Aggregate markup. First, the average elasticity σ is calibrated to aggregate markup in the US. In the benchmark calibration, I set the aggregate markup to 1.15, which is consistent with estimates in recent literature on markup.[42]

Distribution of relative payroll. The second set of moment conditions involves the average and standard deviation of industry-level distribution of *relative payroll*. For each 6-digit sector s, we have information for the firms in about 4 employment-based size classes, which contains total payroll, number of firms and total employment. Within the sector s, I define the *relative payroll* as the ratio of the average payroll of firms in a given size class c to the average payroll of all firms in the industry

$$rel\_payroll_{c,s} = \frac{\frac{payroll_{c,s}}{N_{c,s}}}{\frac{\sum_c payroll_{c,s}}{\sum_c N_{c,s}}}$$

where $payroll_{c,s}$ and $N_{c,s}$ are the total payroll and number of firms for size class c in industry s. I pool this statistic within the industry k.[43] Then the distribution of *relative payroll* within industry k is given as

$$G_k(rel\_payroll < a) = \frac{\sum_{c,s\in k} N_{c,s} \cdot I(rel\_payroll_{c,s} < a)}{\sum_{c,s\in k} N_{c,s}},$$

which corresponds to the fraction of firms within industry k with relative

[42] I choose the midpoint of estimates in recent literature on markup. See De Loecker et al. (2018), Edmond et al. (2018) and Barkai (2017) for example.

[43] The industry-level corresponds to approximately 3 digit NAICS sector.

payroll smaller than a. The distribution of relative payroll G_k is a measure of firm size distribution for industry k. Intuitively, larger firms tend to have larger relative payroll. Besides, if there is no size heterogeneity across firms within each sector s, relative payroll would be one for all firms.

Panel A of Table 1.2 reports the average and standard deviation of industry-level distribution of relative payroll in the data. On average, about 66% of firms have *relative payroll* smaller than 0.5. About 82% of firms have payroll less than the industry average payroll. A little more than 1% of firms have payroll more than 10 times the industry average.[44]

Marketing intensity. The last set of moment conditions includes the aggregate marketing intensity and dispersion of marketing intensity across industries. The aggregate marketing intensity is around 3 percent of GDP before mid-1990s. Besides, Panel B of Table 1.2 shows that there is a large dispersion of marketing intensity across industries. For instance, the industry at 90 percentile of the distribution has a marketing intensity 80% higher than the median. Additionally, there is more dispersion in the bottom half of the distribution. Median-to-bottom ratio $\frac{p50}{p10}$ is 40% higher than the top-to-median ratio $\frac{p90}{p50}$.

In the model, given the returns to scale of marketing technology, the dispersion of marketing intensity is driven by the preference parameter u_k and the quality differentiation ξ_k. Figure 1.9 plots marketing intensity against u_k and ξ_k. Two observations can be made. Firstly, for the industry where the appeal is more important (u_k is smaller), marketing intensity is higher. Secondly, the impact of ξ_k on marketing intensity depends on the level of u_k. When u_k is small, complete information allocation would be incentive-compatible. Thus, there is no signaling competition between firms and ξ_k affects marketing intensity only through aggregate markup. A higher ξ_k implies fewer high-quality firms, which would lower the aggregate markup, and hence marketing intensity is higher. However, the magnitude of this channel is small. On the other hand, when u_k becomes larger, IC conditions become binding and the amount of quality differentiation ξ_k matters. A higher ξ_k indicates an industry with more homogeneous products, thus a lower level of marketing intensity.

Model fit. In Panel B of Table 1.1, I report the choices of calibrated parameters for the benchmark calibration. The average elasticity σ is 13.52 which matches the aggregate markup of 1.15 exactly. The returns to scale of mar-

[44] In Edmond et al. (2018), they report the firm size distribution based on relative sales in 2012 and find a similar pattern.

Table 1.2: Moments

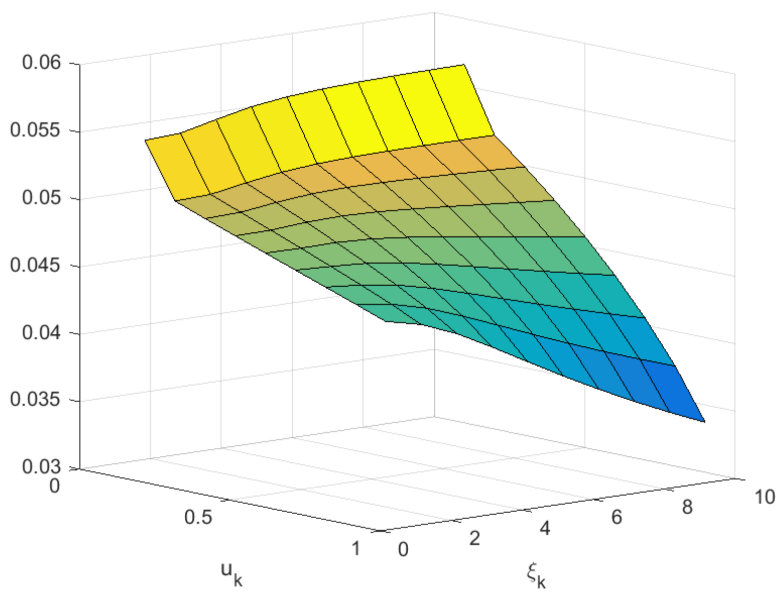
Panel A: Distribution of relative payroll

	data	benchmark	high μ	high β
avg fraction of firms with rel payroll				
< 0.5	0.662	0.694	0.603	0.705
< 1	0.819	0.895	0.842	0.894
< 5	0.962	0.985	0.976	0.985
< 10	0.987	0.990	0.986	0.991
< 100	1	0.998	1	0.998
dispersion of firm size distr. across industry				
< 1	0.133	0.110	0.112	0.116
< 5	0.019	0.009	0.014	0.009
< 10	0.008	0.006	0.010	0.006

Panel B: Dispersion of marketing intensity

	data	benchmark	high μ	high β
std dev	0.017	0.011	0.010	0.014
$p50/p10$	2.56	2.39	2.08	2.37
$p90/p50$	1.79	1.42	1.32	1.42
$(p50/p10)/(p90/p50)$	1.43	1.68	1.57	1.67

Figure 1.9: Illustration: u_k, ξ_k and marketing intensity



keting is equal to 0.054 which reproduces aggregate marketing intensity of 3 percent.

For the Lognormal distribution of ξ_k, the location parameter ξ_μ and shape parameter ξ_σ are calibrated to 0.28 and 2.21. Panel A of Table 1.2 shows that the model could match the average and dispersion of industry-level distribution of relative payroll in the data. For example, in the benchmark calibration, on average about 1% of firms have payroll larger than 10 times of industry average. Across industries, the standard deviation of fraction of firms with relative payroll smaller than industry average is 0.11. In the data, the corresponding moments are 1.3% and 0.133 respectively.

The shape parameters of Beta distribution are calibrated to $u_a = 0.047$ and $u_b = 0.009$, which implies the average of u_k is close to 1 and there is large dispersion in u_k across industries.[45] Panel B of Table 1.2 compares the model-generated dispersion of marketing intensity across industries with the data. The model can capture the large dispersion of marketing intensity across industries. For instance, in the model, marketing intensity of the median industry is about 2.4 times of the industry at 10 percentile, where the median-to-bottom ratio is 2.56 in the data.

Finally, in Table 1.1 and 1.2, two alternative sets of calibration results are reported. I re-calibrate the model to match an aggregate markup of 1.25 and an aggregate marketing intensity of 4% separately. Compared to the benchmark calibration, the high-markup calibration generates larger dispersion in firm size distribution and smaller variation in marketing intensity.

1.6 Results

In this section, I present the main quantitative results. Section 1.6.1 compares the model-generated elasticity between MPCR and markup with the data. The calibrated model can generate a similar level of MPCR-markup elasticity and its co-movement with aggregate marketing intensity.

In section 1.6.2, I evaluate to what extent access to marketing can undo the information frictions. Compared to complete information allocation, information frictions reduce productivity by about 3 percent. Higher returns to marketing

[45]For Beta distribution, $E[u] = \frac{u_a}{u_a+u_b}$ and $var[u] = \frac{k}{(k+1)^2[(k+1)u_b+1]}$ where $k = \frac{u_a}{u_b}$.

Table 1.3: MPCR-markup elasticity

	Data	benchmark	high μ	high β
	(1)	(2)	(3)	(4)
	$log(\frac{XAD}{COGS})$	$log(\frac{WL_a}{WL_p})$	$log(\frac{WL_a}{WL_p})$	$log(\frac{WL_a}{WL_p})$
$log(markup)$	1.393***	1.673***	1.171***	1.751***
	(0.079)	(0.117)	(0.073)	(0.116)

Notes: First column reports the regression results from the Compustat data for the data before 1995. For second column onwards, I report the regression results from model-generated data, where I randomly draw 470 industries. I apply industry-year FE for the Compustat data and industry FE for the model-generated data. Standard errors are clustered at industry-year or industry level and listed in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1.4: Co-movement between MPCR-markup elasticity and aggregate marekting intensity

Change in β	0	$\uparrow$ 10%	$\uparrow$ 20%	$\uparrow$ 30%
MPCR-markup elasticity				
benchmark	1.67	1.73	1.79	1.84
high μ	1.17	1.22	1.26	1.30
high β	1.75	1.81	1.87	1.93
implied marketing intensity, percentage				
benchmark	3.00	3.29	3.58	3.86
high μ	3.02	3.32	3.62	3.90
high β	4.04	4.41	4.77	5.11

fuel the signaling competition and generate larger losses.[46] In section 1.6.3, I discuss the impact of marketing on aggregate productivity. The gains from quality information revealed by marketing tend to dominate the loss due to markup dispersion.

1.6.1 MPCR-markup elasticity

I begin with MPCR-markup elasticity. Since information concerning the elasticity pattern is not utilized in the calibration process, we could compare the model-generated MPCR-markup elasticity with the data.

The first column of Table 1.3 shows the correlation pattern between MPCR and markup in the data. If markup increases by one percent, MPCR would rise by approximately 1.4 percent. From the second column onwards, I report

[46]I compare the allocation in LCSE with complete information allocation for different levels of returns to marketing β. Technical changes that increase returns to marketing can generate positive welfare impact by reallocating resources towards higher quality firms.

the regression results from model-generated data. The level of elasticity varies across different calibrations. A one percent increase in markup is associated with a 1.17 to 1.75 percent increase in MPCR. The calibrated model with higher returns to marketing tends to have a higher elasticity. When marketing is less costly, high-quality firms would spend more on marketing to differentiate from lower-quality firms. This channel would result in a larger dispersion in MPCR for a given amount of markup dispersion. On the other hand, the calibrated model with higher markup tends to have a lower elasticity. The level of information distortion is not sensitive to the change in average markup whereas the markup dispersion is. Thus, with a larger dispersion in markup, the MPCR-markup elasticity becomes smaller.

Then I investigate whether the model could generate the co-movement between MPCR-markup elasticity and aggregate marketing intensity as in the data. In Table 1.4, I illustrate how the aggregate marketing intensity and MPCR-markup elasticity evolve with changes in returns to marketing β. Across three parameterizations of the model, higher returns to marketing increase the marketing intensity and MPCR-markup elasticity simultaneously. Therefore, the model could generate the co-movement pattern shown in the data. However, the model cannot match the level of increase in the MPCR-markup elasticity. A 30 percent increase in returns of marketing technology increases the MPCR-markup elasticity by about 10 percent in the model, whereas the increase is about 25 percent in the data.[47]

1.6.2 Can access to marketing undo information frictions?

I next evaluate the productivity losses due to information frictions by comparing the allocation in LCSE and corresponding complete information allocation. In Table 1.5, I show the percentage losses in aggregate productivity due to information frictions. In the benchmark calibration, the existence of information frictions results in a 2.7 percent loss in productivity. This figure becomes larger for the alternative set of calibrations.

In the model, aggregate productivity could be decomposed into two components — a *between-firm* component and a *between-activity* component. The *between-firm* component summarizes the allocative efficiency of marketing labor and production labor across firms, whereas the *between-activity* component summarizes the allocative efficiency of marketing labor and production labor in

[47] This is from comparing the elasticity before and after 1995.

Table 1.5: Percentage losses of aggregate productivity

	benchmark	high μ	high β
compared to complete information			
$log(Z/Z^{CI}) * 100$	-2.669	-2.990	-3.345
between-activity	-2.668	-2.988	-3.343
between-firm	-0.002	-0.002	-0.002
compared to first best			
$log(Z/Z^{\star}) * 100$	-3.093	-3.806	-3.790

Table 1.6: Decomposition of *between-firm* component

	benchmark	high μ	high β
compared to complete information			
between-firm	-0.0018	-0.0024	-0.002
change $q_k(Q_n)$	0.0003	0.0004	0.0006
change $X_k(Q_n)$	-0.0021	-0.0028	-0.0027

the aggregate. Across all sets of calibration, we find the loss is almost entirely driven by the *between-activity* component. As equation 1.25 and 1.29 show, the *between-firm* component is a geometric mean of industry efficiency which depends on joint distribution of firm size and information distortion. Then I could further decompose the *between-firm* component into the change in relative size and the change in information distortion. Conceptually, the presence of information frictions would have two counteracting effects on the *between-firm* component. The first effect is that it will reallocate market share towards high-quality firms and increase industry-level efficiency Z_k. On the other hand, it will also imply a dispersion of $\frac{X_k}{X_k(Q_n)}$ and hence lower the industry-level efficiency. The small loss could be a result of these two opposite forces.

Table 1.6 illustrates the results of the decomposition. In the first row of the table, I report the same productivity losses due to *between-firm* component as in Table 1.5. In the second and third row, I compute the hypothetical *between-firm* component of aggregate productivity by changing $\frac{X_k}{X_k(Q_n)}$ and $q_k(Q_n)$ separately. We can find that both of these two channels have small impact on productivity compared to complete information allocation. Since markup is a deterministic function of $q_k(Q_n)$, the low level of reallocation also suggests that the presence of information frictions cannot lead to a sizable change in the markup distribution.

Given that information frictions lead to sizable productivity losses, how does the loss depend on returns to marketing? The upper panel of Table 1.7 shows

Table 1.7: Productivity losses and rising returns of marketing

Change in β	0	$\uparrow$ 10%	$\uparrow$ 20%	$\uparrow$ 30%
productivity loss compared to complete info				
benchmark	-2.669	-2.928	-3.178	-3.418
high μ	-2.990	-3.283	-3.567	-3.841
high β	-3.345	-3.637	-3.916	-4.183
implied marketing intensity, percentage				
benchmark	3.00	3.29	3.58	3.86
high μ	3.00	3.30	3.59	3.88
high β	4.04	4.41	4.77	5.11

that the productivity losses become larger with higher β. A 30% increase in β leads to more than 20% increase in productivity losses. Intuitively, higher returns to marketing imply that it is easier for lower quality firms to mimic high-quality firms. Therefore, high-quality firms would spend more on marketing to differentiate themselves, which results in more resource misallocation. One thing to note is that I evaluate the productivity losses from information frictions by comparing the allocation in LCSE with complete information allocation for different returns to marketing β. However, technical changes that increase returns to marketing can generate positive impact on productivity by reallocating resources to higher quality firms.[48]

1.6.3 The impact of marketing on aggregate productivity

So far, I compare the LCSE with the complete information allocation and find that information frictions result in a sizable loss in productivity. In this section, I study the impact of marketing on aggregate productivity where I compare the LCSE with an allocation that quality information is not revealed.

The model suggests that in a world without marketing, the channel to convey the quality information is not available. The excessive spending on marketing, however, provides useful information concerning the firm's quality to the consumer. Thus, this channel is usually interpreted as *signalling-efficiency* channel (Bagwell, 2007). On the other hand, the revealed information results in markup dispersion and hence resource misallocation between firms. To this

[48]Higher returns to marketing also generate a negative productivity effect on *between-activity* component. Consider a one-industry model with no information frictions, the *between-activity* component is $\left(\frac{\alpha}{\alpha+\beta(1-u)}\right)^{\alpha}\left(\frac{\beta(1-u)}{\alpha+\beta(1-u)}\right)^{\beta(1-u)}$, which is decreasing with β. Besides, there is an additional scale effect $L^{\alpha+\beta(1-u)}$. The gains from scale effect would dominate the loss from the *between-activity* component when L is large enough.

Table 1.8: Productivity gains from marketing

	benchmark	high μ	high β
compared to no information allocation			
$log(Z/Z^{NI}) * 100$	98.85	71.57	102.86

end, I consider an alternative allocation when quality information is not available. Specifically, within each industry, I impose the restriction that all firms have to deliver the same level of appeal. The consumer infers that each variety is of industry-average level quality. I label this allocation as *no-information* allocation.[49]

Table 1.8 presents the comparison between the benchmark calibration with the *no-information* allocation. The gains from marketing are huge — almost 100 percentage points. It becomes larger if it involves a larger firm size dispersion when information reveals. Consider a CES aggregator ($\epsilon = 0$) for an intuition. Firm sales are proportionate to $Q^{\frac{1-1/\sigma}{1-(1-1/\sigma)(\alpha+\beta(1-u))}}$ when there is complete information. Higher returns to marketing generate more dispersion in firm size. Thus the *no-information* allocation that equates size for different quality firms is more costly. On the contrary, higher markup (lower σ) reduces the size dispersion and hence the *no-information* allocation is less costly.

Quality information also generates a negative impact on productivity, which is resource misallocation due to markup dispersion. However, as shown in the last row of Table 1.5, loss from markup dispersion is small. The gains from quality information dominate the loss from markup dispersion.

The model could over-estimate the gains from marketing for many reasons. Firstly, it assumes there is no quality information available for every industry. As Nelson (1974) noted, industries are different in terms of the nature of goods/services produced. For some industries (*search goods*), the quality of a good can be determined before purchase. For other industries (*experience goods*), the quality can only be evaluated after consumption. Thus, we should expect different industries to have different degrees of information frictions. For industries with low extent of information frictions, access to marketing

[49] When the appeal $\Phi_k^\star$ is fixed within each industry, each firm would produce the same amount of output $Y_k^\star$. The quality of each variety reveals only after consumption. Prior to consumption, there is no source of quality information for consumers and consumers infer each variety is of industry-average quality. I consider the *no-information* allocation as the allocation that maximizes the *prior-consumption* utility of consumer and participation constraints of firms hold. Note this may not be a *pooling equilibrium* in our setting. A *pooling equilibrium* also needs to satisfy the IC condition for the lowest quality firms.

provides no additional quality information to the consumers. Then the gains from marketing would become lower. Secondly, LCSE implies a minimum loss in the process of information revelation. The deviation from LCSE would lower the gains from marketing. Thirdly, as in Edmond et al. (2018), markup is only related to firm size (quality) in the model. If marketing generates dispersion in the markup which is not related to size (quality) then it would involve a larger loss from misallocation.

There are many other channels that marketing could increase productivity beyond the mechanism in the model. For example, marketing could allow firms to reach customers and inform them of the existence of a firm (Arkolakis, 2010). It is not possible to study all of them within a single framework. Nevertheless, the result remains as one of the first attempts to study the question quantitatively.

1.7 Implication and Discussion

In this section, I discuss some of the model's implications. First of all, the model sheds light on the recent debate of the rise of aggregate markup in the economy. If I estimate the aggregate markup as the *cost-weighted* one in the model, the aggregate markup is virtually flat since the late 1980s. The same is true for the markup distribution. Consistent with data, a rise of returns to marketing cannot generate sizable changes in the markup distribution in the model.

Secondly, using cross-country data on marketing spending, Corrado and Hao (2014a) document a positive correlation between media advertising intensity and the level of economic development. Besides, for a given level of advertising intensity, there is a considerable variation of income per capita across countries. Our model provides a framework to understand these two phenomena.

Lastly, I show that the MPCR-markup elasticity only depends on the aggregate returns of marketing technology. Thus, using a fraction of marketing spending to measure MPCR would generate identical results in terms of level of the elasticity and its co-movement with aggregate marketing intensity.

1.7.1 Trend of market power in the data

In the model, aggregate markup that matters for resource allocation and productivity is production-cost weighted firm-level markup

$$\mu = \sum_{k,n} \mu_k(Q_n) \frac{WL_{kp}(Q_n)}{WL_p} M_k(Q_n).$$

I then examine how a rise in returns to marketing affects aggregate markup and markup distribution in the economy.

Table 1.9 shows the aggregate markup and distribution of markup in the model. The rising returns to marketing alone cannot generate a sizable change in the markup distribution. Figure 1.10 shows the trend of market power in the US since 1980. Two aggregate markups are reported. We can see that the *sales-weighted* one keeps increasing since 1980 whereas the *cost-weighted* one flattens out. Thus, whether there is a rise in the aggregate markup since late 1980s crucially depends on the way we aggregate firm-level markup. Figure 1.11 presents the *cost-weighted* markup distribution. A similar trend is observed that since the late 1980s, the markup distribution becomes stable.

The model cannot generate the rapid rise of *sales-weighted* markup. De Loecker et al. (2018) decompose the rise of *sales-weighted* markup and find that the rise of *sales-weighted* markup is driven by *reallocation* component since late 1980s.[50] This observation poses a challenge on using the models with variable markups to explain the rise of markup since reallocation of sales generally indicates a reallocation of production cost as well. Thus, *sales-weighted* and *cost-weighted* markup tend to move in the same direction.

1.7.2 Marketing and economic development

Marketing spending is correlated with the level of economic development. Figure 1.12 plots the media advertising intensity against GDP per capita for multiple countries from 1981 to 2011. Two observations can be made from the figure. Firstly, advertising intensity is positively correlated with income level across countries, and for a given country, it increases along the path of economic development. Secondly, there is a considerable variation in the income level for a given level of advertising intensity.

[50] Using Census data in the US, Autor et al. (2017) document the rise of aggregate markup if firm-level markups are aggregated using value-added as weights.

Figure 1.10: Aggregate markup trend in Compustat data

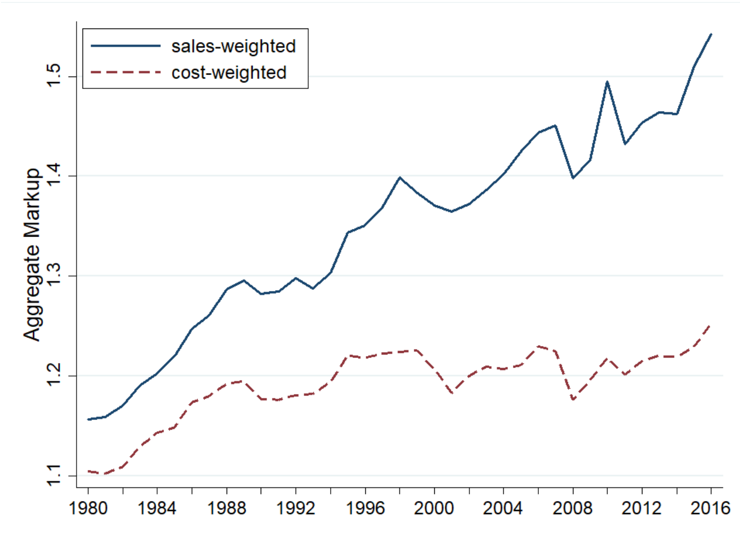


Figure 1.11: Markup dispersion in Compustat data

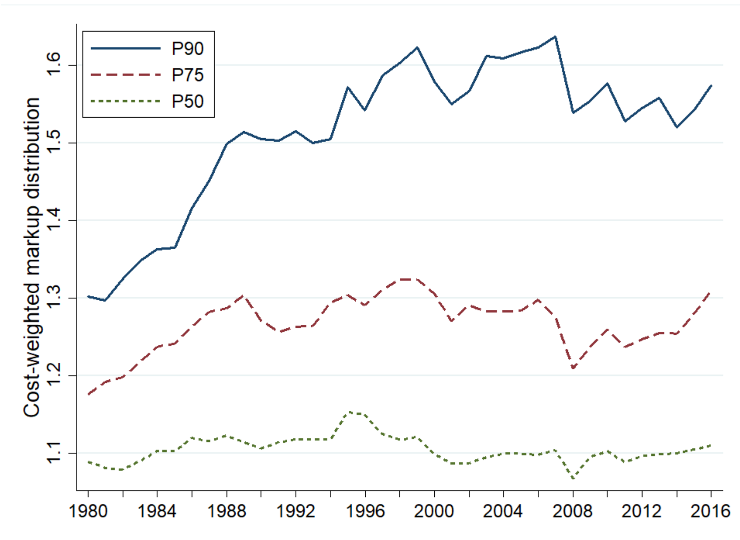
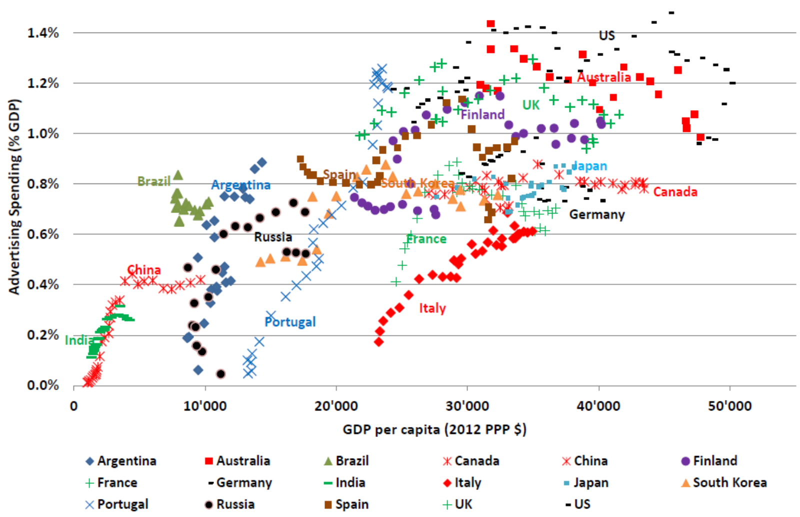


Table 1.9: Aggregate markup and markup distribution

Increase in β	0	↑ 10%	↑ 20%	↑ 30%
cost-weighted markup distribution				
aggregate markup	1.1522	1.1525	1.1528	1.1530
p25 markup	1.0856	1.0857	1.0858	1.0860
p50 markup	1.1462	1.1465	1.1468	1.1471
p75 markup	1.1939	1.1941	1.1943	1.1946
p90 markup	1.2429	1.2434	1.2439	1.2445
sales-weighted markup distribution				
aggregate markup	1.1566	1.1569	1.1572	1.1575

Figure 1.12: Advertising and economic development



Source: Figure 5 in Corrado and Hao (2014b).

In the model, aggregate marketing intensity is given by

$$\frac{WL_a}{PC} = \frac{\sum_k \beta(1-u_k)X_k \frac{WL_{kp}}{WL_p}}{\mu},$$

which depends on four elements. The first one is the returns to scale of marketing β. Along the process of economic development and technology adoption, new channels and platforms for marketing emerge, which increase the returns to marketing activities. The second component is production-cost share of each industry $\frac{WL_{kp}}{WL_p}$. Since appeal tends to be more important for service industries, structural transformation that reallocates resources towards services would increase the aggregate marketing intensity. Thirdly, aggregate marketing intensity is negatively correlated with aggregate markup in the economy. Lastly, the country with more intense signaling competition, i.e., higher X_k, tends to spend more on marketing. Therefore, given the returns to marketing activities, a high-income country that produces high-quality products could have a relatively low level of marketing intensity due to a higher level of markup or lack of competition among firms.

1.7.3 Marketing vs Advertising

In the firm-level data, I use advertising expenses to measure the marketing expense of the firm. Given there is a shift of marketing practices from *Outbound* marketing towards *Inbound* marketing, the advertising intensity is actually decreasing since 2000. In this section, I show that using a fraction of marketing spending to measure MPCR would generate exactly the same cross-sectional elasticity pattern with markup. Thus, this measurement issue will not affect the results on the MPCR-markup elasticity and its co-movement with aggregate marketing intensity.

Suppose appeal is produced by combining two elements of marketing input using a Cobb-Douglas production function: $\Phi = QM^{\beta}$ where $M = L_{a1}^{\chi} L_{a2}^{1-\chi}$.[51] Then the cost function of appeal is given by $C(\Phi) = \frac{W}{\chi^{\chi}(1-\chi)^{1-\chi}} \left(\frac{\Phi}{Q}\right)^{\frac{1}{\beta}} = WL_{a1} + WL_{a2} = WL_a$. The presence of two components within marketing function implies a level shift in the cost function of Φ. If we re-do the algebra in the modeling section, this shift has no impact on the distribution of firm size q and the extent of information distortion X. Intuitively, the level shift of the cost function of appeal will not change either the dispersion of sales across firms or the incentive to mimic other firms. Even if we use a fraction of marketing spending to measure MPCR, the MPCR-markup elasticity depends only on the information distortion X, which is determined by the aggregate returns of marketing technology β. Thus, measuring MPCR using advertising expense would result in an identical level of MPCR-markup elasticity

$$\frac{WL_{a1}(Q_n)}{WL_p(Q_n)} = \frac{\chi\beta(1-u)}{\alpha} X(Q_n).$$

Similarly, for the co-movement pattern, the change of χ within marketing technology is irrelevant for both the MPCR-markup elasticity and aggregate marketing intensity.

1.8 Conclusion

Empirically, this chapter documents that (i) aggregate marketing intensity in the US increased sharply around the mid-1990s, (ii) there is a positive correlation between firm-level MPCR and markup, (iii) the *cross-sectional* MPCR-markup elasticity *co-moves* closely with aggregate marketing intensity.

To explain these facts, I develop a model with heterogeneous firms and endogenous variable markups where firms engage in marketing to signal their quality. The existence of information frictions generates a positive correlation between MPCR and markup. Technical changes that increase the returns to

[51] Same argument goes through with CES production function.

marketing fuel the signaling competition and generate co-movement between MPCR-markup elasticity and aggregate marketing intensity. I use the calibrated model to quantify the impact of marketing and information frictions on aggregate productivity. Firstly, I find that access to marketing cannot undo the information frictions completely. Compared to complete information allocation, information frictions reduce the productivity by about 3 percent, and the loss increases with returns to marketing. Moreover, the quality information revealed by marketing can generate substantial gains in productivity.

1.9 Appendix

1.9.1 Measurement of marketing spending

I estimate marketing spending in the US since 1987 using data from different sources. The marketing expenditure consists of three components: advertising expenses, purchased marketing research and marketing consulting services, and marketing labor compensation. I estimate these three components separately and add them up to estimate the total marketing spending in the US.

Advertising expenses. Advertising expense data are extracted from IRS tax data, both at the industry-level and aggregate level. For the aggregate level, the advertising expenses are reported by the legal form of organizations, including corporations, sole proprietors and partnerships. For partnerships, there is no advertising data available and it is estimated using data on business receipts. I take a conservative view to estimate the advertising expense for the partnerships using the advertising intensity of sole proprietors, which is lower than that of corporations.

IRS also reports the advertising expense at the major industry-level, approximately 3-digit NAICS level. The data before 1998 are reported using SIC code. I harmonize the industry code by using the SIC and NAICS crosswalk.[52]

Marketing research and consulting. The second component of marketing spending is purchased marketing research and marketing consulting services. The total spending on those activities is estimated using the revenue of the corresponding industries, specifically NAICS 541613 for marketing consulting services and NAICS 541910 for marketing research. The aggregate revenue of these two industries are estimated using survey-based measures from the Service Annual Survey (SAS).

Note that SAS only reports the estimated revenue for the management consulting sector (NAICS 54161) as a whole. The revenue of marketing consulting services is estimated using the employment share of NAICS 541613, which is available from County Business Patterns (CBP) from 1998 onwards.[53] Additionally, the SAS data after 1998 is based on employer firms, whereas the data before 1998 covers both employer and nonemployer firms. The revenue of

[52]I construct the SIC-NAICS crosswalk using `https://www.bls.gov/ces/sic2tonaics.htm` and `https://www.bls.gov/cew/datatoc.htm`.

[53]Available at `https://www.census.gov/programs-surveys/cbp/data/datasets.html`.

nonemployer firms is estimated using employment share from CBP and receipts from Nonemployer statistics program.[54] Finally, to estimate the industry-level purchase of the output from these two industries, I use the 2002 benchmark input-output table.

Marketing labor compensation. The last component of marketing spending is the own-account component, where the own-account component is measured by the compensation of marketing workers. I first identify a list of marketing occupations and then estimate the labor compensation associated with those marketing occupations for each industry.

Specifically, the labor compensation is estimated using a two-step approach. In the first step, I use the RAS method to estimate the employment by occupation and industry, where I treat CPS employment data by occupation as row totals and BEA/BLS industry employment as column totals.[55] Then I convert the employment by industry and occupation to compensation using relative CPS wage by occupation controlled by industry-level compensation data. As a result, the aggregate compensation I compute is consistent with the data published by BEA.

Table 1.10 shows the list of occupations that is marketing related. The occupations code is based on OCC2010 from IPUMS CPS. Figure 1.13 breaks down three components of the marketing expenditure. Although advertising expenses are the largest among all three components, its importance is diminishing over time. The increase of marketing labor compensation and marketing service contribute to most of the rise since the mid-1990s. The diminishing importance of advertising reflects the shift in the marketing practices of the firms — from *Outbound* towards *Inbound*.

I follow the approach in Corrado and Hao (2014a) to construct marketing spending by aggregating data from different sources. I discuss the potential double-counting problem and how I deal with it in my measure. The advertising expenses available from IRS contain only direct promotional expenses that are *purchased* from other firms and industries. It will include the purchased services by the advertising industries (NAICS 5418). The IRS advertising expenses should not include the purchased marketing consulting and market research services as they are not direct promotional expenses and should be recorded elsewhere in the tax returns.[60] In addition, despite the fast growth

[54] `https://www.census.gov/programs-surveys/nonemployer-statistics/data/datasets.html`. For the data before 1998 with SIC code, I construct the comparable series.

[55] I use the IPUMS-CPS data as the initial value for each cell.

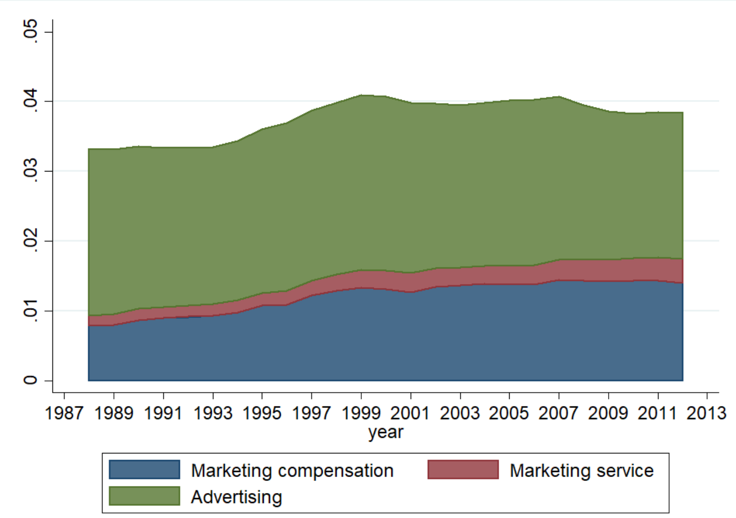
[60] See the discussion of tax breaks on market research at `https://smallbusiness.chron.`

Table 1.10: Marketing occupations

Occupations	occ2010
managers in marketing, advertising, and public relations	30
meeting and convention planners	720[56]
market research analysts and marketing specialists	730[57]
web developers	1000[58]
database administrators	1060
market researchers	1800[59]
public relations specialists	2825
technical writers	2840
writers and authors	2850
media and communication workers, nec	2860
broadcast and sound engineering technicians and radio operators, and media and communication equipment workers, all other	2900
photographers	2910
television, video, and motion picture camera operators and editors	2920
customer service representatives	5240

Notes: The marketing occupation listed is primarily adopted from Corrado and Hao (2014a). Occupations are based on OCC2010 from Ipums CPS data. I estimate the marketing compensation for the private, non-media and non-marketing related industries, which exclude BEA industry 38, 39, 50, 59. For occupation 720, 730, 1000, 1800, those occupation groups contain non-marketing occupations. Specifically, fundraiser is a sub-category of occupation 720; market research analysts and marketing specialists is a sub-category of occupation 730; web developers is a sub-category of occupation 1000; market researchers is a sub-category of occupation 1800. I estimate the share of marketing related compensation within these groups using the BLS/OES data from 2012 to 2017 where data for those sub-categories are available.

Figure 1.13: Marketing expenditure decomposition since 1987



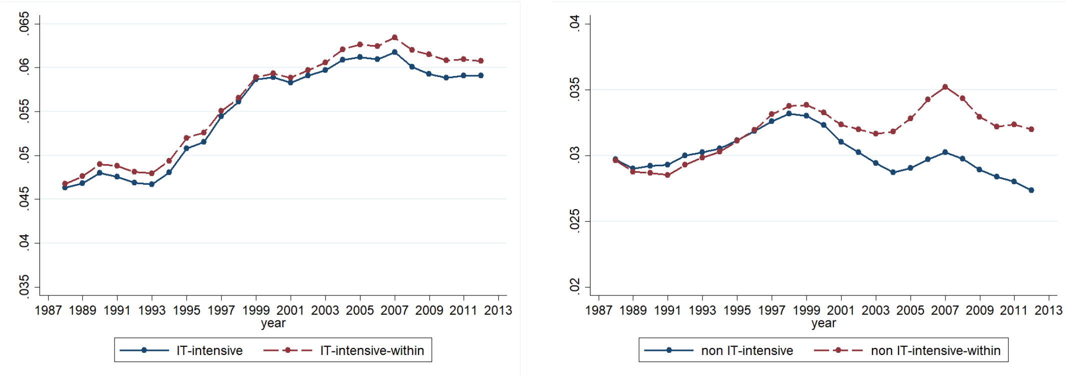
Notes: This figure shows a decomposition of aggregate marketing intensity in US since 1987, where aggregate marketing expenditure is the sum of advertising expense, purchased marketing service and marketing labor compensation

of the marketing consulting and marketing research component, Figure 1.13 shows its contribution to the rising aggregate marketing intensity is limited. Besides, the IRS advertising expenses do not include own-account marketing labor compensation as it is recorded as salaries of employees in the tax records. However, there is a potential double-counting problem as marketing labor compensation in the media and marketing related industries may already be counted as purchased advertising expenses for other industries. To avoid this double-counting problem, I only estimate the marketing labor compensation for the private, non-agricultural, non-media, non-marketing related industries.

In Figure 1.14, I show the aggregate marketing intensity and its within-industry component for both IT-intensive and non IT-intensive industries. For the IT-intensive industries, the within-industry component of aggregate marketing intensity increased by about 25% from 1994 to 2007. On the other hand, the same intensity only increased by about 15% for non IT-intensive industries. Thus, the result suggests the rise of aggregate marketing intensity is related to the information technology.

com/tax-breaks-market-research-34403.html.

Figure 1.14: Aggregate marketing intensity: IT intensive vs non IT-intensive



(a) IT-intensive (b) non IT-intensive

Notes: The left (right) panel shows aggregate marketing intensity and its within-industry component for IT-intensive (non IT-intensive) industries. The IT intensity of each industry is measured by the ratio between IT investment and total investment. I take the yearly average of the IT investment ratio in the 1980s and classify industries according to this measure. Industry-level investment data is available through BEA.

1.9.2 Firm-level data and markup estimation

I use the Compustat Fundamental Annual file for the firm-level data, where I observe sales, capital stock information, operating expenses, and industry classification. The choice of data is driven by two reasons. First, the Compustat data covers a wide range of sectors over a substantial period of time. Besides, it reports detailed components of the firm's operating expenses, which allows for the estimation of markup and marketing spending.

Markup estimation. I use the production-based approach in De Loecker and Warzynski (2012) to identify markups. The advantage of the production approach is that we only have to assume cost minimization for each firms. It does not require researchers to specify how the firms are competing and make assumptions on demand structure.

$$\mu_{it} = \theta^v_{it} \cdot (\alpha^v_{it})^{-1} \tag{1.36}$$

where μ_{it} is the markup, θ^v_{it} is the output elasticity of variable input, and α^v_{it} is the revenue share of variable input. The revenue share α^v_{it} is readily available from the data, and we have to estimate the output elasticity of variable input θ^v_{it}. Following De Loecker et al. (2018), I consider COGS as the variable input.

To estimate the θ^v_{it}, we consider an industry-specific production function

$$q_{it} = f(k_{it}, v_{it}; \theta) + \omega_{it} + \epsilon_{it} \tag{1.37}$$

where q_{it} measures the logarithm of firm's output, v_{it} is the logarithm of COGS, k_{it} is the logarithm of capital stock, ω_{it} is the logarithm of productivity, and ϵ_{it} is the unanticipated shock or classical measurement error in output. Then output elasticity is measured as $\theta^v_{it} = \frac{\partial f(k_{it}, v_{it}; \theta)}{\partial v}$. I consider two production function specifications. For the benchmark results, I consider an industry-specific Cobb-Douglas production function, i.e., $f = \theta_k k_{it} + \theta_v v_{it}$. I also estimate an industry-specific translog production function as a robustness check.[61]

Since the productivity ω_{it} is unobserved, OLS estimation involves the endogeneity of regressors. I use the standard two-stage approach to estimate the production function. In the first stage, the measurement error and unanticipated shocks to output are purged. Specifically, I calculate predicted real sales $\hat{q_{it}}$ using a non-parametric regression

$$q_{it} = \phi(k_{it}, v_{it}) + \epsilon_{it}$$

where $\phi = f + h(k, v)$ and h is the control function.[62] Suppose the productivity follows an AR(1) process $\omega_{it} = \rho\omega_{it-1} + \zeta_{it}$, where residual ζ_{it} captures contemporaneous productivity shock. In the second stage, I estimate coefficient of production function θ using moment conditions:

$$E[\zeta_{it}(\theta) Z_{it}] = 0$$

where $\zeta_{it}(\theta)$ is obtained as the residual when projecting $\omega_{it}(\theta)$ on its lag $\omega_{it-1}(\theta)$, $\omega_{it} = \hat{q_{it}} - f(k_{it}, v_{it}; \theta)$, and Z_{it} includes k_{it}, v_{it-1} and all its interaction terms. The identifying assumption is that lagged input choice (v_{it-1}) and predetermined capital stock (k_{it}) are independent to the contemporaneous productivity shock ζ_{it}.

To estimate the production function, I download the Fundamental Annual Compustat file from WRDS for the period between 1950 and 2017. I exclude observations with negative or missing values in sales, COGS, operating expenses, or gross plants, property, and equipment (PPE). I use NAICS code reported in the data to generate the industry classification that is consistent with BEA 65 industry classification.[63] Firms with missing industry codes are

[61] I also consider a three-input version of production function where I add SGA as an input of production function. The empirical facts are also robust to this three-input production function.

[62] We could use the control function h to control for the endogeneity of productivity under the assumption that a firm's demand for variable input is an invertible function of the firm's productivity and capital stock.

[63] The industry code is at 3-digit NAICS level, which is available at `https://apps.bea.gov/national/FA2004/Details/xls/detailnonres_stk1.xlsx`.

dropped. All financial variables are deflated with the appropriate deflators.[64] The resulting sample contains 227,287 observations that span from 1951 to 2017. I use this sample to estimate the industry-level production function and markup of firms.

Advertising spending. I use the advertising spending reported by Compustat as a proxy for firm-level marketing spending. Ptok et al. (2018) document that the advertising spending reported by the Compustat is highly correlated with total marketing spending of firms.[65] Therefore, I consider the advertising spending as a valid proxy to study the cross-sectional relationship between markup and MPCR. To be consistent with the data on aggregate marketing spending, I consider the sample between 1987 and 2013, which leaves 141,092 observations in the sample. I further restrict the sample by keeping observations with non-missing values in advertising spending, which leaves 46,260 observations. It is the sample that I use to establish the relationship between markup and MPCR.[66]

1.9.3 Marketing as a fixed cost

Marketing spending is counted as a fixed cost in this chapter. I document two empirical facts regarding the relationship between MPCR and markup. Firstly, I find that firms with higher markups tend to have higher MPCR, i.e., they spend more on marketing relative to production. Secondly, I find the cross-sectional elasticity between MPCR and markup becomes larger since the mid-1990s, which co-moves with aggregate marketing intensity closely.

During the sample period, we also observe a rise of fixed cost, where fixed cost is measured by SGA expenses.[67] Figure 1.19 shows the trend of elasticity between SGA-production cost ratio and markup over time. Although firms with higher markup tend to have a higher fixed cost share, the SGA-production cost ratio does not become more sensitive to changes in markup.

[64]Following Traina (2018), I deflate the sales and COGS using the GDP deflator and construct real capital stock using the perpetual inventory method. In addition, within a given firm, missing observations of sales, COGS, operating expenses, gross PPE, and net PPE are interpolated using their neighboring values.

[65]Ptok et al. (2018) acquire firm-level total marketing spending from an alternative data source (*Advertising Age*), and find that the correlation between advertising spending reported by the Compustat and the total marketing spending is about 0.8.

[66]Among the observations with non-missing values in advertising spending, 45,502 observations have positive advertising spending.

[67]The rise of SGA's share in sales and operating expenses are documented by many researchers. See De Loecker et al. (2018) and Traina (2018) for example. Marketing spending is only a small fraction of SGA expenses of firms. See section 1.9.6.

Intuitively, if technology has changed over time to induce a higher fixed cost of operation, firms will increase the markup to avoid making losses. However, this only implies a positive correlation between fixed cost share and markup and does not necessarily induce a stronger elasticity between fixed cost share and markup. In addition, Figure 1.6 and 1.19 highlight the difference between marketing spending and total fixed cost, and suggest that the model used to explain the co-movement pattern should work for marketing but not fixed cost as a whole.

I develop a model where firms spend on marketing to signal their quality. Technical changes that increase the returns to scale of marketing technology will lower the cost of producing appeal, which intensifies the signaling competition and generates the co-movement between aggregate marketing intensity and MPCR-markup elasticity. This mechanism is less likely to work for the fixed cost as a whole since most components of fixed cost do not have a signaling role. In section 1.4.2, I show that, with complete information, there is no signaling competition between firms, and higher returns to marketing do not affect the MPCR-markup elasticity.

1.9.4 MPCR-markup elasticity: alternative regression

In the main text, I run the regression of logarithm of MPCR on logarithm of markup. However, the COGS appears in the denominator of both the dependent variable and regressor. Thus, if advertising spending and sales are noise information across firms, the regression would generate a positive MPCR-markup elasticity as well. In this section, I consider an alternative specification

$$log\,(XAD_{is}) = \rho_0\ log(\mu_{is}) + \rho_1 log(COGS_{is}) + \delta_s + \epsilon_{is}$$

and compare the estimated ρ_0 from this specification to the estimated ρ in the main regression.

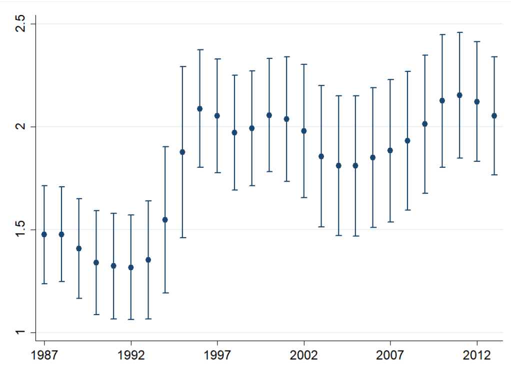
Table 1.11 shows the regression results from the alternative specification. The estimated ρ_0 is similar to the estimated ρ in the main regression. Furthermore, in Figure 1.15, the estimated ρ_0 from the alternative specification also increased sharply around the mid-1990s, which results in same co-movement pattern with aggregate marketing intensity.

Table 1.11: MPCR-markup elasticity: alternative regression specification

	(1) $log(XAD)$	(2) $log(\frac{XAD}{COGS})$	(3) $log(XAD)$	(4) $log(\frac{XAD}{COGS})$
$log(COGS)$	1.084***		1.043***	
	(0.0103)		(0.0116)	
$log(markup_{CD})$	1.815***	1.720***		
	(0.0562)	(0.0556)		
$log(markup_{Translog})$			1.363***	1.333***
			(0.0959)	(0.0942)
N	41983	41983	41713	41713

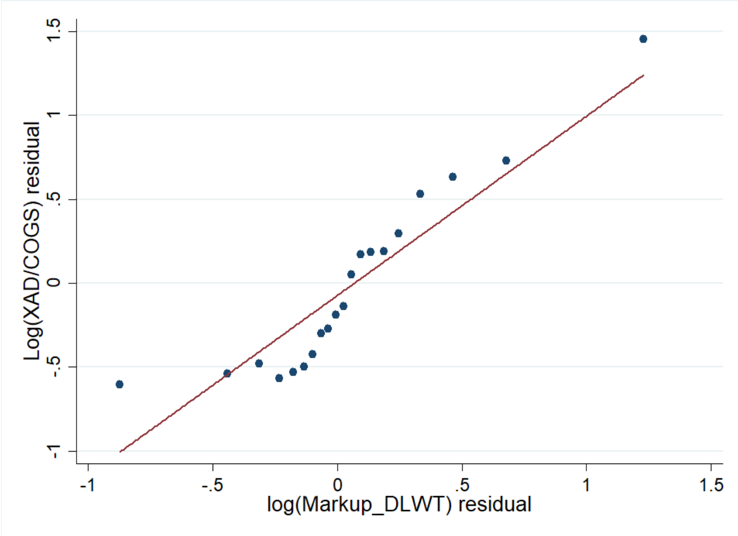
Notes: This table compares the MPCR-markup elasticity estimated from two specifications. I apply industry-year FE and standard errors are clustered at industry-year level and listed in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 1.15: MPCR-markup elasticity: alternative specification



Notes: The left (right) panel shows the estimated ρ_0 for year-on-year regressions. For each year, I conduct the regression by combining data within a three-year moving window. Markup is estimated using an industry-specific Cobb-Douglas production function in the left panel or Translog production function in the right panel.

Figure 1.16: Binscatter plot between log(MPCR) and log(markup) with Translog production function



Notes: This binscatter plot shows the correlation between markup and marketing-production cost ratio. Both variables are residualized using industry-year fixed effects. Marketing cost is measured by advertising expense and production cost is measured by $COGS$. Markup is estimated using industry-specific Translog production function.

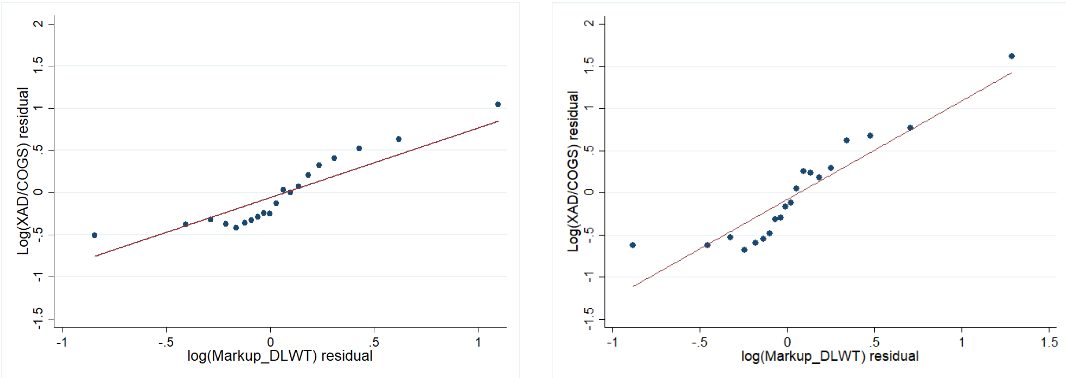
1.9.5 MPCR-markup elasticity: Translog production function

In the main text, I show the positive correlation between MPCR and markup and the co-movement between MPCR-markup elasticity and aggregate marketing intensity using markup estimated from an industry-specific Cobb-Douglas production function. In this section, I conduct the same exercise using industry-specific translog production. As shown in Figure 1.16, 1.17 and 1.18, the choice of production function in the markup estimation does not change the empirical findings.

1.9.6 SGA-production cost ratio vs markup

De Loecker et al. (2018) and Traina (2018) give evidence of a positive correlation between SGA expenses and markup. However, as emphasized by Ptok et al. (2018), SGA is not a measure of marketing spending of the firms. Specifically, only 2 out of 29 items that constitute SGA relate directly to the marketing spending. Thus, we should expect the co-movement between MPCR-markup elasticity and aggregate marketing intensity not happen if we measure the marketing using SGA expenses.

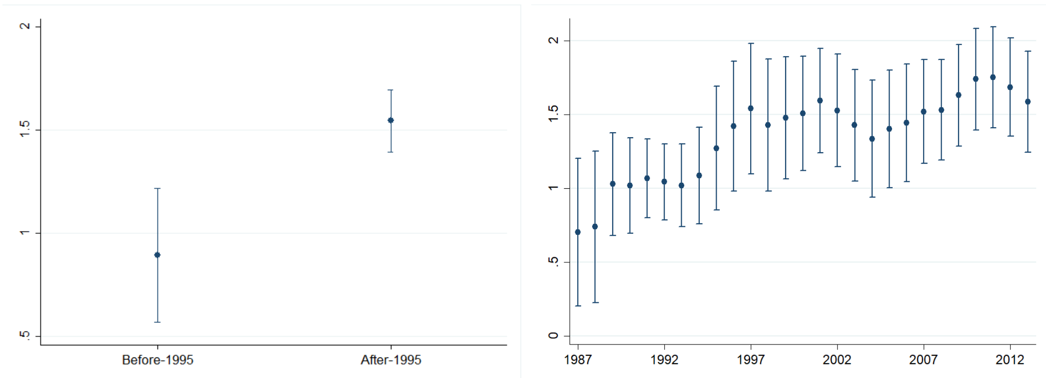
Figure 1.17: Binscatter plot between log(MPCR) and log(markup) with Translog production function: before and after mid 1990s



(a) Before 1995 (b) After 1995

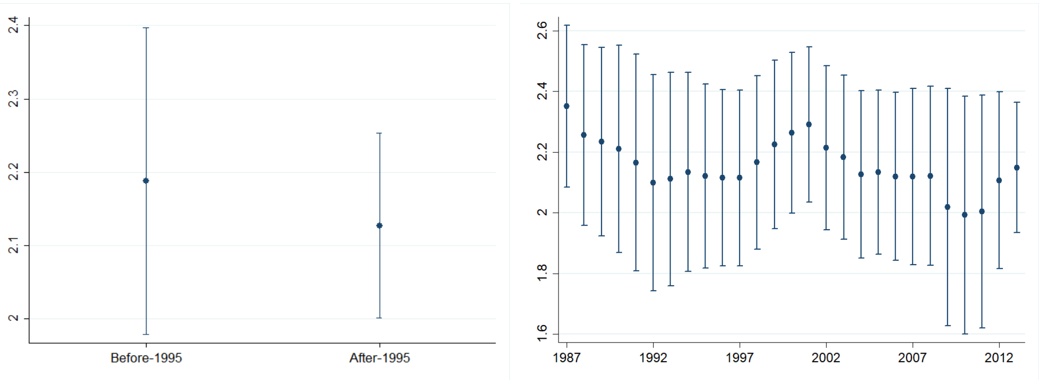
Notes: The left (right) panel shows a binscatter plot between markup and marketing-production cost raio before (after) 1995. Both variables are residualized using industry-year fixed effects. Marketing cost is measured by advertising expenditure and production cost is measured by $COGS$. Markup is estimated using industry-specific Translog production function.

Figure 1.18: MPCR-markup elasticity with Translog production function



Notes: This figure shows the elasticity between log(MPCR) and log(markup), where the elasticity is obtained from a regression where I regress log(MPCR) on log(markup) and industry-year fixed effects. The left panel shows the result for two separate regressions: before 1995 and 1995. The right panel shows the result for year-on-year regressions. For each year, I conduct the regression by combining data within a three-year moving window. Markup is estimated using industry-specific Translog production function.

Figure 1.19: MPCR-markup elasticity: SGA as a measure of marketing spending



Notes: This figure shows the elasticity between log(MPCR) and log(markup), where the elasticity is obtained from a regression where I regress log(MPCR) on log(markup) and industry-year fixed effects. The left panel shows the result for two separate regressions: before 1995 and after 1995. The right panel shows the result for year-on-year regressions. For each year, I conduct the regression by combining data within a three-year moving window. Marketing spending for each firm is measured using SGA expenses.

Figure 1.19 shows the MPCR-markup elasticity where I use SGA to measure marketing spending. Consistent with the expectation, there is no significant trend in the MPCR-markup elasticity. The level of *cross-sectional* MPCR-markup elasticity *does not* co-move with aggregate marketing intensity.

1.9.7 Proof of Propositions and Lemmas

1.9.7.1 Proof of Lemma 1

Proof. The profit function of a firm in industry k with variety ω is given as

$$\Pi_k(\omega) = P_k(\omega)C_k(\omega) - WC_k(\omega)^{\frac{1}{\alpha}} - W\left(\frac{\Phi_k(\omega)}{Q_k(\omega)}\right)^{\frac{1}{\beta}} \tag{1.38}$$

Then the marginal rate of substitution of signal $\Phi_k(\omega)$ for demand $C_k(\omega)$ is given as

$$\frac{-\frac{\partial \Pi_k(\omega)}{\partial \Phi_k(\omega)}}{\frac{\partial \Pi_k(\omega)}{\partial C_k(\omega)}} = \frac{\frac{1}{\beta}W\left(\frac{\Phi_k(\omega)}{Q_k(\omega)}\right)^{\frac{1}{\beta}-1}\frac{1}{Q_k(\omega)}}{P_k(\omega) - \frac{1}{\alpha}WC_k(\omega)^{\frac{1}{\alpha}-1}} \tag{1.39}$$

which is decreasing with quality of firms for given point in the space of signals and demand when $P_k(\omega) > \frac{1}{\alpha}WC_k(\omega)^{\frac{1}{\alpha}-1}$. $\square$

1.9.7.2 Proof of Lemma 2

Proof. I prove the Lemma 2 in two steps. In the first step, I show analytically that the extent of information distortion depends on the level of u in a world with no markup dispersion. In the second step, I show the same logic applies to the model with Kimball aggregator.

When there is no markup dispersion, the complete information allocation is incentive-compatible when all types have no incentives to mimic the others. Specifically, it implies the following inequality holds

$$1 \geq exp((j-i)\Delta)^{\frac{\left(1-\frac{1}{\sigma}\right)}{1-(\alpha+\beta(1-u))\left(1-\frac{1}{\sigma}\right)}} \frac{1-\alpha\left(1-\frac{1}{\sigma}\right)-\left(1-\frac{1}{\sigma}\right)\beta(1-u)exp\left((j-i)\Delta\right)^{\frac{1}{\beta}}}{1-\alpha\left(1-\frac{1}{\sigma}\right)-\left(1-\frac{1}{\sigma}\right)\beta(1-u)} \tag{1.40}$$

where $i, j \in \{1, 2, ..., N\}$. Intuitively, the gain from mimicking other types should be lower compared to truthtelling. I denote the right hand side of the inequality as $f(x)$ where $x = j - i$, and study the maximum of $f(x)$.

Some algebra shows that f has following properties: (i) $f(0) = 1$, (ii) f achieves maximum when $x = x^\star$ and (iii) $f'(x) > 0$ if $x < x^\star$ and $f'(x) < 0$ if $x > x^\star$. The maximum value of f is given by

$$f(x^\star) = \left[\frac{1-\alpha\left(1-\frac{1}{\sigma}\right)}{1-\alpha\left(1-\frac{1}{\sigma}\right)+\beta u\left(1-\frac{1}{\sigma}\right)}\right]^{\frac{\left(1-\frac{1}{\sigma}\right)\beta}{1-(\alpha+\beta(1-u))\left(1-\frac{1}{\sigma}\right)}+1} \left[\frac{1}{1-u}\right]^{\frac{\left(1-\frac{1}{\sigma}\right)\beta}{1-(\alpha+\beta(1-u))\left(1-\frac{1}{\sigma}\right)}} \tag{1.41}$$

where $x^\star = \frac{\beta}{\Delta} ln\left(\frac{1-\alpha\left(1-\frac{1}{\sigma}\right)}{\left[\left(1-(\alpha+\beta(1-u)\left(1-\frac{1}{\sigma}\right)\right)+\beta\left(1-\frac{1}{\sigma}\right)\right](1-u)}\right)$. We can see $f(1) > 1$ when u is large. It implies that IC conditions will not satisfy when u is large enough. On the other hand, when $u \longrightarrow 0$, $f(x^\star) \longrightarrow 1$ and $x^\star \longrightarrow 0$. Thus, IC condtions will statisfy when u is small enough.

For Kimball aggregator, the incentive to mimic higher quality firms depends on the level of quality. At each level of quality Q_n, the gain from mimicing Q_{n+1} type of firms compared to the truthtelling is given as

$$\begin{aligned} \frac{\Pi(Q_{n+1}; Q_n)}{\Pi(Q_n; Q_n)} = & \frac{\Upsilon'(q(Q_n))q(Q_n)}{\Upsilon'(q(Q_{n-1}))q(Q_{n-1})} \\ & \cdot \frac{1-\alpha\mu(q(Q_n))^{-1}-\beta(1-u)exp\left(\frac{\Delta}{\beta}\right)\mu(q(Q_n))^{-1}}{1-\alpha\mu(q(Q_{n-1}))^{-1}-\beta(1-u)\mu(q(Q_{n-1}))^{-1}} \end{aligned} \tag{1.42}$$

where I denote $\Pi(Q_i; Q_j)$ as profit of type Q_j mimicing type Q_i. Given Δ is small,

$$\frac{\Pi(Q_{n+1};Q_n)}{\Pi(Q_n;Q_n)} \approx exp(\Delta)^{\rho_n} \frac{1-\alpha\left(1-\frac{1}{\sigma_n}\right)-\left(1-\frac{1}{\sigma_n}\right)\beta(1-u)exp\left(\frac{\Delta}{\beta}\right)}{1-\alpha\left(1-\frac{1}{\sigma_n}\right)-\left(1-\frac{1}{\sigma_n}\right)\beta(1-u)} \tag{1.43}$$

where σ_n is the demand elasticity for type Q_n firms and $\rho_n = \frac{\left(1-\frac{1}{\sigma_n}\right)}{1-(\alpha+\beta(1-u))\left(1-\frac{1}{\sigma_n}\right)}$. The right hand side of equation 1.43 can be further approximated as

$$\frac{\Pi(Q_{n+1};Q_n)}{\Pi(Q_n;Q_n)} \approx 1+u\Delta\rho_n-\rho_n\Delta^2\left(\frac{1-u}{4\beta}+\left(\frac{3}{4}-u\right)\rho_n\right). \tag{1.44}$$

Given a small Δ, the complete information allocation would (not) be incentive compatible when u is small (large).

□

1.9.7.3 Solving the LCSE

In this section, I establish the argument that LCSE has a simple *complete-information distortion* feature. I start by looking at the maximization problem of quality Q_2

$$max \;\; \Pi_k(P(Q_2),\Phi(Q_2),b=Q_2,Q_2)$$

subject to the *incentive compatibility* (IC) condition

$$\Pi_k(P(Q_1),\Phi(Q_1),b=Q_1,Q_1) \geq \Pi_k(P(Q_2),\Phi(Q_2),b=Q_2,Q_1)$$

Setting up the Lagrangian for type Q_2 in industry k

$$L_{k2} = \Pi_k(Q_2;Q_2)+\lambda_{k2}(\Pi_k(Q_1;Q_1)-\Pi_k(Q_2;Q_1)) \tag{1.45}$$

The optimality condition of $P(Q_2)$ implies

$$\frac{\partial\Pi_k(Q_2;Q_2)}{\partial P(Q_2)} = \lambda_{k2}\frac{\partial\Pi_k(Q_2;Q_1)}{\partial P(Q_2)} \tag{1.46}$$

Since the marginal returns of $P(Q_2)$ does not depends on types of firms, thus $\frac{\partial\Pi_k(Q_2;Q_2)}{\partial P(Q_2)} = 0$. The optimality condition of $\Phi(Q_2)$ implies

$$\frac{\partial\Pi_k(Q_2;Q_2)}{\partial\Phi(Q_2)}-\frac{\partial\Pi_k(Q_2;Q_1)}{\partial\Phi(Q_2)} = (\lambda_{k2}-1)\frac{\partial\Pi_k(Q_2;Q_1)}{\partial\Phi(Q_2)} \tag{1.47}$$

The left hand side of the above equation is positive. Then I further show that $\lambda_{k2} \in (0,1)$, which is equivalent to show $\frac{\partial \Pi_k(Q_2;Q_1)}{\partial \Phi(Q_2)} < 0$. I prove this by contradiction. Suppose $\frac{\partial \Pi_k(Q_2;Q_1)}{\partial \Phi(Q_2)} > 0$. Then there exists a new pair of allocation that makes Q_1 indifferent but Q_2 strictly better off. Thus, $\frac{\partial \Pi_k(Q_2;Q_1)}{\partial \Phi(Q_2)}$ can not be positive at optimal. Similarly, when $\frac{\partial \Pi_k(Q_2;Q_1)}{\partial \Phi(Q_2)} = 0$, an increase in $\Phi(Q_2)$ will be a profitable move for Q_2, and Q_1 has no incentive to mimic. Thus, $\frac{\partial \Pi_k(Q_2;Q_1)}{\partial \Phi(Q_2)} < 0$ and $\lambda_{k2} \in (0,1)$. We can rewrite the Lagrangian as

$$\begin{aligned} L_{k2} &= \Pi_k(Q_2;Q_2) - \lambda_{k2}\Pi_k(Q_2;Q_1) + \lambda_{k2}\Pi_k(Q_1;Q_1) \\ &= (1-\lambda_{k2})\left(P(Q_2)C(Q_2) - WC(Q_2)^{\frac{1}{\alpha}} - W\left(\frac{\Phi(Q_2)}{Q^{\star}_{k2}}\right)^{\frac{1}{\beta}}\right) + \lambda_{k2}\Pi_k\left(Q_1;Q_1\right) \end{aligned} \tag{1.48}$$

Thus, Q_2 firms would maximize the profits as in the complete information case but with a higher quality $Q^{\star}_{k2}$, where $Q^{\star}_{k2}$ is given as

$$\lambda_{k2}\left[\left(\frac{1}{Q_2}\right)^{\frac{1}{\beta}} - \left(\frac{1}{Q_1}\right)^{\frac{1}{\beta}}\right] = (1-\lambda_{k2})\left[\left(\frac{1}{Q^{\star}_{k2}}\right)^{\frac{1}{\beta}} - \left(\frac{1}{Q_2}\right)^{\frac{1}{\beta}}\right].$$

Thus, $Q^{\star}_{k2} > Q_2$ and $\frac{WL_{ka}(Q_2)}{P_k(Q_2)C_k(Q_2)} = \frac{\beta(1-u_k)}{\mu_k(Q_2)}X_k(Q_2)$.

For general Q_n and Q_{n+1}, the same argument follows. The difference is that for $n > 1$, the incentive to mimic higher quality firm is larger due to the curvature of Kimball aggregator. Let $X_k(Q_n) = \left(\frac{Q^{\star}_{Kn}}{Q_n}\right)^{\frac{1}{\beta}}$ be the information distortion of quality Q_n, then $X_k(Q_n)$ is positively correlated with Lagrangian multiplier λ_{kn}

$$\left[1 - \frac{1}{X_k(Q_n)}\right] = \frac{\lambda_{kn}}{1-\lambda_{kn}}\left[e^{\frac{\Delta}{\beta}} - 1\right] \tag{1.49}$$

As the incentives to mimic higher quality firms increase with quality type Q_n, λ_{kn} will be higher which translates into a rising $X_k(Q_n)$. I prove this result in Proposition 2.

1.9.7.4 Proof of Proposition 2

Proof. I prove the Proposition 2 in two steps. In the first step, I show that if $\frac{\Pi(Q_{i+1};Q_i)}{\Pi(Q_i;Q_i)} > 1$, then $\frac{\Pi(Q_{i+m+1};Q_{i+m})}{\Pi(Q_{i+m};Q_{i+m})} > 1$ where m is an integer and $m > 1$. Namely, if type Q_i firms have incentive to mimic Q_{i+1} firms, then type Q_{i+m} firms have incentive to mimic Q_{i+m+1} firms. In the second step, I show that if

$X(Q_i) > 1$ and $Q_j > Q_i$, then $X(Q_j) > X(Q_i) > 1$.

First, when Δ is small, I study the dynamics of $\frac{\Pi(Q_{i+1};Q_i)}{\Pi(Q_i;Q_i)}$ using the approximation result in 1.44. I denote the right hand side of equation 1.44 as RHS_n and then differentiate it with respect to Q_n. The dynamics of RHS_n depends on the level of u

$$\frac{\partial RHS_n}{\partial Q_n} = \frac{\partial RHS_n}{\partial \rho_n}\frac{\partial \rho_n}{\partial Q_n} \propto -\left(\Delta u - \Delta^2 \frac{1-u}{4\beta} - 2\rho_n \Delta^2 \left(\frac{3}{4} - u\right)\right). \quad (1.50)$$

First, when $u \le \frac{\Delta}{4\beta+\Delta}$, $\frac{\partial RHS_n}{\partial \rho_n} < 0$ and $RHS_n < 1$. Thus, IC conditions will not be binding when u is small. When $\frac{\Delta}{4\beta+\Delta} < u < \frac{3}{4}$, there could be a $\rho^\star > 0$ such that $\frac{\partial RHS_n}{\partial \rho_n}|_{\rho^\star} = 0$ and $RHS|_{\rho^\star} > 1$.[68] RHS_n is increasing with Q_n when $Q_n < Q^\star$ and decreasing towards one when $Q_n > Q^\star$.[69] When $u \ge \frac{3}{4}$, $RHS_n > 1$ and RHS_n is decreasing towards one. Thus, if $RHS_i > 1$, then $RHS_{i+m} > 1$ for a finite $m > 0$.

I prove the second step by contradiction. From the results in the first step, if $X(Q_i) > 1$, then $X(Q_{i+m}) > 1$. Suppose there is an integer $m > 0$ such that $X(Q_{i+m}) = 1$. Without loss of generality, suppose $X(Q_{i+m-1}) > 1$, then $RHS_{i+m-1} > 1$ which contradicts with $X(Q_{i+m}) = 1$. Therefore, IC conditions are binding for all firms with $Q > Q_i$. Combining the results in equation 1.34 and 1.35, it implies

$$\begin{aligned} 1 = \frac{\Pi(Q_{n+1};Q_n)}{\Pi(Q_n;Q_n)} &\approx exp(\Delta)^{\rho_n}\left(\frac{X(Q_{n+1})}{X(Q_n)}\right)^{\beta(1-u)\rho_n} \\ &\cdot \frac{1-\alpha\left(1-\frac{1}{\sigma_n}\right) - \left(1-\frac{1}{\sigma_n}\right)\beta(1-u)exp\left(\frac{\Delta}{\beta}\right)X(Q_{n+1})}{1-\alpha\left(1-\frac{1}{\sigma_n}\right) - \left(1-\frac{1}{\sigma_n}\right)\beta(1-u)X(Q_n)}. \end{aligned} \quad (1.51)$$

Equation 1.51 implicitly determines $X(Q_{n+1})$ as a function of $X(Q_n)$ and $X(Q_{n+1})$ is increasing in $X(Q_n)$, i.e., $\frac{\partial X(Q_{n+1})}{\partial X(Q_n)} > 0$. Denote the function $X(Q_{n+1}) = F(X(Q_n))$. Suppose the graph of F is not always above 45-degree line in the space of $X(Q_n)$ and $X(Q_{n+1})$. Then there exists a steady state $X^\star$ such that $X^\star = F(X^\star)$, which implies

$$1 = \frac{\Pi(Q_{n+1};Q_n)}{\Pi(Q_n;Q_n)} \approx exp(\Delta)^{\rho_n}\frac{1-\alpha\left(1-\frac{1}{\sigma_n}\right) - \left(1-\frac{1}{\sigma_n}\right)\beta(1-u)exp\left(\frac{\Delta}{\beta}\right)X^\star}{1-\alpha\left(1-\frac{1}{\sigma_n}\right) - \left(1-\frac{1}{\sigma_n}\right)\beta(1-u)X^\star}. \quad (1.52)$$

However, as I show in the first step, RHS_n decreases with quality of firms, the

[68] When $\rho^\star > \rho_{max}$, $RHS_n > 1$ and RHS_n is decreasing towards one.

[69] $Q^\star$ is implicitly determined by $\rho^\star$. When $Q_n \longrightarrow \infty$, $RHS_n \longrightarrow 1$.

right hand side of 1.52 can not be a constant as $Q \longrightarrow \infty$. Thus, the graph of F is above 45-degree line which implies $X(Q_{n+1}) > X(Q_n)$.

$\square$

1.9.8 Robustness check

1.9.8.1 Importance of u and ξ

In this section, I study the role of u and ξ in shaping the results. To illustrate this, consider an one-industry version of the model which highlights the mechanism of the model. Specifically, I calibrate β, u, ξ and σ to match the aggregate marketing intensity, average firm size distribution, aggregate markup and the productivity losses from information frictions as in the benchmark calibration. Starting from this allocation, I change u and ξ separately to evaluate the importance of these two parameters in driving the results.

Table 1.12 reports the productivity loss from information frictions for different u and ξ. The first column reports the productivity loss from this calibrated one-industry model. Recall that $1 - u$ determines how much consumers value the appeal in the utility function. It is clear that as appeal becomes more important, the productivity loss becomes smaller. When u is small enough, the quality information is no longer valuable for the consumer. Thus, complete information allocation would also be incentive-compatible. Similarly, the tail parameter ξ determines the amount of differentiation in the industry. As ξ becomes larger, the amount of differentiation becomes lower and hence the loss from signaling is lower. Besides, the MPCR-markup elasticity also depends on the level of u and quality differentiation between firms. An increase in u or ξ tends to increase the MPCR-markup elasticity.

Table 1.12: Importance of u and ξ

productivity loss compared to complete info			
u	$u = 0.963$	$u = 0.5$	$u = 0.9$
	-2.66	-0.23	-1.97
ξ	$\xi = 3.51$	$\xi = 2$	$\xi = 5$
	-2.66	-2.86	-2.47
MPCR-markup elasticity			
u	$u = 0.963$	$u = 0.5$	$u = 0.9$
	1.858	0.436	1.371
ξ	$\xi = 3.51$	$\xi = 2$	$\xi = 5$
	1.858	0.552	5.498

1.9.8.2 Step size of quality ladder Δ

In the calibration section, I set the step size of the quality ladder $\Delta = 0.02$. In this subsection, I re-calibrate the model and jointly estimate Δ along with parameters $\{\beta, u_a, u_b, \xi_\mu, \xi_\sigma, \sigma\}$. Table 1.13 shows that the parameterization with calibrated Δ yields similar results.

1.9.8.3 Oligopolistic competition

An alternative specification with variable demand elasticity is the model of oligopolistic competition with nested CES demand as in Atkeson and Burstein (2008). This section presents the results using this demand system and show that the result is robust to this alternative specification.

In the model, a representative consumer derives utility from the aggregate consumption C, which combines a continuum of sectoral consumption good C_s using a CES preference with elasticity θ [70]

$$C = \left(\int_0^1 C_s^{\frac{\theta-1}{\theta}} ds \right)^{\frac{\theta}{\theta-1}} \tag{1.53}$$

The sectoral consumption good C_s is produced by aggregating the finite num-

[70] The sector in this model corresponds narrowly-defined industries, thus $\gamma > \theta > 1$. For more details of this model, please refer to Atkeson and Burstein (2008) and Edmond et al. (2015).

Table 1.13: Results with calibrated Δ

Panel A: Calibrated parameters

		benchmark	calibrated Δ
σ	average elasticity	13.52	13.30
β	returns to scale of marketing	0.054	0.052
u_a	shape parameter of Beta distr. u_k	0.047	0.052
u_b	shape parameter of Beta distr. u_k	0.009	0.016
ξ_μ	location parameter of Lognormal distr. ξ_k	0.28	0.27
ξ_σ	shape parameter of Lognormal distr. ξ_k	2.21	2.48
Δ	step size of quality ladder	0.020	0.019

Panel B: Results

Increase in β	0	$\uparrow$ 10%	$\uparrow$ 20%	$\uparrow$ 30%
productivity loss compared to complete info				
benchmark	-2.67	-2.93	-3.18	-3.42
calibrated Δ	-2.24	-2.45	-2.66	-2.86
implied marketing intensity, percentage				
benchmark	3.00	3.29	3.58	3.86
calibrated Δ	3.05	3.35	3.64	3.93
MPCR-markup elasticity				
benchmark	1.67	1.73	1.79	1.84
calibrated Δ	1.43	1.48	1.53	1.57

ber of varieties with each sector using a CES preference with elasticity γ

$$C_s = \left(\sum_{i=1}^{n} (Q_s(i)^u \Phi_s(i)^{1-u} C_s(i))^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}} \tag{1.54}$$

where $\Phi_s(i)$ is the firm appeal of firm i in sector s, and $C_s(i)$ is the consumption from the firm.

Each firm in the sector s competes on quantity (Cournot competition), chooses quantity $C_s(i)$ and appeal $\Phi_s(i)$ to maximize the profit. They internalize the impact of their decisions on sectoral consumption C_s. The FOC associated with LCSE resembles the one in the model with Kimball aggregator

$$\frac{WL_{sp}(Q_s(i))}{P_s(i)Y_s(i)} = \frac{\alpha}{\mu_s(i)} \tag{1.55}$$

$$\frac{WL_{sa}(Q_s(i))}{P_s(i)Y_s(i)} = \frac{\beta(1-u)}{\mu_s(i)} X(Q_s(i)) \tag{1.56}$$

where $X(Q_s(i)) = \left(\frac{Q_s(i)^\star}{Q_s(i)}\right)^{\frac{1}{\beta}}$. Compared to the monopolistic competition model, the difference lies in how markup depends on firm size. In the oligopoilstic competition model, markup is determined by the within-sector sales share $\omega_s(i)$

$$\mu_s(i) = \frac{1}{1 - \left[\frac{1}{\gamma}(1-\omega_s(i)) + \frac{1}{\theta}\omega_s(i)\right]} \tag{1.57}$$

On the other hand, with monopolistic competition, markup is determined by the relative size $q_k(Q)$. Nevertheless, they both capture the same idea that demand is less elastic for larger firms.

In this environment, LCSE allocation can still be solved by starting the complete information allocation. Specifically, I begin with the sales share of lowest quality firms $\omega_s(Q_1)$ and $X_s(Q_1) = 1$, and solve $\{\omega_s(Q_n), X_s(Q_n)\}$ iteratively. We repeat the procedure until $\sum \omega(Q_n)M(Q_n) = 1$.

Calibration. I first set the number of industries to be 1000 and the number of firms within each industry to be $N_s = 910$.[71] I choose $\theta = 1.03$ and $\gamma = 8.86$. I set $u = 0.963$ which corresponds to the calibration of one-sector model and choose $\beta = 0.046$ and $\xi = 5.8$ to match the aggregate marketing intensity and average distribution of firm size.

Panel A of Table 1.14 compares the firm size distribution in this model with

[71] It corresponds the number of 6-digit sectors and median number of firms within each sector in the mid-1990s.

Table 1.14: Oligopolistic competition

Panel A: Moments

	data	benchmark	oligopoly
avg fraction of firms with rel payroll			
< 0.5	0.662	0.694	0.685
< 1	0.819	0.895	0.873
< 5	0.962	0.985	0.984
< 10	0.987	0.990	0.993
< 100	1	0.998	0.999

Panel B: Results

Increase in β	0	$\uparrow$ 10%	$\uparrow$ 20%	$\uparrow$ 30%
productivity loss compared to complete info				
benchmark	-2.67	-2.93	-3.18	-3.42
oligopoly	-3.00	-3.31	-3.63	-3.94
implied marketing intensity, percentage				
benchmark	3.00	3.29	3.58	3.86
oligopoly	3.00	3.28	3.59	3.90
MPCR-markup elasticity				
benchmark	1.67	1.73	1.79	1.84
oligopoly	2.18	2.21	2.24	2.27

the data and benchmark calibration. The oligopoly model implies a firm size distribution close to the benchmark calibration. Panel B of Table 1.14 reports the productivity loss compared to complete information allocation. The loss from information frictions is slightly higher compared to the level in the benchmark calibration, and the loss from information frictions is increasing with returns to marketing technology. The model also implies a positive MPCR-markup elasticity and its co-movement with aggregate marketing intensity. The main difference is that the oligopolistic competition involves a level of elasticity higher than the benchmark. Since the level of information distortion remains roughly the same, the reason for this difference is the lack of variation in markup. Given $\mu_i(s) = \frac{1}{1-\left[\frac{1}{\gamma}(1-\omega_i(s))+\frac{1}{\theta}\omega_i(s)\right]}$, for an average industry, the variation in markup is lower compared to the model with monopolistic competition. Since there is a finite number of firms within each sector, the same level of markup dispersion as in the monopolistic competition model would generate an aggregate markup which is too large.

References

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy*, 118(6):1151–1199.

Atkeson, A. and Burstein, A. (2008). Pricing-to-Market, Trade Costs, and International Relative Prices. *American Economic Review*, 98(5):1998–2031.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2017). The Fall of the Labor Share and the Rise of Superstar Firms. Technical report, National Bureau of Economic Research.

Bagwell, K. (2007). The economic analysis of advertising. *Handbook of industrial organization*, 3:1701–1844.

Bagwell, K. and Ramey, G. (1988). Advertising and limit pricing. *The Rand Journal of Economics*, pages 59–71.

Baqaee, D. R. and Farhi, E. (2018). Productivity and Misallocation in General Equilibrium. Technical report, National Bureau of Economic Research.

Barkai, S. (2017). Declining labor and capital shares. *University of Chicago*.

Belleflamme, P. and Peitz, M. (2015). *Industrial Organization: Markets and Strategies*. Cambridge University Press.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. (2010). What's Advertising Content Worth? Evidence from A Consumer Credit Marketing Field Experiment. *The Quarterly Journal of Economics*, 125(1):263–306.

Brandt, L., Van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9):2784–2820.

Cho, I.-K. and Kreps, D. M. (1987). Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics*, 102(2):179–221.

Cho, I.-K. and Sobel, J. (1990). Strategic Stability and Uniqueness in Signaling Games. *Journal of Economic Theory*, 50(2):381–413.

Corrado, C., Hulten, C., and Sichel, D. (2005). Measuring capital and technology: an expanded framework. In *Measuring capital in the new economy*, pages 11–46. University of Chicago Press.

Corrado, C. A. and Hao, J. X. (2014a). Brands As Productive Assets: Concepts, Measurement, and Global Trends. WIPO Economic Research Working Papers 13, World Intellectual Property Organization - Economics and Statistics Division.

Corrado, C. A. and Hao, J. X. (2014b). Intangible Investment Worldwide: Determinants and Properties. In *Proc. Joint Statistical Meetings 2014 - Business and Economic Statistics Section*.

De Loecker, J., Eeckhout, J., and Unger, G. (2018). The Rise of Market Power and the Macroeconomic Implications. Technical report, National Bureau of Economic Research.

De Loecker, J. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–71.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, Markups, and the Gains from International Trade. *American Economic Review*, 105(10):3183–3221.

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). How Costly are Markups? Technical report, National Bureau of Economic Research.

Goldfarb, A. (2014). What is Different about Online Advertising? *Review of Industrial Organization*, 44(2):115–129.

Goldfarb, A. and Tucker, C. E. (2011). Privacy Regulation and Online Advertising. *Management science*, 57(1):57–71.

Gopinath, G. and Itskhoki, O. (2010). Frequency of Price Adjustment and Pass-Through. *The Quarterly Journal of Economics*, 125(2):675–727.

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

Haskel, J. and Westlake, S. (2018). *Capitalism without Capital: The Rise of the Intangible Economy*. Princeton University Press.

Hottman, C. J., Redding, S. J., and Weinstein, D. E. (2016). Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364.

Klenow, P. J. and Willis, J. L. (2016). Real Rigidities and Nominal Price Changes. *Economica*, 83(331):443–472.

Nelson, P. (1974). Advertising as Information. *Journal of Political Economy*, 82(4):729–754.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2):307–342.

Ptok, A., Jindal, R. P., and Reinartz, W. J. (2018). Selling, General, and Administrative Expense (SGA)-based Metrics in Marketing: Conceptual and Measurement Challenges. *Journal of the Academy of Marketing Science*, 46(6):987–1011.

Ramey, G. (1996). D1 Signaling Equilibria with Multiple Signals and A Continuum of Types. *Journal of Economic Theory*, 69(2):508–531.

Rauch, F. (2013). Advertising Expenditure and Consumer Prices. *International Journal of Industrial Organization*, 31(4):331–341.

Stigler, G. J. and Becker, G. S. (1977). De Gustibus Non Est Disputandum. *The American Economic Review*, 67(2):76–90.

Traina, J. (2018). Is Aggregate Market Power Increasing? Production Trends Using Financial Statements. Technical report, National Bureau of Economic Research.

Chapter 2

Zombie Firms, State Subsidies, and Resource Misallocation

2.1 Introduction

Resource misallocation lowers aggregate productivity (Hsieh and Klenow, 2009). A growing body of recent research focuses on one particular channel that results in resource misallocation — zombie congestion.[1] The existence of zombie firms not only distorts resource allocation across operating firms but also changes the composition of the operating firms through the entry selection. Given the low productivity nature of zombie firms, the excessive resource employed by zombie firms seems puzzling. In this chapter, I investigate the role of government subsidies in explaining this phenomenon and quantitatively examine the effectiveness of different policies in resolving zombie problem and boosting aggregate productivity.

I use the firm-level data from the manufacturing sector in China between 1998 and 2007. Following the literature, I classify a firm as a zombie if its profit net of government subsidy is negative for two consecutive periods and its age is greater than five years.[2] The measure intends to identify those firms with persistent problems in making positive profits without the help of government. During the sample period, zombie firms are prevalent, but the share of zombie firms is diminishing. Zombie rate declines from 16% in 1999 to 5% in 2007, and about half of the decline can be attributed to the rise of the private sector

[1]See Blattner et al. (2018), Kwon et al. (2015), and McGowan et al. (2018) for example.

[2]The zombie measure used in this paper follows recent literature on zombie firms. See McGowan et al. (2018) and Chang et al. (2020) for example.

in the economy. Besides, zombie firms are larger in terms of inputs employed but less productive. In 1999, an average zombie firm used 40% more labor inputs than a non-zombie firm, but its TFP is more than 50% lower than that of non-zombie. The magnitude of gaps in size and productivity remains even after controlling for ownership, industries, and provinces.

Using the unique subsidy information in the dataset, I provide direct evidence on the existence of tax-like wedges that reallocate resources towards zombie firms. A zombie firm is ten percentage points more likely to receive a positive subsidy from the government, and the average subsidy rate for zombie firms is much higher than that of non-zombie firms. In fact, rescuing firms in financial distress is one explicit goal of the state subsidies in China, especially for large firms which are vital in terms of local employment and social stability (Lee et al., 2014; Chen et al., 2008).

The difference in the average subsidy rate between zombie and non-zombie firms not only reflects the underlying joint distribution of subsidy rate and productivity, but also the measure I use to identify zombies. In the data, firms with lower productivity tend to receive a higher subsidy rate. This negative correlation increases the average subsidy rate for a group of low productivity firms. In addition, the zombie measure based on profit net of subsidy would also disproportionately identify firms with low productivity and high subsidy rate as zombies. These forces together result in a large gap in subsidy rate between zombies and non-zombies.

Motivated by the features of zombies, I develop a model with heterogeneous firms to account for these features, and then use the model to quantify the impact of zombies on aggregate productivity. The model resembles the one in Restuccia and Rogerson (2008), where the firm draws a joint pair of productivity and subsidy rate upon entry. The existence of subsidy rate generates the dispersion of marginal products across firms and allows us to capture the fact that zombie firms are larger yet less productive. Besides, zombie measure based on profit net of subsidy endogenously identifies firms with low productivity and high subsidy rate as zombies. Following Yang (2019), I assume productivity and subsidy rate are jointly normally distributed, and calibrate the model to match the moments regarding zombie rate, relative mean size, and size dispersion of zombie firms in 1999.

With the calibrated model, I conduct three quantitative exercises. Firstly, I study the productivity gain from removing dispersion in subsidy rate completely. The aggregate productivity can increase by more than 280% if we move to a frictionless economy. This number tends to be larger than the one estimated in Hsieh and Klenow (2009). The reason is that they conservatively choose a low value of elasticity of substitution between firms, which results in a lower gain from reallocation. Besides, I find that the gain in productivity is primarily driven by the reduction of intensive margin misallocation. On the extensive margin, the positive selection effect that increases average productivity of operating firms is almost fully counteracted by the negative variety effect. As noted by Fattal Jaef (2018), in the model with frictions, negative correlation between subsidy rate and productivity reduces the labor demand for high productivity firms and hence wage rate in the economy, which makes entry profitable for more firms.

Empirically, the rising share of the private sector reduces the dispersion of the subsidy rate. I then use the model to evaluate the gain in productivity associated with the rise of private sector. Reducing the dispersion of subsidy to match half of the decline in the zombie rate, the aggregate productivity increases by about 40%. This quantitative exercise illustrates that the rise of private sector can induce a large productivity gain even when there is no productivity gap between private firms and SOEs.

Lastly, I use the model to evaluate the effectiveness of policies that increase the exit rate of zombies. This exercise is of particular interest given that the Chinese government recently took the initiative to deal with zombie problem by closing those firms.[3] Increasing the zombie exit rate can significantly reduce the zombie rate in the economy, while its productivity effect is limited. Compared with the large productivity gain from reducing the dispersion of subsidy, this result implies that policies should tackle the resource misallocation in the economy directly.

Related literature

This chapter belongs to a large growing body of literature on resource misallocation, including seminal papers such as Hsieh and Klenow (2009) and Restuccia and Rogerson (2008). The correlation between size and productivity I stress in this chapter has been shown by Bartelsman et al. (2013) as a robust measure for resource misallocation across countries. They assume tax-like wedges to explain the observed level of correlation between productivity

[3]See the recent report from Xinhua News: `http://english.www.gov.cn/statecouncil/ministries/201910/18/content_WS5da91079c6d0bcf8c4c15556.html`.

and size. I contribute to this literature by providing direct evidence on the existence of such wedges — state subsidy. In addition, I show that the empirical distribution between productivity and subsidy rate differs for zombies and non-zombies. It offers direct support for the model where each firm draws a pair of productivity and subsidy rate to enter the market, e.g., Restuccia and Rogerson (2008) and Yang (2019).

The majority of existing literature that studies zombie firms focuses on the effects of zombie congestion in Japan during the 1990s. Caballero et al. (2008) find that zombie congestion reduces both investment and employment growth for healthy firms in the industry. Kwon et al. (2015) show that without zombie lending, the aggregate productivity would have grown about one percentage point faster annually during the 1990s. This paper examines the effectiveness of policies that increase the zombie exit rate, and finds a relatively limited productivity effect.

Recent empirical studies on zombie firms in China stress the role of state subsidy in the formation of zombie firms (He et al., 2018; Chang et al., 2020). I quantitatively assess the productivity effect of dispersion in subsidy rate. In addition, I document that the rising share of private firms reduces the dispersion of subsidy rate across firms, which could lead to sizable productivity gains. This complements the existing studies focusing on the productivity gap between SOEs and private firms, such as Song et al. (2011).

The chapter is organized as follows. Section 2.2 shows empirical findings. Section 2.3 presents the model, which is calibrated in section 2.4. Section 2.5 discusses the results from numerical exercises. Section 2.6 concludes.

2.2 Empirical findings

2.2.1 Data and measurement

I use firm-level data from the manufacturing sector in China. It contains all non-state-owned firms (non-SOEs) with more than 5 million yuan in revenue and all state-owned firms (SOEs) from 1998 to 2007.[4] From the data, I extract detailed information on firm operations, including asset, interest payment, short-term and long-term liability, profit, and subsidy.

[4]See Appendix for the detailed discussion of the data set.

I combine information on reported value of output, intermediate input, and employment to estimate firms' TFP, which is calculated using the index approach following Brandt et al. (2012)

$$ln\ TFP_{it} = q_{it} - \overline{q_t} - \widetilde{S_{it}}(l_{it} - \overline{l_t}) - (1 - \widetilde{S_{it}})(k_{it} - \overline{k_t})$$

where q_{it}, l_{it} and k_{it} are the logarithms of real value added, employment and real capital stock of firm i at year t. The overlined variables correspond to the industry average. The weight on the labor input is $\widetilde{S_{it}} = (S_{it} + \overline{S_t})/2$, where S_{it} is the labor share of firm i and $\overline{S_t}$ is the industry average labor share. This approach allows us to compare each firm with a hypothetical industry average firm. The advantage of this approach is that it allows technological differences across firms within the same industry.[5]

Past studies of zombie firms used different measurements to identify zombies. Broadly, those measures can be classified into two categories. The first approach focuses on the lending aspect of zombies. It involves the construction of a hypothetical minimum interest payment and compares it with actual interest payment of the firm (Caballero et al., 2008), or earnings of the firm (Fukuda and Nakamura, 2011). The second approach is to identify zombies as firms with persistent financial difficulties using profitability measures, such as measures proposed by McGowan et al. (2018) and the Chinese government.

I follow the second approach and classify a firm as a zombie if (i) its profit net of subsidy is negative for two consecutive periods, and (ii) its age is greater than five years. The first requirement detects whether firms have persistent problems in making positive profits without the help of government. As emphasized by recent papers that study zombie firms in China, state subsidy is an important friction that helps low productivity firms to survive (Chang et al., 2020; He et al., 2018). Using profit net of subsidy can better identify those firms that rely on the government to make profits. The second restriction rules out young firms at the start of their life cycle as it can take a while to make profits. The choice of measurement is driven by data availability. Section 2.2.4 discusses alternative measures of zombie firms in details.

In theory, zombie firms should be those firms that cannot survive in a friction-

[5]Given I do not observe firm-level price, the measure of TFP is revenue-based. Relying on a very restrictive model specification, Hsieh and Klenow (2009) (HK) interpret the dispersion in revenue-based TFP (TFPR) as firm-level distortions. However, as emphasized by Haltiwanger et al. (2018), the strict assumptions underlying HK's method do not hold in the data. Besides, empirical literature finds that physical TFP (TFPQ) is highly correlated with TFPR (Foster et al., 2008, 2016; Eslava et al., 2013).

Figure 2.1: Zombie rate in China

Notes: This figure plots the zombie rate in China during the sample period. Labor (capital) zombie rate is calculated as the fraction of total amount of labor (real capital stock) employed by zombie firms.

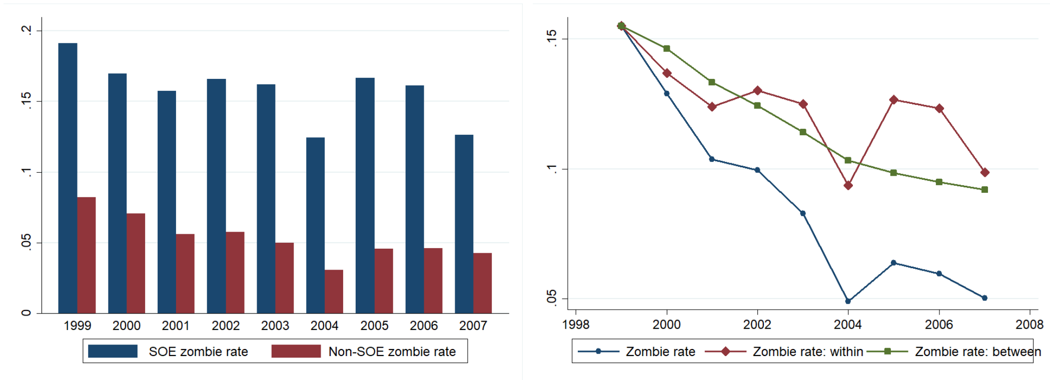
less environment. The measure is admittedly imperfect. In principle, I would commit two types of classification errors. Some firms with poor performance will be categorized as non-zombies if they receive debt forgiveness or subsidized loans from the bank or government. Thus, the measure would underestimate the extent of zombies in the economy. On the other hand, I may also potentially misclassify a healthy firm as a zombie. However, given the fast growth of the Chinese economy during the sample period, the probability that a healthy firm suffering persistent profitability problems is low. Therefore, the measure is more likely to underestimate the extent of zombies in China.

2.2.2 Characteristics of zombie firms

Figure 2.1 illustrates the zombie rate in China during the sample period. Since zombie firms must have negative net of subsidy profits for two consecutive periods, the measurement starts from 1999. Three measures are presented in the figure: zombie rate, labor zombie rate, and capital zombie rate. The zombie rate is calculated as the fraction of firms that are zombies according to our measure. Labor (capital) zombie rate is calculated as the share of the total amount of workers (real capital stock) employed by zombie firms. Several facts arise from the figure.

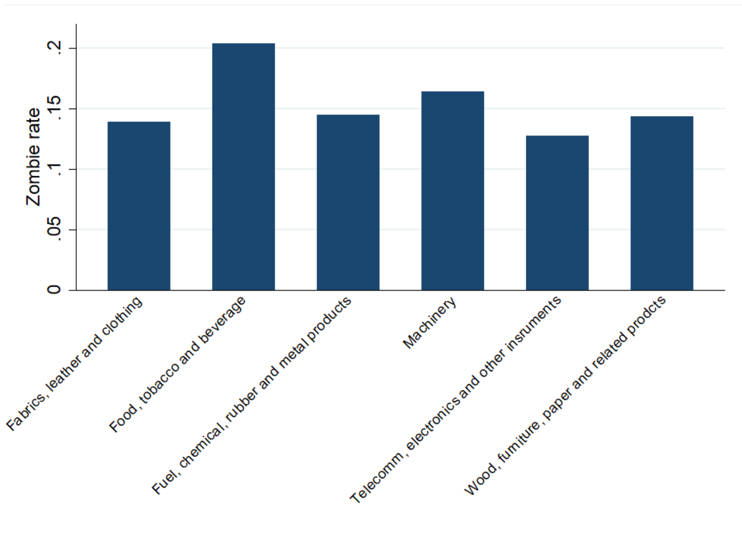
Firstly, zombie firms are prevalent. In 1999, the fraction of zombie firms is about 16%. Zombie firms are not just SOEs. The left panel of Figure 2.2

Figure 2.2: Zombie rate: SOEs vs Non-SOEs



Notes: The left panel of the figure plots the zombie rate for SOEs and non-SOEs in China during the sample period. The right panel of the figure conducts the between-group and within-group decomposition for aggregate zombie rate, where firms are classified into two groups based on SOE status.

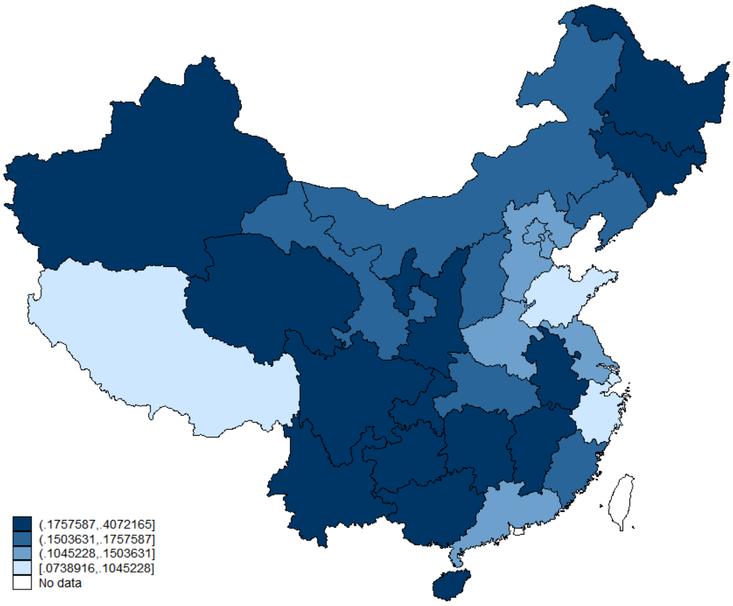
Figure 2.3: Zombie rate by industry groups in 1999



reports the zombie rate for SOEs and non-SOEs. Although the zombie rate is higher for SOEs, the zombie rate for non-SOEs still exceeds 8% in 1999. The high zombie rate is not driven by a few industries or a few regions. Figure 2.3 presents the zombie rate by industry groups in 1999. Food, tobacco and beverage sector tends to have a relatively higher zombie rate. Figure 2.4 depicts the geographic variation of zombie rates. The zombie rate is higher (lower) in inland (coastal) areas.

Secondly, across all three measures of zombie rate shown in Figure 2.1, the extent of zombies in the economy is diminishing. The share of zombie firms declines from 16% in 1999 to 5% in 2007. The left panel of Figure 2.2 shows that the zombie rates for both SOEs and non-SOEs are decreasing over time.

Figure 2.4: Zombie rate by provinces in 1999



The right panel of Figure 2.2 decomposes the decline of zombie rate into within-group and between-group components. The within-group component keeps the fraction of SOEs fixed as in 1999 and allows the zombie rate for SOEs and non-SOEs to evolve. The between-group component keeps the zombie rate for two groups fixed as in 1999 and allows the fraction of SOEs to change. The decomposition illustrates that, by the year 2007, about half of the decline in zombie rate is driven by the rapid rise of non-SOE sector.[6] Similar decompositions are conducted across industries and provinces, where I find the decline of zombie rate is almost entirely driven by the within-industry (province) component.[7]

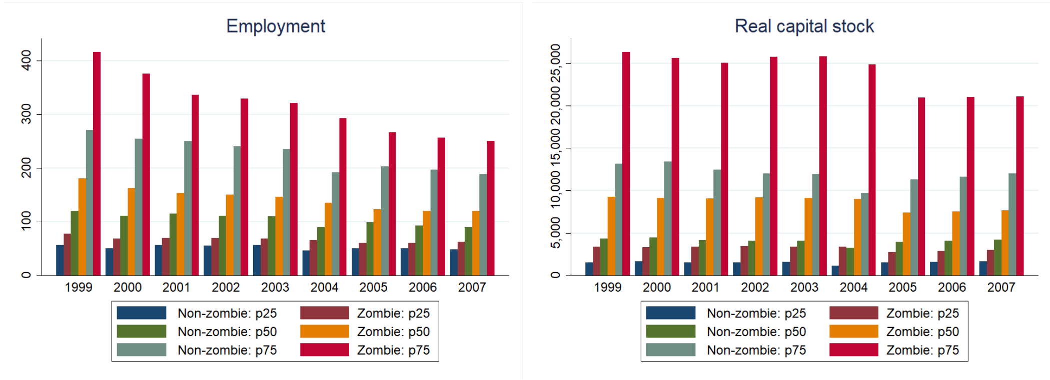
Lastly, Figure 2.1 shows that zombie firms are larger since zombie firms employ more labor and capital inputs than non-zombies. In 1999, the fraction of labor utilized by zombie firms is about 21%, which implies that an average zombie firm is 40% larger than a non-zombie firm in terms of workforce. The average size difference is not driven by a few outliers. Figure 2.5 plots the size distribution of zombies and non-zombies in terms of employment and capital stock. Across different quartiles, zombie firms are larger than non-zombies. In Appendix, I report the regression-based size gap between the zombie and non-zombie firms, a size gap of similar magnitude remains even if I control for ownership, industries, and provinces.[8]

[6]In the sample, the fraction of non-SOEs increases from 33% in 1999 to more than 90% in 2007. See Figure 2.20 in Appendix.

[7]See Figure 2.21 in Appendix.

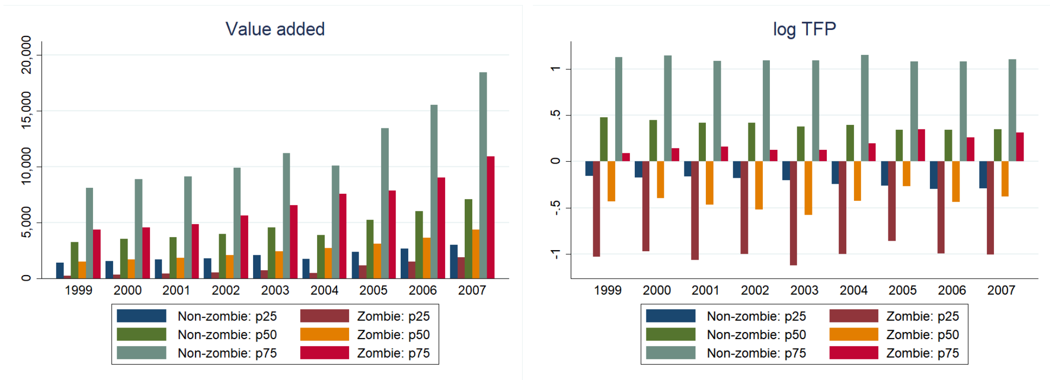
[8]See Figure 2.22 in Appendix.

Figure 2.5: Size distribution: zombies vs non-zombies



Notes: This figure plots different quartiles in the size distribution of zombie and non-zombie firms, where size is measured by employment (real capital) in the left (right) panel.

Figure 2.6: Output and productivity: zombies vs non-zombies



Notes: This figure plots different quartiles in the output and productivity distribution of zombie and non-zombie firms, where output (productivity) is measured by value added (logarithm of TFP).

Zombie firms employ more resources, yet they produce much less output compared to non-zombie firms. The left panel of Figure 2.6 reports the size distribution in terms of value added. In contrast with Figure 2.5, non-zombie firms are larger in terms of value added across quartiles. The right panel of Figure 2.6 further plots the distribution of productivity. In 1999, the median productivity of non-zombies is more than 100% higher than that of zombies. A productivity gap of similar magnitude remains even if I control for ownership, industries, and provinces.

2.2.3 Zombies and state subsidies

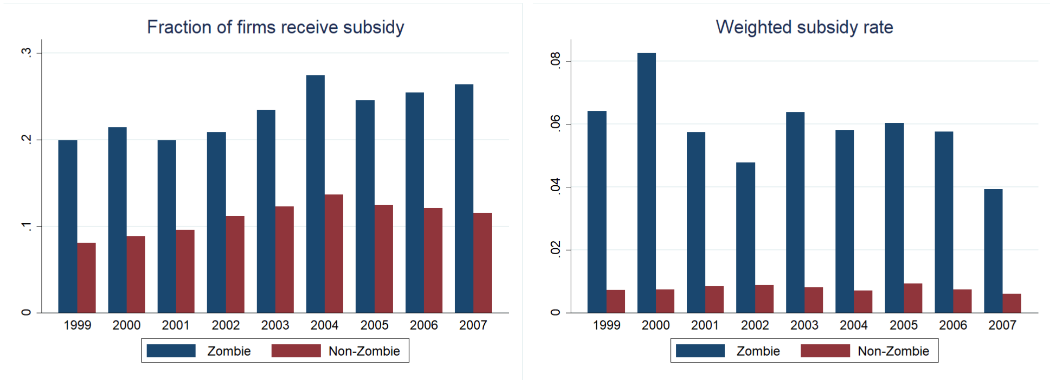
Zombie firms are larger yet less productive than non-zombies. A frictionless model where size and productivity are perfectly correlated cannot explain this fact. In the literature, researchers assume tax-like wedges to explain the observed level of correlation between productivity and size (Bartelsman et al., 2013). However, there is little evidence on the existence of such wedges. The unique feature of Chinese data is that I observe the amount of state subsidy each firm receives. I calculate the firm-level subsidy rate as the ratio of subsidy to value added.

The left panel of Figure 2.7 reports the fraction of firms that receive the subsidy. Zombie firms are more likely to receive subsidies. In the right panel, I plot the value added weighted average of firm-level subsidy rate. The subsidy rate for zombie firms is about six percentage points higher than that of non-zombies in 1999. The subsidy is not intended to rescue SOEs or a few industries only. The left panel of Figure 2.8 shows that zombie firms are ten percentage points more likely to receive subsidy after controlling for ownership, industries, and provinces. Conditional on receiving a subsidy, the subsidy rate of zombie firms is 50% higher than non-zombie firms in 1999 (the right panel of Figure 2.8).

With the firm-level data on subsidy rate and productivity, I could further explore the underlying relationship between these two variables. The left panel of Figure 2.9 shows a binscatter plot between subsidy rate and TFP in 2000, where both variables are residualized by ownership, industry and province dummies.[9] It is clear that subsidy rate and TFP are negatively correlated at the firm level. The dispersion of subsidy rates will reallocate resources from high productivity firms to low productivity firms, resulting in the fact that low

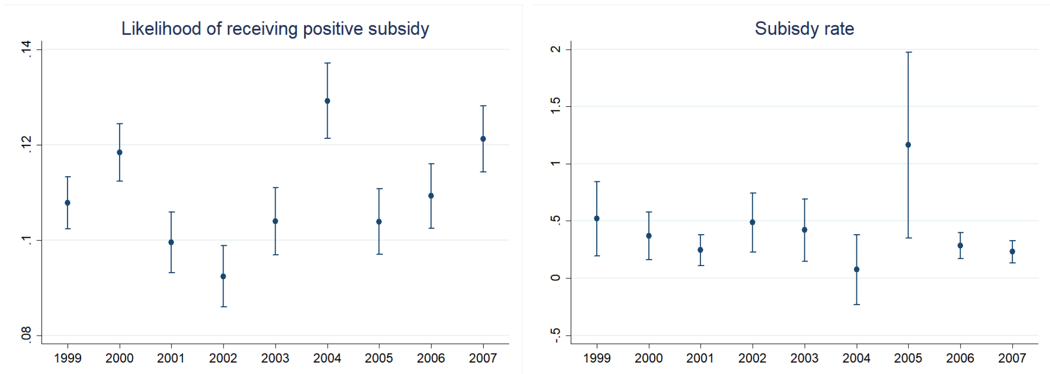
[9] I report the result for the year 2000 to facilitate comparison with alternative zombie measures in section 2.2.4. The similar pattern is documented across all sample periods.

Figure 2.7: Subsidy: zombies vs non-zombies



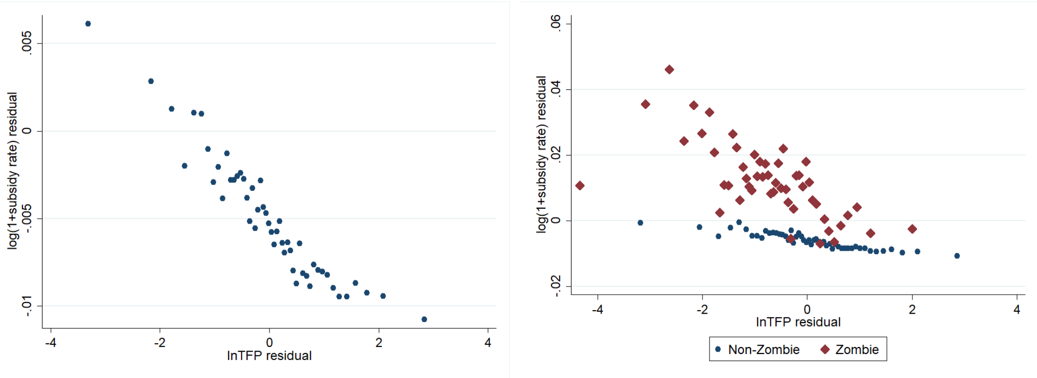
Notes: The left panel of this figure plots the fraction of firms receiving positive subsidy for zombie and non-zombie firms. The right panel shows the value added weighted average of subsidy rate for zombie and non-zombie firms.

Figure 2.8: Regression-based subsidy gap for zombie firms



Notes: The left panel of this figure plots the additional likelihood of zombie firms receiving positive subsidy. I obtain the estimates as β_1 from the regression $D_i^S = \beta_1 D_i^Z + FEs + \epsilon_i$, where D_i^S is the dummy for receiving positive subsidy and D_i^Z is the dummy for zombie firms. Fixed effects are included to control for ownership, industries and provinces. The right panel shows the magnitude of subsidy rate difference for zombie and non-zombie firms conditional on receiving positive subsidy. I obtain the estimates as β_1 from the regression $\frac{S_i}{VA_i} = \beta_1 D_i^Z + FEs + \epsilon_i$, where $\frac{S_i}{VA_i}$ is subsidy rate. Both the point estimate and its 95% confidence interval are reported.

Figure 2.9: Subsidy rate and productivity



Notes: The left panel of this figure shows the binscatter plot between logarithm of subsidy rate and TFP, where both variables are residualized by ownership, industry and province dummies. Subsidy rate is calculated as the ratio between subsidy and value added of the firm. The right panel plots these two residuals separately for zombie and non-zombie firms. Both panels are produced by using the data in 2000.

productivity firms could employ more inputs into production.

In the right panel of Figure 2.9, I plot these residualized variables separately for zombie and non-zombie firms. Similar to previous findings, zombie firms tend to have higher subsidy rates and lower productivity. Given the negative correlation between subsidy rate and TFP shown in the left panel of Figure 2.9, a group of low productivity firms would have a higher average subsidy rate. However, this does not explain the difference in the level of subsidy rate residuals across two groups of firms. The measure based on the profit net of subsidy tends to disproportionately identify firms with low productivity and high subsidy rate as zombies. As I will discuss in section 2.3.2, this pattern is consistent with a model where firm simultaneously draws a pair of productivity and subsidy rate to enter the market.

The subsidy rate reported in Figure 2.7 is likely to be an underestimate of the extent of subsidy in China since it mainly involves the transfer of monetary assets from the government to firms directly. Some other forms of subsidy that do not imply a direct transfer of resources are not included, such as tax-breaks and low-interest loans. Aghion et al. (2015) find that more than 40% of firms receive a tax holiday between 1998 and 2007, i.e., they paid less than the statutory corporate income tax rate or statutory value added tax rate. In addition, subsidies may not come from the government directly. During the sample period, almost all the banks are state-owned, the zombie lending channel is an alternative way to rescue firms. A report by the government official Xinhua news agency reveals that up to 88% of Chinese listed firms were granted state subsidies in 2014.

2.2.4 Alternative measures of zombies

In this section, I contrast my measure of zombie firms with two alternative measures in the literature.[10] The first measure is the one proposed by the Chinese government, where firms with negative profit for three consecutive years are classified as zombie firms. The second measure is proposed by Fukuda and Nakamura (2011) (FN hereafter), where a zombie firm at period t needs to satisfy both "profitability criterion" and "evergreen lending criterion".

The "profitability criterion" means the earnings before interests and taxes (EBIT) of the firm is smaller than the minimum of required interest payment $R^*_{i,t}$ at period t

$$R^*_{i,t} = rs_{t-1}BS_{i,t-1} + \left(\frac{1}{5}\sum_{j=1}^{5} rl_{t-j}\right) BL_{i,t-1}$$

where $BS_{i,t}$, $BL_{i,t}$ are short-term, long-term loan for firm i at the end of year t; rs_t, rl_t are average short-term and long-term prime rate in year t. $R^*_{i,t}$ measures the minimum interest payment firm has to pay when there is no interest subsidies in the lending market. If EBIT is lower than $R^*_{i,t}$, then it implies the firm cannot make a profit if there are no interest subsidies. The "evergreen lending criterion" requires that debt of the firm is over half of total assets in period $t-1$ and borrowings increase in period t. In Chinese data, information on bank loan is not available, thus I use liability of the firm to measure firm's borrowing.

Figure 2.10 reports zombie rate across different classifications. Compared with the benchmark measure, FN method identifies more zombie firms, and the method proposed by Chinese government identifies fewer zombie firms. Since $R^*_{i,t}$ tends to be larger than actual interest payment, FN method imposes a higher threshold on the earning of firms and a lower requirement on the persistence of financial difficulty, which results in a higher zombie rate.[11]

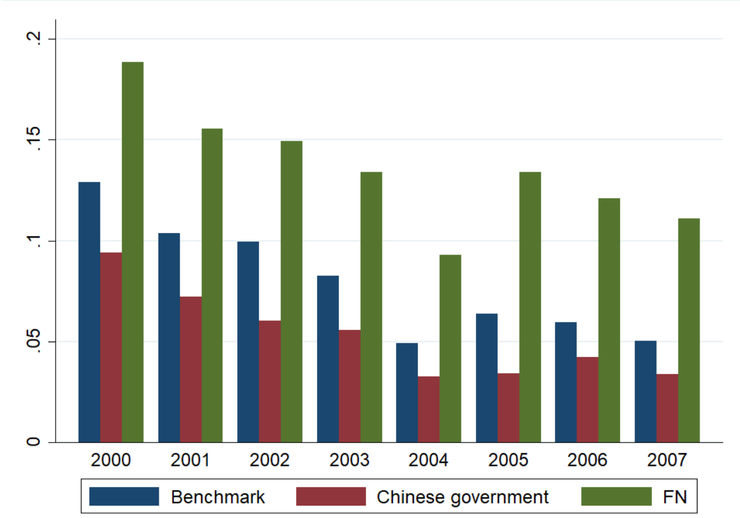
Compared with benchmark measure, the alternative measures generate similar implications on the size and productivity gap between zombie and non-zombie firms, but they have different implications on subsidy rate.[12] The two upper

[10] In Figure 2.24, I also report the zombie rate when there is no age restriction in the benchmark measure.

[11] I use the gross profit to measure firm earning. Thus, the "profitability criterion" would hold if gross profit is smaller than gap between $R^*_{i,t}$ and actual interest payment.

[12] See Figure 2.25 in Appendix for the size and productivity gap between zombie and non-zombie firms.

Figure 2.10: Alternative measures of zombie firms



Notes: This figure plots the zombie rate for alternative measures. Two alternative measures are considered: Chinese government's measure and the measure proposed by Fukuda and Nakamura (2011) (FN).

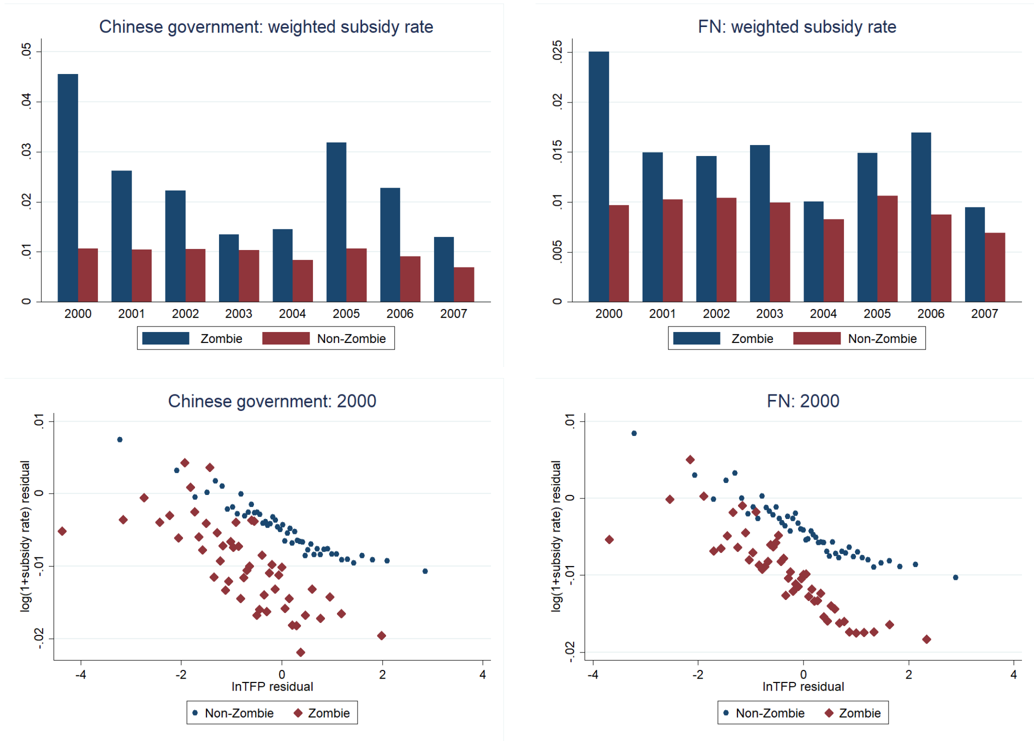
panels of Figure 2.11 show weighted subsidy rate for zombie and non-zombies for these alternative measures. On average, zombie firms receive a higher subsidy rate. However, the difference in subsidy rate between zombies and non-zombies becomes smaller compared to the right panel of Figure 2.7. This is because these two alternative measures only focus on the profit of firms (Chinese government's measure in particular). It will identify firms with lower subsidy rate and productivity as zombie firms.[13] The two lower panels of Figure 2.11 confirm this point by presenting the binscatter plot for the residuals of logarithm of TFP and subsidy rate. This selection mechanism tends to reduce the subsidy rate of zombies.

I prefer the benchmark measure over these alternative measures for two reasons. Compared with profit-based measures, profit net of subsidy is a more precise concept to identify zombies as firms that rely on government subsidies to make profits should be counted as zombie firms. Besides, since the firm-level data does not contain information on bank loans, using firm's liability to construct the minimum interest payment $R^*_{i,t}$ will overestimate the amount of interest firms need to pay as some types of liabilities are not interest-bearing, payable for instance. Huang and Chen (2017) find that, for more than 70% firms, actual interest payment is smaller than $R^*_{i,t}$.[14]

[13]FN method also relies on firm's financing structure and interest subsidies, which confound the comparison with benchmark measure.

[14]This would imply more than 70% of firms will be classified as zombies according to measure proposed by Caballero et al. (2008).

Figure 2.11: Subsidy rate and zombie firms: alternative zombie measures



Notes: The two upper panels plot the weighted subsidy rate for zombie and non-zombie firms using two alternative measures. Subsidy rate is calculated as the ratio between subsidy and value added of the firm. The two lower panels show the binscatter plot between logarithm of subsidy rate and TFP, where both variables are residualized by ownership, industry and province dummies. I plot these two residualized variables separately for zombie and non-zombie firms, where zombie firms are identified using alternative measures. Both panels are produced by using the data in 2000.

To summarize, I document the following facts. Firstly, zombie firms are prevalent, but the share of zombies is diminishing between 1998 and 2007. Secondly, zombie firms are larger in terms of inputs employed but much less productive. Thirdly, zombie firms on average have a higher subsidy rate than non-zombies. The difference in average subsidy rate depends on the underlying joint distribution of subsidy rate and productivity, and the definition of zombies. In the next section, I outline a model that can account for these features and then use it to quantify the impact of zombie firms on productivity.

2.3 Model

In this section, I consider a model of firm dynamics to quantify the productivity effects from distortion generated by zombie firms. The model builds on the standard industry dynamics model, i.e., Restuccia and Rogerson (2008). The model is stationary. Each firm i draws their productivity a_i and subsidy rate s_i from a joint distribution $G(a,s)$ upon entry. The individual draw of $\{a_i, s_i\}$ is fixed during the life time of the firm. Operating firms maximize their profit in each period and exit the market at exogenous rates.

2.3.1 Model setup

Firms are producing homogenous output, which is set to be the numeraire. In period t, firm i with productivity a_i and subsidy rate s_i chooses capital k_{it} and labor l_{it} to maximize per-period profit

$$\pi_t\left(a_i, s_i\right) = max\ (1+s_i) f\left(k_{it}, l_{it}\right) - w_t l_{it} - r_t k_{it} - \kappa, \tag{2.1}$$

where $f\left(k_{it}, l_{it}\right) = a_i k_{it}^{\alpha} l_{it}^{\gamma}$, κ is the fixed cost of operating in each period. The production function features decreasing returns to scale, i.e., $\alpha + \gamma < 1$. The introduction of the subsidy rate s could increase the size of low productivity firms and capture the fact that zombie firms are larger and less productive. In addition, the subsidy rate s generates dispersion of marginal product of input across firms, which would both result in resource misallocation among operating firms and change the entry threshold for potential entrants.

The value of a firm with $\{a_i, s_i\}$ in period t is given by the flow profits of operation in period t plus the discounted expected value of the firm in the next period conditional on survival

$$V_t\left(a_i, s_i\right) = \pi_t\left(a_i, s_i\right) + \beta\left(1 - \eta(a_i, s_i)\right) V_{t+1}(a_i, s_i). \tag{2.2}$$

The continuation value of firm i depends on its exit rate. I assume the exit rate $\eta(a_i, s_i)$ depends on the productivity-subsidy draw of the firm. At steady state, the value of the firm is given by

$$V\left(a_i, s_i\right) = \frac{\pi(a_i, s_i)}{1 - \beta\left(1 - \eta\left(a_i, s_i\right)\right)}. \tag{2.3}$$

There is a unit measure of potential entrants in each period. To enter the market, each potential entrant draws a pair of productivity a and subsidy rate s from distribution $G(a,s)$ and enters if $V(a,s) \geq 0$. I denote the corresponding entry rule as $Y(a, s)$ such that

$$Y(a, s) = \begin{cases} 1 & \text{if } V(a, s) \geq 0 \\ 0 & \text{if } V(a, s) < 0. \end{cases}$$

The timing within each period can be summarized as follows: (i) potential entrants draw productivity and subsidy rate from distribution $G(a, s)$ and decide to enter or not; (ii) incumbent firms exit according to exogenous exit rates; (iii) successful entrants and staying incumbents maximize their per-period profits.

Let $n_t(a, s)$ be the measure of operating firms with the draw $\{a, s\}$ at period t. The following law of motion needs to be satisfied

$$n_{t+1}(a, s) = (1 - \eta(a, s))n_t(a, s) + Y_{t+1}(a, s)g(a, s) \tag{2.4}$$

for all a and s. In period $t + 1$, the mass of firms with $\{a, s\}$ would be those continuing firms plus new entrants. In the steady state,

$$n(a, s) = \frac{Y(a, s)g(a, s)}{\eta(a, s)}. \tag{2.5}$$

It is clear that $n(a, s) = 0$ if $V(a, s) < 0$.

Lastly, I formulate the demand side of the model. Assume there is a representative household in the economy. They are endowed with one unit of labor and supply inelastically. The problem for household is as follows

$$\max_{C_t, K_{t+1}} \sum_{t=0}^{\infty} \beta^t u(C_t)$$

$$st.\ C_t + K_{t+1} = (r_t + 1 - \delta)K_t + w_t + \Pi_t - T_t$$

where Π_t, T_t are the profits from all firms and net lump sum tax respectively.[15] At the steady state, the Euler equation can be simplified as

$$1 = \beta(1 + r - \delta) \tag{2.6}$$

which pins down the rental price of capital r in the equilibrium.

The stationary equilibrium in this model is a set of prices $\{r, w\}$, choices for

[15] The government would impose a lump sum tax T_t on household each period and redistribute it to the subsidized firms.

household $\{C, K\}$, choices for firms $\{k(a,s), l(a,s), Y(a,s)\}$, the measure of firms $n(a,s)$, such that (i) household maximizes the utility; (ii) firms maximize profits; (iii) the entry threshold $Y(a,s) = 1$ if it satisfies $V(a,s) \geq 0$; (iv) law of motion for the measure $n(a,s)$ holds; (v) all markets clear.

2.3.2 Zombie firms in the model

Consistent with zombie classification in the empirical section, I classify a firm as a zombie firm if its profit net of subsidy is negative

$$\pi_t^{NetS}(a_i, s_i) = \pi_t(a_i, s_i) - s_i f(a_i, s_i) < 0. \tag{2.7}$$

In steady state, the above equation can be simplified as

$$\pi^{NetS}(a_i, s_i) = ((1-\alpha-\gamma)(1+s_i) - s_i)\, a_i^{\frac{1}{1-\alpha-\gamma}} (1+s_i)^{\frac{\alpha+\gamma}{1-\alpha-\gamma}} \phi(r, w) - \kappa < 0 \tag{2.8}$$

where $\phi(r, w)$ is a function of steady state rental rate of capital and wage. Equation 2.8 shows that firms with low productivity and high subsidy rate are more likely to be zombie firms.[16]

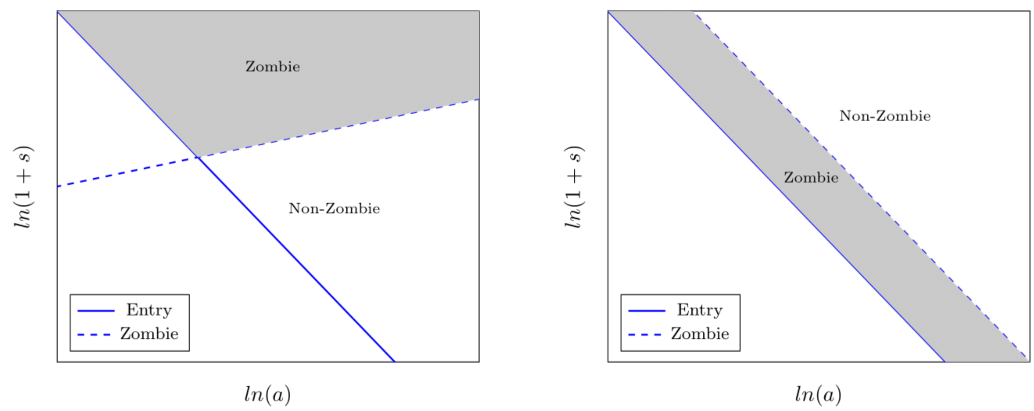
This definition is motivated by the method used to identify zombies in the data. But it differs from the one in the empirical section in two aspects. Firstly, as there is no firm-level productivity shock in the model, an operating firm with a negative profit net of subsidy cannot be "recovered". However, in the data, I restrict the zombie firms to have a negative profit net of subsidy for two consecutive periods to rule out the temporary shock to firms' earnings. Secondly, the model abstracts away the learning dynamics at the start of a firm's life cycle. Thus, I do not require the firms to be older than five years as in the empirical section.[17]

Figure 2.12 contrasts the zombie classification based on profit net of subsidy and profit alone. The grey area in the left panel illustrates the distribution of zombie firms based on equation 2.8 and entry rule $Y(a,s)$. Thus, zombie firms tend to be those with high subsidy rate. In addition, a firm with high productivity could also be counted as a zombie if it receives a large subsidy. This is consistent with the fact that there is an overlap in the productivity distribution of zombie and non-zombie firms as shown in the right panel of

[16]Firms with negative subsidy rate cannot be classified as zombies.

[17]In Appendix, I consider an alternative calibration of the model to match the moments in the sample where all the firms with age smaller than five years are dropped.

Figure 2.12: Zombie classification



(a) Profit net of subsidy

(b) Profit

Figure 2.6. The right panel of Figure 2.12 plots the distribution of zombie firms when we classify firms based on profit alone.[18] Based on profit alone, it is less clear whether zombie firms would have higher subsidy rates or not. This is consistent with the pattern in Figure 2.11. Classifying zombie firms using profit-based measures, the difference in average subsidy rate between zombie and non-zombie firms is lower.[19]

The exit rate of zombies is of particular interest since they are usually characterized as low productivity firms that are not exiting the market due to various distortions. I assume the exit rate depends on the zombie status of the firm

$$\eta(a_i, s_i) = \begin{cases} \eta_z & \text{if } i \text{ is a zombie firm} \\ \eta_n & \text{if } i \text{ is a non-zombie firm.} \end{cases}$$

In section 2.5.3, I explore the productivity gains from policies that increase zombie exit rate η_z.

To evaluate the impact of zombie firms on the economy, I aggregate the firm-level output and input choices to derive an aggregate production function

$$Y = \left(\int a^{\frac{1}{1-\alpha-\gamma}} \left(\frac{1+s}{(1+\bar{s})} \right)^{\frac{\alpha+\gamma}{1-\alpha-\gamma}} n(a,s) \right)^{1-\alpha-\gamma} K^{\alpha} = AK^{\alpha} \tag{2.9}$$

[18] In the model, only firms with positive profit will enter the market. We can think of firms with profit smaller than a constant as zombies.

[19] In a model-based environment, Yang (2019) classifies the firms with low-productivity that should exit when there is no friction as zombie firms. It implies a vertical "zombie" selection line in Figure 2.12.

where aggregate productivity A is a subsidy-weighted average of firm-level productivity which converts aggregate capital stock K to aggregate output Y; $\overline{s}$ is the output-weighted average of firm-level subsidy rate.[20]

With the expression for A, we could find the gain in aggregate productivity if firm-level distortions are removed

$$\frac{A^\star}{A} = \frac{\left(\int a^{\frac{1}{1-\alpha-\gamma}} n^\star(a,s)\right)^{1-\alpha-\gamma}}{\left(\int a^{\frac{1}{1-\alpha-\gamma}} \left(\frac{1+s}{(1+\overline{s})}\right)^{\frac{\alpha+\gamma}{1-\alpha-\gamma}} n(a,s)\right)^{1-\alpha-\gamma}} \tag{2.10}$$

$$= \underbrace{\frac{E\left[a^{\frac{1}{1-\alpha-\gamma}}|\Omega\right]^{1-\alpha-\gamma}}{E\left[a^{\frac{1}{1-\alpha-\gamma}} \left(\frac{1+s}{(1+\overline{s})}\right)^{\frac{\alpha+\gamma}{1-\alpha-\gamma}} |\Omega\right]^{1-\alpha-\gamma}}}_{\text{Intensive-margin}} \underbrace{\frac{E\left[a^{\frac{1}{1-\alpha-\gamma}}|\Omega^\star\right]^{1-\alpha-\gamma}}{E\left[a^{\frac{1}{1-\alpha-\gamma}}|\Omega\right]^{1-\alpha-\gamma}}}_{\text{Selection}} \underbrace{\left(\frac{N^\star}{N}\right)^{1-\alpha-\gamma}}_{\text{Variety}} \tag{2.11}$$

where $A^\star$ is the aggregate productivity in the frictionless economy, Ω ($\Omega^\star$) denotes the set of operating firms in the economy with (without) frictions and N ($N^\star$) is the total measure of operating firms in the economy with (without) frictions.[21]

The gain in aggregate productivity can be decomposed into three components: Intensive-margin, Selection and Variety component. The Intensive-margin component shows the increase in average productivity if we remove the dispersion in subsidy rates across firms without changing the composition of firms. Next, the Selection component captures the gain in average productivity if we allow composition of firms to change. Last, the Variety component illustrates the change in total mass of operating firms if we move to the frictionless economy. Despite firms are producing a homogeneous product, the technology that features decreasing returns to scale effectively makes firms become imperfect substitutes in production. Each additional firm is valuable as it helps to overcome the decreasing returns to scale faced by other firms. In section 2.5.1, I evaluate the productivity gain from removing distortions using this decomposition result.

[20] $(1+\overline{s}) = \int (1+s)\frac{y(a,s)}{Y} n(a,s)$. In addition, there is no aggregate labor input in the production function as I assume total labor supply equals to one.

[21] $N = \int n(a,s)$ and $N^\star = \int n^\star(a,s)$.

2.4 Calibration

In this section, I outline the choices of model parameters and calibration procedure. The strategy is to first set a number of generic parameters using the values in the existing literature, which include capital and labor output elasticity α, γ, discount rate β and capital depreciation rate δ. Then I calibrate the remaining parameters to match the moments from data in 1999.

For the generic parameters, I first set $\alpha + \gamma = 0.85$ following Restuccia and Rogerson (2008). $\alpha + \gamma$ is the returns to scale of the production function. The introduction of the curvature in the production function is necessary to generate profits and firm dynamics in our model. Next, I set $\alpha = \gamma = 0.425$ since aggregate labor share in the national account of Chinese data is roughly 50%. Then, $\beta = 0.96$ implies that the real interest rate is approximately four percent, which is consistent with Chinese data during the sample period. I choose annual depreciation rate $\delta = 0.1$ according to Bai et al. (2006).

For the parameters specific to our setting, I first set exogenous exit rates for zombie and non-zombie firms to match the data in 1999, i.e., $\eta_n = 0.15$ and $\eta_z = 0.18$. The exit rate is calculated as the fraction of zombie and non-zombie firms exiting the sample between 1999 and 2000. Despite the poor performance of zombie firms, the exit rate for zombie firms is only three percentage points higher than that of non-zombies.[22]

In terms of the joint distribution $G(a, s)$, I follow Yang (2019) and assume that $ln(a)$ and $ln(1 + s)$ are jointly normally distributed as

$$\begin{pmatrix} ln(a) \\ ln(1+s) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_a \\ \mu_s \end{pmatrix}, \begin{pmatrix} \sigma_a & \rho_{as}\sqrt{\sigma_a \sigma_s} \\ \rho_{as}\sqrt{\sigma_a \sigma_s} & \sigma_s \end{pmatrix} \right).$$

I normalize $\mu_a = 1$ and then calibrate the five remaining parameters $\{\mu_s, \sigma_a, \rho_{as}, \sigma_s, \kappa\}$ to match a set of moment conditions using the Method of Simulated Moments. Table 2.1 and 2.2 report the calibrated parameters and corresponding moments, where firm size is measured by employment.

The calibration jointly estimates five parameters to match the five moments in the data, I outline the intuition for the choices of parameters. Firstly, since firms with higher subsidy rates are more likely to be classified as zombies, the mean of subsidy distribution μ_s should be relatively low to match the zombie

[22]Figure 2.26 in Appendix plots the measured exit rate for zombie and non-zombie firms between 1999 and 2007.

Table 2.1: Parameterization

Parameter	Definition	Value
α	Capital output elasticity	0.43
γ	Labor output elasticity	0.43
β	Discount rate	0.96
δ	Depreciation rate	0.10
μ_a	Mean of productivity distr.	1.00
μ_s	Mean of subsidy distr.	-0.40
σ_a	Variance of productivity distr.	0.91
ρ_{as}	Correlation coefficient	-0.69
σ_s	Variance of subsidy distr.	0.48
κ	Operating cost	23.09
η_z	Exit rate for zombie firms	0.18
η_n	Exit rate for non-zombie firms	0.15

Table 2.2: Moments

Moment	Data	Model
Zombie rate	0.16	0.16
Relative mean size of zombie and non-zombie	1.41	1.41
S.D. of (log) size	1.26	1.26
Relative S.D. of (log) size of zombie and non-zombie	1.04	1.01
Profit share of output	0.11	0.09
Exit rate of non-zombie firms	0.15	0.15
Exit rate of zombie firms	0.18	0.18

Figure 2.13: Productivity distribution: zombies vs non-zombies

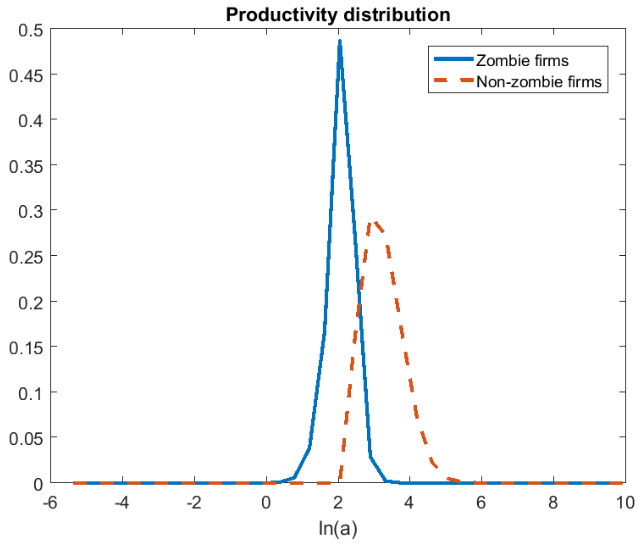
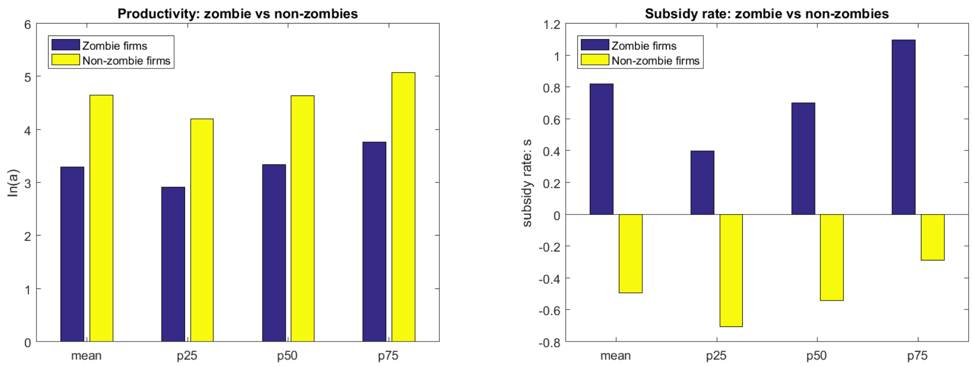


Figure 2.14: Productivity and subsidy rate: zombies vs non-zombies



rate in the data. Secondly, zombie firms are less productive and yet larger than the non-zombies, it requires the correlation coefficient ρ_{as} to be negative such that low productivity firms are more likely to receive more subsidies. Thirdly, the variance of productivity and subsidy distribution allow us to match both the overall dispersion of firm size and relative dispersion of zombie firms. Lastly, fixed cost κ drives the overall profit share of output. The model matches the moments on zombie rate, relative size of zombie firms and overall dispersion of firm size perfectly. It slightly underestimates the relative size dispersion of zombie firms and overall profit share of output. One thing to note is that the zombie definition in the data rules out young firms. Since the model is not well suited to study the strong learning dynamics at the beginning of the firm's life cycle, in the Appendix, I re-calibrate the model to match the relevant moments in the sample where all the firms with age smaller than five years are dropped.

Figure 2.13 shows the productivity distribution of zombie and non-zombie firms in the model. As in the data, I find the zombie firms are less productive than non-zombies. In the left panel of Figure 2.14, I compare the average productivity and different quartiles in the output-weighted distribution of productivity for zombie and non-zombie firms. The productivity gap between non-zombies and zombies is somewhat larger compared with data in Figure 2.6.[23] Given the fact that zombie firms are much less productive and yet larger than the non-zombies, we would expect they receive much more subsidies. The right panel of Figure 2.14 reports the average subsidy rate and different quartiles in the output-weighted distribution of subsidy rate. On average, zombie firms would receive a subsidy rate of 80%, and non-zombie firms would receive a negative subsidy rate of about 50%. This figure is larger than the subsidy rate I find in the data in Figure 2.7 and 2.8. As a small subsidy rate difference can not be reconciled with the fact that zombie firms are much less productive and yet larger in size. I interpret the subsidy in the model as a generic distortion wedge that increases the size of zombie firms. It is clear that the government can give various forms of subsidies including cheap credit and land. The subsidy information in the data would significantly underestimate the true extent of support from the government.

2.5 Results

This section conducts three quantitative exercises using the calibrated model. Section 2.5.1 reports the productivity gain from removing distortions completely and compares the magnitude of productivity gain with the existing literature. Section 2.5.2 shows that rising private sector in the Chinese economy induces a reduction in the dispersion of subsidy rate. This reduction could generate sizable productivity gain even the private firms have no productivity advantage over SOEs. In section 2.5.3, I explore the effectiveness of government policies that increase the zombie exit rate.

2.5.1 Productivity gain from removing distortions

In this section, I use the calibrated model to evaluate productivity gains from removing dispersion in subsidy rate completely. Table 2.3 reports the decompo-

[23] I use the index approach to estimate firm productivity in the data. The productivity gap between zombies and non-zombies is what matters for the resource misallocation.

Table 2.3: Productivity gain from removing distortions

	Intensive-margin	Selection	Variety	Total
Benchmark	283.47%	25.25%	-19.69%	285.74%
$\sigma_s = 0.24$	97.54%	23.53%	-19.02%	97.60%
$\sigma_s = 0.048$	14.58%	16.60%	-14.24%	14.58%

Notes: The table reports the percentage change in aggregate productivity and its decomposition from removing dispersion of subsidy rate completely. Each cell is calculated as $100\,(C-1)$, where C stands for either individual components in equation 2.11 or $\frac{A^\star}{A}$. As a result, last column is not the sum of first three columns.

sition of productivity gain. The first row shows that, for the benchmark economy, the aggregate productivity can increase by more than 280% if removing the resource misallocation completely. The Intensive-margin component alone contributes to nearly 100% in the productivity gain.

The gain in aggregate productivity tends to be larger than the estimates in the literature. Hsieh and Klenow (2009) equalize the marginal products within the four-digit industry and find the aggregate productivity can increase by about 115% using the same Chinese firm-level data.[24] Yang (2019) emphasizes the importance of selection and finds the aggregate productivity losses can be 40% higher than the estimate in Hsieh and Klenow (2009). The underlying reason that drives the difference between my estimate and these studies is the returns to scale in the production function, which is effectively the elasticity of substitution between firms in Hsieh and Klenow (2009).

Higher substitutability between firms would induce a larger gain if distortions are removed. In the benchmark result of Hsieh and Klenow (2009), they conservatively assume an elasticity of substitution of 3 across firms, which implies the returns to scale $\alpha + \gamma = 0.5$.[25] This is in contrast with values around 0.85 in models with heterogeneous firms, e.g., Restuccia and Rogerson (2008). In fact, Yang (2019) also illustrates that productivity gain is more than 200% if using a scale parameter of 0.8.

The first row of Table 2.3 shows that the contribution of Selection component is almost fully counteracted by a negative Variety component. As emphasized by Fattal Jaef (2018) and Yang (2019), in a model with entry and exit, it is unclear whether the magnitude of total gains in TFP from removing distortions will be larger than the Intensive-margin component or not. The entry and exit of firms will induce a change in the composition of operating firms (Selection

[24]The figure is for 1998. See table IV in Hsieh and Klenow (2009).

[25]It also implies that each firm has a markup of 50%.

component) and a change in the mass of operating firms (Variety component). Whether the Variety component counteracts the Selection component depends on the underlying distribution of productivity and distortion. In the calibrated model, if we remove the dispersion of subsidy rate across firms completely, the mass of operating firms will be smaller, offsetting the gain due to selection. Since the subsidy rate is negatively correlated with productivity in the benchmark economy, high productivity firms will have low subsidy rates, which will reduce their labor demand and hence wage rate in the economy, making it profitable for more firms to enter.

The last two rows in Table 2.3 show that the dispersion of subsidy rate is crucial in driving the productivity gain. The second (third) row reduces the dispersion of subsidy rate to be half (one-tenth) of the benchmark economy. Total productivity gain will reduce to 97.6% and 14.58%. This reduction is primarily driven by the Intensive-margin component alone. The Selection and Variety components still offset each other.

2.5.2 Productivity gain from rising private firms

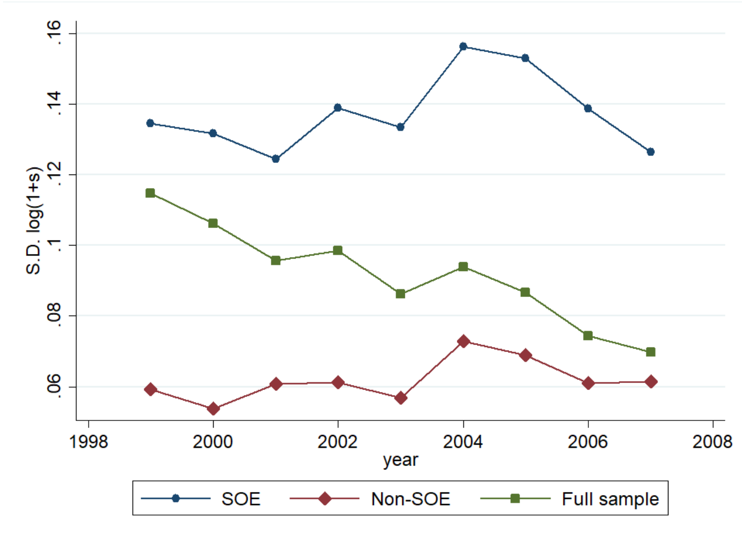
In the model, the dispersion of subsidy rate is the key in driving the loss in aggregate productivity. The left panel of Figure 2.16 plots the aggregate productivity against the dispersion of subsidy σ_s. I normalize the aggregate productivity in the benchmark economy to be one. By decreasing the dispersion σ_s, the aggregate productivity increases significantly. A 10% decrease in the dispersion from the benchmark estimate increases aggregate productivity by 14%.[26] The decline of subsidy dispersion would also reduce the fraction of firms receiving a high subsidy rate, which contributes to a lower zombie rate as shown in the right panel of Figure 2.16.

I then ask the question whether the decline in zombie rate observed in the data can be driven by a change in the dispersion of subsidy rate. Figure 2.15 shows the dispersion of subsidy rate in the data. Between 1999 and 2007, there is a strong decline in dispersion in subsidy rates. Besides, the decline in the full sample is driven by the rise of private firms in the economy as the dispersion within SOEs and non-SOEs exhibit little variation over time.

Figure 2.2 in the empirical section shows that, between 1999 and 2007, the decline in zombie rate explained by the rise of non-SOEs is about 5%. If I

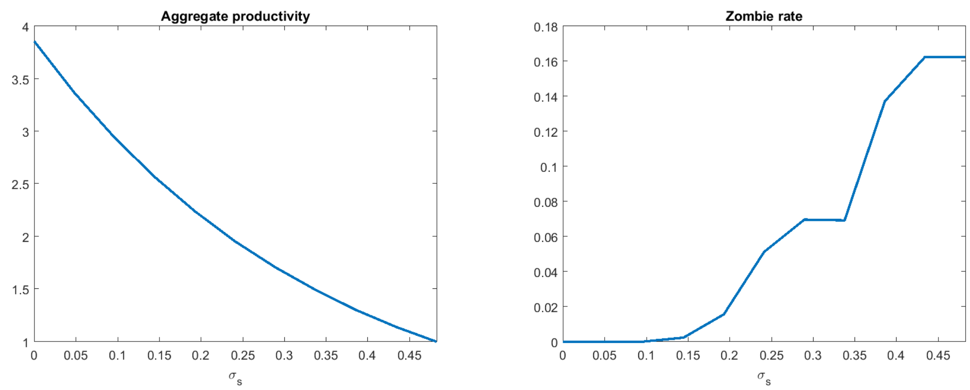
[26]The gain is more significant as dispersion approaches zero. The extreme case $\sigma_s = 0$ corresponds to the total gain in aggregate productivity in the first row of Table 2.3.

Figure 2.15: Subsidy dispersion in the data



Notes: The figure plots the dispersion of subsidy rate in the data for SOEs, non-SOEs and full sample.

Figure 2.16: Aggregate productivity, zombie rate and dispersion of subsidy

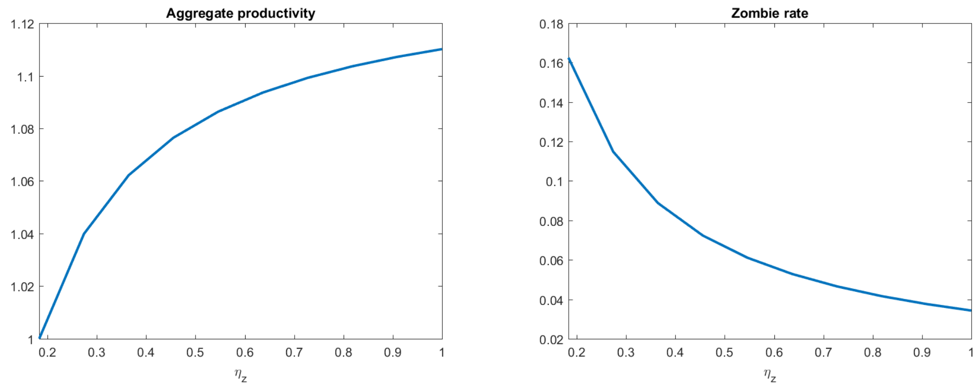


reduce the dispersion of subsidy to match the 5% decline in zombie rate, the aggregate productivity increases by about 40%. This quantitative exercise illustrates that the rise of the private sector can induce a large productivity gain even when there is no productivity gap between private firms and SOEs.

2.5.3 How effective is the policies that increase the exit rate of zombies?

Zombie firms are usually characterized as the low productivity firms that do not exit the market due to various frictions. In this section, I explore the importance of zombie exit rate in driving the aggregate productivity of the economy.

Figure 2.17: Aggregate productivity and exit rate of zombie firms



The right panel of Figure 2.17 plots zombie rate in the economy if we increase zombie exit rate η_z. A higher zombie exit rate reduces zombie rate significantly. The left panel of Figure 2.17 shows the corresponding gain in aggregate productivity. Increasing zombie exit rate by 50% from the benchmark level raises aggregate productivity by 4%. The gain in aggregate productivity is concave in the zombie exit rate. In the extreme case where we force zombie firms to have a 100% exit rate, aggregate productivity only increases by 11%.[27] The result implies that, without changing the underlying frictions in the economy (dispersion of subsidy rate across firms), a higher exit rate of zombie firms only generates modest effects on productivity.

This exercise shows that the recent policies of the Chinese government that aim to address zombie firms by closing them will be effective in reducing the zombie rate. However, its productivity effect is relatively limited. I calibrate the benchmark economy to match data in 1999. If the extent of distortions diminishes over time, the productivity effects of a higher zombie exit rate would be even smaller. This is in contrast with the result in section 2.5.2, where I find that the rise of private firms tends to reduce the dispersion of subsidy rate in the economy and induces a large productivity gain. Therefore, policies that attempt to solve the zombie problem should directly tackle the distortion in the economy that helps zombie firms survive. One caveat of the result is that I focus on the steady state comparison of models with different zombie exit rates and neglect the transition dynamics. Increasing the zombie exit rate may induce a larger productivity effect when transition is taken into consideration, which I reserve for future research.

[27]A 100% exit rate implies that each zombie firm operates for one period.

2.6 Conclusion

This chapter studies the distorting effects of zombie firms on the economy. Using the firm-level data from the manufacturing sector in China between 1998 and 2007, I document that zombie firms are larger in terms of inputs employed but less productive. Zombie firms are more likely to receive a positive subsidy from the government, and the average subsidy rate for zombie firms is higher than that of non-zombie firms. This difference in average subsidy rate between zombie and non-zombie firms depends on both the underlying joint distribution of subsidy rate and productivity and the definition of zombies.

I outline a model with heterogeneous firms to account for the features of zombies, and then use the calibrated version of the model to quantify the impact of zombies on aggregate productivity. Quantitative exercise highlights the importance of returns to scale in production technology in driving the productivity gain from removing distortions. In addition, I show that the rise of private sector tends to reduce the dispersion of subsidy rate, which can lead to a sizable productivity gain even when there is no productivity difference between private firms and SOEs. Lastly, while the policies that increase the exit rate of zombies can substantially reduce the zombie rate, its productivity effect is relatively limited.

2.7 Appendix

2.7.1 Firm-level data from China

The firm-level data used in this paper is from the annual surveys conducted by China's National Bureau of Statistics (NBS). It includes all state-owned enterprises (SOEs) and non-state-owned enterprises (non-SOEs) with more than 5 million RMB in revenue from 1998 to 2007.[28] The sample is a representative summary of the whole industrial sector in China. Compared with 2004 industrial census for all firms, it covers more than 90% of gross output of the industrial sector. The industry sector contains mining, manufacturing, and public utilities. This paper focuses on the manufacturing sector. The raw data is presented annually. Following Brandt et al. (2012), I convert the data to a panel by linking the firms over time using identifying information such as firm ID, firm name and address.[29]

The data reports extensive firm-level information on balance sheet variables and operating characteristics, such as short-term and long-term assets and liabilities, output, input and taxes. The benchmark measure of zombie firms requires firm-level data on profit, subsidy, and age. I measure the firm's profit using the accounting profit reported by the firm directly. I construct firm age using the information on the year of birth of the firm.

The unique feature of Chinese data is that it contains firm-level information on the amount of subsidies firms receive from the government. Following the accounting standard, the subsidy information mainly involves the transfer of monetary assets from the government to firms. The subsidy information in the data contains support from the government for many production activities. It could reflect subsidies towards capital expenses, labor expenses, and expenses on other activities such as R&D. The subsidies that do not involve direct transfer of assets from the government to firms are not included, such as tax-breaks and low-interest loans.

For the alternative zombie measure, I extract information on a firm's short-term and long-term liability to construct the minimum required interest that

[28]The 5 million RMB is not a "hard" rule, and firms with sales below the threshold for a year are not automatically removed from the sample. See Brandt et al. (2014) for a discussion of the data set.

[29]See `https://feb.kuleuven.be/public/N07057/CHINA/appendix/` for the algorithm to match firms over time.

firms need to pay when there are no interest subsidies. The implied minimum interest payment tends to overestimate the amount of interest firms need to pay since there are types of liability that are not interest-bearing, payable for instance.

To compare the size of zombie and non-zombie firms, I use employment and real capital stock. The measurement of employment is directly observable in the data. I construct the real capital stock using the method outlined by Brandt et al. (2012), in which I infer each firm's initial capital stock and calculate annual investment as the change in nominal capital stock between years. Together with price deflator for investment, the real capital stock is then calculated using the perpetual inventory method.[30] Using the two-digit output and input deflator provided by Brandt et al. (2012), I calculate the firm-level real value added as the real output net of real intermediate inputs and indirect taxes.

I classify firms into SOEs and non-SOEs using a two-stage approach. In the first stage, I classify firms using the information on registration types. Following Nie et al. (2016), I categorize the state-owned and collective firms as SOEs, and classify private and two types of foreign firms as non-SOEs.[31,32] For the firms that cannot be assigned in the first stage, I use the additional information on paid-in capital to categorize firms as SOEs if the state or collective capital comprises the largest share in a firm's paid-in capital.[33] I extract a firm's location information using the geographic code in the data.

2.7.2 Alternative calibration

The zombie definition in the model does not involve any age restriction since our model is not suitable to explain the strong learning dynamics at the beginning of a firm's life cycle. I consider an alternative calibration where I jointly choose five parameters $\{\mu_s, \sigma_a, \rho_{as}, \sigma_s, \kappa\}$ to match the relevant moments in the sample where all firms with age smaller than 5 years are dropped. Table 2.4 and 2.5 report the choices of parameters and moments I use in the estimation.

[30] See `https://feb.kuleuven.be/public/N07057/CHINA/appendix/` for the algorithm to construct firm-level real capital stock.

[31] State-owned firms at the township level are usually registered as collective firms. See Lu et al. (2010) for the detailed discussion of collective firms in China.

[32] Foreign firms include those from Hong Kong, Macau, and Taiwan, and those from foreign countries.

[33] Only firms with registration type 159 (other limited liability companies), 160 (joint-stock companies), and 190 (other companies) are classified in the second stage.

Table 2.4: Alternative calibration: parameterization

Parameter	Definition	Value
μ_s	Mean of subsidy distr.	-0.69
σ_a	Variance of productivity distr.	0.91
ρ_{as}	Correlation coefficient	-0.69
σ_s	Variance of subsidy distr.	0.56
κ	Operating cost	17.53

Table 2.5: Alternative calibration: moments

Moment	Data	Model
Zombie rate	0.21	0.20
Relative mean size of zombie and non-zombie	1.26	1.26
S.D. of (log) firm size	1.29	1.29
Relative S.D. of (log) size of zombie and non-zombie	1.01	1.02
Profit share of output	0.11	0.09

Table 2.6: Decomposition of productivity gain from removing distortions: alternative calibration

	Intensive-margin	Selection	Variety	Total
$\sigma_s = 0.56$	378.56%	22.65%	-16.92%	387.65%
$\sigma_s = 0.28$	121.53%	23.35%	-18.87%	121.69%
$\sigma_s = 0.056$	17.19%	12.51%	-11.12%	17.19%

Notes: The table reports the percentage change in aggregate productivity and its decomposition from removing dispersion of subsidy rate completely for alternative calibration. Each cell is calculated as $100\,(C-1)$, where C stands for either individual components in equation 2.11 or $\frac{A^\star}{A}$. As a result, last column is not the sum of first three columns.

Figure 2.18: Aggregate productivity, zombie rate and dispersion of subsidy: alternative calibration

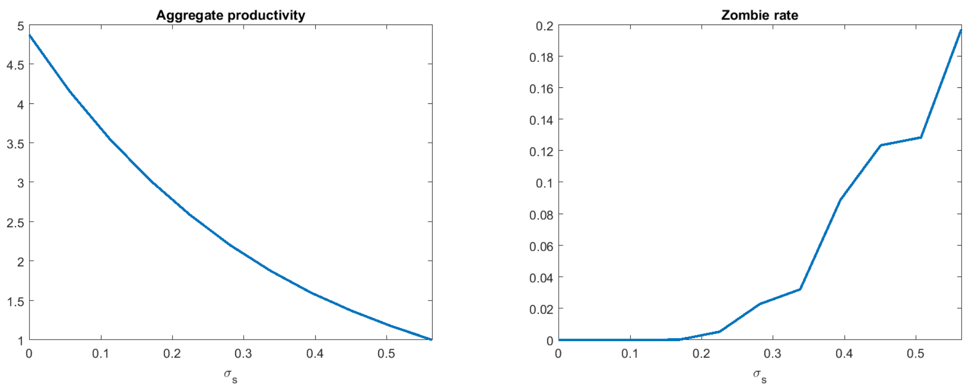
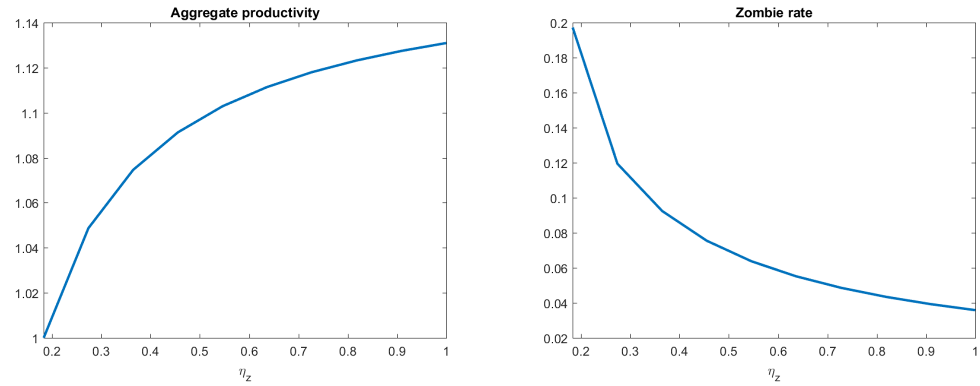


Figure 2.19: Aggregate productivity and exit rate of zombie firms: alternative calibration



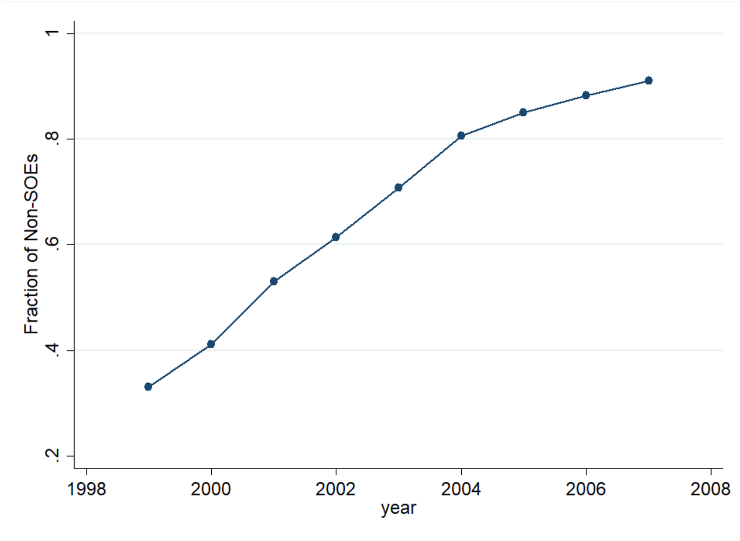
Restricting the sample to older firms results in a larger zombie rate and smaller relative size of zombies, which leads to a lower mean and larger dispersion of subsidy distribution.

Table 2.6 reports the productivity gain from removing subsidy dispersion completely. Since calibrated dispersion in subsidy rate is larger than the benchmark economy, removing dispersion of subsidy induces an even larger productivity gain, and it is primarily driven by the Intensive-margin component. The left panel of Figure 2.18 depicts the relation between aggregate productivity and dispersion of subsidy. After decomposing the zombie rate by SOE status in the restricted sample, I find that the decline of zombie rate explained by the rise of the private sector is still around 5%. Reducing the dispersion to match the 5% decline in zombie rate will induce a 12% increase in aggregate productivity. Lastly, I explore the effectiveness of policies that increase the zombie exit rate. Figure 2.19 shows that increasing zombie exit rate will significantly reduce the amount of zombie firms in the economy while its productivity effect is relatively limited.

2.7.3 Supplementary figures

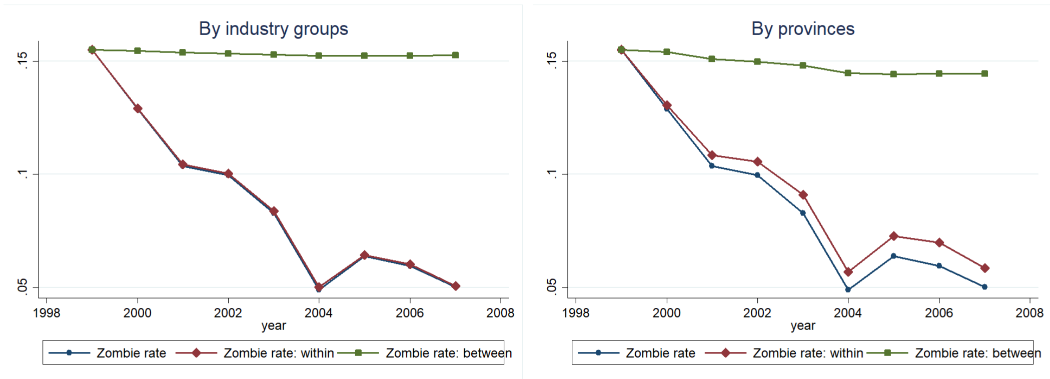
This section provides supplementary figures. Figure 2.20 illustrates the rapid rise of private sector in the economy. The share of non-SOEs increases from 33% in 1999 to more than 90% in 2007. Figure 2.21 conducts the between-within decomposition for the aggregate zombie rate across provinces and industries. In both decompositions, the decline of the zombie rate is driven by the within-group component. Figure 2.22 and 2.23 report the average size and productivity gap between zombie and non-zombie firms after controlling for

Figure 2.20: Rising share of non-SOE firms



Notes: This figure plots the fraction of non-SOEs in the sample. Ownership is determined by using information on registry type and paid-in capital.

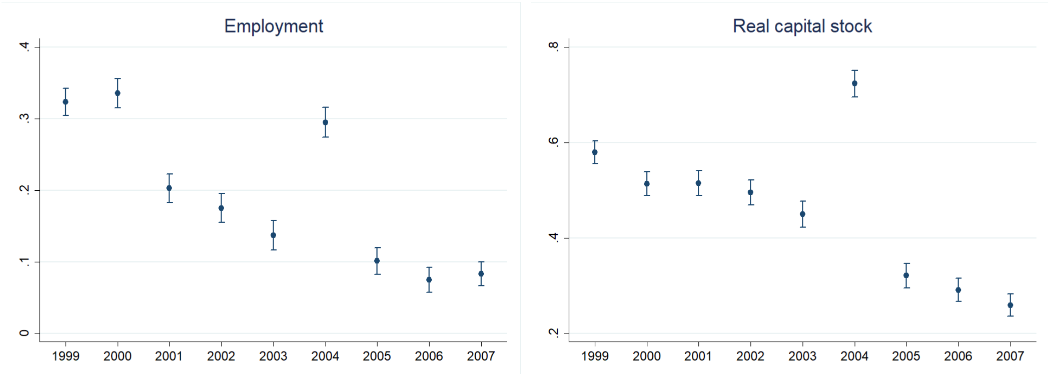
Figure 2.21: Decomposing the decline of zombie rate by industries and provinces



Notes: The figure shows the between-within decomposition of the aggregate zombie rate by industries (left panel) and provinces (right panel).

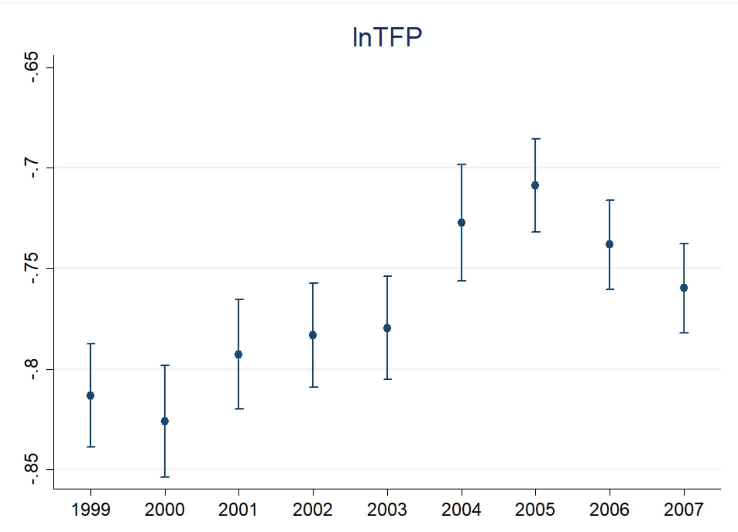
ownership, industries and provinces. The general pattern that zombie firms are larger yet mush less productive remains. Figure 2.24 reports the zombie rate if no age restriction is imposed in the benchmark measure. Figure 2.25 shows the regression-based average size and productivity differences between zombie and non-zombie firms when I use two alternative measures to identify zombies. Zombie firms are still much larger in terms of inputs employed yet much less productive. Figure 2.26 plots the annual exit rate from the sample for zombie and non-zombie firms.

Figure 2.22: Regression-based size gap between zombies and non-zombies



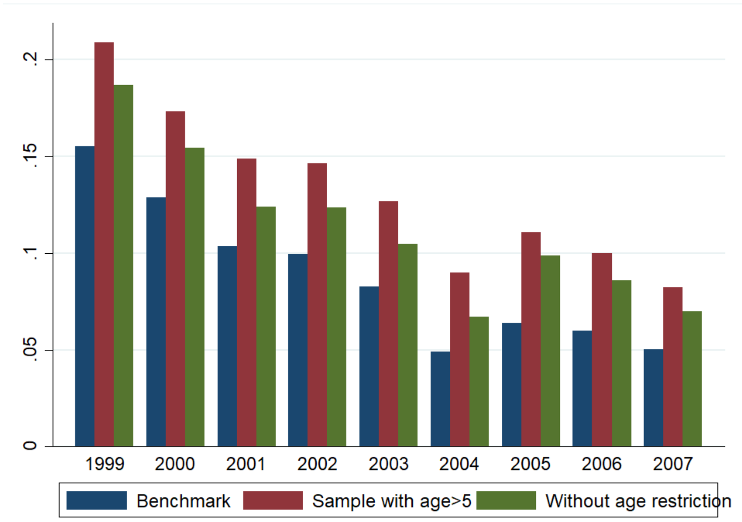
Notes: This figure reports the regression-based average size differences between zombie and non-zombie firms. I obtain the estimates as β_1 from the regression $y_i = \beta_1 D_i^Z + FEs + \epsilon_i$, where y_i is the logarithm of firm size measures and D_i^Z is the dummy for zombie firm. Fixed effects are included to control for ownership, industries and provinces. Both the point estimate and its 95% confidence interval are reported.

Figure 2.23: Regression-based productivity gap between zombies and non-zombies



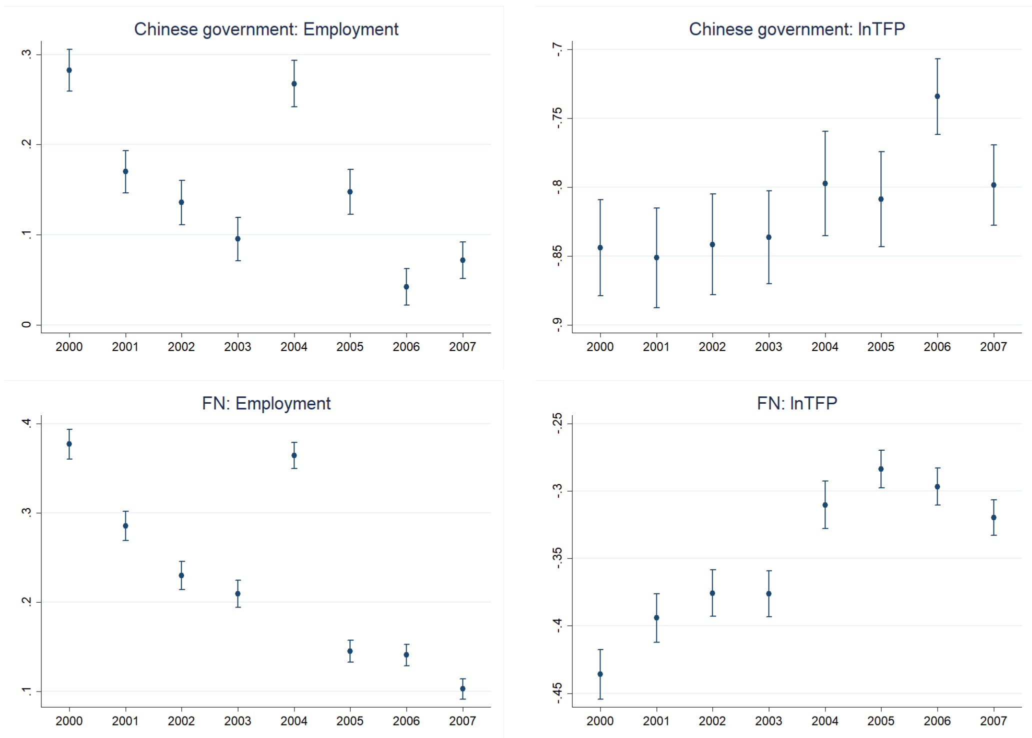
Notes: This figure reports the regression-based average productivity differences between zombie and non-zombie firms. I obtain the estimates as β_1 from the regression $log(TFP_i) = \beta_1 D_i^Z + FEs + \epsilon_i$, where D_i^Z is the dummy for zombie firm. Fixed effects are included to control for ownership, industries and provinces. Both the point estimate and its 95% confidence interval are reported.

Figure 2.24: Zombie rate and age restriction



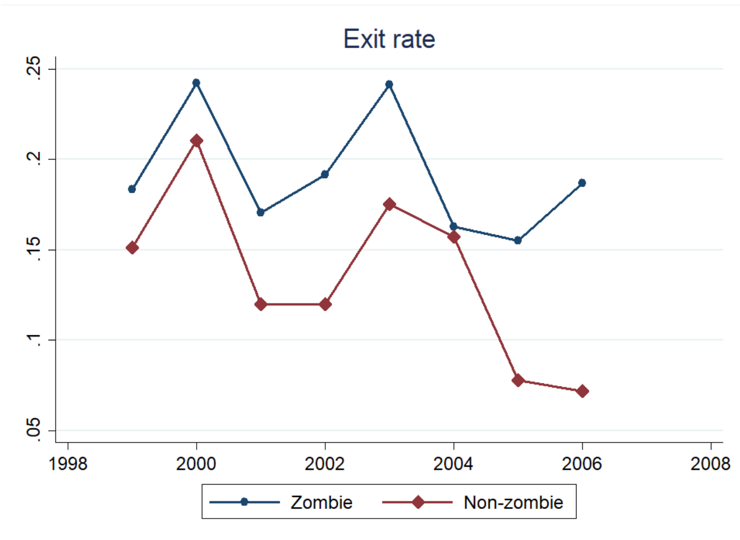
Notes: This figure compares the benchmark zombie rate with two alternatives. The first alternative is the zombie rate in the subsample where all firms with age smaller than 5 years are dropped. The second alternative is to identify zombie firms without age restriction.

Figure 2.25: Size and productivity gap between zombies and non-zombies: alternative zombie measures



Notes: This figure reports the regression-based average size and productivity differences between zombie and non-zombie firms for two alternative zombie measures. I obtain the estimates as β_1 from the regression $y_i = \beta_1 D_i^Z + FEs + \epsilon_i$, where y_i is the logarithm of firm size and productivity measure, and D_i^Z is the dummy for zombie firm. Fixed effects are included to control for ownership, industries and provinces. Both the point estimate and its 95% confidence interval are reported.

Figure 2.26: Measured exit rate



Notes: This figure plots the annual exit rate from the sample for zombie and non-zombies.

References

Aghion, P., Cai, J., Dewatripont, M., Du, L., Harrison, A., and Legros, P. (2015). Industrial policy and competition. *American Economic Journal: Macroeconomics*, 7(4):1–32.

Bai, C.-E., Hsieh, C.-T., and Qian, Y. (2006). The Return to Capital in China. Technical report, National Bureau of Economic Research.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, 103(1):305–34.

Blattner, L., Farinha, L., and Rebelo, F. (2018). When Losses Turn into Loans: The Cost of Undercapitalized Banks. Technical report.

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2012). Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, 97(2):339–351.

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2014). Challenges of Working with the Chinese NBS Firm-level Data. *China Economic Review*, 30:339–352.

Caballero, R. J., Hoshi, T., and Kashyap, A. K. (2008). Zombie Lending and Depressed Restructuring in Japan. *American Economic Review*, 98(5):1943–77.

Chang, Q., Zhou, Y., Liu, G., Wang, D., and Zhang, X. (2020). How does government intervention affect the formation of zombie firms? *Economic Modelling*.

Chen, X., Lee, C.-W. J., and Li, J. (2008). Government Assisted Earnings Management in China. *Journal of Accounting and Public Policy*, 27(3):262–274.

Eslava, M., Haltiwanger, J., Kugler, A., and Kugler, M. (2013). Trade and market selection: Evidence from manufacturing plants in colombia. *Review of Economic Dynamics*, 16(1):135–158.

Fattal Jaef, R. N. (2018). Entry and exit, multiproduct firms, and allocative distortions. *American Economic Journal: Macroeconomics*, 10(2):86–112.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Foster, L., Haltiwanger, J., and Syverson, C. (2016). The slow growth of new plants: Learning about demand? *Economica*, 83(329):91–129.

Fukuda, S.-i. and Nakamura, J.-i. (2011). Why Did 'Zombie' Firms Recover in Japan? *The World Economy*, 34(7):1124–1137.

Haltiwanger, J., Kulick, R., and Syverson, C. (2018). Misallocation measures: The distortion that ate the residual. Working Paper 24199, National Bureau of Economic Research.

He, Q., Li, X., and Zhu, W. (2018). Political connection and the walking dead: Evidence from china's privately owned firms. *International Review of Economics & Finance*.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Huang, S. and Chen, Y. (2017). The distribution features and classified disposition of china's zombie firms. *China Industrial Economics*, 34(3):24–43.

Kwon, H. U., Narita, F., and Narita, M. (2015). Resource Reallocation and Zombie Lending in Japan in the 1990s. *Review of Economic Dynamics*, 18(4):709–732.

Lee, E., Walker, M., and Zeng, C. (2014). Do Chinese Government Subsidies Affect Firm Value? *Accounting, Organizations and Society*, 39(3):149–169.

Lu, J., Tao, Z., and Yang, Z. (2010). The costs and benefits of government control: Evidence from china's collectively-owned enterprises. *China Economic Review*, 21(2):282–292.

McGowan, M., Andrews, D., and Millot, V. (2018). The Walking Dead? Zombie Firms and Productivity Performance in OECD Countries. *Economic Policy*, 33(96):685–736.

Nie, H., Jiang, T., Zhang, Y., and Mingyue, F. (2016). Report on China's Zombie Enterprises: Current Situation, Reasons and Countermeasures. *Macroeconomic Management (In Chinese)*, 9:63–88.

Restuccia, D. and Rogerson, R. (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics*, 11(4):707–720.

Song, Z., Storesletten, K., and Zilibotti, F. (2011). Growing like China. *American Economic Review*, 101(1):196–233.

Xinhua News (2019). China clears up over 95% of 'zombie' firms: official. http://english.www.gov.cn/statecouncil/ministries/201910/18/content_WS5da91079c6d0bcf8c4c15556.html. Online; accessed 25-July-2020.

Yang, M.-J. (2019). Micro-level misallocation and selection. *Working paper. University of Utah.*

Chapter 3

The Rise of Information Inputs in Production

3.1 Introduction

The last few decades witnessed the rapid development of information technology. This chapter documents the concurrent rises in wage premium of information workers and information intermediate input, and tries to understand the forces that drive these changes.

It is reasonable to think that advances in information technology benefit workers who work with information more intensively. We classify occupations into info and non-info occupations based on an info score (information score) obtained from O*NET. The occupation-level info score indicates the degree of data and information analyzing activities needed in an occupation. Using data from IPUMS CPS, we show that the average hourly wage is higher for occupations with higher info score, and that wage premium related to info score have been increasing since 1980. In particular, the relative wage of info workers to non-info workers increases by about 30% from 1980 to 2010. Together with the wage rise, the employment size of info workers also increases relative to that of non-info workers.

The rise of information technology reduces the transaction cost, which could induce firms to outsource more intermediate inputs (Abramovsky and Griffith, 2006). We group industries into info sector (information sector) and non-info sector (non-information sector) based on the median info score of each industry's workforce. Industries with median info score higher than the national median are classified as info sector, and the rest are in the non-info sector. We find that there is a strong reallocation of output towards info sector. In addition, intermediate shares on output from info sector increase in both sectors, while intermediate shares associated with output from non-info sector decline.

Motivated by the rising importance of information-intensive inputs, we propose an accounting framework to understand the drivers of these phenomena. The model features two sectors, info and non-info sectors, and two types of workers, info and non-info workers. Each sector uses capital stock, info and non-info workers, and info and non-info intermediate inputs to produce. We allow technical changes to be sector-factor specific, and we interpret the productivity associated with intermediate input as outsourcing efficiency.

Drawing on the estimates of elasticities in the literature, we infer the productivity changes between 1980 and 2014. We find that, for both sectors, the technologies associated with non-info labor grow fastest, with an annual growth rate of 2.8 percent in info sector and 2.3 percent in non-info sector. In addition, the technologies augmenting info inputs do not exhibit sizeable growth over time.

To understand the role of differential rates of technical change, we examine two partial-equilibrium counterfactual experiments. In the first experiment, we isolate the contribution of different sources of technical change to the observed intermediate shares by allowing one set of technologies to vary over time and keeping the other technologies and prices fixed at the 1980 level. From this experiment, we find that, for three out of four intermediate shares we study, labor-augmenting technical change contributes to the observed change in intermediate shares. On the other hand, outsourcing-specific technical change can only explain one of the changes.

In our second experiment, we fix the labor-augmenting technologies at the 1980 values and allow factor output shares to evolve as in the data. The result shows that without these technical changes, the wage premium of info workers would have increased more in both sectors. Since info inputs and non-info inputs are complements, the technical change that is biased towards non-info inputs increases the demand and expenditure share for info inputs. In the absence of such technical change, the rising expenditure share on info input can only be explained with a stronger increase in factor prices of info inputs.

Our paper is related to the literature that studies the impact of technical change on labor market outcomes. Several recent papers document that the rise in skill premium has progressively slowed since the 1990s, where skill premium is measured as the relative wage rate of college graduates over high school graduates (Valletta, 2016; Autor, 2017). Autor (2017) argues that the flattening of skill premium is driven by deceleration of skill demand commencing from 1992. Our results suggest deceleration does not happen for the workers in info occupations. In fact, the majority of the rise in info worker wage premium happens after the mid-1990s. Besides, rising returns to info occupations remain even after controlling for education. We also compare our info score with routine-nonroutine measures from Acemoglu and Autor (2011) in explaining the residual wage variation. Although our info score has limited explanatory power in 1980, it explains more variation than the routine and non-routine measures combined in recent years.

In a closely related paper, Gallipoli and Makridis (2018) construct a similar IT intensity measure and show that the relative wage rate of IT-intensive occupations is increasing. Different from our info score, they combine 12 scores from the O*NET dataset to measure IT intensity, including the one we use for our info score. We view our measure captures the meaning of information-intensive activities more precisely. In addition, the wage premium of info workers using our measure exhibits faster growth over time. Gallipoli and Makridis (2018) also link the rise of IT-intensive jobs with structural change towards the service sector, while we study the evolution of input-output structure over time.

Our accounting exercise resembles the one in Barany and Siegel (2019). Following Autor et al. (2003), they infer sector-factor specific technical changes in an environment where ICT capital substitutes routine workers. They find that technologies augmenting routine labor grow fastest, and it explains a large fraction of the differential growth rates of labor productivity across industries. This is consistent with our finding that non-info labor augmenting technology has the strongest growth since non-info occupations have a large overlap with routine occupations. We differ by also incorporating intermediate input into this framework, and study the role of outsourcing-specific technical change in explaining the observed change in intermediate shares.

This chapter proceeds as follows. Section 3.2 presents the empirical facts. Section 3.3 develops the accounting framework which we use to infer the sector-factor specific technical change. Section 3.4 presents the quantitative results. Section 3.5 concludes.

3.2 Empirical facts

3.2.1 Data and measurement

First, we measure occupational information intensity by directly taking the score on "Analyzing data or information" from O*NET database.[1] This score is reported for each occupation on a scale of 0 to 7.[2] A higher score indicates that a higher degree of data and information analyzing skill is required or needed for the occupation.

Our info score is related to IT intensity measure in Gallipoli and Makridis (2018), where they combine 12 scales in Work Activity, Skills, and Knowledge Scales to measure occupation-level IT intensity. However, among 12 scales, many of them are vaguely defined, such as "Updating and using relevant knowledge", "Quality control analysis", and "Management of material resources". We view our particular scale captures the meaning of information-intensive activities more precisely.

We then assign each industry an info score using 1980 data. We first rank all workers in a given industry based on their occupational info scores, taking into account the appropriate weights.[3] Then we assign the info score of the median worker to be the info score of the industry. Table 3.3 in Appendix presents the result. We classify industries with info score above the national median as info sector, and the rest as non-info sector. Based on this classification, the information sector includes Wholesale; Utilities; Information; Finance, Insurance and Real Estate; Professional Business Service; Educational Services, Health Care and Social Assistance; and Government.

3.2.2 Intermediate input from information sector

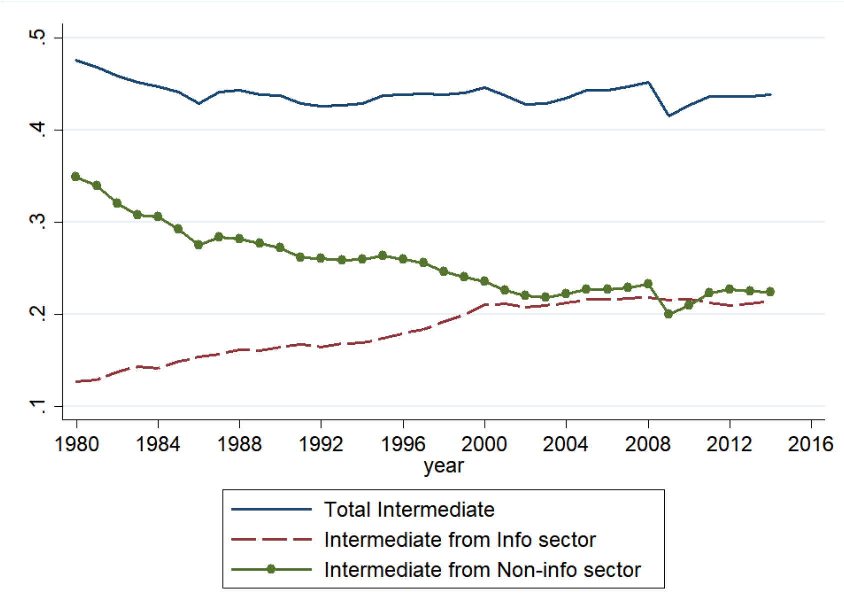
To calculate intermediate input shares, we use the Annual Input-Output Table from BEA. Using an expenditure-side approach, we decompose the aggregate

[1] O*NET database V15.1. "Analyzing data or information" is the only overlap scale between IT-intensity measure in Gallipoli and Makridis (2018) and non-routine analytical task measure in Acemoglu and Autor (2011).

[2] For each question, an importance score and a level score are reported. See the data appendix for details.

[3] The weights are individual weights multiplied by hours of work.

Figure 3.1: Aggregate intermediate share



Notes: This figure plots the aggregate intermediate share and its sectoral decomposition in U.S. Source: BEA Input-Output Table, and authors' calculations.

intermediate share $\frac{I}{G}$ as:

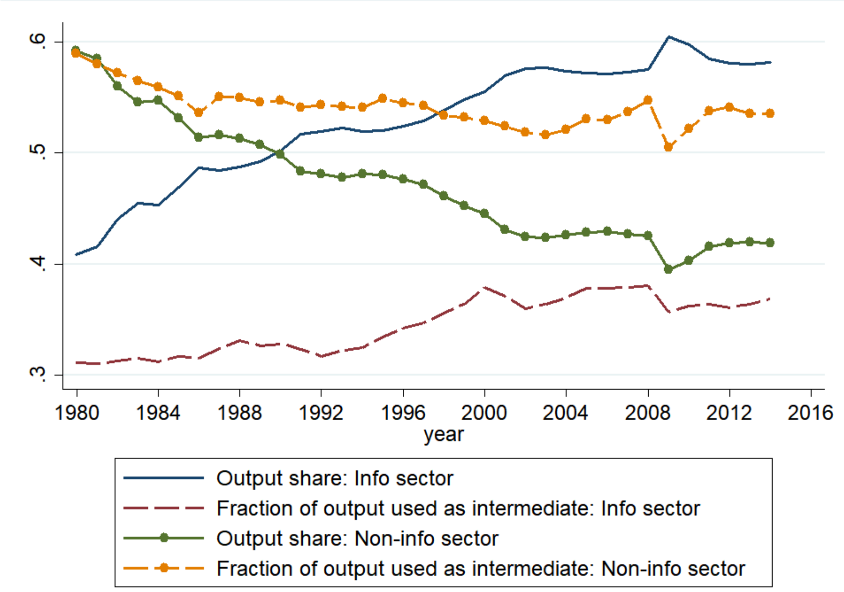
$$\frac{I}{G} = \frac{I_I + I_N}{G} = \frac{I_I}{G_I}\frac{G_I}{G} + \frac{I_N}{G_N}\frac{G_N}{G} \tag{3.1}$$

where I and G are total intermediate inputs and gross output. I_I and I_N are the intermediate input produced by info sector and non-info sector, G_I and G_N are the corresponding sectoral gross output. $\frac{I_I}{G_I}$ ($\frac{I_N}{G_N}$) denotes the fraction of info (non-info) sector's output used as intermediates.

Figure 3.1 presents aggregate share ($\frac{I}{G}$) and its sectoral decomposition ($\frac{I_I}{G}$ and $\frac{I_N}{G}$). The solid line, depicting $\frac{I}{G}$, shows that about 45% of gross output is used as intermediate input, and this ratio is roughly constant over time. The constant ratio masks important sectoral heterogeneity. The intermediate input from info sector ($\frac{I_I}{G}$) increases from about 13% in 1980 to more than 21% in 2014, while the intermediate input from non-info sector ($\frac{I_N}{G}$) declines.

As in equation 3.1, this rise in $\frac{I_I}{G}$ could be due to either a larger output share of info sector (larger $\frac{G_I}{G}$) or a more intensive use of info sector's output as intermediate inputs (higher $\frac{I_I}{G_I}$). Hence, Figure 3.2 plots each sector's output share ($\frac{G_I}{G}$ and $\frac{G_N}{G}$) and the fraction of its output used as intermediate input ($\frac{I_I}{G_I}$ and $\frac{I_N}{G_N}$). During the sample period, info sector's output share $\frac{G_I}{G}$ increases from 40% in 1980 to 58% in 2014, and the fraction of its output used as intermediate inputs $\frac{I_I}{G_I}$ also increases albeit with a smaller magnitude. Therefore, both the

Figure 3.2: Aggregate intermediate share decomposition



Notes: This figure shows the expenditure-side decomposition of aggregate intermediate share. Two solid line plot the output share of info and non-info sector in the gross output. Two dashed line show the fraction of sectoral output used as intermediate input for info and non-info sector. Source: BEA Input-Output Table, and authors' calculations.

rising $\frac{G_I}{G}$ and $\frac{I_I}{G_I}$ contribute to the higher $\frac{I_I}{G}$.

To understand the usage of info intermediate inputs, we also present a production-side decomposition:

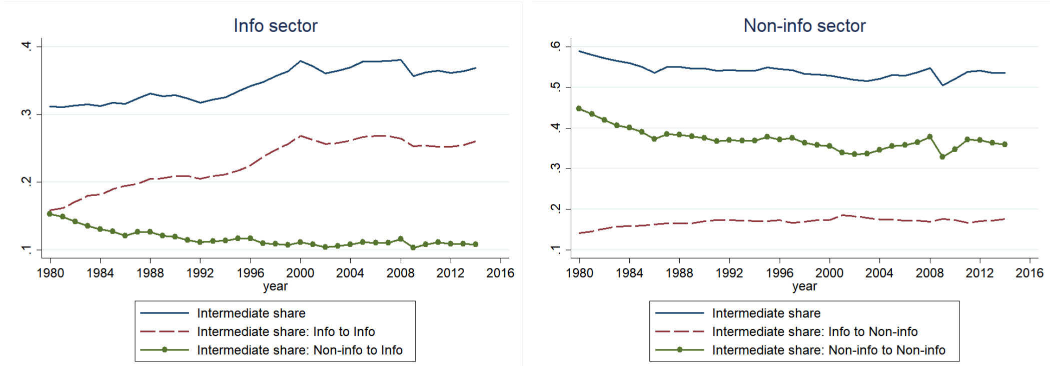
$$\frac{I}{G} = \frac{I_{II} + I_{IN}}{G_I}\frac{G_I}{G} + \frac{I_{NI} + I_{NN}}{G_N}\frac{G_N}{G} \tag{3.2}$$

where I_{JI} and I_{JN} are the info and non-info intermediate inputs used by sector $J \in \{I, N\}$. $\frac{I_{II}+I_{IN}}{G_I}$ and $\frac{I_{NI}+I_{NN}}{G_N}$ are the total intermediate shares in info and non-info sector. Figure 3.3 plots these two total shares, as well as its individual components, $\frac{I_{JI}}{G_J}$ and $\frac{I_{JN}}{G_J}$ where $J \in \{I, N\}$. In both info and non-info sectors, the intermediate share of info goods ($\frac{I_{JI}}{G_J}$) increases, and the intermediate share of non-info goods ($\frac{I_{JN}}{G_J}$) declines. Therefore, both sectors are using more info intermediate inputs over time.

3.2.3 Rise of info labor

During the same period, the return to occupations with higher info score has been increasing. For each occupation, we calculate the mean of real hourly wage by year. Figure 3.4 shows the binscatter plot of this occupational mean against the info score for 1980, 1990, 2000, and 2010. Obviously, the return to

Figure 3.3: Intermediate share by sector



Notes: This figure shows the intermediate share as well as its decomposition for both info and non-info sector. Source: BEA Input-Output Table, and authors' calculations.

Table 3.1: Regression of income on info score

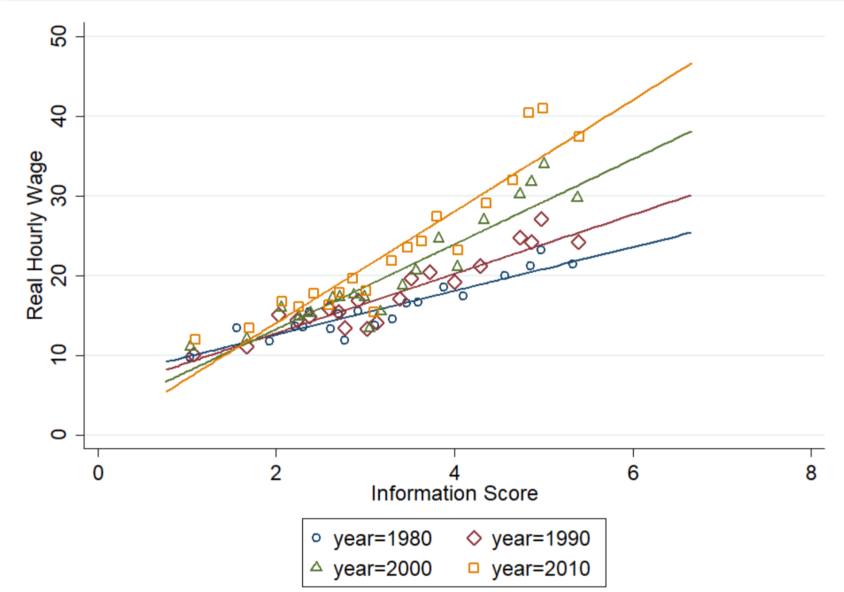
	(1) 1980	(2) 1990	(3) 2000	(4) 2010
info score	0.102***	0.131***	0.157***	0.177***
	(31.366)	(48.708)	(52.620)	(63.924)
experience	Yes	Yes	Yes	Yes
experience squared	Yes	Yes	Yes	Yes
race	Yes	Yes	Yes	Yes
edu	Yes	Yes	Yes	Yes
sex	Yes	Yes	Yes	Yes
N	47960	46728	42907	61462

Notes: This table reports the coefficients from regressing logarithm of individual real hourly wage on info score, controlling for race, gender, education, experience. Results are reported for 1980, 1990, 2000 and 2010. t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source: CPS ASEC from IPUMS, and authors' calculations.

info score has grown. In addition to this occupational level result, we regress the logarithm of individual real hourly wage on info score, controlling for race, gender, education, and experience. Table 3.1 summarizes the results. The coefficient on info score is also increasing over time, which implies the wage premium associated with information-intensive occupations becomes larger.

To illustrate the magnitude of this rise, we classify workers into two types, info workers and non-info workers. Workers in occupations with info score higher than the 1980 national median are classified as info workers and the rest are classified as non-info workers. Figure 3.5 plots the wage premium of info workers over time. The solid line depicts the raw wage premium in the data, while the dashed line shows the wage premium when we control for the composition change of info workers and non-info workers. For both measures, the wage premium of info workers increases by about 30% between

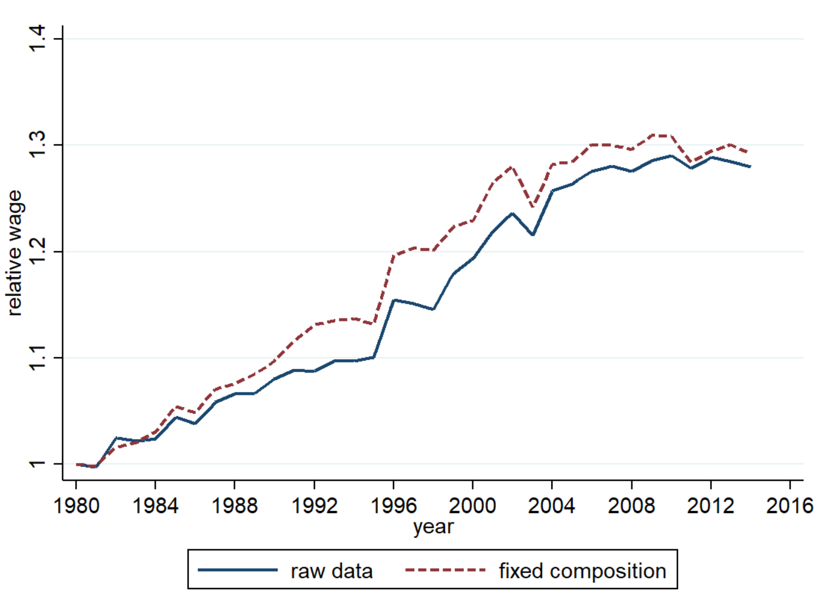
Figure 3.4: Returns to info occupations



Notes: This figure shows the binscatter plot between occupational average real hourly wage and info score for 1980, 1990, 2000 and 2010. Source: CPS ASEC from IPUMS, and authors' calculations.

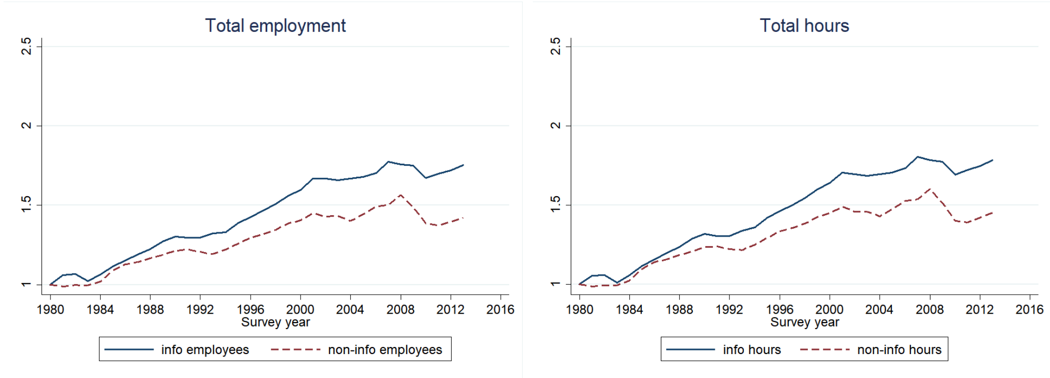
1980 and 2010. Besides, the post-mid-1990s period explains more than 50% of the increase, and the wage premium stagnates around 2008.

Figure 3.5: Wage premium of info worker



Notes: Raw data refers to the average hourly wage of info workers relative to that of non-info workers. We can express this relative wage rate as $\frac{w_{i,t}}{w_{n,t}} = \frac{\sum_c s_{i,c,t} w_{i,c,t}}{\sum_c s_{n,c,t} w_{n,c,t}}$ where $\sum_c s_{i,c,t} = 1$ and $\sum_c s_{n,c,t} = 1$. Here c represents demographic cells. $s_{i,c,t}$ means the share of occupation i and demographic cell c at time t. For both info and non-info labor, we have twelve education categories, two race categories, six age categories and two sex categories. The fixed composition relative wage is calculated by replacing $s_{i,c,t}$ with time average $s_{i,c}$. Source: CPS ASEC from IPUMS, and authors' calculations.

Figure 3.6: Employment and hours worked: info worker and non-info worker

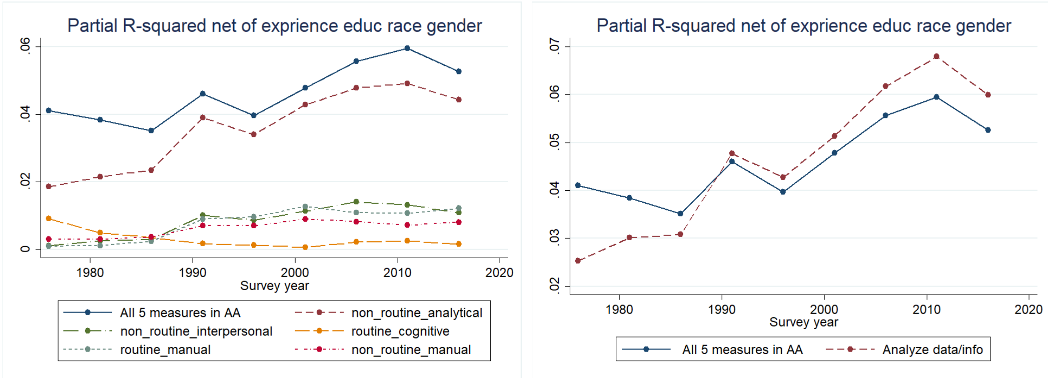


Notes: This figure plots the trend of employment and hours worked for info and non-info worker. We normalize the 1980 values to unity. Source: CPS ASEC from IPUMS, and authors' calculations.

Apart from the increase in relative wage, we also observe a rise in the employment of info workers relative to non-info workers. Figure 3.6 shows the employment and hours worked of info workers and non-info workers. Normalizing the 1980 values to unity, both employment and hours worked of info worker increase by almost 80% since 1980 whereas they only increase by 40% for non-info worker.

Our info score captures an important dimension of wage distribution for the following reasons. Firstly, recent papers document that the growth of college premium has slowed since the 1990s and almost stagnated in the 2000s (Valletta, 2016; Autor, 2017). In contrast, our result shows that the post-mid-1990s period explains the majority of the rise in wage premium of info workers. Secondly, we also compare the info score with the routine and non-routine measures used by Acemoglu and Autor (2011) in explaining residual wage variation. We first regress logarithm of real hourly wage on experience, education, gender and race, collect the residuals, and then regress the residuals on our info score and the routine-non-routine measures. Figure 3.7 presents the R-squared of these residual regressions. The left panel shows that the rising explanatory power of routine and non-routine measures is mostly driven by the non-routine analytical content of the occupations. The right panel contrasts our info score with the five measures in Acemoglu and Autor (2011). The explanatory power of the info score increases over time, and it explains more wage dispersion than the routine and non-routine measures combined in recent years. Lastly, we compare our info score with IT intensity measure in Gallipoli and Makridis (2018). Using a different measure of IT intensity, they find that between 1980 and 2015, the wage premium of IT worker increases

Figure 3.7: Partial R^2 net of experience, education, race and gender



Notes: This figure compares info score with routine-nonroutine measures in Acemoglu and Autor (2011) (AA) in terms of explaining the residual wage variation. The partial R^2 values presented above are calculated as follows. First, we regress logarithm of real hourly wages on a quadratic in experience, and dummies of race, gender and education. The wage residual are then regressed separately on the variable groups of interest. R^2 from each residual regression is plotted. Source: IPUMS CPS ASEC and O*NET, authors' calculations.

by 23%.[4] Compared to their results, the wage premium using our info score measure displays a larger increase over time (Figure 3.5).

Motivated by the two facts (i) the output from info sector is used more intensively as an intermediate input in the production; (ii) there is a sharp increase of the employment and wage premium of info worker relative to non-info worker, we develop an accounting framework with factor-augmenting technical changes to understand the drivers of these observations.

3.3 Theoretical framework

In this section, we present a partial-equilibrium accounting framework with factor-augmenting technical change. Sectoral gross output is produced by combining capital stock, info and non-info labor inputs, and info and non-info intermediate inputs. Under this specification, we then infer the series of factor-augmenting productivities using data on sectoral output growth, quantities and prices of various factor inputs and factor shares. Since we observe the sectoral gross output, this approach does not require a specification on demand side.

[4]See Table 1 in Gallipoli and Makridis (2018).

3.3.1 Sectoral production

We consider two aggregate sectors: info sector (I) and non-info sector (N). Each sector $J \in \{I, N\}$ has the following nested CES production function for gross output:

$$Q_J = \left[(A_J K_J)^{\frac{\sigma_k - 1}{\sigma_k}} + \left(\left[C_{Ji}^{\frac{\sigma_C - 1}{\sigma_C}} + C_{Jn}^{\frac{\sigma_C - 1}{\sigma_C}} \right]^{\frac{\sigma_C}{\sigma_C - 1}} \right)^{\frac{\sigma_k - 1}{\sigma_k}} \right]^{\frac{\sigma_k}{\sigma_k - 1}} \tag{3.3}$$

$$C_{Ji} = \left[(A_{Ji} L_{Ji})^{\frac{\sigma_I - 1}{\sigma_I}} + (\lambda_{JI} M_{JI})^{\frac{\sigma_I - 1}{\sigma_I}} \right]^{\frac{\sigma_I}{\sigma_I - 1}} \tag{3.4}$$

$$C_{Jn} = \left[(A_{Jn} L_{Jn})^{\frac{\sigma_N - 1}{\sigma_N}} + (\lambda_{JN} M_{JN})^{\frac{\sigma_N - 1}{\sigma_N}} \right]^{\frac{\sigma_N}{\sigma_N - 1}}. \tag{3.5}$$

The set of sector-factor specific technical changes is $\{A_J, A_{Ji}, A_{Jn}, \lambda_{JI}, \lambda_{JN}\}$. The upper case I and N refer to sector, and the lower case i and n refer to worker type. M_{JI} and M_{JN} are info and non-info output used by sector J as intermediate input. We interpret the corresponding factor-augmenting technologies λ_{JI} and λ_{JN} as outsourcing-specific technical changes. L_{Ji} and L_{Jn} denote the info and non-info labor employed in sector J with labor-augmenting technologies A_{Ji} and A_{Jn}. Finally, K_J is sectoral capital stock with capital-augmenting technology A_J.

Our nested CES specification is a generalization of the production technology in Chan (2017).[5] Info labor L_{Ji} and info intermediate input M_{JI} are combined to produce info input aggregate C_{Ji} with elasticity σ_I.[6] Similarly, non-info labor L_{Jn} and non-info intermediate input M_{JN} form the non-info input aggregate C_{Jn}, with elasticity σ_N. Then the info aggregate C_{Ji} and non-info aggregate C_{Jn} form a labor-intermediate aggregate with elasticity σ_C. Finally, output in sector J is the CES combination of capital K_J and the labor-intermediate aggregate that contains all other inputs.

[5] Chan (2017) consider a production function with $\sigma_k = \sigma_C = 1$.

[6] For example, firms which need marketing services can combine their own marketing labor and purchased marketing services.

3.3.2 Inferring technologies

In this section, we discuss how we infer the sector-factor specific productivities. Each sector solves the following profit maximization problem

$$\max_{K_J, L_{Ji}, L_{Jn}, M_{JI}, M_{JN}} \quad P_J Q_J - W_{Ji} L_{Ji} - W_{Jn} L_{Jn} - P_I M_{JI} - P_N M_{JN} - R K_J \tag{3.6}$$

subject to equation 3.3 to 3.5. W_{Ji} and W_{Jn} are sector-specific wages of info and non-info labor. P_J is price of sector-J output.[7]

Solving the above problem, we express the relative technologies in relative prices and relative factor shares:

$$\frac{A_{Ji}}{\lambda_{JI}} = \frac{W_{Ji}}{P_I} \left[\frac{W_{Ji} L_{Ji}}{P_I M_{JI}} \right]^{\frac{1}{\sigma_I - 1}} \tag{3.7}$$

$$\frac{A_{Jn}}{\lambda_{JN}} = \frac{W_{Jn}}{P_N} \left[\frac{W_{Jn} L_{Jn}}{P_N M_{JN}} \right]^{\frac{1}{\sigma_N - 1}} \tag{3.8}$$

$$\frac{A_{Ji}}{A_{Jn}} = \left[\frac{W_{Ji} L_{Ji}}{W_{Jn} L_{Jn}} \right]^{\frac{1}{\sigma_C - 1}} \frac{\left[1 + \frac{P_N M_{JN}}{W_{Jn} L_{Jn}} \right]^{\frac{\sigma_C - \sigma_N}{(\sigma_N - 1)(\sigma_C - 1)}}}{\left[1 + \frac{P_I M_{JI}}{W_{Ji} L_{Ji}} \right]^{\frac{\sigma_C - \sigma_I}{(\sigma_I - 1)(\sigma_C - 1)}}} \frac{W_{Ji}}{W_{Jn}} \tag{3.9}$$

$$\frac{A_J}{A_{Ji}} = \frac{R}{W_{Ji}} \left[\frac{R K_J}{W_{Ji} L_{Ji}} \right]^{\frac{1}{\sigma_k - 1}} \left[1 + \frac{P_I M_{JI}}{W_{Ji} L_{Ji}} \right]^{\frac{\sigma_k - \sigma_I}{(\sigma_I - 1)(\sigma_k - 1)}} \left[1 + \frac{W_{Jn} L_{Jn} + P_N M_{JN}}{W_{Ji} L_{Ji} + P_I M_{JI}} \right]^{\frac{\sigma_k - \sigma_C}{(\sigma_C - 1)(\sigma_k - 1)}} . \tag{3.10}$$

Equation 3.7 and 3.8 write the productivity of a worker type (i and n) relative to the corresponding intermediate input as a function of relative prices and expenditure ratios. Similarly, relative technologies of info worker and non-info worker can be calculated from equation 3.9 using wage premium of info workers, output share of both types of labor and intermediate inputs. Finally equation 3.10, given a series of capital-augmenting technologies A_J, delivers the series of factor-augmenting technologies of info labor and hence all remaining technologies.

Using the optimality conditions, sectoral gross output is given by:

$$Q_J = A_J K_J \left[1 + \frac{W_{Ji} L_{Ji} + P_J M_{JI} + W_{Jn} L_{Jn} + P_N M_{JN}}{R K_J} \right]^{\frac{\sigma_k}{\sigma_k - 1}} . \tag{3.11}$$

[7] Since sectors could use different combinations of intermediate goods within M_{JI} and M_{JN}, we consider a case where intermediate prices are sector-factor specific in the robustness check section.

This delivers the series of capital-augmenting technology A_J, given the series of sectoral real output Q_J, capital stock K_J and factor shares.[8]

3.3.3 Data and implementation

This subsection outlines the data and methods we use to measure the relative factor prices and factor output shares. Without loss of generality, we normalize all the quantity measures by the Full-Time Equivalent (FTE) labor force following Barany and Siegel (2019).

We extract industry level nominal and real gross output and its decomposition from World KLEMS data.[9] We group industries into info and non-info sector, and calculate the total labor output share for each sector:

$$\theta_J^L = \frac{\text{Compensation of employees in sector J}}{\text{Gross output in sector J}}. \tag{3.12}$$

The left panel of Figure 3.8 plots the labor output shares. It is higher in info sector. Since the mid-1980s, labor output share is on a decreasing trend, especially for non-info sector. What we measure here is not exactly the labor income share in the literature, but this result is consistent with the recent decline of labor income share (e.g., Karabarbounis and Neiman (2014), Autor et al. (2017)).[10]

KLEMS data does not provide labor compensation by occupations. Thus, we turn to IPUMS CPS data. As explained in section 3.2, we classify occupations into info and non-info occupations, with the corresponding workers being info and non-info workers. The labor income share of occupation o in sector J is

[8]We can infer the initial technologies from initial sectoral prices

$$A_{Jt=0} = \frac{R_{Jt=0}}{P_{Jt=0}} \left[1 + \frac{W_{Ji}L_{Ji} + P_J M_{JI} + W_{Jn}L_{Jn} + P_N M_{JN}}{RK_J}\right]^{-\frac{1}{\sigma_k - 1}}.$$

Since we are concerned with the growth rate of the technologies, without loss of generality, we could also normalize $A_{Jt=0}$ for both sectors.

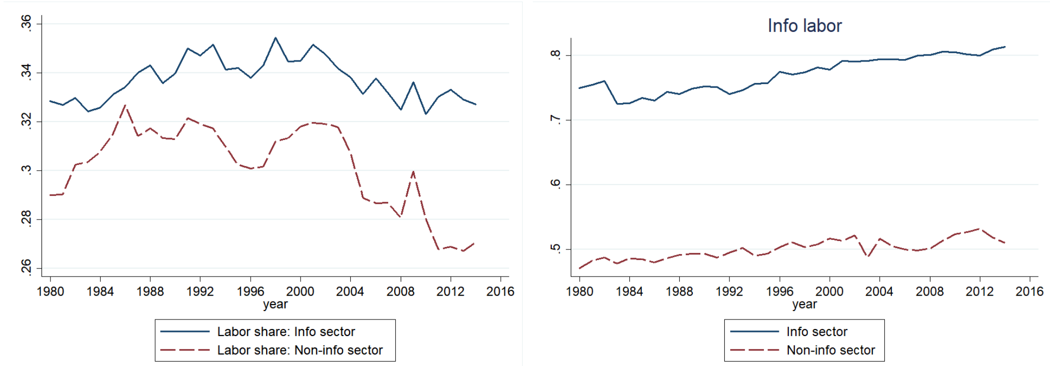
[9]Data is available at `http://www.worldklems.net/data.htm`.

[10]Similarly, we calculate the total intermediate share of each sector J as

$$\theta_J^{Int} = \frac{\text{Intermediate inputs used by sector J}}{\text{Gross output in sector J}}. \tag{3.13}$$

Besides, to get the intermediate share by each sector, we use the data from the Annual Input-Output Table as in section 3.2. Note there is a small discrepancy between intermediate shares reported by World KLEMS and Input-Output Table published by BEA. We harmonize the two data sources by scaling the intermediate shares reported by BEA such that they add up to the one reported by World KLEMS.

Figure 3.8: Labor share



Notes: The left panel plots the fraction of labor income in gross output for both info and non-info sector using data from World KLEMS. The right panel plots the ratio between info worker's income and total labor income for both info and non-info sector using data from IPUMS CPS.

calculated as

$$\theta_{Jo} = \frac{\text{Earnings of occupation } o \text{ workers in sector } J}{\text{Earnings of all sector } J \text{ workers}}. \tag{3.14}$$

Then the output share of info and non-info labor in sector J are $\theta_J^L \theta_{Ji}$ and $\theta_J^L \theta_{Jn}$. The right panel of Figure 3.8 plots θ_{Ji} for both sectors. Info workers account for about 80% of all the labor income in info sector and 50% in non-info sector. In both sectors, income share of info worker is on an increasing trend.

To infer the sector-specific wage for info and non-info workers, we first calculate the share of hours worked by occupation o in sector J, L_{Jo}. Then the sector-occupation specific wage rate W_{Jo} can be obtained using the accounting identity[11]

$$W_{Jo} L_{Jo} = \widehat{P_J Q_J} \theta_J^L \theta_{Jo}. \tag{3.15}$$

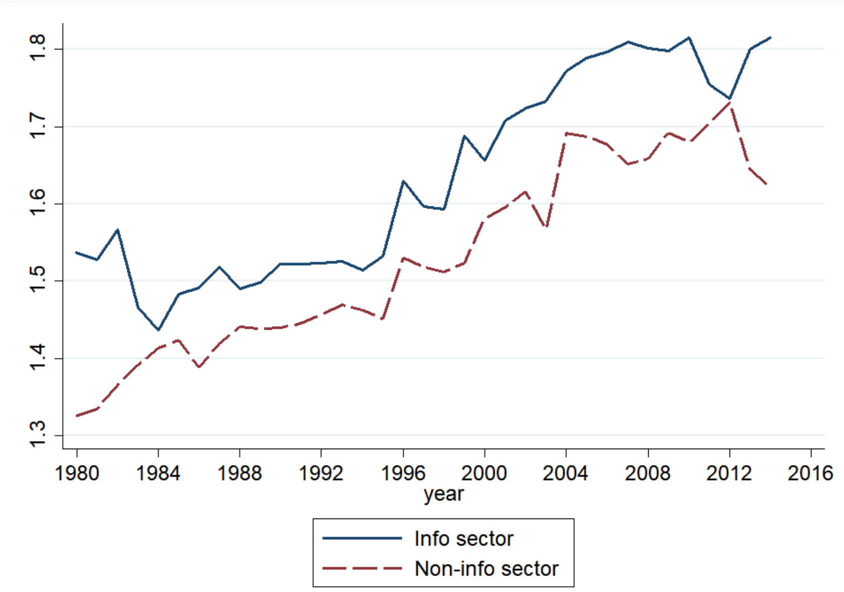
Figure 3.9 illustrates the implied wage premium of info labor in two sectors. As in the Figure 3.5, the wage premium increases in both sectors and the post-mid-1990s period explains the majority of the rise.

Next, we calculate the real quantity and price for sectoral output using the cyclical expansion approach in Herrendorf et al. (2013).[12] Figure 3.10 plots the evolution of quantity and price indices for both sectors, where info sector grows faster than the non-info sector in terms of both measures.

[11] The right-hand side gross output is normalized by FTE workers, i.e., $\widehat{P_J Q_J} = \frac{P_J Q_J}{L_{FTE}}$.

[12] See the data appendix for details.

Figure 3.9: Inferred wage premium of info labor by sector



Notes: This figure plots the inferred wage premium for both info and non-info sector. The wage rate is inferred by using the accounting identity in equation 3.15. Source: CPS ASEC from IPUMS, World KLEMS, and authors' calculations.

Figure 3.10: Quantity and price indices by sector

Notes: The quantity and price indices of info and non-info sector are calculated using cyclical expansion approach as in Herrendorf et al. (2013). Source: World KLEMS, and authors' calculations.

Finally, we calculate the rental price of capital and real capital stock in each sector. From the fixed asset account of BEA, we extract quantity index of net stock of fixed assets. We use the following accounting identity to get the rental price of capital:

$$RK = PQ \sum_{j} \frac{P_J Q_J}{PQ} \left(1 - \theta_J^L - \theta_J^{Int}\right). \tag{3.16}$$

We back out K_J in each sector using the industry-level accounting identity.

3.3.4 Elasticities

We need to parameterize four elasticities. First, we parameterize σ_I and σ_N, the elasticity between a given type of labor and its corresponding intermediate input. Chan (2017) estimates that elasticity between task-specific labor and intermediate input, where he considers both detailed task classes and one aggregate task class.[13] Across different specifications, the average elasticity of substitution is about 2.3. We use this as the benchmark estimate for σ_I and σ_N.[14]

Second, we choose σ_C, the elasticity of substitution between info aggregate and non-info aggregate. Gallipoli and Makridis (2018) estimate elasticity of substitution between IT and non-IT intensive labor and find values of 1.3 in services and 1.6 in manufacturing. However, this does not map directly into the elasticity between info and non-info aggregate. We can infer σ_C using the information on σ_I, elasticity between info and non-info labor and factor shares.[15] We set the benchmark value for σ_C to be 0.5.

[13] Detailed task classes contain thirteen input types: ICT, Legal & Accounting, Engineering, Marketing, Employment & Training, Transportation, Security, Cleaning, Other services, Food, Wood & Paper, Heavy Industry and Tools, Machinery, Goods.

[14] In the specification with detailed input classes, we can classify the input types into info input and non-info input. For example, input types, ICT, Legal & Accounting, Engineering, Marketing, Employment & Training, are classified as info input. The remaining types are classified as non-info input. In both groups, the average elasticity is around 2.3.

[15] If we substitute equation 3.7 and 3.8 into 3.9, we express the ratio between info and non-info labor in terms of prices and technologies. Differentiating the equation with respect to $log\left(W_{Ji}\right)$ holding other prices fixed yields

$$\sigma_C = -\frac{\partial log\left(\frac{L_{Ji}}{L_{Jn}}\right)}{\partial log\left(W_{Ji}\right)} + \left(-\frac{\partial log\left(\frac{L_{Ji}}{L_{Jn}}\right)}{\partial log\left(W_{Ji}\right)} - \sigma_I\right)\left[\frac{P_I M_{JI}}{W_{Ji} L_{Ji}}\right]$$

where $-\frac{\partial log\left(\frac{L_{Ji}}{L_{Jn}}\right)}{\partial log(W_{Ji})}$ measures the elasticity of substitution between info and non-info labor and $\frac{P_I M_{JI}}{W_{Ji} L_{Ji}}$ is around 0.9 for info sector and 1.1 for non-info sector.

Lastly, we set σ_k, the elasticity between capital and labor-intermediate aggregate. Barnes et al. (2008) estimate the firm-level elasticity between capital and other inputs to be 0.4. However, as pointed by Oberfield and Raval (2014), macro-level elasticity tends to be larger than micro-level elasticity.[16] We set $\sigma_k = 0.6$ as the benchmark value. In section 3.4.3.1, we conduct a battery of robustness checks regarding these elasticities.

3.4 Quantitative results

Table 3.2 reports the inferred annual growth rate of sector-factor specific technologies from 1980 to 2014 and two sub-periods, 1980-1995 and 1995-2014. Several patterns emerge. First, over the entire sample period, the productivity of non-info labor grows faster in both sectors, with an annual growth rate of 2.8 percent in info sector and 2.3 percent in non-info sector. Technologies augmenting info labor grows less significantly, with a negative growth rate in info sector. In terms of outsourcing-specific technologies, the increase is also more pronounced for non-info intermediate. Outsourcing-specific technologies for info intermediate do not exhibit strong growth with an annual growth rate of 0.4 percent in non-info sector and a negative growth rate in info sector. Finally, the capital-augmenting technology increases at 1.6 percent annually in info sector, while falls in non-info sector.[17]

The average growth rate over the entire period masks interesting heterogeneity over time. Comparing the two sub-periods, in both sectors, the growth of technology augmenting non-info labor is primarily driven by the rise during 1980-1995. On the other hand, technologies augmenting info input tend to grow faster in the post-1995 era. The results suggest that the advent of the Internet era has unequal impacts across factors.

In an environment without intermediate inputs, Barany and Siegel (2019) conduct a similar accounting exercise to study the drivers of the difference in labor productivity growth across industries.[18] They find that sector-specific routine labor augmenting technical change is the most important driver of sectoral differences. Similar to their findings, we find that non-info labor augmenting

[16] Oberfield and Raval (2014) find that within-plant substitution between capital and labor accounts for about 60% of overall substitution for the manufacturing sector.

[17] The negative growth rate of capital-augmenting technology is consistent with the literature. See Antras (2004) and Barany and Siegel (2019) for example.

[18] They classify occupations into three groups: routine, manual and abstract. They consider three sectors: low-skilled service, goods and high-skilled service sectors.

Table 3.2: Annual growth rate of technologies

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.993	0.989	1.028	1.010	1.016
Non-info sector	1.007	1.004	1.023	1.012	0.977
1980-1995					
Info sector	0.983	0.984	1.042	0.995	1.019
Non-info sector	1.003	1.004	1.041	1.015	0.987
1995-2014					
Info sector	1.001	0.992	1.017	1.023	1.014
Non-info sector	1.010	1.004	1.009	1.010	0.970

Notes: This table reports the inferred growth rate of sector-factor specific technical change, where growth rate refers to $\frac{A_{t+1}}{A_t}$. A number smaller the one means the corresponding productivity is declining.

technology has strongest growth over time since most of non-info occupations belong to the group of routine workers.[19]

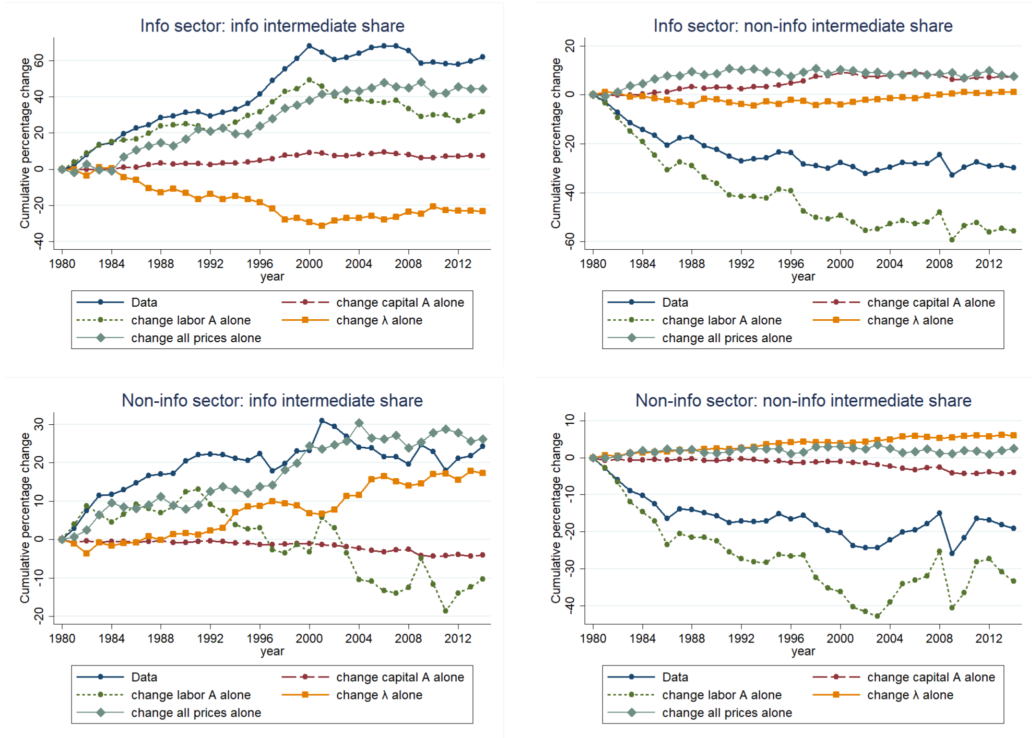
In the data, we observe a rise in wage premium of info workers and an increase in intermediate share from info sector. These patterns could be driven by the sector-factor specific technical changes documented in Table 3.2. In the following sections, we study the role of factor-augmenting technologies in driving the observed patterns. We focus on the role of labor-augmenting technical changes and outsourcing-specific technical changes.

3.4.1 Intermediate shares and technical changes

The evolution of intermediate share depends on both the productivity growth and price changes. To see how each factor contributes to the observed changes in intermediate shares, we perform four counterfactual exercises. In each exercise, we allow one set of factors to vary over time and keep the others fixed at the 1980 level, and then compute the implied intermediate shares. Figure 3.11 shows the cumulative percentage change in intermediate shares for the four exercises: (i) varying capital-augmenting technology A_J; (ii) varying labor-augmenting technologies A_{Ji} and A_{Jn}; (iii) varying outsourcing-specific technologies λ_{Ji} and λ_{Jn}; and (iv) varying all prices R, W and P. The two upper panels of Figure 3.11 represent the change of info and non-info intermediate shares in info sector. The two lower panels show the change of the

[19]More than 70% non-info occupations belong to routine occupations. See Table 3.8 in Appendix.

Figure 3.11: Technical changes and intermediate shares



Notes: This figure shows the cumulative percentage change of intermediate shares since 1980, including the observed data and four counterfactual cases. The evolution of intermediate share depends on both the productivity growth and price changes. In each counterfactual experiment, we allow one set of variables to change while fixing others at the 1980 level and then compute the implied intermediate share. For example, the first case allows capital augmenting technology to evolve over time while fixing all prices and other productivities at the 1980 level.

corresponding intermediate shares in non-info sector.

Figure 3.11 shows that, among four intermediate shares, outsourcing-specifc technical change can only explain the rising info intermediate share in non-info sector. On the other hand, labor-augmenting technical change contributes to the changes in the rest three intermediate shares.

These results depend on the elasticity of substitution between different inputs and the pace of technical changes. For example, in non-info sector, Table 3.2 shows that non-info labor augmenting technology grows faster than the one augments info labor. This unbalanced growth increases the demand and expenditure share on info input due to the complementarity between info and non-info inputs. However, within info inputs, technology augmenting info-labor grows over time, which substitutes expenditure away from info intermediates. Between 1980 and 1990, the complementarity effect dominates the substitution effect, resulting in a rise of info intermediate share. As info labor's productivity grows faster since 1990, substitution effect prevails, which leads to a decline in the info intermediate share.

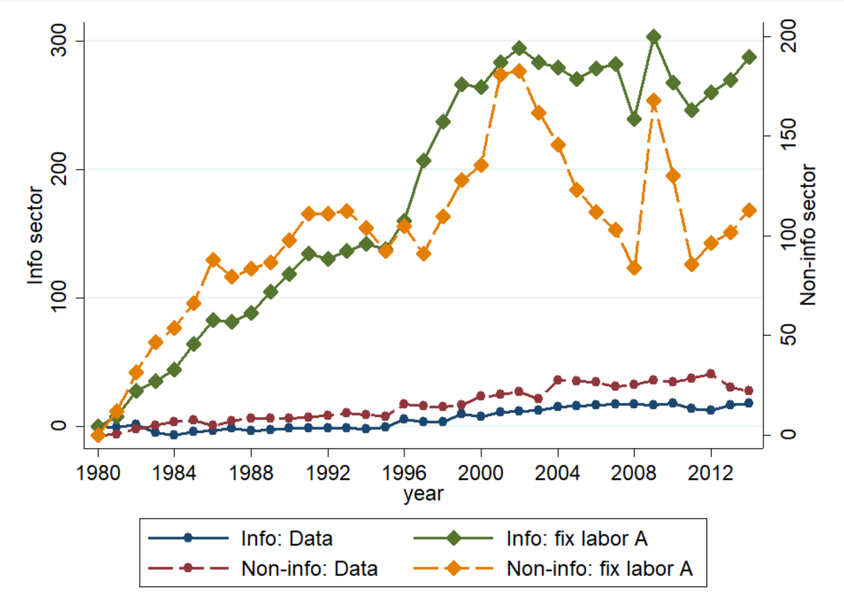
Figure 3.11 also suggests that capital-augmenting technical change does not play an important role in explaining the observed changes in intermediate shares. The implied intermediate shares exhibit little variation over time when we only allow capital-augmenting technology to change while fixing other productivities and prices at the 1980 level. Besides, change in prices contributes to the rising info intermediate share but cannot explain the declining non-info intermediate share.

The above results are based on a partial-equilibrium framework where we keep prices constant while changing productivities. Leaving the demand side unspecified gives us flexibility in accounting for the evolution of sector-factor technologies. However, it loses the general equilibrium effect that changes in technologies generate variations in intermediate shares through affecting equilibrium relative prices. The general equilibrium effect of outsourcing-specific technical change could be important in explaining the sector reallocation. We leave this for future research.

3.4.2 Wage premium and labor-augmenting technical change

Here we explore the role of labor-augmenting technologies in driving the wage premium of info workers. To calculate the implied wage premium in the absence of labor-augmenting technologies, we fix the labor-augmenting technologies as in the 1980's level and allow the intermediate shares and labor output shares to evolve as in the data. Figure 3.12 plots the cumulative percentage change for the observed and counterfactual wage premium in both sectors. Without labor-augmenting technical changes, the wage premium of info workers increases much more dramatically in both sectors. Intuitively, since info inputs and non-info inputs are complements, technical change that is biased towards non-info inputs would increase the demand and expenditure share for info inputs. In the absence of such technical change, the rising expenditure share on info input can only be explained with a stronger increase in info factor prices.

Figure 3.12: Wage premium and labor-augmenting technical change



Notes: This figure shows cumulative percentage change in observed and counterfactual wage premium for info and non-info sector since 1980. Counterfactual wage premium is calculated by fixing labor-augmenting technical change at its 1980 level.

3.4.3 Robustness check

3.4.3.1 Alternative elasticity of substitution

In the benchmark calibration, we set the elasticities $\sigma_I = \sigma_N = 2.3$, $\sigma_C = 0.5$ and $\sigma_k = 0.6$. In this section, we examine the robustness of our results regarding these elasticities.

Elasticity between labor and intermediate input. Here we consider three alternative sets of values: (i) $\sigma_I = \sigma_N = 1.8$; (ii) $\sigma_I = \sigma_N = 2.8$; (iii) $\sigma_I = 2.2$ and $\sigma_N = 2$.[20] In Table 3.4 and Figure 3.13 to 3.15, we report both the annual growth rate of factor-augmenting technologies and cumulative percentage change in counterfactual intermediate shares for alternative sets of values. Panel (a)-(c) in Figure 3.21 plot the corresponding counterfactual wage premium. Across different parameterizations, the general pattern remains unchanged as non-info input augmenting technologies grow faster. In addition, apart from the info intermediate share in non-info sector, labor-augmenting technical change contributes to the observed change in intermediate shares.

Elasticity between info aggregate and non-info aggregate. In our baseline results, we set $\sigma_C = 0.5$. Here we use alternative values of σ_C, 0.4 and

[20]The last set of values is the median elasticity for info input and non-info input in Chan (2017) given the classification in footnote 14.

0.8. In Table 3.5, we report the implied growth rates. Compared with Table 3.2, a greater substitutability between info and non-info aggregate enlarges the gap in growth rates between technologies augmenting info and non-info inputs, and technologies augmenting non-info inputs grow even faster compared with benchmark calibration. Figure 3.16 and 3.17 show the change in counterfactual intermediate shares for these alternative values. When $\sigma_C = 0.8$, labor-augmenting technical change contributes to the observed changes for all four intermediate shares. Panel (d) in Figure 3.21 and panel (a) in Figure 3.22 plot the counterfactual wage premium. One thing to note is that, with a larger σ_C, wage premium would increase much more than the benchmark calibration if we fix the labor augmenting technologies at the 1980 level.

Elasticity between capital and labor-intermediate aggregate. We consider two alternative values of σ_k, 0.4 and 0.8. Table 3.6 shows that alternative values of σ_k change the level of growth rate for capital-augmenting technologies. The implied growth rates remain similar for technologies augmenting info and non-info inputs. Counterfactual intermediate shares are reported in Figure 3.18 and 3.19, and wage premium is reported in Figure 3.22.

3.4.3.2 Sector-specific price for intermediate inputs

In the benchmark framework, we assume that the intermediate input price is the corresponding price of sectoral output. However, since two sectors could use different compositions of intermediate goods within info input M_{JI} and non-info input M_{JN}, the corresponding price indices may not exhibit the same pattern over time. Thus we construct intermediate input price indices that are sector-specific. First, we use the BEA Input-Output Table to estimate the intermediate purchase of info and non-info sector from 65 KLEMS industries. We then construct the sector-factor specific price index using the cyclical expansion approach. With the new series of prices of intermediate input, we redo the accounting exercise. Table 3.7 presents the inferred growth rate of technologies when we allow for sector-factor specific prices of intermediate inputs. Compared with Table 3.2, there are minor changes in the growth rate of technologies augmenting info and non-info inputs. Figure 3.20 and 3.22 depict the resulting counterfactual intermediate shares and wage premium. The general patterns remain unchanged.

3.5 Conclusion

In this chapter, we document the rise of information inputs in production. Specifically, we find (i) output from info sector is used more intensively as an intermediate input and (ii) there is a sharp rise of returns and employment for info workers relative to non-info workers. We develop an accounting framework to understand the nature of technical change that drives these patterns. The accounting exercise shows that technologies augmenting non-info inputs tend to grow faster. The partial-equilibrium counterfactual analysis indicates that labor-augmenting technical change is important in explaining the observed change in wage premium and intermediate shares.

3.6 Appendix

3.6.1 Data appendix

In this section, we discuss the data sources and variables we use in detail.

Occupational info score. We first obtain an occupational information task intensity from O*NET database V15.1. O*NET database provides occupational information on worker attributes and job characteristics. We measure occupational information intensity by directly taking the score on "Analyzing data or information" from the O*NET database. For the question, an importance score and a level score are reported. Acemoglu and Autor (2011) use the importance score, but we view the level score is more appropriate for our purpose.[21] The level score is reported for each occupation on a scale of 0 to 7. A higher score indicates that a higher degree of data and information analyzing skill is required or needed for the occupation. The occupation code in the O*NET database is based on SOC2010 occupation code. We use a crosswalk to link each SOC2010 occupation to 2010 classification in the IPUMS data.[22]

Individual level data on employment and wages. We use the March ASEC sample of the CPS data downloaded from IPUMS. We only keep individuals who are between 20 to 64 years old, work for at least 40 weeks in the past year, and work for at least 35 hours each week. This project focuses on wage workers, so we drop all self-employed individuals. The income variable we use is wage income.

For occupations, we use the "occ10ly" variable. According to IPUMS, this provides consistent occupational codes for the respondents during the previous calendar year using the 2010 Census Bureau occupational classification system. This allows us to match the occupational info score from O*NET data.

For industry classification, we make use of the "ind90ly" variable. It provides consistent occupational codes for the respondents during the previous calendar year using the 1990 industrial classification system. We convert this into the 2003 industrial classification system and use the crosswalk between 2003 codes

[21] The O*NET website gives an example to explain the difference. See `https://www.onetonline.org/help/online/scales`. Speaking skill is important to both lawyers and paralegals, but a higher level of speaking skill is required to become a lawyer. The correlation between level score and importance score is 0.97.

[22] The crosswalk is available via `https://view.officeapps.live.com/op/view.aspx?src=https://www2.census.gov/programs-surveys/demo/guidance/industry-occupation/2010-occ-codes-with-crosswalk-from-2002-2011.xls`.

and NAICS two-digit sector.[23] Finally, we match the NAICS code to the 15 sector classification used by BEA.

BEA Input-Output data. We classify industries into info and non-info sector based on 15 sector classification used by BEA.[24] With occupational info score from O*NET and employment information in CPS, we first calculate the median info score of each sector's workforce in 1980. Industries with median info score higher than the national median are classified as the info sector. The rest are classified as the non-info sector. Table 3.3 presents the results.

We study the intermediate shares by constructing industry-by-industry direct requirements table. Specifically, we first extract the annual make and use table between 1980 and 2014.[25] Annual use table shows the uses of commodities by intermediate and final users. Annual make table shows the production of commodities by industry.

We follow the notation of BEA, and extract the following information to construct the industry-by-industry direct requirements matrix.[26] q is a commodity-by-one column vector, which reports the amount of output by commodities. g is an industry-by-one column vector, which reports output by industries. U is a commodity-by-industry matrix, where each column shows for a given industry the amount of different commodities it uses. W is an industry-by-commodity matrix, where each column shows for a given commodity the amount of output produced by different industries. e is a commodity-by-one column vector, which shows the total final demand purchases for each commodity.

Since supply and use table are reported at more disaggregated industry level, we aggregate the matrices and vectors $\{g, W, U\}$ along the industry dimension so that they are consistent with info and non-info classification. The resulting matrices and vectors are denoted as $\{\widetilde{g}, \widetilde{W}, \widetilde{U}\}$. Given the definition of those

[23] We use the sample where both the 1990 and the 2003 classification are available, to construct a crosswalk between 1990 and 2003 codes. The crosswalk between 2003 industry codes and NAICS codes can be found at the IPUMS USA webpage: `https://usa.ipums.org/usa/volii/indcross03.shtml`.

[24] The 15 sectors are approximately at two-digit NAICS level. The crosswalk between 15 BEA industries and two-digit NAICS code is available at `https://www.bea.gov/sites/default/files/2018-04/GDPbyInd_VA_1947-2017.xlsx`.

[25] Annual make and use table are available at `https://www.bea.gov/industry/input-output-accounts-data`. For the year after 1997, we extract the version of make and use table with 71 industries. For the year before 1997, we extract the make and use table with 65 industries. Both industry classifications are based on NAICS code.

[26] See Horowitz et al. (2006) for the concepts and methods of the US Input-Output accounts. The derivation of the industry-by-industry direct requirements table follows closely with BEA's method in Horowitz et al. (2006).

variables, we can find the following accounting identity

$$q = \widetilde{U}i + e$$
$$\widetilde{g} = \widetilde{W}i$$

where i is a unit vector containing only 1's.[27] Then we define $D = \widetilde{W}\widehat{q}^{-1}$ and $B = \widetilde{U}\widehat{\widetilde{g}}^{-1}$, where D is an industry-by-commodity matrix where each column reports the proportion of the total output of that commodity produced in each industry and B is a commodity-by-industry matrix where each column shows the amount of a commodity used by an industry per dollar of output of that industry.[28] Then we can express output vectors q and $\widetilde{g}$ as

$$q = B\widetilde{g} + e$$
$$\widetilde{g} = Dq.$$

Combining the two equations, we can find

$$\widetilde{g} = DB\widetilde{g} + De.$$

DB is the industry-by-industry (two-by-two) direct requirements matrix, where each column shows the amount of industry-level output used by an industry per dollar of output of that industry. We use DB as the measure of intermediate shares of two sector. In addition, with the information on $\widetilde{g}$ and DB, we can also perform the expenditure side decomposition on aggregate intermediate share.

Additional data used in the accounting exercise. For the accounting exercise, we use industry level nominal and real gross output and its decomposition from World KLEMS data.[29] Since World KLEMS is available at 65 NAICS industry level, we aggregate the nominal values across industries to estimate the info and non-info sector's labor output share and intermediate output share.

We construct sectoral output and price indices using the cyclical expansion approach in Herrendorf et al. (2013). We have the data of gross output $P_{st}Q_{st}$ and real gross output Q_{st} at 65 NAICS industry level. We follow the KLEMS data to set year 2009 as the base year and normalize $QI_{s2009} =$

[27] The introduction of unit vector is used for summation of matrices across columns.

[28] $\widehat{q}$ and $\widehat{\widetilde{g}}$ indicate the square matrices in which the elements of the vectors q and $\widetilde{g}$ appear on the main diagonal and zeros elsewhere.

[29] Data is available at `http://www.worldklems.net/data.htm`.

$P_{s2009} = 1$, where QI_{s2009} and P_{s2009} refer to the chain-weighted quantity index and chain-weighted price index for industry s in base year. By definition, we can calculate the chain-weighted quantity and price index using $\frac{QI_{st}}{QI_{s2009}} = \frac{Q_{st}}{Q_{s2009}}$ and $\frac{P_{st}}{P_{s2009}} = \frac{P_{st}Q_{st}}{P_{s2009}Q_{s2009}}\frac{Q_{s2009}}{Q_{st}}$. Then we approximate the change of chain-weighted quantity index for broad info and non-info sector as $\frac{QI_{Jt}}{QI_{Jt-1}} = \sqrt{\frac{\sum_{s\in J} P_{st-1}Q_{st}}{\sum_{s\in J} P_{st-1}Q_{st-1}}\frac{\sum_{s\in J} P_{st}Q_{st}}{\sum_{s\in J} P_{st}Q_{st-1}}}$. Finally, we normalize $QI_{J2009} = 1$ and $P_{J2009} = 1$, and it yields the real quantity $Q_{Jt} = QI_{Jt}Q_{J2009} = QI_{Jt}\{\sum_{s\in J} P_{s2009}Q_{s2009}\}$ and price $P_{Jt} = \frac{\sum_{s\in J} P_{st}Q_{st}}{QI_{Jt}\{\sum_{s\in J} P_{s2009}Q_{s2009}\}}$. To infer rental price of capital, we extract chain-type quantity index for net stock of fixed assets from BEA fixed assets account. We obtain the Full-Time Equivalent (FTE) labor force from BEA to normalize all quantity measures.

In the robustness check section, we construct sector-specific prices for intermediate inputs in two steps. Firstly, using BEA supply and use table, we can construct a direct requirements table of info and non-info sector from 65 KLEMS industries. The resulting direct requirements matrix can be shown as $R = D^{\star}B$.[30] R is a (65×2) matrix where the columns show the intermediate share of info and non-info sector from 65 NAICS industries. Since World KLEMS provides the price deflator for each of the 65 industries' output, we use the same cyclical expansion approach to construct the intermediate prices for each sector.

[30] $D^{\star} = W^{\star}\widehat{q}^{-1}$. We aggregate W along the industry dimension so that it is consistent with 65 NAICS industry level in World KLEMS, and we denote the resulting matrix as $W^{\star}$.

3.6.2 Supplementary tables and figures

Table 3.3: Mean and median information score by sector: 1980

Industry	Mean	Median
1. Agriculture	2.55	1.80
2. Mining	3.04	2.76
3. Utilities	3.39	3.46
4. Construction	2.86	2.70
5. Manufacturing	3.05	2.82
6. Wholesale Trade	3.37	3.46
7. Retail Trade	3.06	2.82
8. Transportation and Warehousing	2.93	2.70
9. Information	3.57	3.91
10. Finance, Insurance, Real Estate	3.62	3.70
11. Professional Business Service	3.62	3.45
12. Educational Services, Health Care and Social Assistance	3.48	3.63
13. Arts, Entertainment, Recreation, Accommodation and Food Services	2.67	2.82
14. Other Services, except Government	2.93	2.82
15. Government	3.72	3.70

Notes: This table reports industry level mean and median info score weighted by the corresponding occupation share, with weights calculated from IPUMS CPS ASEC. We use the weights associated with individual multiplied by hours worked. The median for the whole sample is 3.05. Numbers are rounded to two digits. We classify industries into info sector if median info score of the industry is larger than national median. The rest are classified as non-info sector.

Table 3.4: Annual growth rate of technologies with alternative σ_I and σ_N

Panel A: $\sigma_I = \sigma_N = 1.8$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.990	0.992	1.029	1.010	1.016
Non-info sector	1.005	1.006	1.024	1.012	0.977
1980-1995					
Info sector	0.979	0.989	1.047	0.991	1.019
Non-info sector	1.001	1.006	1.045	1.014	0.987
1995-2014					
Info sector	0.999	0.994	1.014	1.025	1.014
Non-info sector	1.008	1.005	1.007	1.011	0.970

Panel B: $\sigma_I = \sigma_N = 2.8$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.994	0.987	1.028	1.010	1.016
Non-info sector	1.007	1.004	1.023	1.012	0.977
1980-1995					
Info sector	0.985	0.982	1.039	0.996	1.019
Non-info sector	1.004	1.004	1.040	1.016	0.987
1995-2014					
Info sector	1.001	0.991	1.019	1.022	1.014
Non-info sector	1.010	1.003	1.010	1.010	0.970

Panel C: $\sigma_I = 2.2$, $\sigma_N = 2$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.992	0.989	1.028	1.010	1.016
Non-info sector	1.006	1.004	1.024	1.012	0.977
1980-1995					
Info sector	0.982	0.985	1.044	0.993	1.019
Non-info sector	1.003	1.005	1.043	1.014	0.987
1995-2014					
Info sector	1.000	0.992	1.016	1.024	1.014
Non-info sector	1.009	1.004	1.008	1.011	0.970

Notes: This table reports the inferred growth rate of sector-factor specific technical change for alternative values of σ_I and σ_N, where growth rate refers to $\frac{A_{t+1}}{A_t}$. A number smaller the one means the corresponding productivity is declining.

Table 3.5: Annual growth rate of technologies with alternative σ_C

Panel A: $\sigma_C = 0.4$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.994	0.991	1.024	1.006	1.016
Non-info sector	1.009	1.006	1.022	1.011	0.977
1980-1995					
Info sector	0.985	0.987	1.037	0.990	1.019
Non-info sector	1.007	1.008	1.039	1.013	0.987
1995-2014					
Info sector	1.002	0.993	1.014	1.019	1.014
Non-info sector	1.010	1.005	1.009	1.010	0.970

Panel B: $\sigma_C = 0.8$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.977	0.973	1.066	1.048	1.016
Non-info sector	0.989	0.986	1.033	1.022	0.977
1980-1995					
Info sector	0.962	0.964	1.088	1.039	1.019
Non-info sector	0.969	0.971	1.060	1.034	0.987
1995-2014					
Info sector	0.989	0.981	1.050	1.055	1.014
Non-info sector	1.004	0.999	1.013	1.014	0.970

Notes: This table reports the inferred growth rate of sector-factor specific technical change for alternative values of σ_C, where growth rate refers to $\frac{A_{t+1}}{A_t}$. A number smaller the one means the corresponding productivity is declining.

Table 3.6: Annual growth rate of technologies with alternative σ_k

Panel A: $\sigma_k = 0.4$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.995	0.991	1.030	1.012	1.012
Non-info sector	1.005	1.002	1.021	1.010	0.989
1980-1995					
Info sector	0.986	0.987	1.045	0.998	1.013
Non-info sector	1.001	1.003	1.040	1.013	0.998
1995-2014					
Info sector	1.002	0.993	1.019	1.024	1.011
Non-info sector	1.007	1.002	1.007	1.008	0.982

Panel B: $\sigma_k = 0.8$

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.987	0.983	1.022	1.004	1.028
Non-info sector	1.013	1.010	1.030	1.019	0.943
1980-1995					
Info sector	0.974	0.976	1.032	0.986	1.037
Non-info sector	1.008	1.010	1.047	1.020	0.953
1995-2014					
Info sector	0.997	0.989	1.014	1.019	1.021
Non-info sector	1.017	1.011	1.016	1.017	0.936

Notes: This table reports the inferred growth rate of sector-factor specific technical change for alternative values of σ_k, where growth rate refers to $\frac{A_{t+1}}{A_t}$. A number smaller the one means the corresponding productivity is declining.

Table 3.7: Annual growth rate of sector-factor augmenting technologies with sector-factor specific intermediate prices

	Info L A_{Ji}	Info M λ_{JI}	Non-info L A_{Jn}	Non-info M λ_{JN}	Capital K A_J
1980-2014					
Info sector	0.993	0.987	1.028	1.012	1.016
Non-info sector	1.007	1.001	1.023	1.011	0.977
1980-1995					
Info sector	0.983	0.984	1.042	0.994	1.019
Non-info sector	1.003	0.998	1.041	1.012	0.987
1995-2014					
Info sector	1.001	0.990	1.017	1.026	1.014
Non-info sector	1.010	1.003	1.009	1.011	0.970

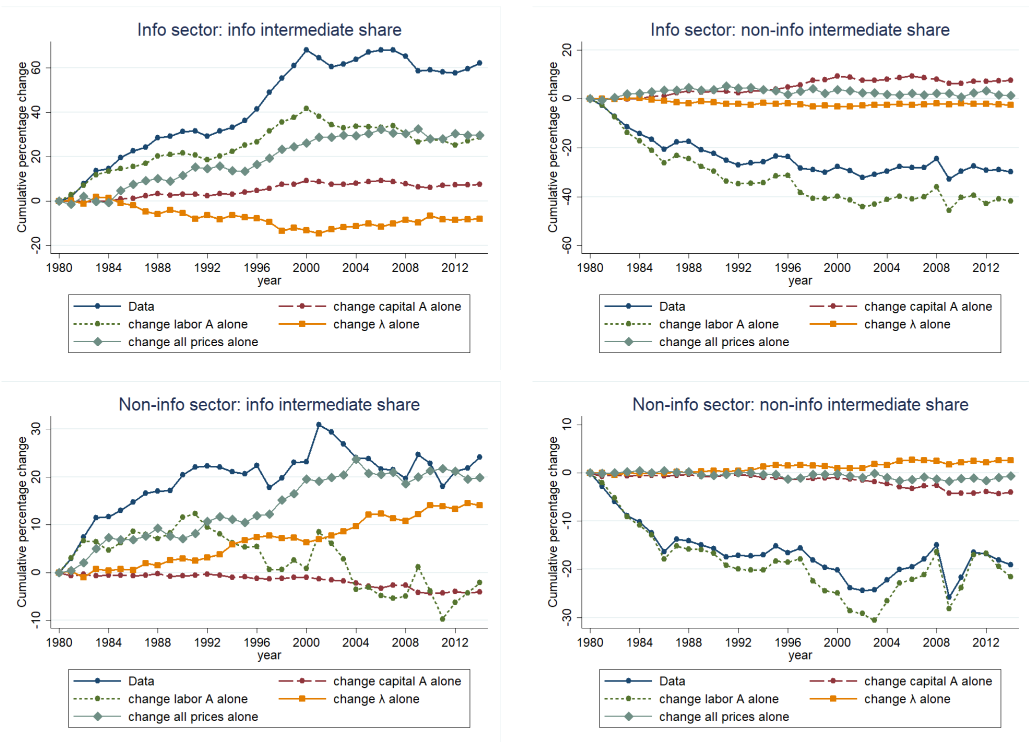
Notes: This table reports the inferred growth rate of sector-factor specific technical change for the setting with sector-factor specific prices for intermediate inputs, where growth rate refers to $\frac{A_{t+1}}{A_t}$. A number smaller the one means the corresponding productivity is declining.

Table 3.8: Distribution of occupations

	Abstract	Routine	Manual
Info	30.6%	16.5%	2.3%
Non-info	4.2%	36.2%	9.7%

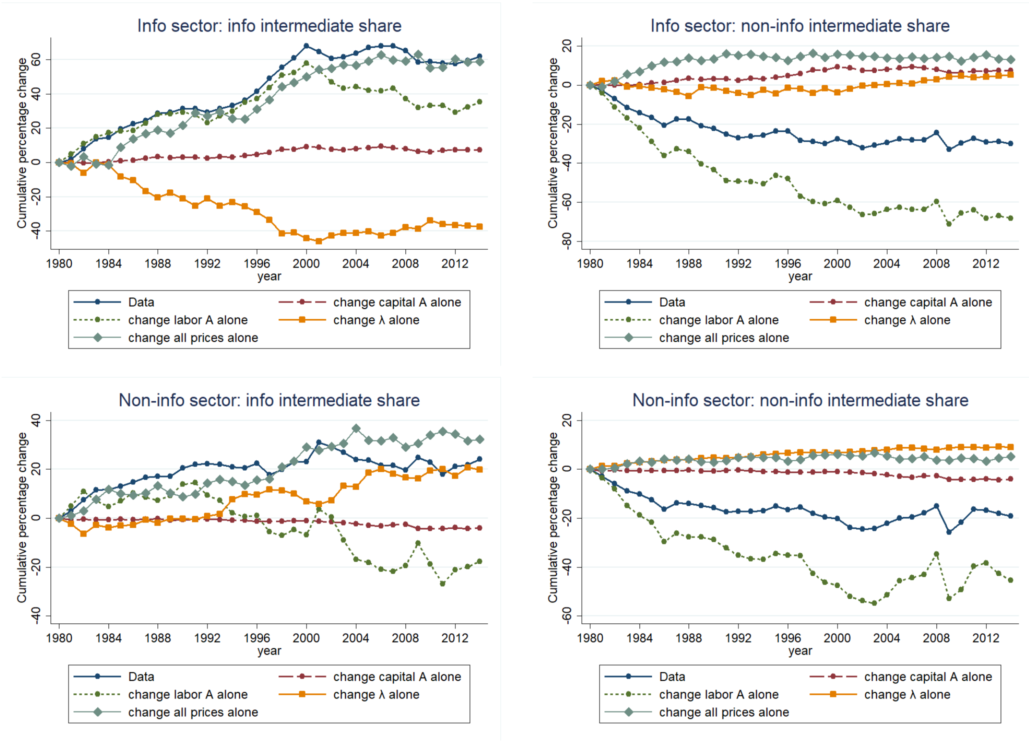
Notes: The table shows the distribution of occupations in different classification cells. Info and non-info refer to the classification in this chapter, where abstract-routine-manual refers to the classification in Barany and Siegel (2019).

Figure 3.13: Technical changes and intermediate shares: $\sigma_I = \sigma_N = 1.8$



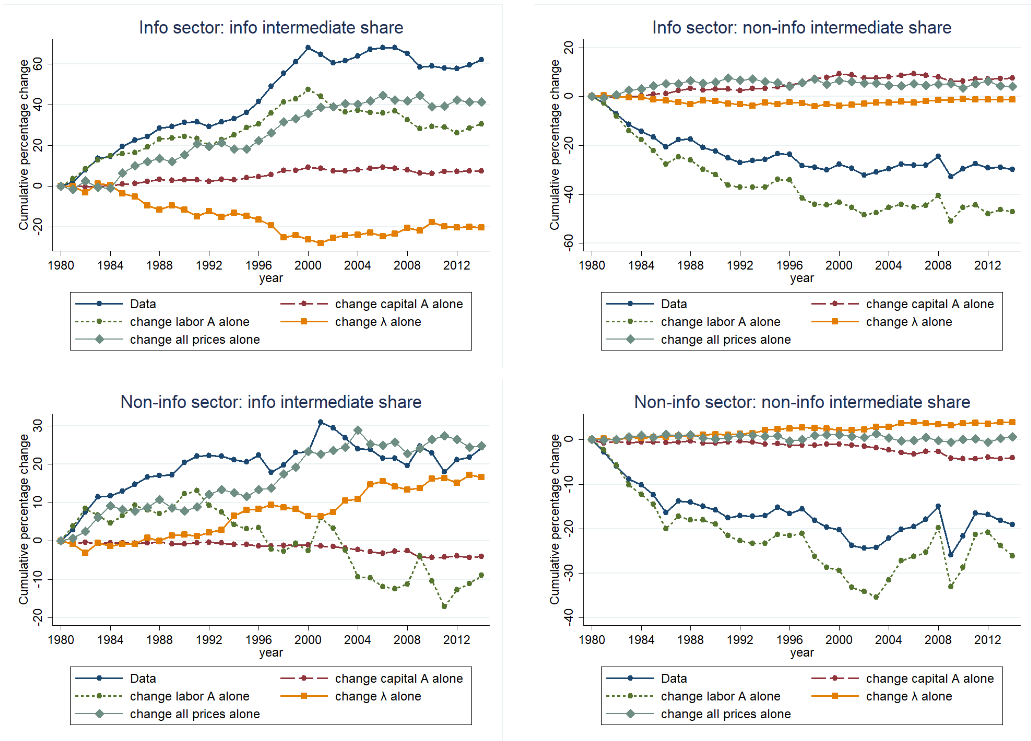
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_I and σ_N, i.e., $\sigma_I = \sigma_N = 1.8$.

Figure 3.14: Technical changes and intermediate shares: $\sigma_I = \sigma_N = 2.8$



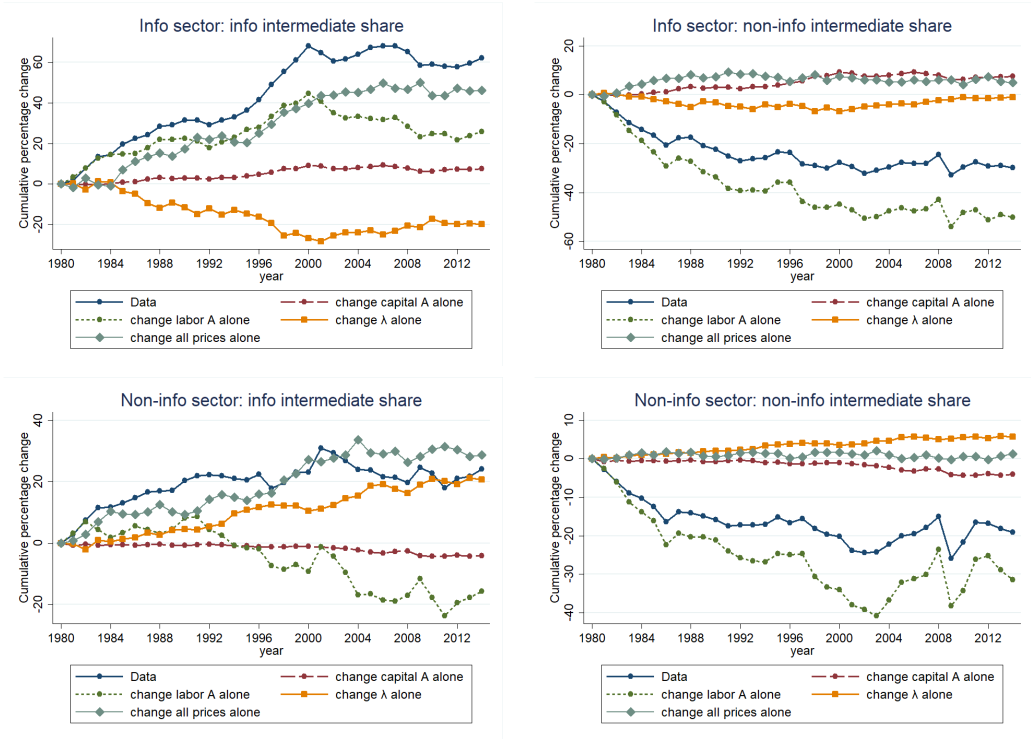
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_I and σ_N, i.e., $\sigma_I = \sigma_N = 2.8$.

Figure 3.15: Technical changes and intermediate shares: $\sigma_I = 2.2$ and $\sigma_N = 2$



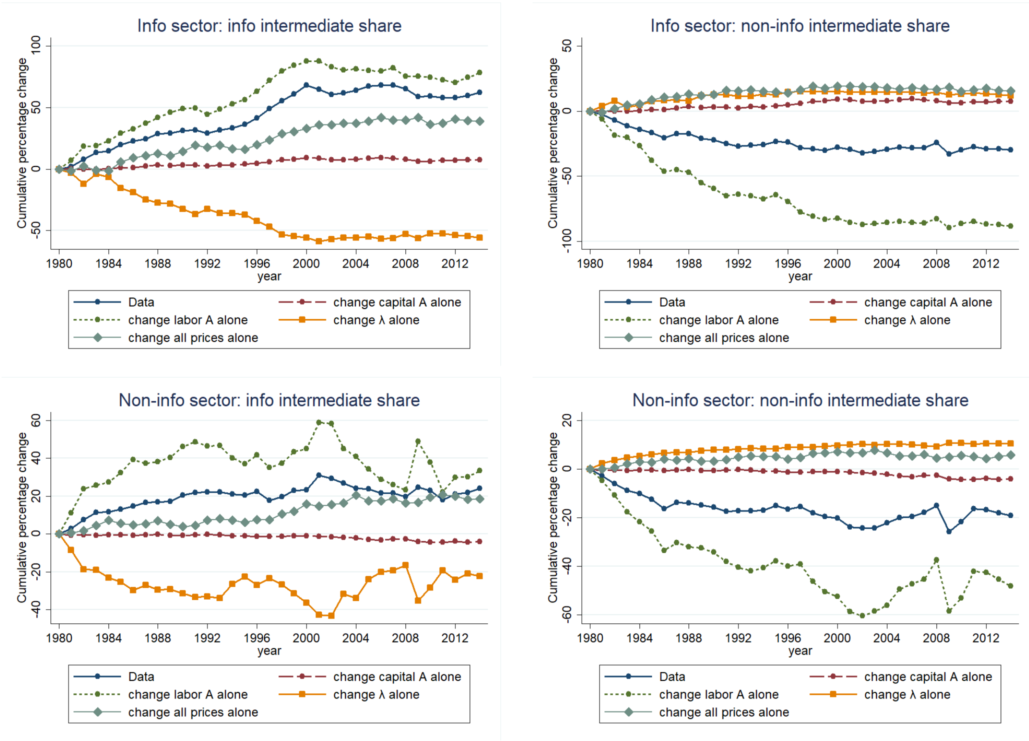
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_I and σ_N, i.e., $\sigma_I = 2.2$ and $\sigma_N = 2$.

Figure 3.16: Technical changes and intermediate shares: $\sigma_C = 0.4$



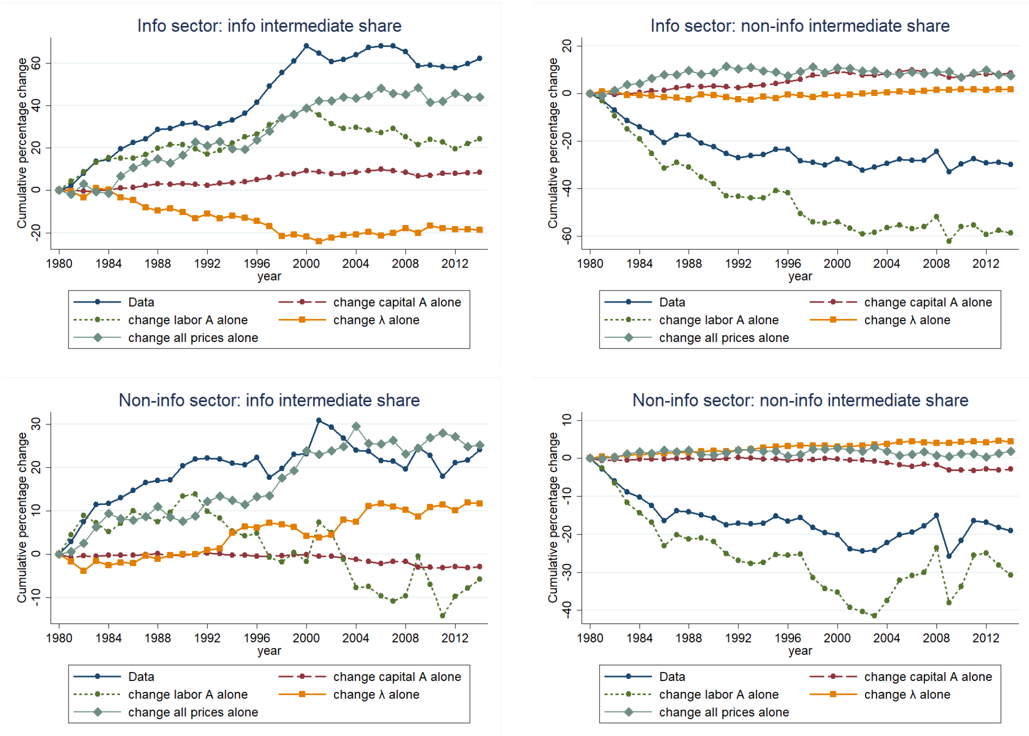
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_C, i.e., $\sigma_C = 0.4$.

Figure 3.17: Technical changes and intermediate shares: $\sigma_C = 0.8$



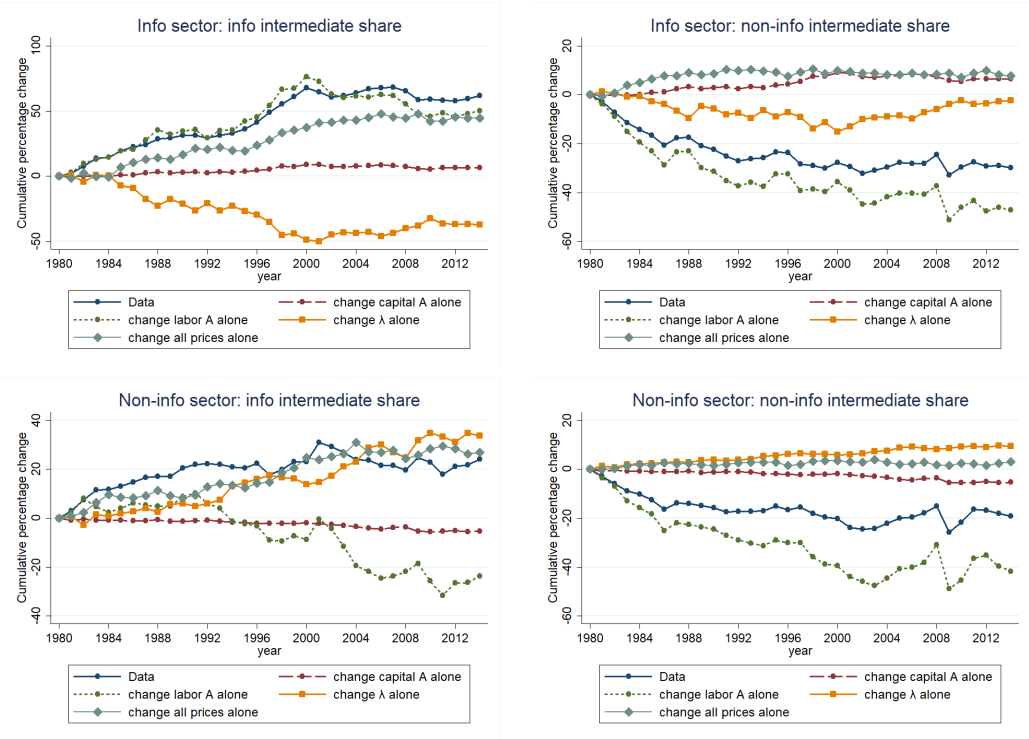
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_C, i.e., $\sigma_C = 0.8$.

Figure 3.18: Technical changes and intermediate shares: $\sigma_k = 0.4$



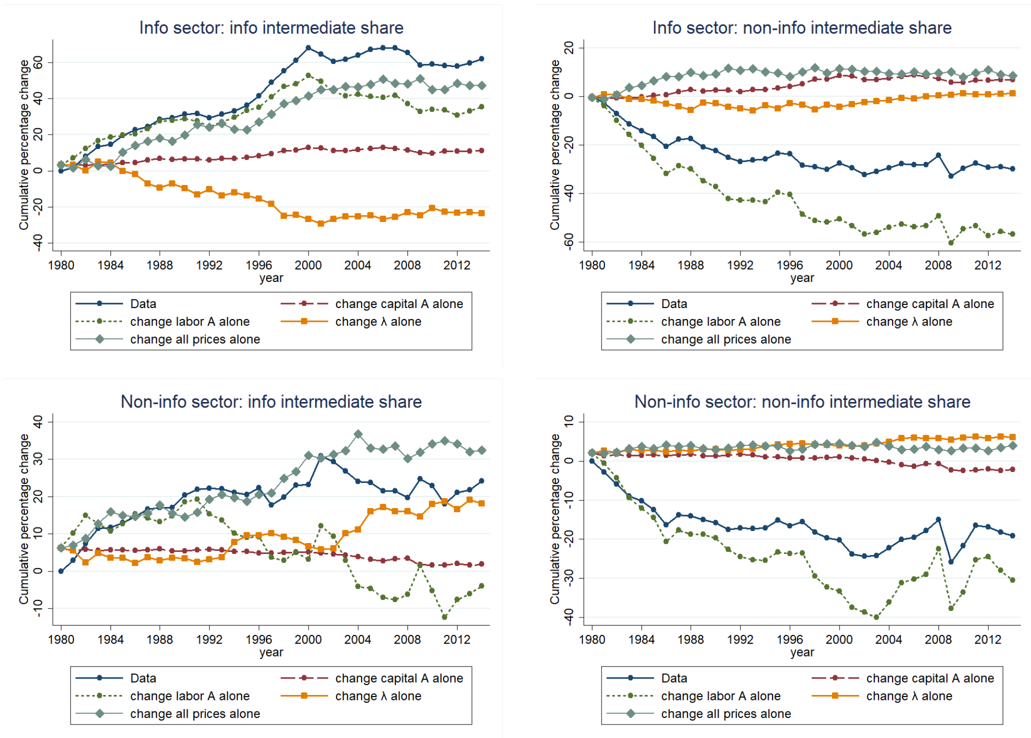
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_k, i.e., $\sigma_k = 0.4$.

Figure 3.19: Technical changes and intermediate shares: $\sigma_k = 0.8$



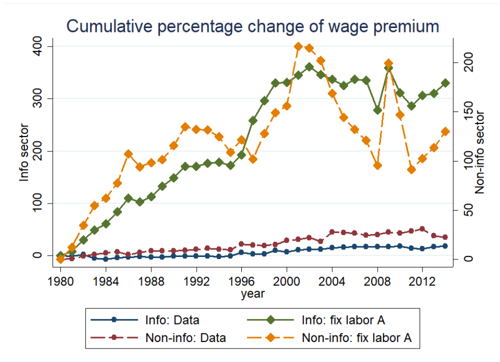
Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We consider alternative values of σ_k, i.e., $\sigma_k = 0.8$.

Figure 3.20: Technical changes and intermediate shares: sector-factor specific intermediate prices

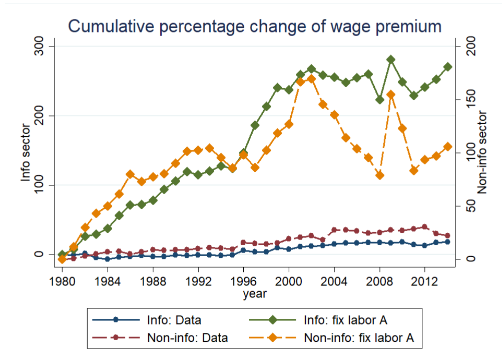


Notes: This figure shows the cumulative percentage change of intermediate shares since 1980. Both the observed data and four counterfactual cases are reported. We allow sector-factor specific prices of intermediate inputs.

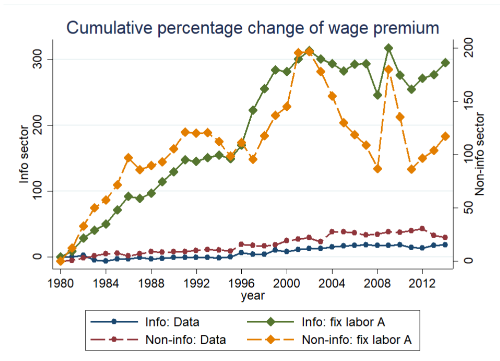
Figure 3.21: Wage premium and labor-augmenting technical change with alternative elasticities



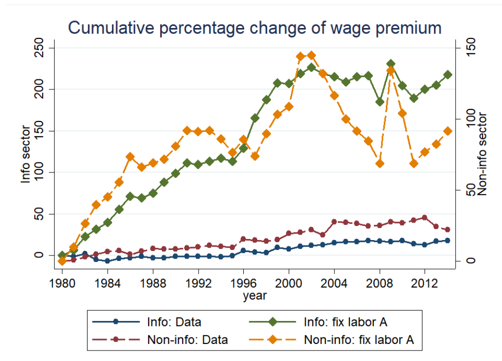
(a) $\sigma_I = \sigma_N = 1.8$



(b) $\sigma_I = \sigma_N = 2.8$



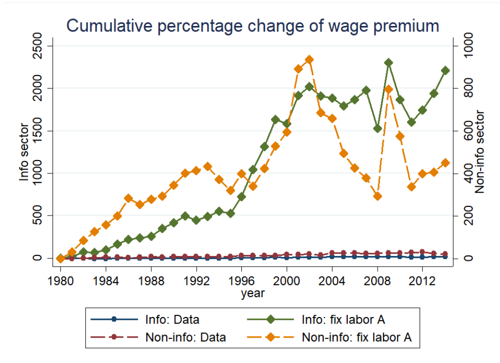
(c) $\sigma_I = 2.2$ and $\sigma_N = 2$



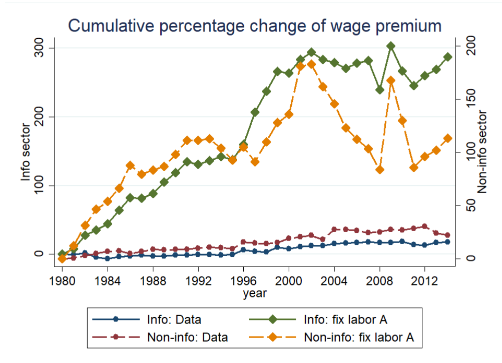
(d) $\sigma_C = 0.4$

Notes: This figure shows cumulative percentage change in observed and counterfactual wage premium since 1980 for alternative values of elasticities. Counterfactual wage premium is calculated by fixing labor-augmenting technical change at its 1980 level.

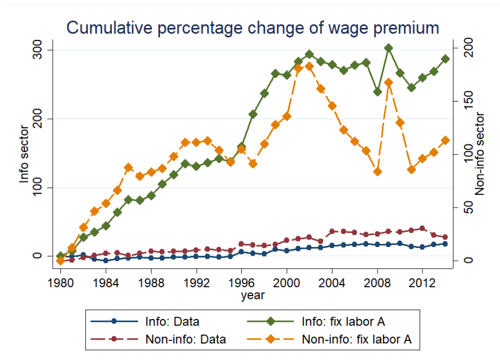
Figure 3.22: Wage premium and labor-augmenting technical change with alternative elasticities (continued)



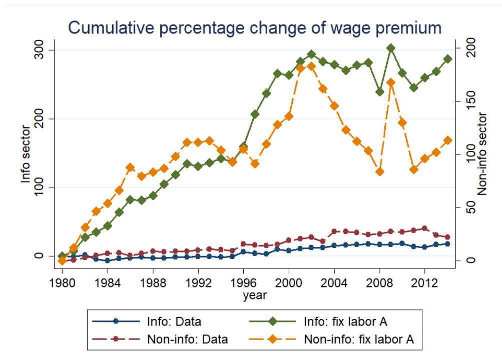
(a) $\sigma_C = 0.8$



(b) $\sigma_k = 0.4$



(c) $\sigma_K = 0.8$



(d) sector-factor specific int prices

Notes: This figure shows cumulative percentage change in observed and counterfactual wage premium since 1980. Panel (a) to (c) consider alternative values of elasticities. Panel (d) considers sector-factor specific intermediate prices. Counterfactual wage premium is calculated by fixing labor-augmenting technical change at its 1980 level.

References

Abramovsky, L. and Griffith, R. (2006). Outsourcing and offshoring of business services: How important is ict? *Journal of the European Economic Association*, 4(2-3):594–601.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Antras, P. (2004). Is the us aggregate production function cobb-douglas? new estimates of the elasticity of substitution. *The BE Journal of Macroeconomics*, 4(1).

Autor, D., Dorn, D., Katz, L. F., Patterson, C., Van Reenen, J., et al. (2017). *The fall of the labor share and the rise of superstar firms*. National Bureau of Economic Research.

Autor, D. H. (2017). How long has this been going on? a discussion of 'recent flattening in the higher/education wage premium: Polarization, skill downgrading, or both?'by robert g. valletta. In *NBER Conference on Research and Income in Wealth "Measuring and Accounting for Innovation in the 21st Century," Washington, DC (March 10–11, 2017)*.

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.

Barany, Z. and Siegel, C. (2019). Engines of sectoral labor productivity growth. In *2019 Meeting Papers*, number 872. Society for Economic Dynamics.

Barnes, S. N., Price, S., and Barriel, M. (2008). The elasticity of substitution: evidence from a uk firm-level data set.

Chan, M. (2017). How substitutable are labor and intermediates? *Unpublished working paper. University of Minnesota.*

Gallipoli, G. and Makridis, C. A. (2018). Structural transformation and the rise of information technology. *Journal of Monetary Economics*, 97:91–110.

Herrendorf, B., Rogerson, R., and Valentinyi, A. (2013). Two perspectives on preferences and structural transformation. *American Economic Review*, 103(7):2752–89.

Horowitz, K. J., Planting, M. A., et al. (2006). Concepts and methods of the us input-output accounts. Technical report, Bureau of Economic Analysis.

Karabarbounis, L. and Neiman, B. (2014). The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1):61–103.

Oberfield, E. and Raval, D. (2014). Micro data and macro technology. Technical report, National Bureau of Economic Research.

Valletta, R. G. (2016). Recent flattening in the higher education wage premium: Polarization, skill downgrading, or both? Technical report, National Bureau of Economic Research.