

On the Cost of Capital, Profits and the Diffusion of Ideas

Has van Vlokhoven



On the Cost of Capital, Profits and the Diffusion of Ideas

Has van Vlokhoven

Academic dissertation for the Degree of Doctor of Philosophy in Economics at Stockholm University to be publicly defended on Wednesday 27 May 2020 at 14.00 in Nordenskiöldsalen, Geovetenskapens Hus, Svante Arrhenius väg 12.

Abstract

Estimating the Cost of Capital and the Profit Share Compensation of the factor of production capital is not directly observed since most firms own part of their capital stock. I develop a new method to estimate capital compensation. I show how firms' input choices reveal the user cost of capital when firms minimize costs and produce according to a homogeneous production function. Subtracting estimated capital compensation together with all other observed costs from sales gives economic profits. Estimating the model using Compustat data, I find that the cost of capital has been declining, and that the profit share has been increasing over the past fifty years from around 4% to around 8% of sales. The increase in the profit share coincides with the observed fall in the labor share, while I estimate the capital share to be falling as well. Therefore, the fall in the labor share is not due to an increased capital intensity, but due to an increase in profits.

Profits and the Marginal Product of Capital Around the World The extent to which marginal products of capital are equalized across countries is informative of how well international capital markets function. I estimate the marginal product of capital across a wide range of countries while allowing for imperfect competition. I find that richer countries have a higher marginal product of capital than poorer countries, but that this is entirely driven by differences in depreciation rates. Thus, in terms of output net of depreciation there is no gain by reallocating capital from poor to rich countries or vice versa. Furthermore, I find that profits have increased globally, but that the rise in profits is more pronounced in rich countries.

The Life Cycle of Profits Old firms make more profits than young firms, and nowadays profits are more back-loaded than thirty years ago. I study to what extent this changing life-cycle pattern of profits explains the observed rise in profits and fall in firm entry. I build a quantitative life cycle model with oligopolistic competition and an occupational choice between being an entrepreneur and being a worker. All else equal, the more back-loaded profits are, the lower the value of the firm due to discounting, and therefore the fewer agents choose to be an entrepreneur. In equilibrium, aggregate profits rise to a level such that agents are indifferent between occupations. I find that the observed change in the life-cycle pattern of profits explains about two-thirds of the rise in profits, and more than fully explains the fall in firm entry.

Diffusion of Ideas in Networks and Endogenous Search I study the diffusion of technology when the decision to search for productivity-enhancing technologies depends on the network of interactions between agents. Agents have the option to engage in costly learning from their first-degree connections. The more productive an agent's connections, the more willing it is to learn. Hence, the network affects the reservation productivity at which agents choose to learn and affects therefore aggregate productivity. I find that the denser the network, the higher learning effort and therefore the higher total factor productivity and the lower inequality. However, the effect of the network on the share of agents that learn in equilibrium is ambiguous. Furthermore, I find that nodes that are central in terms of their closeness to other nodes tend to exert more learning effort and have a higher productivity.

Keywords: *Cost of capital, Profit share, Capital share, Labor share, Markups, Marginal products, International capital markets, Firm life cycle, Firm entry, Technology diffusion, Search, Networks, Productivity distribution.*

Stockholm 2020
<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-179924>

ISBN 978-91-7911-122-9
ISBN 978-91-7911-123-6
ISSN 0346-6892

Department of Economics

Stockholm University, 106 91 Stockholm



ON THE COST OF CAPITAL, PROFITS
AND THE DIFFUSION OF IDEAS

Has van Vlokhoven

On the Cost of Capital, Profits and the Diffusion of Ideas

Has van Vlokhoven

©Has van Vlokhoven, Stockholm University 2020

ISBN print 978-91-7911-122-9

ISBN PDF 978-91-7911-123-6

ISSN 0346-6892

Cover picture: Quinten Matsijs, "The Moneylender and His Wife", 1514

Distributor: Institute for International Economic Studies

Printed in Sweden by Universitetsservice US-AB, Stockholm 2020

Doctoral dissertation
Department of Economics
Stockholm University

Abstracts

Estimating the Cost of Capital and the Profit Share Compensation of the factor of production capital is not directly observed since most firms own part of their capital stock. I develop a new method to estimate capital compensation. I show how firms' input choices reveal the user cost of capital when firms minimize costs and produce according to a homogeneous production function. Subtracting estimated capital compensation together with all other observed costs from sales gives economic profits. Estimating the model using Compustat data, I find that the cost of capital has been declining, and that the profit share has been increasing over the past fifty years from around 4% of sales to around 8% of sales. The increase in the profit share coincides with the observed fall in the labor share, while I estimate the capital share to be falling as well. Therefore, the fall in the labor share is not due to an increased capital intensity, but due to an increase in profits. Furthermore, I find that the increase in profits is due to reallocation between firms, but not due to reallocation between industries. Finally, I find an upward trend in the returns to scale, which combined with the rise in profits implies that markups have been increasing.

Profits and the Marginal Product of Capital Around the World The extent to which marginal products of capital net of depreciation differ across countries is informative of there being frictions in international capital markets. I estimate the marginal product of capital across a wide range of countries while allowing for imperfect competition and non-constant returns to scale technology. I find that richer countries have a higher marginal product of capital than poorer countries, but that this is entirely driven by differences in depreciation rates. Thus, in terms of output net of depreciation, there is no efficiency gain from reallocating capital from poor to rich countries or vice versa. Furthermore, I find that profits have increased globally, but that the rise in profits is more pronounced in rich countries.

The Life Cycle of Profits Old firms make more profits than young firms and, nowadays, profits are more back-loaded over the firm's life cycle than thirty years ago. I study to what extent this changing life-cycle pattern of profits can explain the observed rise in profits and fall in firm entry. I build a quantitative life cycle model with oligopolistic competition and an occupational choice between being an entrepreneur or a worker. All else equal, the more back-loaded profits are, the lower is the value of the firm due to discounting and therefore, fewer agents choose to be an entrepreneur. This fall in entry decreases competition and, in turn, leads to an increase in profits until agents are again indifferent between occupations. I find that the observed change in the life-cycle pattern of profits can explain about two-thirds of the rise in profits, and can explain more than fully the fall in firm entry.

Diffusion of Ideas in Networks and Endogenous Search I study the diffusion of technology when the decision to search for productivity-enhancing technologies depends on the network of interactions between agents. Agents have the option to engage in costly learning from their first-degree connections. The more productive an agent's connections, the more willing it is to learn. Hence, the network affects the reservation productivity at which agents choose to learn and therefore affects aggregate productivity. I find that the denser the network, the higher learning effort and therefore the higher total factor productivity and the lower inequality. However, the effect of the network on the share of agents that learn in *equilibrium* is ambiguous. Furthermore, I find that nodes that are central in terms of their closeness to other nodes tend to exert more learning effort and have a higher productivity.

To my parents

Acknowledgments

I am a product of my environment. During my life I have been influenced and inspired by many people, often without me being aware of it. Therefore, it is impossible to thank everyone here, but I will attempt to thank those that were important during the last years.

First of all, I am deeply indebted to my supervisor Timo Boppart. Without him I would not have been able to obtain my PhD. Throughout the entire PhD, his door has always been open and he always seemed excited to discuss my research. I admire Timo for his patience and his passion for economics.

I want to thank my second supervisor, Per Krusell, for his academic rigor which prevented me from being sloppy and forced me to be more precise in my thinking. Mitch Downey has been extremely helpful with navigating the job market. And I thank Pete Klenow for being my host at Stanford University.

Furthermore, I thank Miguel Ampudia for letting me join him on a research project while I was working at the European Central Bank. Starting doing research before the PhD gave me a head start.

I had the fortune to share the PhD experience with many other students. The moments with you made it much more enjoyable. Thank you Anna, Fredrik, Benni, Domenico, Elle, Serena, Alla, Xueping, Sreyashi, Karin, Alberto, Kasper, Louise, Markus, John, Mattias, Divya, Magnus, Jonna, Richard, Josef, and outside Stockholm, Max and Vicke.

The administrative staff has always been helpful, and Christina Lönnblad and Ulrika Gålnander helped me with putting this dissertation together.

Finally, I want to thank my family for always supporting me.

Has van Vlokhoven
Stockholm, Sweden
April 2020

Contents

Introduction	1
Chapter 1 Estimating the Cost of Capital and the Profit Share	5
Chapter 2 Profits and the Marginal Product of Capital Around the World ..	101
Chapter 3 The Life Cycle of Profits	119
Chapter 4 Diffusion of Ideas in Networks and Endogenous Search	141
Sammanfattning (Swedish Summary)	203

Introduction

The average Swede living today is about twice as rich as her parents and four times as rich as her grandparents. Going back further in time, a person living in Sweden today is about twenty times as rich as the average Swede that lived 150 years ago. This historically unprecedented income growth has had dramatic effects on the ways in which we live our lives. It has led to much more freedom and an increase in possibilities. Nowadays, people take it for granted to take an airplane that can take them within a day to the other side of the world, whereas a century ago many people would never leave their own country—unless, perhaps, to fight a war. In addition, life expectancy has almost doubled in the last 150 years from being around 45 years in 1870 to being 83 years in Sweden today.

This rise in welfare is due to productivity growth. Productivity grows for two reasons: i) innovation and ii) innovations diffusing over the entire economy, also making other producers more productive. In recent time, productivity growth has slowed down. To understand this slow down, it is imperative to better our understanding of innovation and technology diffusion. This is also key for understanding the patterns of economic development we observe and will be helpful for designing policies that can make low-income countries grow faster.

Let me first focus on innovation. A substantial fraction of innovations is done by firms. And when an innovation is not done by a firm, but let us say by a university instead, a firm is still often needed to transform the innovation into a product or service that is useful to a consumer. Thus, to understand innovation and productivity growth, we need to better understand the reasons for why firms innovate. The main incentive for engaging in innovation is to generate profits. This means that it is essential for economists to understand to what extent firms are making profits and how this has changed over time. The first three chapters of this thesis study this in detail. In the first chapter, I develop a new method to estimate profits, and use this method to estimate how profits have changed during the last fifty years in the United States. I find that profits, as a share of output, have roughly doubled during this time period. In the second chapter, I find that other countries around the world have also experienced an increase in profits. Chapter three provides an explanation for why profits have been increasing. I find that profits nowadays appear at a later stage of the firm's life than they used to. Due to discounting, this lowers the value of the firm and makes it less attractive to be an

entrepreneur. Thus, fewer people decide to become an entrepreneur and therefore competition declines and aggregate profits increase. This mechanism can explain about two-thirds of the rise in profits and is consistent with the observed decline in firm entry.

As mentioned above, for innovations to have a substantial impact on aggregate productivity they need to diffuse across the entire economy. Take as an example the idea of the assembly line that was introduced by Henry Ford. If this innovation had only stayed within Ford Motor Company, it would have had a limited effect on aggregate productivity. Instead, the idea of the assembly line has diffused around the world and has revolutionized manufacturing and beyond. The fourth and final chapter of this thesis studies how new innovations diffuse from one firm to other firms to which it is connected. In particular, it studies which network properties are beneficial for diffusion when the effort put into learning is endogenous and depends on the network. I find that the denser the network (i.e., the more connections there are between firms), the faster diffusion and therefore the higher aggregate productivity.

I will now summarize each chapter in more detail.

The first chapter, **Estimating the Cost of Capital and the Profit Share**, estimates how much profits firms are making. Profits are equal to output minus all costs. However, capital costs are not directly observed as they are not reported on the income statement of the firm. Therefore, we need to estimate capital costs first in order to estimate economic profits. In this chapter, I develop a new method to estimate capital costs. This method uses that firms' input choices reveal the cost of capital when firms minimize costs and produce according to a homogeneous production function. Using this method, I find that capital costs as a share of output have been weakly declining during the last 50 years in the United States. Subtracting these estimated capital costs, together with all other observed costs, from output, I find that profits have roughly doubled from being around 4% of output in the 1960s to being around 8% of output today.

Knowing how much profits firms are making is not only important for understanding the productivity dynamics we observe but is also crucial for understanding inequality. In the last few decades, the labor share of income in the US has been declining. This means that a lower share of income goes to workers and a higher share to the owners of the firms and of the capital stock. This fall in the labor share could either be due to capital becoming more important in production (e.g., due to automation) or due to an increase in firms' market power. That the capital share has been declining while the profit share has been increasing means that the fall in the labor share is due to a rise in firms' market power and not due

to capital becoming more important in production.

Considering the distribution of profitability across firms, I find that the entire distribution has shifted to the right. Thus, it are not only the most profitable firms that have become more profitable, but also the median firm has become more profitable over time. Nonetheless, the rise in profits is due to economic activity reallocating from firms with a low profit share to firms with a high profit share. Finally, I find that large firms have a higher profit share than small firms and that this relationship has become stronger over time.

The second chapter, **Profits and the Marginal Product of Capital Around the World**, uses the method developed in the first chapter, to study how the profit share has evolved around the world. I find that the profit share shows an inverted U-shape in Europe between 1990 and 2015, with an overall increase of around 2 percentage points. Profits have also been increasing in Asia, Latin America and North America. This does not mean that profits in all countries have been increasing. For instance, in Canada the profit share has not increased. The global profit share has been increasing by around 2 percentage points from 1990 to 2015, which is somewhat less than the increase in the United States. Overall, richer countries have experienced a somewhat faster increase in profitability than poor countries.

Furthermore, this chapter studies the extent to which marginal products of capital are equalized across countries. This is important for understanding the functioning of international capital markets. When the marginal product of capital net of depreciation differs across countries, international capital markets do not function well, and global output could be increased by reallocating capital from countries with a low marginal product of capital to countries with a high marginal product of capital. I estimate the marginal product of capital across countries while allowing for imperfect competition and the returns to scale to be different from one. I find that richer countries have a higher marginal product of capital than poorer countries, but that this is entirely driven by differences in depreciation rates. Thus, international capital markets seem to be working well, and there is no gain, in terms of output net of depreciation, from reallocating capital from poor to rich countries or vice versa.

In the third chapter, **The Life Cycle of Profits**, I document that over time profits have become more back-loaded over the firm's life cycle. A firm younger than ten years today makes about as much profits on an annual basis as a young firm was doing thirty years ago. However, an old firm today makes much more profits than it used to make thirty years ago. There are two reasons for this changing life-cycle pattern of profits. Young firms today are only slightly larger in terms of their sales

than they used to be, while older firms have become much larger than older firms thirty years ago used to be. Second, young firms have started to make less profits relative to their size while old firms nowadays have a profit share that is about the same as it used to be.

I next build a quantitative model to understand to what extent this changing life-cycle pattern of profits can explain the rise in profits. An entrepreneur will start a business when the value of having that business (i.e., the discounted sum of profits) exceeds the entry costs. All else equal, as profits nowadays appear at a later stage than they used to, the value of the firm is lower due to discounting. This makes it less attractive to start a business and will therefore lead to less firm entry. This lowers the competition among firms and, in turn, leads to an increase in profits. I find that the observed change in the life-cycle pattern of profits can explain about two-thirds of the rise in profits found in chapter 1, and more than fully explains the fall in firm entry that is observed.

Finally, the fourth chapter, **Diffusion of Ideas in Networks and Endogenous Search**, studies the diffusion of technology. New ideas tend to spread gradually and agents that are directly connected to early adopters are more likely to adopt themselves. This means that how the network of interactions between agents looks like affects the speed of diffusion. Furthermore, also search effort depends on the network and on the productivity distribution. When one is connected to high-productive agents, one is willing to put more effort into learning and adopting the technologies these high-productive agents are using than when one is only connected to low-productive agents. This chapter studies theoretically which network properties are beneficial for diffusion when the decision to search for productivity-enhancing technologies depends on the network of interactions between agents.

Agents have the option to engage in costly learning from their first-degree connections. The more productive an agent's connections, the more willing it is to learn. Hence, the network affects the reservation productivity at which agents choose to learn and therefore affects aggregate productivity. I find that the denser is the network (i.e., the more connections there are between firms), the higher is the learning effort and therefore the higher is total factor productivity and the lower is inequality. However, the effect of the network on the share of agents that learn in equilibrium is ambiguous. Finally, I find that nodes that are central in terms of their closeness to other nodes tend to exert more learning effort and have a higher productivity.

Chapter 1

Estimating the Cost of Capital and the Profit Share*

Compensation of the factor of production capital is not directly observed since most firms own, rather than rent, their capital stock. This means that also economic profits are not directly observed. However, being able to distinguish between capital compensation and economic profits is crucial for understanding several recent macroeconomic trends. It is essential for understanding whether the fall in the aggregate labor share of income (Elsby et al., 2013; Karabarbounis and Neiman, 2014) is due to production becoming more capital intense or due to a rise in market power, and it is informative on whether antitrust policy should be enhanced. This distinction is also key for understanding the productivity slowdown (Gordon, 2016) and the fall in firm entry (Decker et al., 2014) as the incentive for entrepreneurs to create new ideas, products and firms is to generate economic profits. Finally, telling capital compensation apart from economic profits is essential for calibrating models. In this chapter, I develop a new method to distinguish between capital compensation and economic profits, and quantify the evolution of capital compensation and economic profits in the US.

Through the lens of a simple but general model, capital compensation is revealed by cross-sectional variation in firms' input choices. To see this, suppose for now that production is constant returns to scale and that firms minimize costs. In the presence of a markup, this makes that nominal output, PY , is equal to the markup, μ , times total costs, which is the sum of expenditure on inputs other than capital, $P^X X$, and capital compensation, $R \cdot P^K K$,

$$PY = \mu (P^X X + R \cdot P^K K) . \quad (1)$$

*I thank Timo Boppart, Mitch Downey, Andreas Ek, Émilien Gouin-Bonenfant, Basile Grassi, John Hassler, Pete Klenow, Per Krusell, Vike Norén, Christina Patterson, Yimei Zou, members of the Stockholm macro discussion group, and seminar and conference participants at Stockholm University, Tilburg University, University of Copenhagen, the Swedish Ministry of Finance, Banca d'Italia Conference on Recent Trends in Firm Organization and Firm Dynamics, and the 8th National PhD Workshop in Finance for helpful comments.

Here, μ and the (user) cost of capital, R , are unobserved while the other variables are typically observed. The effect of varying the nominal capital stock, $P^K K$, on nominal output, while holding other inputs constant, equals the markup times the cost of capital, μR . By cost minimization, this also equals the valued marginal product of capital.¹ Furthermore, the effect on nominal output of varying expenditure on other inputs than capital, while holding capital constant, equals the markup. In Section II, I bring the above equation to a regression framework. Thus, regressing nominal output on input expenditure other than capital and on the capital stock gives as coefficients the markup and the markup times the cost of capital, respectively. Dividing these two estimates with each other gives the cost of capital. I run a modified version of this regression year-by-year which yields a time-varying estimate of the cost of capital. Multiplying this estimate of R with the capital stock gives total capital compensation, which subtracted together with all other observed costs from nominal output gives economic profits. Importantly, in Section II I relax the assumption of constant returns to scale and instead assume that the production function is homogeneous of a constant degree, possibly different from one. This yields a similar relationship between nominal output and inputs as above, but with the markup replaced by the price-average cost ratio.

I estimate the cost of capital and the profit share using Compustat data for the United States from the 1960s until today. The capital stock, and therefore the capital share, includes physical capital plus externally purchased intangibles, but does not include internally developed intangibles. Costs for internally developing intangibles, such as *R&D* and marketing, are part of operating expenses and are therefore part of the costs subtracted from sales to obtain economic profits (together with the other operating expenses and estimated capital compensation).² I find that the user cost of capital has been declining from around 25% in the 1960s and 1970s to around 20% today. The capital stock relative to sales has increased somewhat over time but not enough to compensate for the fall in the cost of capital and therefore, capital compensation as a share of sales has been falling from around 8% to around 7%. Economic profits as a share of sales have doubled from being around 4% in the 1960s and 1970s to being around 8% today.

The rise in profits coincides with a fall in the labor share (Elsby et al., 2013; Karabarbounis and Neiman, 2014).³ Several explanations for the fall in the labor

¹To be precise, the markup times the cost of capital equals the valued marginal product of capital divided by the price of capital P^K . See Section II for details.

²See Koh et al. (2016) for a discussion of how the treatment of intellectual property products affects the measurement of the labor share.

³See Rognlie (2015) for a discussion of the role of housing in measuring the labor share. Correcting for the self-employed and housing, Cetty et al. (2019) and Gutiérrez and Piton (2019) find that the

share have been put forward. These explanations fall into two categories. The first category attributes the fall in the labor share to changes in technology that have increased the importance of capital in production (such as (dis)embodied capital-biased technical change).⁴ The second category attributes the fall in the labor share to increased market power of firms in the product or labor market.⁵ If the fall in the labor share were due to production becoming more capital intense, the fall in the labor share would coincide with a rise in the capital share, whereas if market power has risen the fall in the labor share would be accompanied by a rising profit share. As I find that the profit share has been increasing while the capital share has been declining, the fall in the labor share is not due to an increased capital intensity, but due to increased market power.

The econometric model studied here is a random coefficient model as the markup, μ , and the cost of capital, R , can vary across firms. Therefore, variation in inputs should not be due to variation in the markup or the cost of capital (i.e., the coefficients), but should be due to other factors such as variation in technology (e.g., factor-augmenting productivity) or the price of inputs other than capital (e.g., wages). I address this econometric challenge in the following way. To deal with variation in the markup, I divide both sides in equation (1) by $P^X X$, such that the regressor becomes the capital stock divided by expenditure on inputs other than capital. According to standard economic theory, variation in this capital-input ratio is not due to variation in the markup because the markup distorts the first-order condition of all inputs in the same way. Therefore, variation in markups does not lead to a bias when using this modified specification. On the other hand, the cost of capital only affects the first-order condition with respect to capital and therefore does affect the relative input choice. To deal with variation in the cost of capital, I first allow the cost of capital to depend on the capital-input ratio up to a first order. Second, I include observables that control for the extent to which firms are financially constrained and find that this does not affect the estimates to any considerable extent, suggesting that the remaining, unobserved, variation in capital costs is limited. Third, using across industry variation leads to similar results as using within industry variation (my main specification). This is reassuring since the source of variation in the capital-input ratio might be substantially different comparing firms within industries with each other versus comparing different industries. Fourth, I specify a structural model, and matching

decline in the labor share is mainly a US phenomenon. I sidestep the issue of housing and the self-employed by focusing on the corporate sector.

⁴See, e.g., Karabarbounis and Neiman (2014), Grossman et al. (2017), Autor and Salomons (2018), Acemoglu and Restrepo (2018), Hubmer (2018), Martinez (2019) and Moll et al. (2019).

⁵See, e.g., Barkai (2017), Gutiérrez and Philippon (2017), De Loecker et al. (2018) and Gouin-Bonenfant (2018).

moments of the data, I find that the bias due to variation in the cost of capital is reasonably small. Finally, to validate the cost of capital estimator, I regress the estimated cost of capital on the observed depreciation rate across industry-year pairs and, as expected, I find a correlation that is slightly below one.

One attractive feature of my approach to estimating the valued marginal product of capital and the cost of capital is that I do not have to specify the production function. The fall in the labor share coincides with a drop in the relative price of the investment good, which affects the capital and labor share differently depending on the elasticity of substitution between capital and labor (Karabarbounis and Neiman, 2014). Furthermore, the fall in the labor share might be related to the rise in the skill premium, and therefore, capital-skill complementarity (Krusell et al., 2000) or a task-based production function (Acemoglu and Autor, 2011) might be the appropriate production structure. It is generally challenging to estimate production functions and the elasticity of substitution between different inputs.⁶ The assumption that the production function is homogeneous of a constant degree comprises the above mentioned production functions, and I do not need to estimate a specific functional form or elasticities of substitution. Moreover, firms are allowed to produce multiple products using varying technologies. Also the requirements on the data are limited as (firm-specific) input or output prices and expenditure on each specific input (such as different types of labor) are not needed to be observed, but only data on total expenditure on inputs other than capital, nominal output and the nominal capital stock are required. Those are usually available in a typical firm-level data set. Finally, another advantage of this approach is that firms are allowed to have different production technologies of which recent evidence shows that this is the relevant case (David and Venkateswaran, 2019; Doraszelski and Jaumandreu, 2018; Raval, 2019a).

This paper estimates the cost of capital and the profit share using micro-data on firm inputs and output, whereas the existing literature studies the evolution of the cost of capital and the profit share using a required rate of return approach (Hall and Jorgensen, 1967; Barkai, 2017). The required rate of return approach uses that, according to theory, the user cost of capital, R , equals the sum of the interest rate and depreciation rate minus expected inflation of the capital good, which are obtained from aggregate data.⁷ Applying the required rate of return approach to Compustat data, I find that the estimated profit share is similar across the two

⁶See for instance Oberfield and Raval (2014) and Karabarbounis and Neiman (2014) on estimating the elasticity of substitution between capital and labor, and Katz and Murphy (1992) on estimating the elasticity of substitution between high- and low-skilled labor.

⁷The required rate of return approach also takes into account the differential treatment of capital expenditure by the tax authorities. This is also included in my measure of the cost of capital as input choices reveal the cost of capital including taxation.

methods in both level and trend after 1985, but is very different during the 1960s and the 1970s. The required rate of return approach finds the profit share to be as high in the 1960s and 1970s as it is today, which is also found to be the case in aggregate data (Karabarbounis and Neiman, 2018; Barkai and Benzell, 2018). That profits are high during the 1960s and 1970s is rather surprising given that the labor share was at a high level during this time period, and has led to concerns about whether the required rate of return approach measures capital costs accurately (Karabarbounis and Neiman, 2018). Instead, I find the profit share to be (much) lower in the 1960s and the 1970s compared to today, which is consistent with the fall in the labor share being associated with a rise in profits. One potential explanation for the discrepancy between the two methods is that expected inflation is needed for the required rate of return approach, which is approximated with realized inflation. During the period from the 1960s until the early 1980s, the inflation rate was first rapidly increasing and then rapidly declining. However, this does not necessarily imply that expected inflation was also changing rapidly, and therefore might lead to a biased estimate for the required rate of return approach.

In this paper, I study the long-run changes in the cost of capital and profit share. To estimate the cost of capital, I assume that the firm's first-order condition for each input hold. Presumably, these hold to some extent in the long run. However, the extent to which these hold is likely to vary over the business cycle, making my approach not suitable for studying how the cost of capital and profit share change over the business cycle. Furthermore, variation in the cost of capital across firms should be limited and the extent to which this holds depends on the data studied. In this light, the method developed here might not be appropriate for studying small private firms. Using Compustat data partly alleviates this concern as the sample comprises mainly publicly listed firms, which tend to be larger and have access to the capital market. Indeed, in a recent paper David and Venkateswaran (2019) find that the scope for variation in the cost of capital is limited in this data set.

On the other hand, care should be taken when extrapolating my results to the rest of the economy as Compustat is not a representative sample. In some industries, a larger share of firms is public than in other industries, and public firms tend to be larger than private firms. I deal with the representativeness of the data in three ways. To correct for the differential industry composition, I weight the industry profit share with the economy-wide share of value added for that industry to obtain aggregate profits. If anything, this leads to a faster increase in the profit share. To correct for size-based selection I reweight observations

based on the size distribution obtained from the Census. Because I find that larger firms are more profitable than smaller firms and that this relationship has become stronger over time, this attenuates the rise in the economy-wide profit share. However, the rise in the size-based reweighted profit share is still substantial by being around three percentage points. Third, I estimate the profit share using economy-wide industry level data from the Bureau of Economic Analysis. This yields an estimate that is very similar to the estimate based on Compustat data.

It has been documented that a large part of the fall in the labor share is due to reallocation toward firms with a low labor share (Autor et al., 2017; Kehrig and Vincent, 2018), potentially due to offshoring labor-intensive tasks (Elsby et al., 2013). However, based on existing evidence it is not clear whether these firms have a low labor share because they have a high capital share (e.g., due to automating more tasks) or a high profit share. Estimating the profit share using micro-data allows me to study how the distribution of the profit share across firms has changed over time. I find that larger firms have a higher profit share and that the rise in the profit share is due to reallocation between firms, though mainly not due to reallocation between industries.⁸ Furthermore, I find that the median profit share has risen at the same rate as the average profit share.

When calculating economic profits I do not subtract entry costs as I do not observe them. Thus, the rise in what I call economic profits could potentially be explained in its entirety by a rise in entry costs. I argue that this is not the case because the life-cycle pattern of profits has changed over time. In a simple model of entry, a potential entrant enters the market when the present value of profits is larger than the cost of entry. In a stationary equilibrium, firms of all ages are observed and hence total profits observed at any point in time exceed the present value of profits when discounting is positive and therefore exceed the entry costs (Atkeson and Kehoe, 2005).⁹ The discrepancy between total and discounted profits is larger the more back-loaded profits are over the life cycle. I find that profits have indeed become more back-loaded. Young firms today make as much or less profits than young firms in the 1980s but older firms have become more profitable. Therefore, the present value of the future stream of profits (or, equivalently, entry costs) has grown at a lower rate than total profits. This suggests that at least part of the rise in profits is not due to a rise in entry costs. Furthermore, depending on the discount factor, the value of entering the market might, in fact, have declined, which could explain the observed fall in firm entry (Decker et al., 2014).

The returns to scale do not need to be estimated in order to estimate the cost of

⁸That the rise in the profit share is due to reallocation across firms is also found for markups by De Loecker et al. (2018) and Baqaee and Farhi (2020).

⁹This is the case as long as profits are not heavily front-loaded.

capital and the profit share, and are allowed to vary over time and across firms. However, knowing the returns to scale makes it possible to estimate the markup as the markup is the price-average cost ratio times the returns to scale.¹⁰ I follow Syverson (2004) and estimate the returns to scale using cost shares based on my estimate of the cost of capital. I find that the returns to scale have been increasing from just below one in the the 1960s and 1970s to around 1.05 toward the end of the sample period. Combining this increase in the returns to scale with the increase in the profit share, the cost-weighted average markup has increased from around 1.05 to around 1.15. This estimate of the markup relies on additional assumptions compared to my estimate of the profit share, such as that real quantities are observed. However, an advantage compared to the literature on estimating markups using production data (De Loecker and Warzynski, 2012; De Loecker et al., 2018) is that firms are allowed to differ in their factor-augmenting technologies. The rise in markups that I find is similar, though somewhat smaller in magnitude, as what De Loecker et al. (2018) find using the same data.

In my main specification, I assume that labor markets are competitive. However, it could be that the wage is a markdown on the marginal product of labor, which would be another source of profits besides imperfect competition in the product market. I relax this assumption and show that it is possible to also estimate the markdown with my framework. Allowing for imperfect competition in the labor market does not affect the estimated profit share. Furthermore, I find that the markdown has been relatively constant at around 1.05. This means that firms have some labor market power but that, at least in this sample, the rise in profits and the decline in the labor share cannot be explained by firms having experienced an increase in their labor market power.

Section I discusses the literature in more detail. Section II derives the estimator of the cost of capital. The data I use to estimate the evolution of capital compensation and the profit share is discussed in Section III. Section IV shows the main results and Section V compares my approach with the required rate of return approach in the literature to estimating the cost of capital. Section VI shows that my results are robust to different specifications and Section VII studies heterogeneity in profitability across firms and discusses the life-cycle pattern of profits. In Section VIII, I estimate the returns to scale and the markup. Section IX finds that changes in markups are negatively correlated with changes in concentration at the industry level. Section X discusses alternative applications of the cost of capital estimator and Section XI concludes.

¹⁰That is because the returns to scale equal average costs divided by marginal costs.

I Related Literature

This paper is foremost related to the literature on estimating the cost of capital. Hall and Jorgensen (1967) derive a formula for the user cost of capital from the condition that the price of a new capital good equals the discounted value of all future services derived from this capital good. This makes that the cost of capital equals the interest rate minus expected inflation of the capital good plus the depreciation rate, corrected for the tax treatment of capital.

Barkai (2017), Barkai and Benzell (2018) and Eggertsson et al. (2018) use this required rate of return approach to estimate the evolution of the cost of capital, the capital share, and therefore the profit share in the US economy. Barkai (2017) finds that the capital share has been declining and that the profit share has been increasing since the late 1980s. This coincides with the fall in the labor share and therefore indicates that the labor share has fallen due to an increase in profits. Karabarbounis and Neiman (2018), Barkai and Benzell (2018) and Eggertsson et al. (2018) extend this analysis further back in time and find that the profit share was declining beforehand and that the profit share in the 1960s and 1970s was about equally large as it is today.¹¹ This evolution of the profit share is not consistent with the narrative that relates profits to the labor share as the labor share was high during the 1960s and 1970s. Moreover, the labor share was roughly constant during this period and therefore the large swings in the profit share, as inferred using the required rate of return approach, suggest large changes in the role of capital in the production technology.¹² One possibility for this surprising behavior of the profit share is that the required rate of return approach mismeasures the cost of capital. The main challenge for the required rate of return approach is to estimate the interest rate (including a risk premium) and the expected inflation of the capital good.¹³

My contribution to this literature is that I develop a complementary method to estimate the cost of capital and therefore the profit share. In my framework, I do not need to specify the individual components of the cost of capital, as opposed to Hall and Jorgensen (1967), because I estimate the entire cost of capital directly using firm-level data, whereas the papers cited above estimate each individual component separately from different data sources (e.g., the interest rate from

¹¹Karabarbounis and Neiman (2018) call it factorless income instead of economic profits. See Rognlie (2018) for an excellent discussion of Karabarbounis and Neiman (2018).

¹²Furthermore, Karabarbounis and Neiman (2018) show that the required rate of return approach implies that there is a tight negative relationship between the profit share and the real interest rate over the business cycle.

¹³For instance, it is not clear how the risk premium has evolved over time. Farhi and Gourio (2018) find a rising risk premium.

financial markets data and the depreciation rate from national accounts data). However, a drawback of my method is that heterogeneity in the cost of capital across firms might lead to a biased estimate.

Another potential problem with estimating the profit share using the required rate of return approach is mismeasurement of the capital stock as the cost of capital needs to be multiplied by the capital stock to obtain total capital compensation.¹⁴ In contrast, with my approach, unobserved capital does not lead to a biased estimate of the profit share as long as the unobserved capital stock is proportional to the observed capital stock. The reason is that I estimate the cost of capital using data on capital. If the capital stock were to be undermeasured, this means that I would estimate a too high cost of capital. However, when this too high cost of capital is multiplied with the too low measured capital stock to obtain total capital compensation, these two errors cancel out exactly. On the other hand, idiosyncratic measurement error leads to attenuation bias for my method, while this is not an issue for the the required rate of return approach. Given that using variation across industries leads to similar results as using variation across firms within an industry, and that there is plausibly less measurement error in the industry-average capital stock, attenuation bias is not a major concern.

Second, this paper contributes to the literature on estimating markups using production data.¹⁵ De Loecker et al. (2018) estimate markups in the same data set as I use, using the production approach (Hall, 1988; De Loecker and Warzynski, 2012). They find that markups have increased since the 1980s. The markup equals the output elasticity with respect to an input times nominal output divided by expenditure on that input. The difficulty is in consistently estimating the output elasticity. Raval (2019b) finds contradicting estimates of the markup using different inputs and argues that the reason for this is that the typical estimator of the output elasticity does not allow for differences in factor-augmenting productivity across firms. I add to this literature by developing a method to estimate the markup allowing for differences in factor-augmenting technology. I find that the markup increases over time, although at a lower rate than what is found by De Loecker et al. (2018).¹⁶ Furthermore, both methods lead to different results for how the distribution of markups has changed over time. I find that the median markup has increased at a similar rate as the average markup while De Loecker et al.

¹⁴See Hall (2001), Atkeson and Kehoe (2005), McGrattan and Prescott (2005), Corrado et al. (2009) and Eisfeldt and Papanikolaou (2013) for a discussion of the role of unobserved intangibles, such as brand names, patents and organizational capital.

¹⁵See Basu (2019) and Syverson (2019) for a discussion.

¹⁶Traina (2018) uses the same method as De Loecker et al. (2018) but looks at different cost variables within the same data set (i.e., Compustat) and finds a smaller increase in markups than what is found by De Loecker et al. (2018).

(2018) find that the median markup has not increased. I show in Section VII that this discrepancy in distributional implications can be explained by an increased heterogeneity in factor-augmenting technology over time.

Third, this paper contributes to the literature on explaining the rise in profits and markups. Explanations include consumer inertia (Bornstein, 2018), an increase in common ownership (Azar and Vives, 2019), IT improvements leading to a fall in the firm-level costs of spanning multiple markets (Aghion et al., 2019), falling interest rates (Liu et al., 2019) and a decline in knowledge diffusion between frontier and laggard firms (Akcigit and Ates, 2019). I find that profits have become more back-loaded over the firm's life cycle. All else equal, this lowers the present value of profits and for the entry condition to continue to hold, aggregate profits/markups need to increase in response.

II Estimating the Cost of Capital

Capital compensation is the product of the (user) cost of capital, R , and the nominal capital stock, $P^K K$,

$$\text{Capital Compensation}_t \equiv R_t \cdot P_t^K K_t, \quad (2)$$

where t denotes time. The price of capital, P_t^K , is the price paid to acquire one unit of real capital K (e.g., the price of buying a computer). The cost of capital, R , refers to the cost of employing one unit of nominal capital during a period, and does, for instance, take into account that capital does not fully depreciate within a period, that acquiring capital leads to services in the future that have to be discounted appropriately and the way capital expenditure is treated by the taxation authority.¹⁷ As the capital stock is observed, estimating capital compensation comes down to estimating the cost of capital R .¹⁸ In order to do so I rely on basic economic theory. I assume cost minimization and that the production function is homogeneous of a constant degree. For notational simplicity, I here consider the static firm problem in which the firm rents capital. See Appendix A that the dynamic firm problem leads to the same first-order conditions, except that expectations of the variables are taken.

Suppose that firm i at time t produces real output Y_{it} using as inputs capital, K_{it} , and M other inputs, X_{it}^m , according to the production function $Y_{it} =$

¹⁷Equivalently, the cost of capital can be denoted in real terms. This would neither affect the derivation of the estimator nor the results. I decided to define R in nominal terms because I observe the nominal capital stock and not the real capital stock.

¹⁸In Section V, I discuss systematic mismeasurement of the capital stock and show that this has no effect on my estimate of the profit share.

$F_{it}(X_{it}^1, \dots, X_{it}^M, K_{it})$. These inputs could, for instance, be (different types of) labor and materials. Different types of capital are also allowed, but for notational simplicity I only consider one capital good here.

Assumption 1 (Cost minimization). *The firm chooses inputs to minimize costs subject to output at time t being equal to some scalar \bar{Y}_{it} and takes input prices, $P_{it}^{X^m}$ and P_{it}^K , and R_{it} as given,*

$$\begin{aligned} \min_{\{X_{it}^1, \dots, X_{it}^M, K_{it}\}} \quad & \sum_{m=1}^M P_{it}^{X^m} X_{it}^m + R_{it} P_{it}^K K_{it} \\ \text{s.t.} \quad & F_{it}(X_{it}^1, \dots, X_{it}^M, K_{it}) = \bar{Y}_{it}. \end{aligned} \quad (3)$$

I allow for firm-specific input prices. However, I assume that the firm takes these input prices as given. This does not preclude input providers from charging a markup over their marginal cost, but it does assume that the price cannot be a function of the quantity demanded. This precludes for instance monopsonistic competition and bulk discounts.¹⁹ Note that I have not made any assumptions about the price-setting behavior in the output market.

The following Lagrangian is associated with the cost minimization problem:

$$\begin{aligned} \mathcal{L}_{it}(X_{it}^1, \dots, X_{it}^M, K_{it}, \lambda_{it}) = \sum_{m=1}^M P_{it}^{X^m} X_{it}^m + R_{it} P_{it}^K K_{it} + \\ \lambda_{it} (\bar{Y}_{it} - F_{it}(X_{it}^1, \dots, X_{it}^M, K_{it})) , \end{aligned} \quad (4)$$

where λ_{it} is the Lagrange multiplier, which equals the marginal cost. Define the markup as the output price, P_{it} , divided by marginal costs: $\mu_{it} = \frac{P_{it}}{\lambda_{it}}$.

Taking the derivative of the Lagrangian with respect to capital, K_{it} , gives

$$R_{it} P_{it}^K = \lambda_{it} \frac{\partial F_{it}(\cdot)}{\partial K_{it}}.$$

Realizing that λ_{it} equals the price divided by the markup and denoting the price times the marginal product as the valued marginal product, $VMPK_{it}^m$, gives, after rewriting, that the cost of capital equals the valued marginal product of capital divided by the price of capital and the markup,

$$R_{it} = \frac{VMPK_{it}}{\mu_{it} P_{it}^K}. \quad (5)$$

¹⁹In Section VI, I relax the assumption that the firm takes the wage as given and estimate the markdown in the labor market together with the cost of capital. I find that this leads to a similar estimate for the profit share as in the baseline.

The price of capital, P_{it}^K , shows up in this equation because the cost of capital, R_{it} , is defined in nominal terms. $\frac{VMPP_{K_{it}}}{P_{it}^K}$ is the increase in output (valued at prices) when the nominal capital stock increases by one. When there is no risk of confusion, I will refer to this ratio as the marginal product of capital.

To estimate the cost of capital, I first estimate the markup and the valued marginal product of capital divided by the price of capital (both scaled by an identical constant), and then divide these two estimates with each other. Subsequently multiplying this estimate with the capital stock gives total capital compensation and hence allows me to estimate economic profits.

Taking the derivative with respect to the other inputs, X_{it}^m , gives a similar first-order condition, namely that the markup times the price of an input equals the valued marginal product of that input,

$$\mu_{it} P_{it}^{X^m} = VMPP_{X_{it}^m}, \quad m \in \{1, \dots, M\}. \quad (6)$$

This first-order condition will be used when estimating the valued marginal product of capital and the markup.

In order to obtain estimates of the (scaled) markup and the (scaled) valued marginal product of capital divided by the price of capital, I assume that the production function is homogeneous of a constant degree.

Assumption 2 (Homogeneous production function). *The production function F_i is homogeneous of degree ϕ_{it} : $F_{it}(\lambda X_{it}^1, \dots, \lambda X_{it}^M, \lambda K_{it}) = \lambda^{\phi_{it}} F_{it}(X_{it}^1, \dots, X_{it}^M, K_{it})$ for $\lambda > 0$.*

The production function is indexed by it to highlight that the production function is allowed to differ across firms and that it can change over time. E.g., firms can differ in their factor-biased technologies and also in their elasticities of substitution between different inputs. Homogeneous production functions include, but are not limited to, frequently used production functions such as Cobb-Douglas, CES, task-based production functions as in Acemoglu and Autor (2011), and capital-skill complementarity (Krusell et al., 2000). Note that here there is only one type of capital, but this model could easily be extended to allow for multiple types of capital, such as structures and equipment. Then, the cost of capital would be the weighted average cost of capital of structures and equipment. To estimate the cost of capital in this extended model, observing only the total capital stock is sufficient, without needing to impose that the different types of capital are perfect substitutes. This is for the same reason, which I show momentarily, why it is only necessary to observe total operating expenses and not each individual expense.

A commonly used production function that is not homogeneous is the translog

production function. Furthermore, I have assumed that capital and the inputs X_i^m are factors of production and are not fixed costs such as overhead costs. If there are fixed costs, these should not be included in the production function and therefore not in the estimation of the cost of capital, but still need to be subtracted from revenues to obtain economic profits. I will discuss fixed costs in more detail in Section VI.

Firms are also allowed to produce multiple products according to a product-specific production function. This does not affect the estimation, and data on outputs and inputs do not need to be observed at the product level. The reason is that the final equation, that I bring to the data, is linear. See Appendix B for more details.

Assumption 2 ensures that Euler's theorem holds, which states that output equals the sum of the factors of production multiplied by its respective marginal products divided by the returns to scale (for simplicity I omit the time subscripts):

$$Y_i = \sum_{m=1}^M \frac{MPX_i^m}{\phi_i} X_i^m + \frac{MPK_i}{\phi_i} K_i. \quad (7)$$

Anticipating that I only observe nominal output and capital, I multiply Euler's theorem by the output price P_i and multiply and divide capital with its price, P_i^K , to obtain

$$P_i Y_i = \sum_{m=1}^M \frac{VMPX_i^m}{\phi_i} X_i^m + \frac{VMPK_i}{\phi_i P_i^K} P_i^K K_i. \quad (8)$$

Instead of marginal products, valued marginal products appear on the right-hand side. Using that $VMPX_i^m$ equals $\mu_i P_i^{X^m}$ by the first-order condition (6), and taking $\frac{\mu_i}{\phi_i}$ out of the summation gives

$$P_i Y_i = \frac{\mu_i}{\phi_i} \sum_{m=1}^M P_i^{X^m} X_i^m + \frac{VMPK_i}{\phi_i P_i^K} P_i^K K_i. \quad (9)$$

There are two benefits of plugging in the first-order condition with respect to inputs X^m . The first benefit is that only expenditure on all inputs other than capital is needed and not data on each specific input, while still allowing for a rich production structure. The second benefit is, as I show momentarily, that this makes that firm-specific input prices, $P_i^{X^m}$, do not lead to a bias when bringing this equation to a regression framework. Finally, denoting total expenditure on inputs other than capital by firm i by $P_i^X X_i = \sum_{m=1}^M P_i^{X^m} X_i^m$, and dividing both sides of equation (9) by total input expenditure gives modified Euler's theorem.

Proposition 1 (Modified Euler's theorem). *Suppose that Assumption 1 and 2 hold, then*

$$\frac{P_i Y_i}{P_i^X X_i} = \frac{\mu_i}{\phi_i} + \frac{VMPK_i}{\phi_i P_i^K} \frac{P_i^K K_i}{P_i^X X_i}. \quad (10)$$

As I discuss in more detail later, dividing both sides by input expenditure, $P^X X$, makes that markup dispersion does not lead to a bias when bringing modified Euler's theorem to a regression framework, while a bias would occur when we would not divide by input expenditure.

Modified Euler's theorem shows that there is a tight relationship between nominal output divided by input expenditure (the output-input ratio) and the nominal capital stock divided by input expenditure (the capital-input ratio), governed by the marginal product of capital, the markup and the returns to scale. Nominal output, input expenditure and the capital stock are directly observed, while the other variables are not. If one has prior information about the returns to scale and the markup, one can back out the marginal product of capital using equation (10). Caselli and Feyrer (2007) estimate the marginal product of capital assuming constant returns to scale and perfect competition. Indeed, setting $\phi = 1$ and $\mu = 1$ in equation (10) leads to the same equation as the one that Caselli and Feyrer (2007) use to estimate the marginal product of capital. Thus, modified Euler's theorem generalizes the estimator of the marginal product in Caselli and Feyrer (2007) to a general homogeneous production function and imperfect competition.

Moreover, if there is variation in the capital-input ratio, the marginal product of capital can be estimated together with the markup (i.e., without needing to assume a value for the markup). However, an assumption about the returns to scale still needs to be made. So far, I have not yet used the first-order condition with respect to capital. I will use this first-order condition next in order to estimate the cost of capital and show that no information about the returns to scale is needed in order to estimate the cost of capital.

The Cost of Capital

In order to estimate the cost of capital, I bring modified Euler's theorem (10) to a regression framework,

$$\frac{P_i Y_i}{P_i^X X_i} = \alpha_0 + \alpha_1 \frac{P_i^K K_i}{P_i^X X_i} + \varepsilon_i, \quad (11)$$

where ε_i denotes the error term. Note that Euler's theorem is exact, meaning that for instance, unobserved productivity, which is the typical econometric challenge when estimating production functions, does not lead to a bias here (see Appendix B for a discussion of the differences between this regression and the literature on production function estimation.).²⁰ However, the marginal product of capital, the markup and the returns to scale might vary across firms, meaning that this is a random coefficient model and that the error term consists of heterogeneity in the marginal product of capital, the markup and the returns to scale.

The intercept, α_0 , in equation (11) refers to the average of the markup divided by the returns to scale, $\frac{\mu}{\phi}$. The reason is that for a (hypothetical) firm that employs no capital, total costs are given by $P_i^X X_i$, and thus the price-average cost ratio for that firm equals $\frac{P_i Y_i}{P_i^X X_i}$. The slope, α_1 , refers to the average of the marginal product of capital divided by the returns to scale, $\frac{VMPK}{\phi PK}$. That is, the slope gives the average increase in nominal output when the nominal capital stock increases by 1 while holding all other inputs constant. Under constant returns to scale this equals the valued marginal product of capital.

By the first-order condition with respect to capital (5), the cost of capital, R , equals the valued marginal product of capital divided by the markup and the price of capital. Therefore, to estimate the cost of capital I divide the slope coefficient by the intercept coefficient. Because both coefficients are scaled by the returns to scale parameter, this parameter will cancel out when both coefficients are divided with each other. Thus, the returns to scale do not affect the estimate of the cost of capital.

I now summarize the assumptions needed to obtain this estimate of the cost of capital. Cost minimization gives a relationship between the cost of capital and the marginal product of capital. However, marginal products of capital are not observed while the average product of capital, holding all other inputs constant, is observed. Assuming a homogeneous production function gives that there is a size-independent relationship between the marginal product and average product, and that this relationship is the same across all inputs. Furthermore, firms are assumed to take input prices as given such that the wedge between the marginal product and the cost of an input is the same across all inputs, including capital.

²⁰However, unobserved productivity can lead to a bias if it affects the marginal product of capital or the markup. The marginal product is, for instance, affected if there are unanticipated productivity shocks to which capital cannot respond instantaneously.

Identification

The error term in equation (11) comprises variation in the marginal product of capital divided by the returns to scale and variation in the price-average cost ratio. Here, I discuss which conditions on firm heterogeneity need to be imposed in order to identify the cost of capital.

Proposition 2 (Identification). *Suppose that Assumption 1 and 2 hold and that the cost of capital is equalized across firms, then the cost of capital is identified, up to a first-order approximation, by first estimating regression equation (11) and then dividing the resulting slope coefficient by the intercept coefficient, $\hat{R} = \frac{\hat{\alpha}_1}{\hat{\alpha}_0}$.*

Thus, as long as the cost of capital is equalized across firms, the cost of capital is identified.²¹ This proposition implies that firm heterogeneity in other dimensions does not lead to a bias. Firms are allowed to differ in their markups and returns to scale, but also in dimensions that are not captured by the coefficients, such as differences in input prices, $P_i^{X^m}$. I will first explain why heterogeneity in price-average cost ratios does not lead to a bias while heterogeneity in the cost of capital does. Subsequently, I discuss how I deal with this econometric challenge.

Heterogeneous Price-Average Cost Ratios

The next equation shows the regression equation when the error due to heterogeneity in the price-average cost ratio, η_i , is written out:

$$\frac{P_i Y_i}{P_i^X X_i} = \frac{\bar{\mu}}{\phi} + \frac{\bar{\mu}}{\phi} R \frac{P_i^K K_i}{P_i^X X_i} + \underbrace{\left(\frac{\mu_i}{\phi_i} - \frac{\bar{\mu}}{\phi} \right) \left(1 + R \frac{P_i^K K_i}{P_i^X X_i} \right)}_{\eta_i}. \quad (12)$$

$\frac{\bar{\mu}}{\phi}$ denotes the average price-average cost ratio and I have replaced the marginal product of capital with $\mu_i R$ following from the first-order condition with respect to capital. For there to be no bias, η_i has to be uncorrelated with the capital-input ratio $\frac{P_i^K K_i}{P_i^X X_i}$. Because $\frac{\mu_i}{\phi_i} - \frac{\bar{\mu}}{\phi}$ is zero in expectation, this is fulfilled when the price-average cost ratio is uncorrelated with the capital-input ratio.

Lemma 1. *The markup μ_i does not affect the capital-input ratio $\frac{P_i^K K_i}{P_i^X X_i}$.*

Proof. Taking the ratio of the first-order condition with respect to input m (6) and the first-order condition with respect to capital (5) gives

$$\frac{VMPX_{it}^m}{VMPK_{it}} = \frac{P_{it}^{X^m}}{P_{it}^K R_t}.$$

²¹The error induced due to the first-order approximation is negligible (see Appendix J).

Dividing the left-hand side numerator and denominator by $\left(\frac{P_{it}^X X_{it}}{P_{it}^K}\right)^{\phi-1}$ and using that the derivatives of the production function are homogeneous of degree $\phi - 1$ by Euler's theorem gives

$$\frac{P_{it} \frac{\partial F_{it}\left(\frac{P_{it}^K K_{it}}{P_{it}^X X_{it}}, \frac{P_{it}^K X_{it}^1}{P_{it}^X X_{it}}, \dots\right)}{\partial X_{it}^m}}{\frac{\partial F_{it}\left(\frac{P_{it}^K K_{it}}{P_{it}^X X_{it}}, \frac{P_{it}^K X_{it}^1}{P_{it}^X X_{it}}, \dots\right)}{\partial K_{it}}} = \frac{P_{it}^{X^m}}{P_{it}^K R_t}. \quad (13)$$

μ_i does not show up in this equation and hence $\frac{P_{it}^K K_{it}}{P_{it}^X X_{it}}$ does not depend on the markup (note that $\frac{X_{it}^m}{P_{it}^X X_{it}}$ does not depend on the markup because the markup distorts the first-order condition for each input in the same way). \square

The capital-input ratio is not affected by the markup because the markup distorts the first-order condition with respect to capital in the same way as the first-order condition with respect to the other inputs. Combined with a homogeneous production function, this means that both the capital stock and the other inputs are affected in the same way. Furthermore, the returns to scale do also not affect the capital-input ratio and hence, the price-average cost ratio is not correlated with the capital-input ratio. Thus, the error term induced by price-average cost heterogeneity is not correlated with the regressor and hence, this heterogeneity does not lead to biased results.

That markups do not affect the capital-input ratio is the reason why I have divided by input expenditure to obtain regression equation (11). When I would not have divided by input expenditure, heterogeneity in the price-average cost ratio would lead to a bias as the price-average cost ratio does affect the level of inputs.

Lemma 1 assumes that there are no unanticipated shocks to the markup. When there are unanticipated shocks to markups, there might be a correlation between the markup and the capital-input ratio as it might take longer to adjust capital than it takes to adjust other inputs. I discuss this in more detail in Section VI and show that markup shocks only have a limited effect on my estimate by showing that using lagged values as instruments gives similar results.

Heterogeneous Costs of Capital

Now consider the case of heterogeneous costs of capital. For simplicity, assume that the price-average cost ratio is the same across firms,

$$\frac{P_i Y_i}{P_i^X X_i} = \frac{\mu}{\phi} + \frac{\mu \bar{R}}{\phi} \frac{P_i^K K_i}{P_i^X X_i} + \underbrace{\frac{\mu(R_i - \bar{R})}{\phi} \frac{P_i^K K_i}{P_i^X X_i}}_{\psi_i}, \quad (14)$$

where ψ_i denotes the error due to heterogeneity in the cost of capital, and the average cost of capital is denoted by \bar{R} .

The cost of capital affects the first-order condition of capital and not the other first-order conditions, and therefore the capital-input ratio depends on the cost of capital. A firm facing a high cost of capital will choose to employ less capital compared to a firm with a low cost of capital. Therefore, this firm will have a relatively low capital-input ratio. This negative correlation between the cost of capital and the capital-input ratio makes that ψ_i is not equal to zero in expectation and that ψ_i is correlated with the regressor leading to a biased estimate of both the scaled marginal product of capital and the scaled markup.

Differences in the cost of capital across firms could, for instance, be caused by differential access to capital markets and bank loans, differences in risks, or by differences in depreciation rates. Variation across firms in costs of other inputs than capital, P_i^X , does not lead to a bias because nominal input expenditure shows up in modified Euler's theorem and this expenditure is observed. This is not the case for variation in the cost of capital since R_i is unobserved (which is the reason for doing this estimation exercise in the first place).

As Proposition 2 shows, OLS leads to unbiased results as long as the cost of capital is equal across firms. This would be the case if capital is mobile. This paper is about long-run changes in the cost of capital and profit share. Presumably, capital mobility holds to some extent in the long run providing a justification for the method proposed here. Furthermore, this paper uses Compustat data, which mainly comprises publicly listed firms. Hence, these are firms that have access to the capital market and are therefore likely to face a similar cost of capital. Another important observation is that the main object of interest of this paper is the change in the profit share, and not so much the level of the profit share. If variation in the cost of capital (relative to variation in the capital-input ratio) has remained constant over time, then the bias has remained constant over time allowing me to identify the change in the capital share and profit share, even in the presence of a bias.

Nonetheless, I control for differences in the cost of capital across firms to the extent possible with the available data. The econometric challenge is that the slope coefficient is correlated with the regressor. Therefore, in the regression I allow the slope coefficient to depend on the regressor up to a first-order (thus the capital-input ratio squared enters in the regression). This does not mean that I assume this relationship to be linear (it is not), but it captures the variation in the cost of capital to the extent that it is proportional to the capital-input ratio.

Furthermore, I include additional controls that are meant to capture to what extent firms are financially constrained. Specifically, I use i) financial leverage measured as total liabilities divided by total assets, ii) long-term debt divided by total liabilities to capture that firms that rely more on short-term debt might be more financially constrained, iii) interest payments divided by total liabilities, iv) firm size measured by sales and v) sales growth. The estimate of the cost of capital does not differ to any great extent depending on whether these controls are included (see Figure 27a in Appendix H). This makes it plausible that the remaining variation in the cost of capital is small. Furthermore, because firms might differ in the types of capital they are using I also use the depreciation rate as a control. I demean all controls and multiply them by $\frac{P^K K}{P^X X}$, as these are controls for the cost of capital. In addition, I include the controls not interacted with the capital-input ratio as a control for price-average cost heterogeneity. This leads to the following regression, where Z_i^j refers to the j -th control,

$$\frac{P_i Y_i}{P_i^X X_i} = \alpha_0 + \sum_j v_j \left(Z_i^j - \bar{Z}^j \right) + \left(\alpha_1 + \sum_j \gamma_j \left(Z_i^j - \bar{Z}^j \right) + \gamma \left(\frac{P_i^K K_i}{P_i^X X_i} - \frac{\overline{P^K K}}{\overline{P^X X}} \right) \right) \frac{P_i^K K_i}{P_i^X X_i} + \varepsilon_i. \quad (15)$$

Demeaning the controls entails that the coefficient α_1 refers to the average marginal product of capital divided by the returns to scale. The coefficient on the capital-input ratio squared, γ , will be less than or equal to zero since the higher the cost of capital the lower the capital-input ratio.

To run the regression, variation in the capital-input ratio is needed. This variation should come from other sources than variation in the cost of capital, such as variation in technology (e.g., factor-augmenting productivity) or variation in input prices (e.g., wages). Hsieh and Klenow (2009) find a large variation in revenue productivities across firms, which could potentially be explained by differences in the cost of capital. David and Venkateswaran (2019) decompose this variation in revenue productivities into variation due to markups, differences in

technology and other factors such as the cost of capital. They find for the same data as I am using that about half of the variation in capital revenue productivities within 4-digit industries is due to variation in output elasticities and that about 13% of the variation is due to markup dispersion.²² Likewise, Doraszelski and Jaumandreu (2018) and Raval (2019a) also find heterogeneity in technology in other data sets. In my main specification, I will use variation across firms within a 2-digit industry (examples of which are *chemicals and allied products* and *business services*). As robustness, I use variation in the capital-input ratio across industries. Arguably, there is a great deal of variation in technology across industries, which makes this specification attractive. However, variation in the cost of capital across industries might be large as well, for instance when there are differences in risks across industries. Nonetheless, the type of variation in the capital-input ratio across firms versus across industries is very different and, reassuringly, both estimates are similar, suggesting that the variation in the cost of capital is limited.

Bias Formula in Case of Cobb-Douglas and Log-Normal Distribution

An analytic expression of the bias can be obtained in the case of a Cobb-Douglas production function and parameters being log-normally distributed (and when no controls are included in the regression). Suppose that the production function is Cobb-Douglas, $Y_i = A_i K_i^{\theta_i^K} X_i^{\theta_i^X}$, where θ_i^K and θ_i^X refer to the output elasticities with respect to capital and other inputs, respectively, and A_i denotes Hicks-neutral productivity. The ratio of the first-order conditions gives that the capital-input ratio equals the product of technology, $\theta_i \equiv \frac{\theta_i^K}{\theta_i^X}$, and the inverse of the cost of capital, $\frac{1}{R_i}$,

$$\frac{P_i^K K_i}{P_i^X X_i} = \theta_i \frac{1}{R_i}. \quad (16)$$

It turns out that the capital-input ratio is well approximated by a log-normal distribution (see Figure 24 in Appendix H). Based on equation (16), one way of generating a log-normally distributed capital-input ratio is that both θ_i and R_i are independent log-normally distributed. Suppose that these variables, as well as μ_i , are indeed log-normally distributed with parameters μ_θ , σ_θ^2 , μ_R , σ_R^2 , μ_μ and σ_μ^2 , respectively. Then, by the properties of the log-normal distribution, $\frac{K}{X}$ is

²²These numbers are based on a specification in which David and Venkateswaran (2019) assume that all firms in an industry pay the same wage. When they instead use the reported wage bill at the firm level (a smaller sample) they find that 62% and 28% of the variation in capital revenue productivities is due to variation in output elasticities and markups, respectively.

log-normally distributed with $\mu_\theta - \mu_R$ and $\sigma_\theta^2 + \sigma_R^2$ as parameters.

The estimator for the valued marginal product of capital is

$$\frac{\text{cov} \left(\frac{VMPK_i}{P_i^K}, \frac{P_i^K K_i}{P_i^X X_i}, \frac{P_i^K K_i}{P_i^X X_i} \right)}{\text{var} \left(\frac{P_i^K K_i}{P_i^X X_i} \right)} = \frac{\text{cov} \left(\theta \mu, \theta \frac{1}{R} \right)}{\text{var} \left(\frac{P_i^K K_i}{P_i^X X_i} \right)},$$

where the equality is obtained by plugging in the first-order conditions. Using the properties of the log-normal distribution for the covariance and variance gives, after rewriting, that

$$\widehat{\frac{VMPK}{P^K}} = E(\mu R) \underbrace{\frac{e^{\sigma_\theta^2 - \sigma_R^2} - e^{-\sigma_R^2}}{e^{\sigma_\theta^2 + \sigma_R^2} - 1}}_{=\text{relative bias}+1}.$$

The latter ratio equals the bias (in relative terms) plus one. This ratio is smaller than or equal to one, meaning that the estimate of the cost of capital is biased downward when there is dispersion in the cost of capital.²³ When the cost of capital is the same across firms ($\sigma_R^2 = 0$), the ratio is one, and hence the estimator is unbiased. When there is no variation in technology ($\sigma_\theta^2 = 0$), the ratio is 0 and hence a marginal product of 0 is estimated. The bias (in relative terms) only depends on the dispersion in (log) technology σ_θ^2 and the dispersion in the (log) cost of capital σ_R^2 . The absolute value of the relative bias is increasing in σ_R^2 and decreasing in σ_θ^2 .

That the marginal product of capital is estimated to be zero when there is no variation in technology is not a general result, but is due to the assumption of a Cobb-Douglas production function. With a Cobb-Douglas production function, variation in the input price, P_i^X , does not lead to variation in the capital-input ratio and hence, all variation should indeed be due to variation in technology. However, when the production function is not Cobb-Douglas, variation in input prices also leads to variation in the capital-input ratio. In the data I use in this paper, there is a large dispersion in wages, thus, plausibly leading to variation in the capital-input ratio.

Finally, Appendix C shows that the bias is reasonably small based on data that is simulated using a structural model that matches moments of the data, and that deviates from a Cobb-Douglas production function.

²³To see that $\frac{e^{\sigma_\theta^2 - \sigma_R^2} - e^{-\sigma_R^2}}{e^{\sigma_\theta^2 + \sigma_R^2} - 1} \leq 1$, multiply both sides by $e^{\sigma_\theta^2 + \sigma_R^2} - 1$ (which is positive) and rewrite to get $e^{\sigma_\theta^2 - \sigma_R^2} - e^{-\sigma_R^2} + 1 - e^{\sigma_\theta^2 + \sigma_R^2} = (e^{\sigma_\theta^2} - 1)(e^{-\sigma_R^2} - 1) + e^{\sigma_\theta^2}(1 - e^{\sigma_R^2}) \leq 0$.

Weighting

In this paper I use the estimated cost of capital to estimate the evolution of economic profits. Profits of a firm i are nominal output minus costs,

$$Profits_i = P_i Y_i - P_i^X X_i - R_i P_i^K K_i,$$

and total profits are the sum of firm-level profits,

$$Profits = \sum_i Profits_i = PY - P^X X - P^K K \underbrace{\sum_i R_i \frac{P_i^K K_i}{P^K K}}_R, \quad (17)$$

where variables without an i -subscript denote totals. The last term is obtained by dividing and multiplying by the total nominal capital stock $P^K K$. This shows that the aggregate cost of capital, R , is the capital-weighted average cost of capital. The regression only identifies the average marginal product of capital, and not a firm-specific marginal product of capital.²⁴ Therefore, inspired by equation (17), I run the regression weighted by capital (not weighting by capital does not affect the results).²⁵

Furthermore, if there is heterogeneity in the cost of capital, then the expectation of the error term in the regression does not equal zero (see equation (14)), and therefore the intercept, $\frac{\mu}{\phi}$, is not identified (this would also be the case when having an instrument (Wooldridge, 1997, 2003)). However, it turns out that the cost-weighted average price-average cost ratio is identified.²⁶ To see this, rewrite modified Euler's theorem (10) to obtain an expression for the firm-specific price-average cost ratio

$$\frac{\mu_i}{\phi_i} \equiv \frac{P_i Y_i}{P_i^X X_i} - \frac{VMPK_i}{\phi_i P_i^K} \frac{P_i^K K_i}{P_i^X X_i}.$$

The cost-weighted average price-average cost ratio then becomes

$$\frac{\mu}{\phi} = \sum_i \frac{P_i^X X_i}{P^X X} \frac{\mu_i}{\phi_i} = \frac{PY}{P^X X} - \frac{P^K K}{P^X X} \sum_i \frac{VMPK_i}{\phi_i P_i^K} \frac{P_i^K K_i}{P^K K}, \quad (18)$$

²⁴Solon et al. (2015) show that when heterogeneous effects are not fully modeled, the population average effect is not identified when the variance of the regressor differs across group. However, this bias is small when the variation of technology does not depend on the cost of capital.

²⁵Note that taking the weighted marginal product of capital is not inconsistent with the costs of capital being equalized across firms as heterogeneity in marginal products can also occur due to heterogeneity in markups.

²⁶The cost-weighted average markup is also the relevant statistic that summarizes the distortions to employment and investment decisions (Edmond et al., 2018).

where the last term is the capital-weighted average marginal product of capital, an object obtained from the capital-weighted regression.²⁷

Thus, my estimate of R is $\frac{\overline{VMPK/(P^K\phi)}}{\mu/\phi} = \frac{\sum_i \frac{VMPK_i}{P_i^K \phi_i} \frac{P_i^K K_i}{P^K K}}{\sum_i \frac{\mu_i}{\phi_i} \frac{P_i^K X_i}{P^K X}}$, which turns

out to be equal to $\sum_i R_i \frac{P_i^K K_i}{P^K K}$ in expectation. To see that these two expressions are equal, replace $\frac{VMPK_i}{P_i^K}$ with $\mu_i R_i$ and rewrite to obtain that the required condition is that $\sum_i \frac{\mu_i}{\phi_i} \frac{P_i^K X_i}{P^K X}$ has to equal $\sum_i \frac{\mu_i}{\phi_i} \frac{R_i P_i^K K_i}{\sum_k R_k P_k^K K_k}$. This holds because the price-average cost ratio does not affect the capital-input ratio.

To summarize, I run a regression weighted by capital to obtain the average marginal product of capital divided by the returns to scale and I use equation (18) to obtain the cost-weighted average price-average cost ratio. Then, the capital-weighted cost of capital is the division of these two objects. Multiplying the cost of capital by the capital stock gives total capital compensation. Subtracting capital compensation together with operating expenses from nominal output gives economic profits. I will report profits as a share of sales, which is equal to the sales-weighted average profit share at the firm level.

III Data

This paper uses Compustat data for the US, which mainly covers publicly listed firms. The main advantage of Compustat is that it goes back to the 1960s and therefore allows me to study the change in the profit share over a long time horizon, while with 29% still covering a substantial share of US private employment (Davis et al., 2006). A disadvantage is that the firms included are not a random sample of the universe of US firms. These are firms that tend to be older and bigger than the typical firm.²⁸ Therefore, care should be taken when extrapolating my results to the rest of the US economy. Nonetheless, in the setting of the present paper there is an additional advantage to using Compustat data, namely that all firms have access to the capital market and therefore, there is most likely less

²⁷Instead, an error term appears in case of the unweighted average price-average cost ratio (denote by N the number of observations):

$$\frac{1}{N} \sum_i \frac{\mu_i}{\phi_i} = \frac{1}{N} \sum_i \frac{P_i Y_i}{P_i^K X_i} - \frac{1}{N} \sum_i \frac{VMPK}{\phi_i P^K} \frac{P_i^K K_i}{P_i^K X_i} + \frac{1}{N} \sum_i \left(\frac{VMPK_i}{\phi_i P_i^K} - \frac{VMPK}{\phi_i P^K} \right) \frac{P_i^K K_i}{P_i^K X_i},$$

where $VMPK/P^K$ is the average marginal product of capital (weighted or unweighted). The last term is the expectation of the error term which is unobserved and not equal to zero when there is heterogeneity in the cost of capital.

²⁸Moreover, the number of firms in Compustat over time shows an inverted U-shape with the maximum being in the late 1990s.

variation in the cost of capital across firms than would be the case for private firms. Indeed, David and Venkateswaran (2019) find that heterogeneous markups and technology, and not heterogeneous costs of capital, could explain the bulk of the variation in revenue products of capital among Compustat firms.

In order to keep US firms, I drop all firms that have a different country code than "US". Furthermore, I only keep firms that report in US dollars. As is standard, I drop the financial industry (sic codes 60–67), utilities (sic code 49), mining (sic codes 10–14) and the miscellaneous category. I drop observations which have a missing or negative value for sales, input expenditures, capital or one of the controls. Furthermore, to drop outliers, for each year I trim the top and bottom percentile of the variables that show up in the regression (i.e., the sales-input ratio, the capital-input ratio and the controls). Given that there are only few firms in the first years of the data, I only consider the years 1961–2017. This leaves me with 16,759 unique firms and 183,778 firm-year observations.

For output I use the variable "SALE" which represents sales net of returned sales and excludes sales taxes and excise taxes.²⁹ For expenditures on inputs, I use the variable "XOPR" which covers operating expenses. Operating expenses consist of costs of goods sold ("COGS") and selling, general and administrative expenses ("SGA"). Costs of goods sold refer to expenses allocated to production while selling, general and administrative expenses are expenses not directly related to product production such as marketing and R&D. Operating expenses include all employee benefits including (corporate) profit sharing and provisions for bonuses and stock options. This is desired as bonuses are part of the costs needed to attract and retain employees, and are therefore not economic profits. Note that I do not require COGS and SGA, nor their components, to be perfect substitutes with each other.

There has been some discussion in the literature on whether SGA should be considered as a variable input into production or as an overhead cost, i.e., should the production function be $F(K, COGS, SGA)$ or $F(K, COGS)$, respectively (see De Loecker et al., 2018; Karabarbounis and Neiman, 2018; Traina, 2018). In my main specification, I treat SGA as a variable input. As a robustness check, I include COGS and SGA separately in the regression to let the data speak of to what extent SGA is a factor of production. I find that the resulting profit share is similar to my baseline estimate.

For the capital stock I use property, plant and equipment net of depreciation ("PPENT"). Since this variable is recorded as the end of period stock, I use the

²⁹0.2% of the observations correspond to a book year that is not equal to 12 months. For these observations I multiply all flow values by 12 divided by the number of actual months covered.

lagged value. Furthermore, this variable represents tangible capital, but some firms also include intangible capital in this variable.³⁰ Therefore, to treat firms in the same way, I add intangibles to the capital stock.³¹ Intangibles refer to externally purchased intangibles while internally developed intangibles do not appear on the balance sheet. Therefore, as a robustness check, I capitalize R&D expenditures and add this to the capital stock. I find that this leads to similar results. To calculate the depreciation rate (which is one of the controls) I use data on the reported value of depreciation and the capital stock (see Appendix E for details).

I run regressions within 2-digit industries, where I use the sic classifier to identify an industry.³² In order to estimate the evolution of the cost of capital, I run the regression on a 5-year moving window. This is because there is not enough data to run the regression year-by-year. Within each 5-year period I include time-trends for the marginal product of capital and the markup. The estimate refers to the middle of a 5-year period. I only consider industries with more than 100 observations in a 5-year period and I cluster standard errors at the firm level.

Model Validation

To get a sense of whether the linear relationship between the sales-input and the capital-input ratio as predicted by the model is a reasonable description of the data, Figure 26 in Appendix H shows the binned scatter plot for the years 1975, 1995 and 2015. Firms across all industries are included. The relationship is close to linear, but there is some concavity. This concavity is consistent with firms with a high capital-input ratio having a lower cost of capital. However, when both the sales-input and capital-input ratio are residualized using the controls, the relationship becomes linear. This suggests that the controls are rich enough to capture the heterogeneity in the cost of capital.

Finally, to validate the cost of capital estimator, I regress the industry-year cost of capital on the average depreciation rate within that same industry-year. Industries with a higher depreciation rate are expected to have a higher cost of capital, all else equal. And since firms can deduct depreciation from their taxable income, increasing depreciation by one increases the cost of capital by somewhat less than one. Table 7 in Appendix G shows that, as expected, when the observed

³⁰The Compustat description for intangibles states that it excludes "Intangibles included in property, plant, and equipment by the company".

³¹To calculate intangibles, I exclude goodwill, such that intangibles refer to "other intangibles" and if intangibles are missing I set them equal to zero. Including goodwill in the capital stock does not alter the results, see Figure 27b in Appendix H.

³²The reason for using the sic classifier and not the naics classifier is that the naics classifier has many missing values for the first years of the data set while the sic classifier has no missing values. For instance, in the 1970s, about 40% of observations have a missing naics.

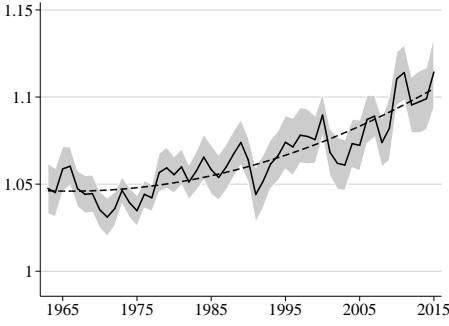


Figure 1: Price-average cost ratio $\frac{\mu}{\phi}$

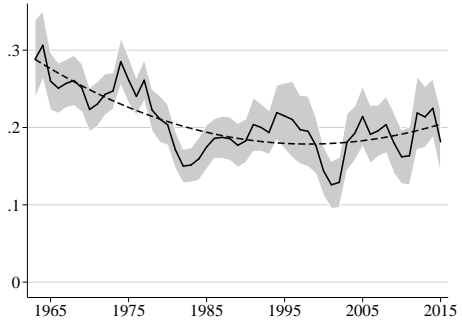


Figure 2: Cost of capital R

depreciation rate increases by one, the estimated cost of capital increases by around 0.8. Note that although I use the depreciation rate as a control when estimating the cost of capital, the regression does not use the *average* depreciation rate. The regression only uses information about the heterogeneity in depreciation rates across firms. Therefore, it is by no means hard-wired that a positive relationship between the industry cost of capital and the depreciation rate emerges.³³

IV Results

Figure 1 shows the evolution of the markup divided by the returns to scale (i.e., the price-average cost ratio). In all graphs, the shaded areas represent 95% confidence intervals based on standard errors clustered at the firm level, and a quadratic best fit is shown in addition. Initially, the price-average cost ratio decreases, but increases after 1970 from just below 1.05 to about 1.1 in 2015. Figure 2 shows that the cost of capital has been decreasing over time by around five to ten percentage points from being around 25–30% in the 1960s to being around 20% toward the end of the sample period. The decline in the cost of capital mainly takes place during the late 1970s and the early 1980s. From the mid-1980s onward the cost of capital is weakly increasing. One explanation for the latter increase is the rise in the depreciation rate due to the composition of capital shifting from structures toward equipment and intangibles.

Figure 3 shows the capital share, which is capital compensation as a share of sales, over time, where capital compensation is the cost of capital, R , times the

³³Furthermore, in Appendix D I show how the resulting profit share at the industry level correlates with market capitalization corrected for asset holdings. I find that when economic profits increase by 1 dollar, market capitalization increases by around 30 dollar. This is consistent with a stationary equilibrium in which the asset-corrected market capitalization equals the discounted sum of future profits and the discount rate is 0.97.

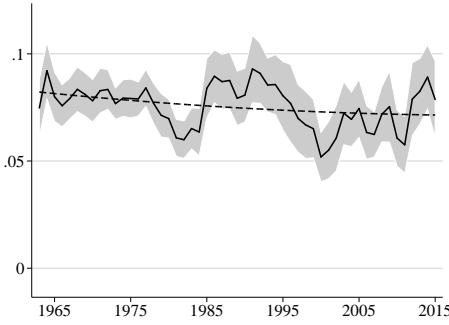


Figure 3: Capital share

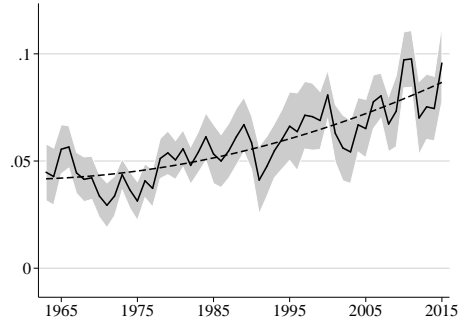


Figure 4: Profit share

capital stock. The capital share has been ranging between 5% and 10%, and has been going up and down during this period.³⁴ Overall, there is a negative trend and the largest part of the decline occurred during the 1990s. Figure 4 shows profits as a share of sales, which is sales minus operating expenses and capital compensation, divided by sales. Profits were just below 5% of sales during the beginning of the sample and increase gradually starting from the 1970s. By 2015 the profit share has almost doubled and is close to 10%. The evolution of the profit share is similar to the evolution of the markup divided by the returns to scale. This is not surprising since the markup divided by the returns to scale is the price divided by the average cost. Therefore, the profit share equals 1 minus the inverse of $\frac{p}{\phi}$. This holds at the industry level but not at the economy level due to a differential weighting. The price-average cost ratio shown in Figure 1 is the price-average cost ratio weighted by input expenditures across industries and the profit share shown in Figure 4 is the sales-weighted average of the industry level profit share (i.e., total profits divided by total sales).

The results so far show profits as a share of sales because there is no good value added data in Compustat. However, for about 10% of the sample, there is data on the wage bill, making it possible to construct value added in the following way. Subtracting the wage bill from operating expenses gives materials, and subtracting materials from sales in turn gives value added. Furthermore, I subtract R&D expenditure from materials in order to mimic the capitalization of intellectual property products by the Bureau of Economic Analysis in the national income and products account. Koh et al. (2016) show that the treatment of intellectual property products is important for the evolution of the labor share. To increase the sample I impute value added for the remaining observations using information

³⁴It is important to keep in mind that the capital share only reflects the cost of owned capital. Rented capital is not part of the capital stock, but is recorded as an operating expense.

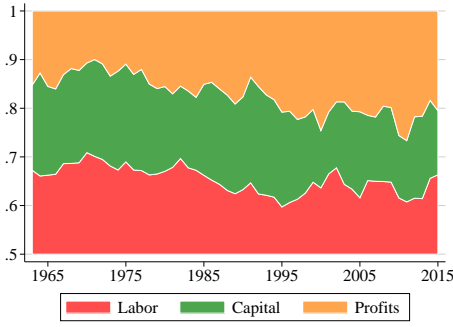


Figure 5: Income shares of value added

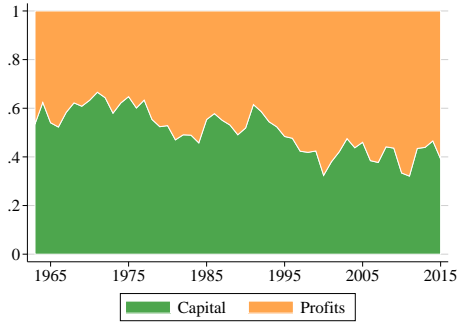


Figure 6: Profit and capital share of gross operating surplus

on the number of employees. For observations that have data on both the wage bill and the number of employees, I calculate the wage per employee. Then, for each industry-year, I calculate the average wage and multiply this by firm-level employment to impute the wage bill for those observations that do not report the wage bill.³⁵ Using this imputed wage bill I impute value added. This gives value added for around 95% of the observations in my sample. I assume that the remaining 5% of observations have the same ratio between sales and value added.

Value added is about 40–45% of sales, and adjusting the factor income shares for this time-varying ratio gives the factor shares as a share of value added, which are shown in Figure 5. The labor share is relatively constant at around 68% during the first 20 years of the sample after which it declines to around 60% in 1995. After 1995, the labor share moves up and down. Thus, overall, there is a negative trend in the labor share. The labor share looks similar in both level and trend to what is found for the corporate sector (Karabarbounis and Neiman, 2014).

One important question in the literature is what has caused the decline in the labor share. Broadly speaking, there are two possible explanations. The first explanation is that due to technological change, the role of labor in production has deteriorated over time (e.g., due to the price of capital changing over time). If this were the case the fall in the labor share would coincide with an increase in the capital share. The other explanation is that firms have increased their market power over time, either in the product or labor market. If this were the case, the fall in the labor share would coincide with an increase in the profit share. Figure 5 also shows the capital and profit share as a share of value added. Coinciding with the fall in the labor share, the profit share increases, while the capital share

³⁵For industry-years for which there is no information on the wage, I use the average wage across all firms within a year. When calculating the average wage, I trim the top and bottom percentiles for each year in order to have a measure that is not sensitive to outliers.

declines. In addition, Figure 6 shows how the residual of the labor share (i.e., gross operating surplus) is distributed between payments to capital and economic profits. At the beginning of the sample period, gross operating surplus is split between around 40% profits and 60% capital compensation. Toward the end of the sample, this split has reversed and around 60% of gross operating surplus is profits. All this together implies that the fall in the labor share is due to firms increasing their market power. In Section VI, I estimate the markdown in the labor market together with the cost of capital and find that the markdown has been roughly constant over time. Therefore, the fall in the labor share can be attributed to an increase in markups in the output market. A fall in the returns to scale would be another explanation for the fall in the labor and capital share and the rise in the profit share, but in Section VIII, I instead find an increasing trend in the returns to scale.

Note that profits as a share of value added have been hovering between 10 and 20%. These numbers are large but turn out to be similar in magnitude (but not necessarily trend) to what is found using the required rate of return approach with aggregate data of the corporate sector (Barkai and Benzell, 2018).

V Comparison with the Required Rate of Return Approach

Barkai (2017) and Barkai and Benzell (2018) are recent papers studying the evolution of the profit share. They follow Hall and Jorgensen (1967) when estimating the cost of capital and subsequently use this to estimate the capital share and the profit share. Hall and Jorgensen (1967) derive the following formula for the user cost of capital, or the required rate of return on capital:

$$R = \left[\frac{D}{D+E} i^D (1 - \tau) + \frac{E}{D+E} i^E + \delta - \mathbb{E}(\pi^k) \right] \frac{1 - itc - z\tau}{1 - \tau}, \quad (19)$$

where D is the value of debt, E is the value equity, i^D is the debt cost of capital, i^E is the equity cost of capital, δ is the depreciation rate, $\mathbb{E}(\pi^k)$ is the expected inflation of the capital good, itc is the investment tax credit, τ is the corporate income marginal tax rate and z is the net present value of depreciation allowances. Equation (19) follows from the condition that the price of a capital good equals the discounted value of all future services derived from this capital good. Note that in the original formulation of Hall and Jorgensen (1967), the discount rate shows up, which is here approximated by $\frac{D}{D+E} i^D (1 - \tau) + \frac{E}{D+E} i^E$.

It is challenging to measure the components in equation (19), of which measuring the equity cost of capital is probably the most challenging. In the literature, several variables have been used to proxy i (some authors do not distinguish between i^D and i^E). For instance, Basu and Fernald (1997) proxy i by the dividend yield on the S&P 500. Barkai (2017) constructs i^D as the yield on AAA corporate bonds and proxies i^E as the yield on the ten-year US treasury plus a 5% equity risk premium.³⁶

The main difference between my approach and the required rate of return approach is that my approach estimates R directly using microdata without needing to specify what the individual components are. For instance, the cost of capital I estimate has the differential tax treatment of expenditure on capital versus expenditure on other inputs incorporated in it without needing to explicitly model the tax code. Corporations can deduct all their expenditure on intermediate inputs from their taxable income, but only part of their expenditure on capital. The cost of capital that I estimate includes corporate income tax levied on the share of capital costs that cannot be deducted, but does not include the tax on economic profits, which is included in the profit share. To see this, denote the corporate income tax rate by τ and the share of capital expenses that can be deducted by ρ . Then the cost of capital is $R = (1 + \tau(1 - \rho))\tilde{R}$ where \tilde{R} denotes the cost of capital excluding tax. If all capital expenses can be subtracted then $R = \tilde{R}$, while if no capital expenses can be subtracted the cost of capital increases by 100τ percent. When deciding on its inputs, the firm takes into account taxation and hence it is the cost of capital including taxes that is equal to the ratio of the marginal product and the markup, which is what is estimated. The advantage of this approach is that ρ does not need to be explicitly known. This is different from the required rate of return approach, in which the fraction of capital costs that can be deducted for tax purposes is needed to be known in order to estimate the cost of capital.

In equation (2) I have defined total capital costs as being linear in capital. Although this definition seems innocuous at first, it limits the way in which capital adjustment costs can be present. Capital adjustment costs are allowed in my framework as long as the total capital costs (including the adjustment cost) are linear in K .³⁷ When all other capital costs are linear in capital, then convex capital adjustment costs violate the assumption of capital costs being linear in capital. When capital costs are not linear, an additional wedge shows up in the first-order condition with respect to capital, making it impossible to estimate the cost of capital without having information on this wedge. This is a limitation of my

³⁶Gutiérrez and Philippon (2017) use the required rate of return approach as well and estimate the equity risk premium using firm-level analyst forecasts.

³⁷The adjustment cost is considered as part of the cost of capital and is included in the estimated R .

approach, as it is for the required rate of return approach (i.e., equation (19) is derived assuming that there are no adjustment costs). However, for the data I am using, David and Venkateswaran (2019) find that adjustment costs only play a limited role in explaining heterogeneity in capital revenue productivities.

Furthermore, to estimate total capital compensation, the cost of capital needs to be multiplied by the capital stock. It is not only difficult to measure the cost of one unit of capital, it is also difficult to measure the capital stock (Corrado et al., 2009).³⁸ The capital composition has shifted from relatively easy to measure structures to relatively difficult to measure equipment and intangibles. Therefore, it is likely that the growth of the capital stock is understated, thus leading to an underestimate of the growth of the capital share and therefore an overestimate of the growth of the profit share when using the required rate of return approach. For the method I propose, systematic mismeasurement of the capital stock has no effect on estimated capital compensation, where systematic mismeasurement of the capital stock is defined as each firm's unobserved capital stock being a common (within industry) fraction of observed capital. This fraction is allowed to vary over time. Systematic mismeasurement does not affect the estimate of total capital compensation because I use the capital stock to estimate the cost of capital and a too low measured capital stock leads to a too high estimated cost of capital R . However, when this too high cost of capital is multiplied with the too low capital stock to obtain total capital compensation, these errors cancel out.

Proposition 3 (Unobserved capital). *If unobserved capital is proportional to observed capital then mismeasurement of capital does not affect the estimated capital compensation $R \cdot P^K K$ and the price-average cost ratio $\frac{\hat{\mu}}{\phi}$.*

Proof. Suppose that unobserved capital is $\iota - 1$ times the observed stock of capital such that the true capital stock (observed plus unobserved) is ι times the observed capital stock K . Following from (18), the estimator for the price-average cost ratio is

$$\begin{aligned} \frac{\hat{\mu}}{\phi} &= \left(\frac{PY}{P^X X} \right) - \frac{\text{cov} \left(\frac{P_i Y_i}{P_i^X X_i}, \frac{P_i^K K_i}{P_i^X X_i} \right)}{\text{var} \left(\frac{P_i^K K_i}{P_i^X X_i} \right)} \left(\frac{P^K K}{P^X X} \right) \\ &= \left(\frac{PY}{P^X X} \right) - \frac{\text{cov} \left(\frac{P_i Y_i}{P_i^X X_i}, \frac{\iota P_i^K K_i}{P_i^X X_i} \right)}{\text{var} \left(\frac{\iota P_i^K K_i}{P_i^X X_i} \right)} \left(\frac{\iota P^K K}{P^X X} \right), \end{aligned}$$

³⁸There is a strand of the literature that attributes the amount of income not accounted for by labor or observed capital to unobserved intangibles, such as brand names, patents and organizational capital (see, e.g., Hall, 2001; Atkeson and Kehoe, 2005; McGrattan and Prescott, 2005; Corrado et al., 2009; Eisefeldt and Papanikolaou, 2013).

where the last equality holds because ι drops out. Thus, the estimate of the price-average cost ratio is independent of ι . The estimator for capital compensation is

$$\begin{aligned} \widehat{R \cdot P^K K} &= \frac{\widehat{VMPK/\phi}}{\widehat{\mu/\phi}} P^K K = \frac{1}{\widehat{\mu/\phi}} \frac{\text{cov} \left(\frac{P_i Y_i}{P_i^X X_i}, \frac{P_i^K K_i}{P_i^X X_i} \right)}{\text{var} \left(\frac{P_i^K K_i}{P_i^X X_i} \right)} P^K K \\ &= \frac{1}{\widehat{\mu/\phi}} \frac{\text{cov} \left(\frac{P_i Y_i}{P_i^X X_i}, \frac{\iota P_i^K K_i}{P_i^X X_i} \right)}{\text{var} \left(\frac{\iota P_i^K K_i}{P_i^X X_i} \right)} \iota P^K K. \end{aligned}$$

Also here ι drops out from the last equation and hence, estimated capital compensation is not affected by systematic mismeasurement of the capital stock. \square

That systematic mismeasurement of the capital stock does not affect my estimate of the capital share and the profit share does not mean that idiosyncratic measurement error does not lead to a bias. Idiosyncratic measurement error leads to attenuation bias and, therefore, the estimated marginal product of capital and capital share would be downward biased. If measurement error has grown over time, then my method will underestimate the growth in the capital share and therefore overestimate the growth in the profit share. If idiosyncratic measurement error has fallen over time, the opposite is true. It is reassuring that using variation across industries gives similar results as, presumably, there is much less measurement error in the industry average than in the firm specific capital-input ratio.

To put my results on the cost of capital and the profit share into perspective, I calculate the profit share in a similar way as Barkai and Benzell (2018) do. I take for the debt cost of capital the yield on AAA corporate bonds and for the equity cost of capital the yield on ten-year government bonds plus a 5% equity premium.³⁹ The value of equity is the number of common shares outstanding times the closing price and the value of debt equals total liabilities. I assume that the expected inflation of the investment good equals the realized inflation and get the inflation rate from the BEA.⁴⁰ Data on the corporate tax rate is taken from Jorgenson and Yun (1991) for the period 1963-1986, and from the OECD tax database for the period thereafter.⁴¹ The present value of capital consumption allowances for tax

³⁹Both yields come from FRED. Specifically, the series Moody's Seasoned Aaa Corporate Bond Yield (AAA) and 10-Year Treasury Constant Maturity Rate (DGS10).

⁴⁰To calculate the inflation rate of the investment good, first calculate the price of the investment good by dividing investment in private nonresidential fixed assets (fixed assets accounts table 4.7, line 37) by the chain-type quantity index (fixed assets accounts table 4.8, line 37) and then calculate the inflation rate of this series. This is done for the corporate non-financial sector.

⁴¹See <http://www.oecd.org/tax/tax-policy/tax-database.htm#oecdctit>.

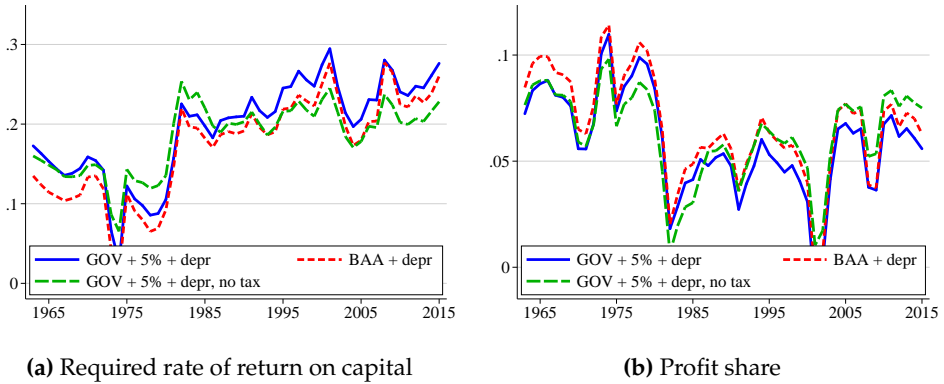


Figure 7: Cost of capital and profit share using the required rate of return approach

purposes is taken from Jorgenson and Sullivan (1981) for the period 1963-1980, and from the tax foundation for the period thereafter.⁴² Values of the investment tax credit come from Jorgenson and Sullivan (1981) for the period 1962-1980. I assume that the investment tax credit stays at its 1980 values in the years thereafter until it is abolished in 1986.

The first panel of Figure 7 shows the evolution of the required rate of return over time (the series denoted by "GOV + 5% + depr"). The cost of capital is decreasing over time during the 1960s and 1970s from around 15% to around 10%. Around 1980 the cost of capital suddenly increases to around 20%, after which there is a weak positive time trend. For the period after 1985 the cost of capital according to this approach is very similar to the cost of capital I estimate, both in level and trend. However, during the 1960s and 1970s I estimate the cost of capital to be more than ten percentage points larger than the required rate of return approach does.⁴³ The required rate of return estimate of the cost of capital is robust to two alternative specifications of the interest rate and tax treatment of capital. One alternative is to take both the debt and equity cost of capital as the yield on corporate BBB bonds.⁴⁴ This leads to almost identical results (this is the series denoted by "BAA + depr"). Another alternative is to not take the differential tax treatment of capital into account and to approximate the interest rate by the government yield plus 5%, plus taking depreciation and inflation of the capital good into account (the series denoted by "GOV + 5% + depr, no tax"). The added

⁴²See <https://taxfoundation.org/oecd-capital-allowances-three-assets-1979-2012/>. I impute the values post-2012 with the 2012 value.

⁴³One potential explanation for why I estimate such a high cost of capital during the 1960s and 1970s is that inflation was high during that period which is likely to lead to an underestimate of the *nominal* capital stock as reported on the balance sheet. This would lead to an overestimate of the cost of capital, R , but does not lead to a biased estimate of the capital share and profit share by Proposition 3.

⁴⁴The yield is the Moody's Seasoned Baa Corporate Bond Yield (BAA) from FRED.

value of the latter approach is that the data on the tax credit and depreciation allowances come from aggregate data and publicly listed firm might differ in their ability to lower their tax burden. This approach gives similar results as well.

The second panel of Figure 7 shows the resulting profit share for the three different specifications. Profits are around 8% of sales during the 1960s and 1970s due to the low cost of capital and then drop quickly to less than 5% in the early 1980s. Afterwards, they steadily rise but to a lower level than what was the case in the 1960s. The profit share resulting from specifying a required rate of return on capital is broadly similar to my estimate for the period after 1985, both in terms of level and trend. However, the results are very different before 1985. The required rate of return approach estimates a high profit share during the 1960s and 1970s and then it drops quickly around 1980. That is, over the entire sample period, the profit share follows a U-shape whereas my estimate of the profit share gives that the profit share is increasing throughout. The reason why the profit share drops so quickly around 1980 according to the required rate of return approach is that this was the Volcker period with high interest rates leading to a high required rate of return as indicated by the first panel of Figure 7. However, a high federal fund rate does not necessarily imply a high cost of capital for firms.

A similar pattern for the profit share emerges when the analysis is done for the entire corporate sector and not restricted to Compustat firms (Barkai and Benzell, 2018). That the profit share is so high during the 1960s and 1970s (while the labor share was also high) has led to concerns whether the required rate of return approach measures the cost of capital accurately (Karabarbounis and Neiman, 2018). For instance, the risk premium might not have been constant over time, or the depreciation rate and inflation expectations are mismeasured, and the level of this measurement error has changed over time. Especially mismeasurement of inflation expectations seems to be a potent explanation for the discrepancy between the two estimates of the profit share, as inflation was volatile during the 1960s, 1970s and early 1980s, coinciding with the period at which the two methods give different results. Instead, both methods give very similar estimates for the period after 1985 when inflation is no longer volatile.

The discrepancy between the two methods is unlikely to be driven by heterogeneity in the cost of capital being larger in the 1960s and 1970s than in the later period. Heterogeneous costs of capital lead to a negative bias in the estimate of the cost of capital, while I estimate the cost of capital to be larger than the required rate of return approach does during the 1960s and 1970s.

Finally, Karabarbounis and Neiman (2018) note that at the business cycle frequency there is a negative correlation between the real interest rate and the

Table 1: Relationship real interest rate and profit share

	Baseline			Required rate of return approach		
	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta_1 \pi_t$	$\Delta_2 \pi_t$	$\Delta_3 \pi_t$	$\Delta_1 \pi_t$	$\Delta_2 \pi_t$	$\Delta_3 \pi_t$
$\Delta_1 r_t$	0.164* (0.0754)			-0.291 (0.182)		
$\Delta_2 r_t$		0.106 (0.0867)			-0.485* (0.183)	
$\Delta_3 r_t$			0.0332 (0.0933)			-0.674*** (0.184)
Obs	49	48	47	49	48	47
R^2	0.042	0.022	0.003	0.044	0.115	0.230

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

profit share according to the required rate of return approach. It is not clear why such a negative relationship would emerge in the real world, and this negative relationship could be purely mechanically as, all else equal, a fall in the real interest rate lowers the required rate of return and therefore leads to a lower estimated capital share and a higher estimated profit share.⁴⁵ Table 1 shows the correlation between the change in the real interest rate and the change in the profit share, according to my method and the required rate of return approach, where the real interest rate refers to the yield on ten-year government bonds minus inflation expectations obtained from the Michigan surveys of consumers. Δ_j refers to the j -th difference, $\Delta_j \pi_t = \pi_t - \pi_{t-j}$, where π refers to the profit share, and likewise for the real rate r . The first three columns show that my method gives a weakly positive relationship between the interest rate and the profit share. The relationship is positive when the one-year difference is considered but disappears for the two- and three-year differences. The last three columns replicate the result that the required rate of return approach leads to a negative relationship between the interest rate and the profit share. My results suggest that this negative relationship obtained by the required rate of return approach is mechanical.

⁴⁵ Although it is, of course, possible that there is a negative relationship between the real interest rate and the profit share. For instance, Liu et al. (2019) provide a possible mechanism through which such a relationship would emerge.

VI Robustness

In this section I perform several robustness checks of my estimate of the profit share.

Using Variation Across Industries

In order to identify the marginal product of capital and the markup (both scaled by the returns to scale ϕ), variation in the capital-input ratio is needed. This variation should be due to variation other than differences in the cost of capital, such as differences in technology (e.g., differences in factor-biased productivities) or input prices across firms. In the above I have dealt with this by controlling for differences in the cost of capital and I have considered variation across firms within an industry. Here, I use variation across industries.⁴⁶ Arguably, there is considerable variation in technology across industries. However, it is not clear whether the variation in the cost of capital is smaller across industries than within industries. In a model with capital mobility and risk differing across industries but being the same across firms within an industry, the cost of capital is the same for firms within an industry but differs across industries. If, in contrast, heterogeneity in the cost of capital predominantly comes from firms differing to the extent in which they are financially constrained then the cost of capital differs across firms within an industry, but the cost of capital is on average the same across industries if the fraction of financially constrained firms is the same across industries. Given that it is likely that in the Compustat data a majority of the firms is not financially constrained, using within industry variation seems the most appropriate, and that is why that is my baseline estimate. However, there is still some value in considering the across industry estimate as it would be a concern if the results were very different.

To use variation across industries, I instrument the capital-input ratio by the

⁴⁶Yet, another alternative would be to estimate a model with firm fixed effects. That is, running a regression of the form

$$\frac{P_{it}Y_{it}}{P_{it}^X X_{it}} = [vmpr_t + vmpr_i] \frac{P_{it}^K K_{it}}{P_{it}^X X_{it}} + \mu_t + \mu_i + \nu_{it},$$

where $vmpr_t$ is the time fixed effect for the marginal product of capital, $vmpr_i$ the firm fixed effect for the marginal product, μ_t the time fixed effect for the markup and μ_i the firm fixed effect for the markup (all scaled by the returns to scale parameter ϕ) and ν_{it} is the error term. Then, the average marginal product at time t (which is the object of interest) would be $vmpr_t$ plus the average of $vmpr_i$. However, it is problematic that in order to identify $vmpr_i$, time variation in the capital-input ratio is needed. To ensure that the fixed effect is really a fixed effect, this regression should be run on a short panel (i.e., a firm with a high cost of capital in the 1980s does not need to have a high cost of capital in the 2000s), but a short panel would not give enough time variation in the capital-input ratio, thus making fixed effects inappropriate.

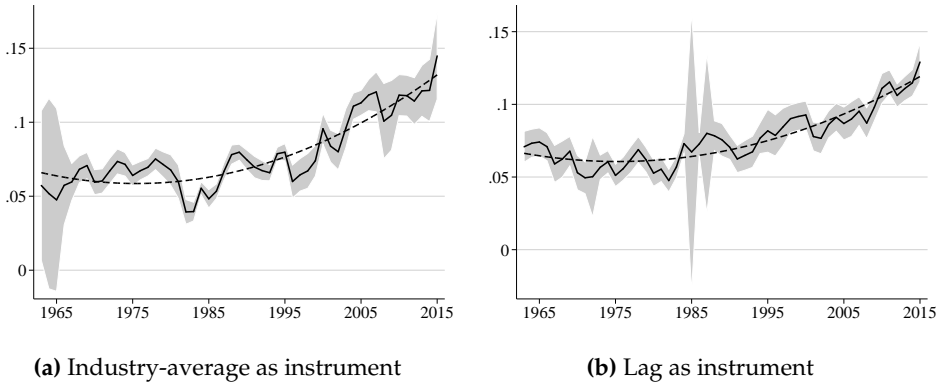


Figure 8: Profit share - robustness

average capital-input ratio within an industry-year.⁴⁷ As already noted above, it is unlikely that the exclusion restriction, which is that the average cost of capital is the same across industries, holds, but there is still some value in using this instrument as robustness. To deal with heterogeneity in risk across industries I use as a control variation in the growth rate of sales across firms within an industry. I also include the same controls as in the baseline, except for the capital-input interaction term. Given that these controls are multiplied by the capital-input ratio, which is endogenous, I use as additional instruments the controls times the average capital-input ratio. Likewise, I interact the instrument with the time trend for the marginal product of capital. It should be clear that I run this regression on the economy level. I only consider industries that in a given year have more than twenty observations since otherwise the average capital-input ratio is noisy. Standard errors are clustered at the industry level.

Figure 8a shows that the resulting profit share is similar in level and trend to the baseline profit share. However, the increase in the profit share starts somewhat later than in the baseline and the increase in profits is larger than in the baseline. An alternative to using the industry capital-input ratio as an instrument is to combine all firms across industries together and run the OLS with controls. Figure 27c in Appendix H shows that combining all industries leads to very similar results as when using the variation within industry, although also in this case the rise in the profit share does emerge later than in the baseline.

⁴⁷Wooldridge (1997, 2003) shows that under reasonable assumptions, 2SLS is a consistent estimator for the slope, even when the endogenous regressor appears in the error term. However, the intercept, i.e. the markup, is not identified since the expectation of the error term does not equal zero. I deal with this using appropriate weights, see above.

Productivity and Markup Shocks, and Adjustment Frictions

One reason why marginal products of capital might be correlated with the capital-input ratio is unexpected productivity and markup shocks, combined with capital not being able to adjust instantaneously. In response to a productivity or markup shock, a firm will change its use of intermediate inputs, affecting the capital-input ratio. That capital cannot adjust immediately will also entail that the shock affects the marginal product of capital. Hence, idiosyncratic productivity and markup shocks lead to heterogeneity in the marginal product of capital and to a correlation between the marginal product and the capital-input ratio. That unanticipated markup shocks lead to a bias is different from Lemma 1 in which anticipated heterogeneity of the markup was discussed and not shocks to the markup.

Note that the correlation between the marginal product and the capital-input ratio due to shocks arises from capital not being flexibly adjustable while other inputs are flexibly adjustable. If it were also to take time to adjust these other inputs, then such a correlation would not necessarily emerge. Adjustment frictions could be the time it takes to hire and train new workers, and the costs associated with changing existing contracts with suppliers or finding new suppliers.

An instrument that deals with unanticipated productivity and markup shocks is the lagged capital-input ratio. The lagged capital-input ratio is not affected by the present shock and therefore captures other variation in the capital-input ratio.⁴⁸ Figure 8b shows that using the lagged capital-input ratio as an instrument leads to a similar estimate of the profit share as the baseline. The level is slightly higher than in the baseline, but the change over time is very similar as in the baseline. That the level is slightly higher is because no controls are included in this regression.

The instrument does not work well when capital adjustment costs are high. This is because in that case the lagged capital-input ratio also captures dispersion in the marginal product of capital in the previous period which will persist into the current period due to the high capital adjustment cost. To deal with adjustment costs, I take five-years sums of the data. That is, I take sales as the sum of sales over five consecutive years and the same for capital and the other inputs.⁴⁹ The logic behind this exercise is that over a longer time horizon, adjustment costs matter less. The resulting profit share, displayed in Figure 27d in Appendix H, is very similar to the baseline.

⁴⁸Appendix I uses simulations to show the validity of this instrument.

⁴⁹If a firm has fewer than five years of consecutive data, I also take the four or three year sum. I drop firms that have less than three years of consecutive data for this exercise.

Overhead Costs

In my main specification, I have treated selling, general and administrative expenses (SGA) as a factor of production. In the literature there is a debate on whether this cost is a factor of production or a fixed overhead cost (De Loecker et al., 2018; Karabarbounis and Neiman, 2018; Traina, 2018). When SGA is a fixed cost it does not enter Euler's theorem, but still needs to be subtracted from sales when estimating economic profits. As a robustness, I include COGS and SGA separately in modified Euler's theorem,

$$P_i Y_i = \frac{VMPK_i}{\phi P_i^K} P_i^K K_i + \frac{\mu_i}{\phi} COGS_i + \frac{\mu_i^{sga}}{\phi} SGA_i. \quad (20)$$

The term μ_i^{sga} denotes the marginal product of selling, general and administrative expenses divided by the price of one unit of these expenses. When these costs are a fixed cost, the marginal product will be zero and the coefficient on SGA will be zero. When these costs are a factor of production, the coefficient will be positive.

Problematic with running OLS on equation (20) is that if SGA is an overhead cost, then due to selection it is likely to be correlated with the markup, μ_i , leading to a bias. The reason is the following. In the data, there is heterogeneity in SGA . If a firm with a high overhead cost does not exit the market, this means that this firm is either larger or charges a higher markup in order to recover these fixed costs. If this firm is larger because it is more productive, this would not lead to a bias because productivity is not an omitted variable in equation (20). However, when this firm charges a higher markup this leads to a bias because markup heterogeneity is part of the error term. Therefore, I add demeaned SGA as a control for both the markup and the marginal product of capital (and I divide the left-hand side and the right-hand side by $COGS$).

Figure 9 shows the results. The first panel shows the marginal product of direct labor and materials divided by the price of one unit of direct labor and material, and the returns to scale (i.e., $\frac{\mu}{\phi}$). This ratio has been increasing from just above 1 to just below 1.15. This is broadly similar to the baseline price-average cost ratio, but the increase is somewhat faster here. The second panel shows the marginal product of selling, general and administrative expenses divided by the price of one unit of these expenses. This value is around one, and therefore these inputs are not a fixed cost, but a factor of production. $\frac{\mu_i}{\phi}$ and $\frac{\mu_i^{sga}}{\phi}$ are of a similar magnitude on average but show different trends with $\frac{\mu_i^{sga}}{\phi}$ decreasing over time. I do not have a good explanation for this difference in trends. The third panel shows that the cost of capital is around 20%, and behaves similarly as in the baseline. Finally,

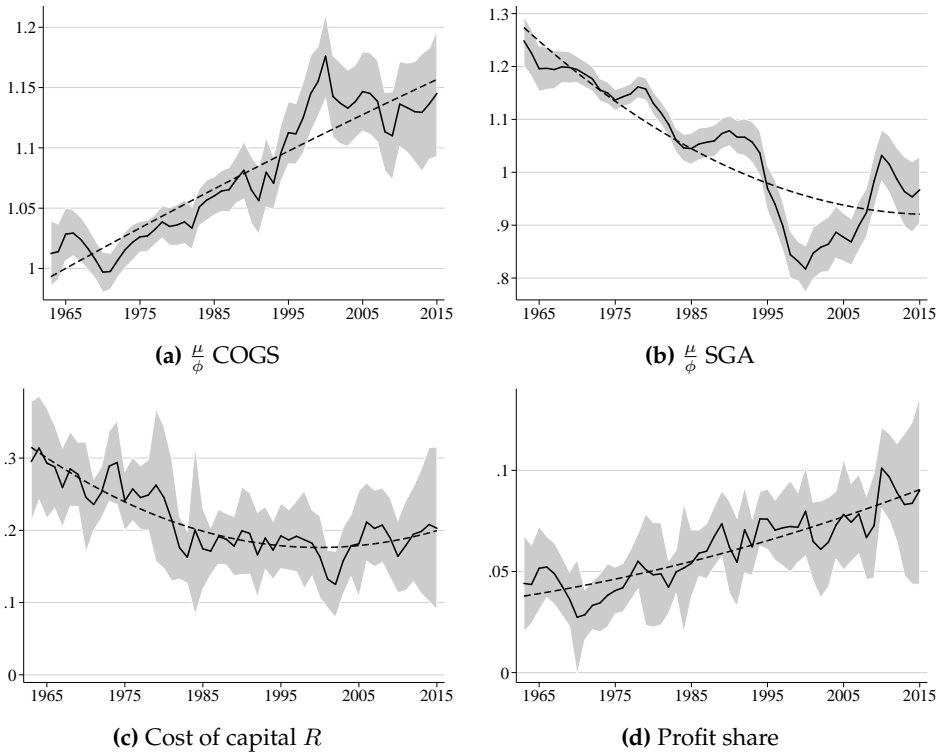


Figure 9: Results when COGS and SGA enter the regression separately

the fourth panel shows that profits under this specification show a similar pattern over time as the baseline does.

Another way of testing whether SGA is a factor of production or an overhead cost is to omit SGA expenditure from modified Euler's theorem. Figure 28 in Appendix H shows that this regression results in a cost of capital that is on average around 50%. This seems unreasonably large and the reason for this large cost of capital is a high estimated marginal product of capital. This suggests that a factor of production is omitted from this regression, and that SGA is indeed a factor of production and should be included in the regression.

The above has given a statistical reason for why SGA is a factor of production. An economic reason for why SGA is a factor of production is the following. Compustat comprises larger firms and one reason why these firms differ in their sales is that they differ in the number of product lines. If each product line needs its own sales and administrative team, and if we consider the production function as the number of varieties that can be produced, then SGA is a factor of production.

Finally, the baseline is my preferred estimate over including COGS and SGA

separately in the regression. The reason is that firms have freedom in allocating expenses to either costs of goods sold or selling, general and administrative expenses, which likely leads to measurement error. Indeed, there are firms that report zero expenses for one of these items.

Capitalizing R&D Expenditure

The capital stock includes intangibles, but these are only intangibles that are externally purchased and do not include internally developed intangibles. If firms within an industry differ in the extent to which they purchase versus internally develop intangibles, this will lead to idiosyncratic measurement error of the capital stock and therefore to attenuation bias of the estimated marginal product of capital and hence an overestimate of the profit share.⁵⁰ Since the share of intangibles has been increasing over time, not measuring intangibles correctly would lead to an overestimate of the rise in the profit share. Therefore, I here include internally developed intangibles in the capital stock, by capitalizing R&D expenditure, and show that this does not alter the conclusions. Another argument for capitalizing R&D is that fully expensing R&D might not be appropriate as R&D expenditure is risky and leads to future services that have to be discounted appropriately.

To calculate the internally developed intangible capital stock, $K_{t+1}^{R\&D}$, I capitalize R&D as follows

$$K_{t+1}^{R\&D} = (1 - \delta^{R\&D})K_t^{R\&D} + R\&D_t,$$

where $R\&D_t$ is the expenditure on R&D. I set the depreciation rate, $\delta^{R\&D}$, equal to 15% (Griliches and Mairesse, 1984).⁵¹ Given that most of these firms have been founded several years before they enter the Compustat data, it is unlikely that they enter with an intangible capital stock of zero. Therefore, I use the first observation of R&D expenditure divided by $\delta^{R\&D}$ as starting value for the intangible capital stock. The underlying assumption is that the firm is in steady state. Furthermore, I set R&D expenditure equal to zero if it is negative in a period, and I interpolate when R&D expenditure is missing. Finally, I subtract R&D expenditure from operating expenses to avoid double counting.

Doing so leads to the estimated profit share displayed in Figure 10. The

⁵⁰ As discussed in Section V, when unobserved intangibles are the same share of the capital stock for each firm then this will not lead to a bias.

⁵¹ The BEA also calculates industry-specific depreciation rates for R&D capital (Li and Hall, 2018). However, these are based on a model that assumes perfect competition and are therefore not applicable to the current setting.

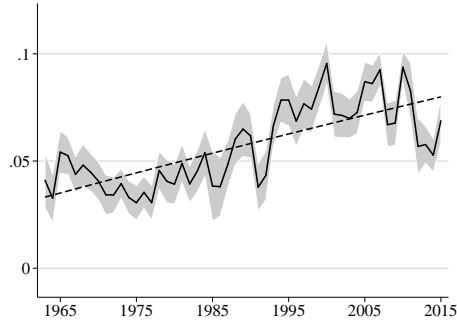


Figure 10: Profit share when R&D is capitalized

estimated profit share is broadly similar to the baseline. The profit share declines in the first decade of the sample and increases afterwards. Different from the baseline is that the profit share levels off after 1995 whereas it continues to increase in the baseline.

Markdown Labor Market

So far I have assumed that firms take input prices as given, although the input price is allowed to differ across observations (e.g., in the case of labor due to compensating differentials or differences in the skill composition across firms). The reason why input prices are allowed to differ across observations while the cost of capital is not, is that total expenditure on inputs is observed while capital compensation is not. However, it is not allowed that there is a wedge other than the markup showing up in the first-order conditions. The reason is that the first-order conditions with respect to inputs X^m are used to infer the markup and then the markup is used to back out the cost of capital. If there were to be an additional wedge, then only the markup times this additional wedge could be estimated. A wedge would, for example, show up when firms have monopsony (or oligopsony) power in an input market.⁵²

Different from the baseline, I here split total input expenditure into expenditure on labor and into expenditure on intermediate inputs. This makes it possible to introduce a wedge in the labor market (which I will also refer to as the markdown), while assuming that the firm takes all other input prices as given. Still, no assumptions about the output market are needed. Splitting up input expenditure allows me to study whether the rise in the profit share is robust to allowing for a

⁵²For a discussion of monopsony power in the labor market see Boal and Ransom (1997). Recent papers that study labor market power are Dube et al. (2020), Azar et al. (2017) and Berger et al. (2019).

markdown in the labor market, and if so, whether the rise in the profit share is due to changing markups in the output market or changing markdowns in the labor market. Allowing for a wedge in the first-order condition of labor leads to the following first-order condition with respect to labor

$$VMPL_{it} = \mu_{it}\mu_{it}^l w_{it}, \quad (21)$$

where μ_{it}^l is the markdown following from labor market power and w_{it} is the wage of firm i at time t . The first-order conditions for capital and the other inputs are as in equations (5) and (6), respectively.

Doing the same manipulations of Euler's theorem (7) as before and plugging in first-order condition (21) gives

$$P_i Y_i = \frac{VMPK_i}{\phi P_i^K} P_i^K K_i + \frac{\mu_i}{\phi} \sum_{m \in M'} P_i^{X^m} X_i^m + \frac{\mu_i \mu_i^l}{\phi} w_i L_i, \quad (22)$$

where M' refers to inputs other than labor (and capital). Hence, regressing output on the capital stock, intermediate input expenditure and labor expenditure gives as coefficients the marginal product of capital, the markup and the markup times the markdown, respectively, all scaled by the returns to scale. To obtain the estimate of the markdown, μ^l , I divide the estimate of $\frac{\mu \cdot \mu^l}{\phi}$ by the estimate of $\frac{\mu}{\phi}$. In order to obtain the estimate of the cost of capital, I divide the estimate of the marginal product of capital by the estimate of the markup.

When running the regression, I divide both sides of equation (22) by total expenditure on intermediate inputs to ensure that markup heterogeneity is not correlated with the regressors. The markdown μ_i^l is correlated with the choice of relative inputs as it does not show up in the first-order conditions other than the first-order condition with respect to labor. Berger et al. (2019) show that the markdown is correlated with firm size in terms of the wage bill. Therefore, I control for markdown heterogeneity using the wage bill. I assume that there is no further heterogeneity in the markdown than what is explained by the wage bill.

Finally, in order to have the error term being equal to zero in expectation, the marginal product of capital has to be weighted by capital and $\frac{\mu \cdot \mu^l}{\phi}$ has to be weighted by the wage bill. Therefore, I run two regressions. In the first regression I weight by capital and obtain the marginal product of capital and in the second regression I weight by the wage bill and obtain $\frac{\mu \cdot \mu^l}{\phi}$. Using equation (22) I use these estimates to obtain the average markup $\frac{\mu}{\phi}$ weighted by expenditure.

Figure 11 shows the results. The first panel shows that the resulting profit share is very similar to the baseline profit share. The second panel shows the markdown

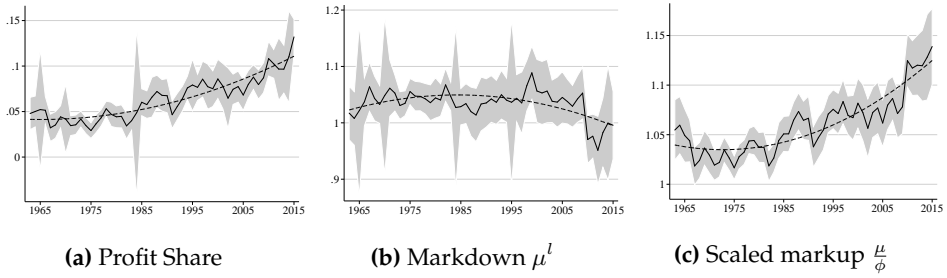


Figure 11: Results with markdown in labor market

which is estimated to be around 1.05 for most of the period. This means that firms have some labor market power. There is no trend in the markdown except for a decline in the last five years. This means that labor market power cannot explain the rise in profits.⁵³ Instead, as the third panel shows, the rise in the profit share is due to an increase in the markup in the product market divided by the returns to scale.

My results on the markdown are very similar to what is found by Berger et al. (2019). They consider changes in local labor market concentration combined with estimates of the degree of labor substitutability. They find that, all else equal, changes in labor market power would have increased the labor share by 2.89 percentage points between 1976 and 2014. With a labor share of around 60%, this means that the markdown has decreased by around 5% which is what I find as well.

Using Economy-wide Weights

Compustat comprises a non-representative subset of all firms, namely mainly publicly listed firms. One way in which the Compustat sample differs from the aggregate economy is in the sectoral composition. For instance, a relatively large number of manufacturing firms are in Compustat. If manufacturing tends to have an above-average profit share, this will lead to a positive bias in the profit share that I estimate. Moreover, a change in the industry composition within Compustat over time could be driving the increase in the profit share I find. If firms in industries with a low profit share tended to be public at the beginning of the sample while industries with a high profit share tended to have relatively more public firms toward the end of the sample, this would mechanically lead to

⁵³This is under the assumption that firms do not have market power in other input markets. My measure of μ^l is the markdown in the labor market relative to the markdown in other input markets. It could potentially be the case that the markdown in all input markets has increased instead of a constant markdown in the labor market.

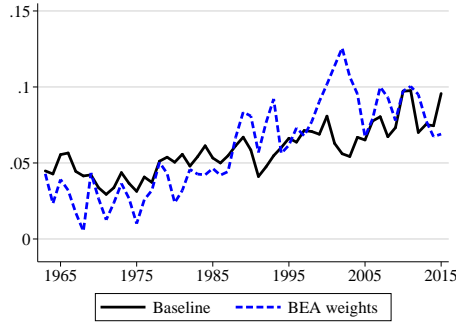


Figure 12: Profit share using BEA value added as sector weights

an increase in the profit share in my data while the economy-wide profit share might not have increased at all.

To correct for this, I calculate the average profit share across industries using the economy-wide size of each industry as weights. This entails that the average profit share is not influenced by the industry composition of Compustat firms. As weights I use industry value added obtained from the Bureau of Economic Analysis (BEA). Figure 12 shows that the profit share using BEA weights is similar to the baseline profit share estimate. If anything, the profit share has risen faster in the weighted sample. Therefore, the rise in the profit share I find is not due to the sectoral composition of publicly listed firms.

Another way of dealing with the representativeness of the data is to estimate the model using economy-wide industry level data from the BEA. This has an advantage that all firms are included in these industry aggregates. Appendix F shows that this yields very similar estimates as the baseline.

VII Heterogeneity

In Section IV I have found that the aggregate profit share has been increasing. Here, I explore how the distribution of the profit share has changed. Is the rise in the profit share due to few firms becoming more profitable or did the entire profit distribution shift to the right?

To calculate the profit share at the firm level, I use the firm-specific cost of capital obtained using the controls in the regression.⁵⁴ Figure 13 shows the unweighted kernel density of the profit share across firms in 1970, 1985, 2000 and 2015. Two observations stand out. The distribution has shifted to the right and the

⁵⁴None of the results in this section change qualitatively when the firm cost of capital is taken as the industry-year average cost of capital.

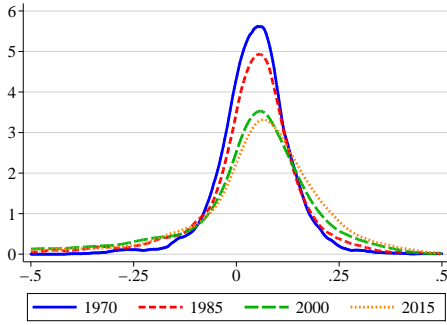


Figure 13: Kernel density profit share

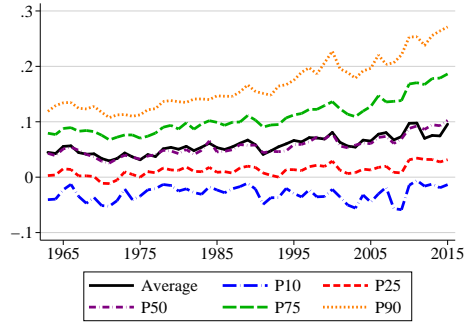


Figure 14: Percentiles profit share distribution

dispersion has increased over time. Figure 14 displays the evolution of some of the percentiles of the profit share distribution over time and tells a very similar message. Here, observations are weighted by sales as the aggregate profit share equals the sales-weighted average profit share. Weighting by sales makes the median comparable to the average. The 25th, 50th, 75th and 90th percentile have all been increasing since 1980 onward. Only the 10th percentile experienced a constant profit share, which was just below 0%. The median profit share coincides almost perfectly with the average profit share. Therefore, the rise in the profit share is due to the entire distribution of firms becoming more profitable, except for the bottom decile. It is true that the profit share of the 90th percentile grows faster than the 75th percentile in absolute terms, which in turn grows faster in absolute terms than the median etc. However, in relative terms the growth rates between the different percentiles are very similar. That the rise in the profit share is present across most of the distribution is different from what De Loecker et al. (2018) find for the markup.⁵⁵ They find that the markup only increased for the upper half of the distribution, and thus that the median has not increased.

One potential reason for this difference in results for the median, between my approach and De Loecker et al. (2018), is that De Loecker et al. (2018) do not account for heterogeneity in factor-augmenting technology. Raval (2019b) and Demirer (2019) show that not accounting for heterogeneity in factor-augmenting technology leads to a biased estimate of the average markup when using the method of De Loecker et al. (2018). Here, I show that not properly accounting for factor-augmenting technology affects the average and median markup in a different way. Rewriting the first-order condition (6) gives that the markup equals

⁵⁵I also find for the markup (which I estimate in Section VIII) that the median has risen at the same rate as the average.

the output elasticity of an input X^m , $\theta_i^{X^m}$, times the inverse revenue share of that input, $\frac{P_i Y_i}{P_i^{X^m} X^m}$. De Loecker et al. (2018) use this equation when estimating the markup, but assume that the output elasticity is the same across firms. This leads to the following relationship between the firm-level markup as estimated by De Loecker et al. (2018), $\hat{\mu}_i^{DLEU}$, and the true markup, μ_i ,

$$\hat{\mu}_i^{DLEU} = \hat{\theta}^{X^m} \frac{P_i Y_i}{P_i^{X^m} X^m} = \frac{\hat{\theta}^{X^m}}{\theta_i^{X^m}} \mu_i, \quad (23)$$

where $\hat{\theta}^{X^m}$ is the estimated output elasticity which is common across firms.⁵⁶ Thus, a firm for which De Loecker et al. (2018) estimate a high markup might in fact be a firm with a low output elasticity. This means that the distribution of the estimated markup is not necessarily informative of the true distribution when there is heterogeneity in the output elasticity. As an example, suppose that $\theta_i^{X^m}$ and μ_i are independent log-normally distributed. This implies that the larger the heterogeneity in the output elasticity, the larger the ratio of the estimated average markup to the median markup relative to the true ratio between the average and median markup,

$$\frac{E(\hat{\mu}^{DLEU})/med(\hat{\mu}^{DLEU})}{E(\mu)/med(\mu)} = e^{\frac{1}{2}\sigma_\theta^2}, \quad (24)$$

where σ_θ^2 is the variance of the logarithm of the output elasticity. The reason is that an increase in the variance of $\theta_i^{X^m}$, all else equal, leads to an increase in the average $\hat{\mu}_i^{DLEU}$ (by Jensen's inequality), while the median is not affected as much.⁵⁷

Thus, an increase in heterogeneity in technology across firms over time can explain why De Loecker et al. (2018) find a larger increase in the average markup compared to the median markup while I do not.

Profits-Size Relationship

Next, I study whether bigger firms are more profitable. Figure 15 shows the point estimates and 95% confidence intervals obtained from regressing the profit share, as a percentage, on log sales. This regression is run by year, and industry fixed effects are included. During the 1960s and early 1970s the profit share was not correlated with firm size. However, after 1975 this started to change and bigger firms became more profitable than smaller firms. The relationship between size

⁵⁶De Loecker et al. (2018) also estimate a translog production function in which they allow for limited heterogeneity in the output elasticity across firms, but they do not allow fully for heterogeneity in output elasticities as they do not allow for heterogeneity in factor-augmenting technology.

⁵⁷I am assuming here that $\hat{\theta}^{X^m}$ is consistently estimated.

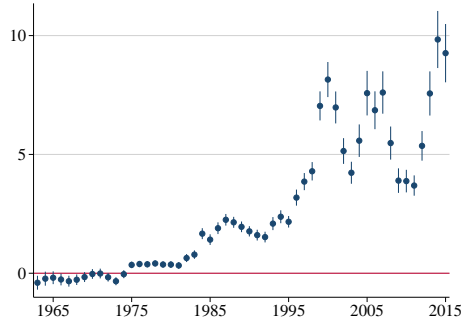


Figure 15: Effect of log firm size on the profit share (p.p.)

Notes: The point estimates denote the percentage point increase in the profit share when sales increases by 1 log point. Industry fixed effects are included. To remove outliers I trim the top and bottom fifth percentile of the dependent and independent variable. The vertical lines denote 95% confidence intervals based on robust standard errors.

and profitability has become stronger over time but has also fluctuated. For instance, the great recession was relatively bad for large firms in terms of profitability. However, after the great recession, the larger firms quickly increased their profitability relative to smaller firms. In 2015 one log point more in sales would imply a ten percentage point higher profit share.

Autor et al. (2017) use Census data and focus on the period after 1981 and show that the fall in the labor share is due to reallocation toward "superstar firms" which are larger firms with a low labor share. Based on existing evidence it is not clear whether these firms have a low labor share because they have a high capital share or because they have a high profit share. Figure 15 shows that these larger firms have a higher profit share (at least in the later period) and that this relationship has become stronger over time.⁵⁸

That larger firms have a larger profit share, and that this relationship has been changing over time, combined with publicly listed firms tending to be larger than private firms, makes it problematic to generalize the estimated profit share among Compustat firms to the entire economy. Extrapolating to private firms, the positive relationship between size and profits toward the end of the sample period suggests that during this period the economy-wide profit share was lower than the profit share among publicly listed firm, at least based on size-dependent selection. On the other hand, during the beginning of the sample period, selection based on firm size does not seem to have led to as much of a discrepancy between the profit share among publicly listed firms and the economy-wide profit share. Therefore,

⁵⁸In contrast, Gutiérrez and Philippon (2019) find that "superstars" have not become more productive over time.

the economy-wide profit share might not have risen as quickly as the rise in the profit share I estimate.

To correct for selection based on firm size, I weight observations according to the sectoral size distribution. The Census Business Dynamic Statistics provide information on the number of firms by binned number of employees from 1977 until 2014 for each industry. In Compustat there are very few observations with fewer than 10 employees whereas the majority of the universe of firms falls into this size class. Including this size class puts a great deal of weight on only a few observations and hence leads to an unreliable estimate of the profit share. Therefore, I calculate the reweighted profit share for different cutoffs. That is, I provide six different estimates. One in which firms of all sizes are included, one in which only the firms with at least 10 employees are included, up to an estimate in which I only consider firms with at least 100 employees. The latter is the least sensitive to having a small sample size for some bins, but its drawback is that it omits firms with fewer than 100 employees. Furthermore, to minimize the reliance on outliers, I only take into account firms with a profit share between minus 100% and plus 50%.

Figure 16 shows the resulting estimates of the profit share corrected for selection based on firm size. As was anticipated based on the minimal effect of firm size on profits, correcting for size-based selection does not affect the profit share around 1980. Because the Census only provides data on the size distribution from 1977 until 2014, it is not possible to make the correction before 1977. First consider the estimate only based on firms with more than 100 employees. The profit share is similar in magnitude as the baseline estimate. The weighting here is based on seven different bins for employment (i.e., from 100 to 250 employees, 250 to 500 employees etc. up to a bin for firms with more than 10,000 employees). That the profit share is increasing after this size-correction means that the rise in the profit share is not only due to the super large firms, but is also present among smaller firms with more than 100 employees.

Each additional estimate of the size-corrected profit share includes one additional bin. All size corrected profit shares based on firms with at least 10 employees display a positive trend in the profit share, although attenuated compared to the baseline estimate. The more small firms there are included in this estimation, the lower the estimated rise in the profit share, but also the less reliable the estimate becomes as there are only few firms in the Compustat data that are small. The size-corrected profit share including all firms is very noisy, but there seems to be also an upward trend for this series. This series is so noisy because for some industry-years there are only a handful of firms with less than 10 employees.

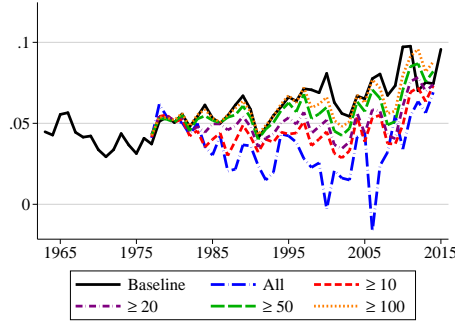


Figure 16: Profit share corrected for selection based on firm size

Notes: The colored lines show the profit share weighted to match the economy-wide size distribution, considering all firms and only firms with more than 10, 20, 50 or 100 employees, respectively.

Moreover, publicly listed small firms are most likely not representative for the typical small firm.

Another way to deal with a changing selection of firms over time is to estimate the profit share for the largest firms only, as most of the largest firms in the economy are represented in Compustat. Figure 27e in Appendix H shows that the profit share of the 500 largest firms in each year is slightly higher than the profit share when taking into account all firms, but that the evolution over time is very similar as the baseline estimate.

Decomposition

An obvious question is whether the rise in the profit share is due to an increase in profits within industries or due to reallocation from industries with low profits to industries with high profits. In order to answer this question, I decompose the change in the profit share into a within industry component and a reallocation component (Haltiwanger, 1997),⁵⁹

$$\begin{aligned} \Delta\pi_t = & \underbrace{\sum_{Continue} m_{jt-1} \Delta\pi_{jt}}_{\text{Within}} + \underbrace{\sum_{Continue} (\pi_{jt-1} - \pi_{t-1}) \Delta m_{jt}}_{\text{Between}} + \underbrace{\sum_{Continue} \Delta m_{jt} \Delta\pi_{jt}}_{\text{Covariance}} \\ & + \underbrace{\sum_{Enter} m_{jt} (\pi_{jt} - \pi_{t-1}) - \sum_{Exit} m_{jt-1} (\pi_{jt-1} - \pi_{t-1})}_{\text{Net entry}}, \end{aligned} \quad (25)$$

⁵⁹Figure 29 in Appendix H shows the profit share for each industry separately.

where π_t is the aggregate profit share, π_{jt} is the profit share of industry j and m_{jt} is the market share of industry j (i.e., sales of industry j divided by total sales). Δ represents one-year differences. The decomposition consists of five terms. The first term represents the within industry effect and measures the increase in profits when the market share of industries would not have changed. The remaining terms together represent the reallocation effect. The second term represents the between industry component which reflects changes in the market share weighted by the deviation of industry profits from aggregate profits. The third term represents the covariance between the change in the market share and the change in profits. This term is positive when growing industries experience a faster increase in their profit share than shrinking industries. The first three terms refer to industries that are in the data for two consecutive periods. There are two additional terms which refer to entering and exiting industries. This is atypical for an industry-level decomposition as all industries are normally present throughout the entire sample. However, when I estimate the profit share I only include industries that have more than 100 observations in a five-year period. Therefore, there are some industries that are present some years, but not in the other years. The last two terms correct for this entry and exit of industries. The fourth term corresponds to the contribution of entering industries by taking the deviation of profits from average profits in the previous year weighted by the industry size of entering industries. The last term represents the effect of exiting industries by taking the deviation of profits in the previous period from aggregate profits weighted by the previous industry size of exiting industries. Net entry can be considered as part of reallocation here, since industries that enter are industries that grow in terms of the number of firms while industries that exit are industries that are shrinking.

Figure 17a shows the cumulative effect of each term in equation (25). The solid black line shows the cumulative change in the profit share with respect to 1963. The majority of the rise in the profit share is attributed to the within-industry effect, although the within effect is negative during the early decades. That the within effect is the dominant factor means that the rise in the profit share is due to industries experiencing an increase in their profits. All other terms are close to zero.

A similar decomposition can also be done at the firm level, simply by replacing the j -subscript in equation (25) with an i -subscript. Doing the decomposition at the firm level allows to decompose the change in the profit share between within-firm effects and reallocation between firms.

Figure 17b shows the resulting decomposition. The within-firm effect falls

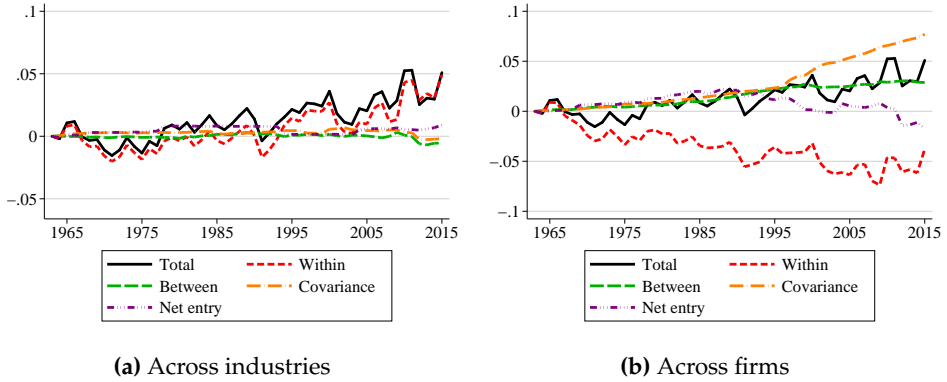


Figure 17: Decomposition profit share

during the entire time period. The accumulated between effect is positive, but constant during the last twenty years. This means that during the period 1963–1995 the market share of profitable firms has been increasing, while during the period 1995–2015 the market share of profitable firms has been constant. The between effect is relatively small compared to the covariance effect. Had it been for the covariance effect alone, the profit share would have been increasing by eight percentage points. The positive covariance effect means that firms that increased their market share also increased their profit share. One explanation for the dominant role played by the covariance term is creative destruction. Some firms are able to increase their market share while increasing their profitability. This suggests that these firms have a dominant marketing strategy or developed a new variety of which they are the sole supplier and which is highly in demand.

Finally, net entry does not contribute permanently to the change in profits in the long run, but its contribution does fluctuate over time. The term "net entry" is misleading as these are firms that enter and exit the data set and do not necessarily enter or exit the market. For instance, an exiting firm could be a firm that goes private or that is acquired. An entering firm could be a firm going from private to public or be the result of a merger. That the reallocation effect (the between plus covariance effect) dominates the within effect is also found by De Loecker et al. (2018) and Baqaee and Farhi (2020) for markups, and Autor et al. (2017) for the labor share.

How can the accumulated within-firm effect be negative while the distribution of the profit share has shifted to the right? The reason is that the decomposition considers changes over time. There have been some firms that were initially large (i.e., a large m_{it-1}) and experienced a decline in their profit share, while some

small firms experienced a (faster) increase in the profit share. This leads to a negative within effect, but can shift the distribution to the right. Furthermore, this did not lead to a drop in the profit share because the firms with a growing profit share experienced an improvement in their market share as well, which also explains the positive covariance term.

Entry Costs and the Life Cycle of Profits

If profits and markups increased, why did firm entry not increase in response? Instead, during the last decades, firm entry in the US has declined (Decker et al., 2014). Does this mean that entry costs have increased over time and that the economic profits I measure reflect only entry costs? I argue here that it is unlikely that the rise in the profit share is entirely driven by a rise in entry costs because the life cycle pattern of profits has changed over time.

Profits tend to be back-loaded over the life cycle of the firm. A young firm hardly makes any profits while older firms do. In a simple model of entry, a potential entrant decides to enter the market when the discounted sum of profits exceeds the entry costs. Suppose that we are in a stationary equilibrium such that firms of all ages are observed. Then total profits observed at any point in time exceed the discounted sum of profits and therefore exceed entry costs (Atkeson and Kehoe, 2005). The discrepancy between observed profits and the present value of profits increases in the extent to which profits are back-loaded, because profits that occur in the distant future are discounted more. Thus, one explanation for the rise in profits is that profits have become more back-loaded over time. All else equal, this would lower the present value of profits, and to make the entry condition continue to hold, aggregate profits need to increase in response. If this were indeed the case, the profit share could have increased without an increase in entry costs.

To test this hypothesis, I obtain the firm founding age from Field and Karpoff (2002) and Loughran and Ritter (2004) (of firms that went public between 1975 and 2018) and calculate average deflated profits by age and see how those have changed over time. Since the founding age is only available for a subset of firms, there is not enough data for the years before 1980 to calculate the life-cycle pattern of profits. Figure 18 shows the resulting life-cycle pattern of profits for firms active in the 1980s, 1990s, 2000s and 2010s, respectively, where I have grouped firms in 5-year age bins. During all four time periods, profits are increasing in firm age, and this relationship is clearly the strongest for the period after 2000. After 2000, firms younger than ten years old were making less or as much profits as firms before 2000, while older firms were making more profits than in the 1980s

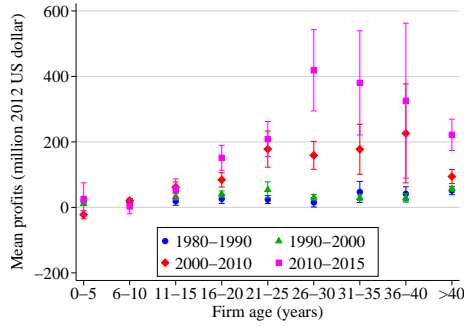


Figure 18: Profits over the firm life cycle

Notes: The point estimates denote average profits in millions of dollars for each age bin (in 2012 dollars, deflated by the GDP deflator), where age is denoted in years. The vertical lines denote 95% confidence intervals.

and 1990s. A potential entrant cares mostly about profits during the beginning of the firm's life because of discounting. That profits were low for the youngest firms after 2000 could, depending on the discount factor, potentially explain why entry has not gone up despite an increase in aggregate profits; the present value of profits has not increased as much as aggregate profits.

VIII Returns to Scale and the Markup Have Been Increasing

In order to estimate the profit share, the returns to scale do not need to be known. However, knowledge of the returns to scale is helpful in understanding the rise in the profit share. When there are decreasing returns to scale, profits are positive even in a competitive market where prices equal marginal costs. The reason is that with decreasing returns to scale average costs are lower than marginal costs (Lucas, 1978). Therefore, a decline in the returns to scale over time could be an explanation for the increase in profits. In this section, I estimate the evolution of the returns to scale and find that the returns to scale are close to one and increasing over time, and therefore do not account for the rise in profits (in a direct sense). Furthermore, having an estimate of the returns to scale allows me to estimate the markup, which I find to be increasing over time.

I estimate the returns to scale using cost shares as in Syverson (2004). A first-order approximation of a production function gives

$$y_{it} = \beta_t^0 + \phi_t v_{it} + \omega_{it}, \quad (26)$$

where small scale letters denote logs; y is the log of real output and ω is the log of productivity. v is the cost-share weighted sum of logged inputs,

$$v_{it} = \alpha_{it}^k k_{it} + \alpha_{it}^{\text{cogs}} \text{cogs}_{it} + \alpha_{it}^{\text{xsga}} \text{sga}_{it},$$

where k is the log of real capital, cogs is the log of real costs of goods sold and sga is the log of real selling, general and administrative expenses. $\alpha_i^k = \frac{R_i P_i^K K_i}{R_i P_i^K K_i + \text{input expenditure}_i}$ is the cost share of capital, and likewise for α_i^{cogs} and α_i^{sga} . Equation (26) states that regressing the log of real sales on v and a constant gives the returns to scale ϕ_t , that are now assumed to be the same across firms within an industry.

I calculate the cost shares using the baseline estimated (firm-specific) cost of capital. Note that the cost shares and therefore technology are allowed to differ across firms. As opposed to what was needed to estimate the cost of capital, this regression requires all variables to be in real terms. No firm-specific deflator is available and therefore I deflate sales, cogs and xsga using the non-financial corporate business sector price index and deflate capital using the non-financial corporate investment price index.⁶⁰

Note that here I consider costs of goods sold and selling, general and administrative expenses separately, because combining them into operating expenses would require that they are perfect substitutes with each other. For my main analysis estimating the profit share, considering operating expenses as one expense does not require both to be perfect substitutes. Moreover, here I need to assume that different types of costs of goods sold (such as materials and labor) are perfect substitutes with each others, whereas this assumption is not needed for estimating the profit share.

Unobserved productivity shows up in equation (26), making OLS biased as input choices (i.e., v) are correlated with productivity. This is also different from my main analysis as productivity does not show up in Euler's theorem. I deal with omitted productivity in equation (26) by using a proxy approach, which is a standard IO technique (see, e.g., Akerberg et al. (2015)). I first purge measurement error by regressing log sales on a third order polynomial of the logs of capital, cogs and sga , for each industry j . Denote the resulting estimate of log sales free of measurement error by ζ_{ijt} . Suppose that productivity follows an AR(1): $\omega_{ijt} = \rho_j \omega_{ijt-1} + \xi_{ijt}$, where ξ_{ijt} is an unanticipated shock to productivity. This

⁶⁰The value added price index is obtained using NIPA table 1.14 from the BEA (Gross Value Added of Domestic Corporate Business in Current Dollars and Gross Value Added of Nonfinancial Domestic Corporate Business in Current and Chained Dollars). The price index is nominal gross value added (line 17) divided by real gross value added (line 41). Both for nonfinancial corporate businesses. See footnote 40 for details on the investment good price index.

gives rise to the following moment condition to obtain the industry-year-specific returns to scale:

$$\mathbb{E}(\xi_{ijt}(\cdot) \mathcal{I}_{ijt-1}) = 0,$$

where \mathcal{I}_{ijt-1} are variables that are in the information set of firm i in industry j at time $t - 1$, and $\xi_{ijt}(\cdot)$ is a function of the parameters ϕ_j , β_j^0 and ρ_j . In order to obtain $\xi_{ijt}(\cdot)$, I first obtain productivity from $\omega_{ijt}(\phi_j, \beta_j^0) = \zeta_{ijt} - \phi_j v_{ijt} - \beta_j^0$, and then $\xi_{ijt}(\phi_j, \beta_j^0, \rho_j) = \omega_{ijt}(\phi_j, \beta_j^0) - \rho_j \omega_{ijt-1}(\phi_j, \beta_j^0)$. To identify the three parameters I use as moments the lag of the log of cogs, ζ_{ijt-1} and a constant. The lag of sga would not be a good moment condition as it includes R&D expenditure and is therefore likely to be correlated with future changes in productivity. I use a five-year moving window to estimate how the returns to scale have changed over time. To calculate the average returns to scale across industries I weight by operating expenses as later I will use the returns to scale to calculate the markup (which is weighted by operating expenses as well). Figure 19 shows that the returns to scale have been increasing over time from a level just below 1 in the 1960s to around 1.05 toward the end of the sample. Therefore, changes in the returns to scale do not explain the rising profit share (in a direct sense).⁶¹

The evolution of the returns to scale over time looks similar to how the relationship between firm size and profitability has evolved over time (see Figure 15). When there is no relationship between firm size and the profit share the returns to scale are approximately one, and when larger firms are relatively more profitable the returns to scale are larger than one. One potential explanation is that, when estimating the returns to scale, I assume that all firms charge the same price and face the same price for their inputs. If there would be price heterogeneity, this would lead to an additional error term in equation (26). Under the assumption of no heterogeneity in prices, that larger firms have a higher profit share means that these firms have a lower average cost than small firms, for instance due to the returns to scale being larger than one. Thus, in a sense, the relationship between the profits-size relationship and the returns to scale I estimate might be ‘mechanical’. However, the proxy approach deals with this unobserved price variation as long as price variation is a monotonic function of productivity (Brandt et al., 2017).

⁶¹Figure 30 in Appendix H shows the returns to scale by industry and the returns to scale have been increasing for almost all industries.

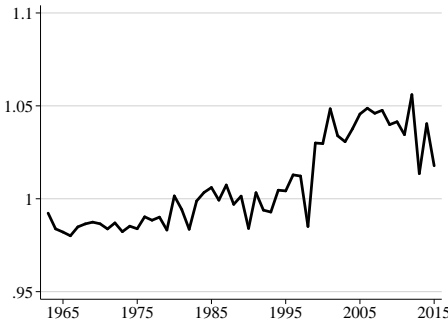


Figure 19: Returns to scale

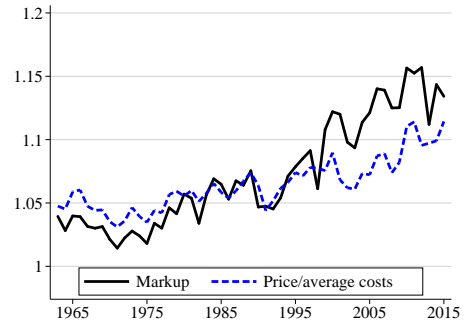


Figure 20: Markup

The markup

Given the estimate of the returns to scale, I can estimate the markup by multiplying the estimate of the price-average cost ratio with the estimate of the returns to scale. The resulting cost-weighted average markup is shown in Figure 20 which also shows for reference the estimated price-average cost ratio. Given that the returns to scale have been increasing over time, it should come as no surprise that the markup has been increasing faster than the price-average cost ratio.

The markup has been increasing from around 1.03 in the 1970s to around 1.15 toward the end of the sample. De Loecker et al. (2018) estimate the evolution of the markup for the same data using the production approach (Hall, 1988; De Loecker and Warzynski, 2012), which states that the markup equals the output elasticity of costs of goods sold times revenue divided by costs of goods sold. Their baseline estimate gives a markup that is about 30 percentage points higher than what I find, but with a broadly similar pattern over time, although the markup increases somewhat faster according to De Loecker et al. (2018). One reason why De Loecker et al. (2018) find a much larger markup than I do is that they do not include selling, general and administrative expenses in the production function, which presumably leads to a too high estimated output elasticity of costs of goods sold. Indeed, when they include selling, general and administrative expenses in the production function they find a markup that is similar in levels to the markup I find. They find a markup that increases from 1980 onward from 1 to 1.3, while I find that the markup starts to increase five to ten years earlier and increases at a lower rate. One difference in estimating the markup between their method and my method is that differences in technology (i.e., factor-biased productivity) across firms are allowed here.

Figure 31 in Appendix H shows the markup by industry and, as is the case for

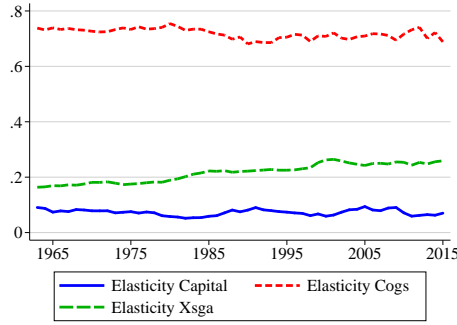


Figure 21: Output elasticities

the profit share, most industries have experienced an increase in markups over time.

Output elasticities Having estimated the returns to scale I am also able to estimate the output elasticities of the different inputs. The output elasticity of an input equals the cost share of that input times the returns to scale.⁶² Figure 21 shows the resulting output elasticities. The elasticity with respect to materials and direct labor is the highest and has been relatively stable over time between 0.7 and 0.75. The output elasticity with respect to capital is the lowest at around 0.08. The output elasticity with respect to selling, general and administrative inputs has been increasing, from around 0.17 to around 0.25. This suggests that the importance of inputs such as R&D and marketing has increased over time. However, this could also be an artifact of the freedom firms have in deciding whether to put an expense under costs of goods sold or under selling, general and administrative expenses (Karabarbounis and Neiman, 2018). If firms today were to classify a larger class of expenses as selling, general and administrative expenses than they used to do, this would mechanically lead to an increase in the estimated corresponding output elasticity.

IX Industry Concentration

Over the last decades, industry concentration in the United States has increased (Autor et al., 2017).⁶³ This could be a sign that a decline in competition has led to a

⁶²To see this, note that the output elasticity with respect to an input X is $E_X = \frac{\partial Y}{\partial X} \frac{X}{Y}$. Plugging in the first-order condition with respect to X and multiplying and dividing by the output price gives $E_X = \mu \frac{P^X X}{PY}$, and further noting that the markup equals the returns to scale times output divided by costs gives $E_X = \phi \frac{P^X X}{\text{costs}}$.

⁶³Rossi-Hansberg et al. (2019) show that local concentration has gone down instead.

fall in the labor share. However, trade costs have fallen rapidly which has increased the geographical size of markets and therefore an increase in concentration does not necessarily imply a fall in competition when customers nowadays have the option to get a product from a more distant producer. Autor et al. (2017) and Barkai (2017) find a negative industry-level correlation between changes in domestic concentration and the labor share. However, the complement of the labor share is not the profit share and therefore this does not necessarily imply that increases in concentration are associated with increases in profits or markups. Here, I study whether changes in industry concentration are correlated with changes in markups.

To study the effect of industry concentration on markups I obtain data from the US Census Bureau on concentration. The census provides data on the share of sales by the 4, 8, 20 and 50 largest firms for a wide variety of industries every 5 years.⁶⁴ In order to obtain the change in industry concentration, a consistent industry classification is needed. Starting with the 1997 census, the Census Bureau switched from the Standard Industrial Classification (SIC) to the North American Industry Classification System (NAICS). Therefore, I split the Compustat sample in two parts and estimate the change in the markup between 1972 and 1995 using the SIC classifier, and estimate the change in the markup between 1997 and 2013 using the NAICS classifier.⁶⁵ Given that the estimated markup is quite erratic at the industry level, I use 5-year averages around these years. For the SIC period I use the 2-digit codes, and the industry concentration is based on the same 2-digit codes or, when not available, the sales-weighted average concentration of the underlying 3-digit industries.⁶⁶ For the NAICS classification there are not as many 2-digit industries as for the SIC classification because manufacturing has only three 2-digit NAICS industries. Therefore, I estimate the markup for manufacturing at the 3-digit level and for the other industries at the 2-digit level.⁶⁷

⁶⁴There is no concentration data for construction.

⁶⁵The SIC and NAICS classifications have changed over time as well. However, the number of (narrowly defined) industries that have been reclassified across the broad industry classifications I use is limited. That is, there were no reclassification across the two digit industries in my data going from SIC1972 to SIC1977. 31 4-digit industries are reclassified to another 2- or 3-digit industry from SIC1977 to SIC1987 out of in total more than 1000 4-digit industries. For NAICS, there were no relevant changes between 1997 and 2002, there were 11 relevant changes between 2002 and 2007 and 2 relevant changes between 2007 and 2012. Again, out of more than 1000 industries this is a relatively small number and unlikely to be driving the change in concentration over time within a broad industry.

⁶⁶The data for the 1972 and 1992 manufacturing censuses is obtained from <https://www.census.gov/econ/concentration.html>. For the other industries I obtain the 1972 censuses from scanned copies of the census books available on <https://archive.org> and <https://www.hathitrust.org>, and the 1992 census data is obtained from https://www.census.gov/prod/www/economic_census.html#min92ind.

⁶⁷Most 3-digit industries other than manufacturing do not have sufficient observations in Compustat to estimate the markup.

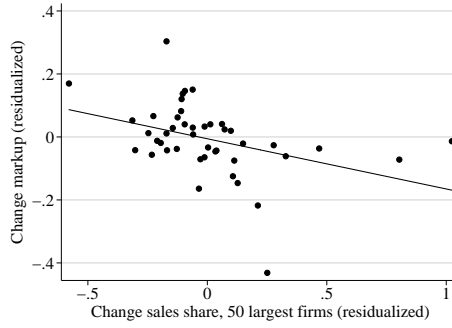


Figure 22: Relationship change in concentration and markups across industries

The concentration data is based on the same 3-digit codes for manufacturing, and for the other industries, concentration is the sales-weighted average concentration of the underlying 3-digit industries.⁶⁸ If no concentration data is available for all firms I use concentration based on firms subject to federal income tax.

Figure 22 shows the relationship between the change in concentration and the change in markups after taking out period fixed effects. In this graph, the change in concentration refers to the sales share of the 50 largest firms. A negative relationship emerges. More formally, I regress the change in the markup during the periods 1972-1995 and 1997-2013 on the change in the concentration ratios during the periods 1972-1992 and 1997-2012 including a dummy for the SIC/NAICS period. The results are shown in Table 2. Changes in concentration measured by the share of sales of the four, eight, twenty and fifty largest firms in an industry have a negative effect, significant at the 5% level. It is surprising that this effect is negative as this means that the more concentrated industries become, the lower the markup becomes.⁶⁹ When all four concentration measures are included in the regression none of them is significant because the concentration measures are highly correlated with each other.

One possible explanation for this result is that competition from abroad has increased which is not incorporated in the concentration measures. Increasing competition from abroad would lead to a fall in the markup and forces small firms to exit, leading to an increase in domestic concentration. If at the same time an increase in competition from abroad leads to more rapid automation, this would imply a negative correlation between changes in domestic concentration and the labor share as Autor et al. (2017) and Barkai (2017) find.

⁶⁸The data of the 2012 census is obtained from American fact finder while the data of the 1997 census is obtained from https://www.census.gov/prod/www/economic_census.html.

⁶⁹This result is different from Grullon et al. (2019) who find a positive relationship between concentration and profit margins but their definition of the profit margin includes capital compensation.

Table 2: Effect of industry concentration on the markup

	(1)	(2)	(3)	(4)	(5)
	$\Delta\mu_t$	$\Delta\mu_t$	$\Delta\mu_t$	$\Delta\mu_t$	$\Delta\mu_t$
$\Delta c_t^4 / c_t^4$	-0.0808* (0.0345)				0.159 (0.194)
$\Delta c_t^8 / c_t^8$		-0.103* (0.0422)			-0.0963 (0.398)
$\Delta c_t^{20} / c_t^{20}$			-0.132* (0.0562)		-0.430 (0.492)
$\Delta c_t^{50} / c_t^{50}$				-0.160* (0.0702)	0.259 (0.354)
Period fixed effects	X	X	X	X	X
Observations	44	44	44	44	44
R^2	0.118	0.144	0.159	0.149	0.180

Notes: Robust standard errors in parentheses. c_t^s refers to the share of sales by the s largest firms in an industry. Source: Census, Compustat and own calculations.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

X Alternative Applications

In this chapter I have developed a new method to estimate the cost of capital and apply it to estimate the evolution of the profit share in the long run. Is this method also useful to answer other questions, such as the effects of a policy reform on competition? The answer is yes, although it depends on the exact nature of the policy reform.

Suppose that we are interested in the effects of a trade liberalization on competition. One option is to use a method similar to De Loecker and Warzynski (2012) and estimate the markup by multiplying the output elasticity of an input with the inverse of the revenue share of that input. Challenging is that the output elasticity needs to be estimated. This is typically done by estimating a production function. Problematic is that in response to the trade reform firms might change their technology. For instance, some firms might decide to automate more tasks. If this were to lead to heterogeneity in the technology used, this would lead to a biased estimate of the output elasticity. In contrast, the method I propose does not require firms to share common technology parameters. However, my method gives biased results when the reform induces heterogeneity in the cost of capital. Naturally, which method to use also depends on what outcome we are interested in. My method estimates the profit share while the method of De Loecker and Warzynski (2012) estimates the markup. However, when the returns to scale are

not affected by the reform then my method is also informative of the change in the markup.

Finally, having estimated the evolution of the profit share over time has implications for the measured Solow residual. In Appendix K, I perform a growth accounting exercise using my estimate of the output elasticities. Allowing for imperfect competition attenuates the productivity slowdown, but by no means makes the productivity slowdown disappear.

XI Conclusions

This chapter develops a new method to estimate the cost of capital and studies how the cost of capital and the resulting profit share have evolved over time. The user cost of capital has been declining from around 25% in the 1960s to around 20% today. This implies a decline in the capital share and an increase in the profit share. This is different from the profit share that is obtained using the required rate of return approach, namely a profit share that is higher in the 1960s and 1970s than it is today. My results imply that the fall in the labor share is associated with an increase in markups and not with a change in capital intensity.

It is not the goal of this paper to get policy recommendations. In order to reach policy recommendations a much more structural approach would be needed. For instance, the rise in profits is not necessarily bad for welfare as it also affects the return to innovation. Instead, the goal of this paper is to study how the patterns of profitability have changed over time, and for this we want to impose as little structure as possible. These facts can then be used to select and discipline models of firm heterogeneity and imperfect competition such that these models can be used to analyze different policies. In this light, I think that especially my results about how the underlying distribution of profitability has changed over time are relevant. I find that the median profit share has gone up at the same rate as the average and that bigger firms have become relatively more profitable over time. Furthermore, I find that the rise in the profit share is a within industry, but across firms phenomenon. Correlating profits with firm age, I find that profits have become more back-loaded over the life cycle of the firm. This affects the value of the firm negatively and can therefore potentially explain why entry has not gone up in response to the increased profits.

References

- Acemoglu, D. and D. Autor (2011) "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, Vol. 4, pp. 1043–1171: Amsterdam: Elsevier-North.
- Acemoglu, D. and P. Restrepo (2018) "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, Vol. 108, No. 6, pp. 1488–1542.
- Akerberg, D. A., K. Caves, and G. Frazer (2015) "Identification Properties of Recent Production Function Estimators," *Econometrica*, Vol. 83, No. 6, pp. 2411–2451.
- Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019) "A Theory of Falling Growth and Rising Rents," Manuscript.
- Akcigit, U. and S. T. Ates (2019) "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory," CEPR Discussion Paper DP13668.
- Atkeson, A. and P. J. Kehoe (2005) "Modeling and Measuring Organization Capital," *Journal of Political Economy*, Vol. 113, No. 5, pp. 1026–1053.
- Autor, D. and A. Salomons (2018) "Is Automation Labor Share–Displacing? Productivity Growth, Employment, and the Labor Share," *Brookings Papers on Economic Activity*, Spring.
- Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017) "The Fall of the Labor Share and the Rise of Superstar Firms," Manuscript.
- Azar, J., I. Marinescu, and M. I. Steinbaum (2017) "Labor Market Concentration," Working Paper 24147, National Bureau of Economic Research.
- Azar, J. and X. Vives (2019) "Common Ownership and the Secular Stagnation Hypothesis," *AEA Papers and Proceedings*, Vol. 109, pp. 322–326.
- Baqee, D. R. and E. Farhi (2020) "Productivity and Misallocation in General Equilibrium," *Quarterly Journal of Economics*, Vol. 135, No. 1, pp. 105–163.
- Barkai, S. (2017) "Declining Labor and Capital Shares," Manuscript.
- Barkai, S. and S. G. Benzell (2018) "70 Years of US Corporate Profits," Manuscript.
- Basu, S. (2019) "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence," *Journal of Economic Perspectives*, Vol. 33, No. 3, pp. 3–22.

- Basu, S. and J. G. Fernald (1997) "Returns to Scale in US Production: Estimates and Implications," *Journal of Political Economy*, Vol. 105, No. 2, pp. 249–283.
- Berger, D. W., K. F. Herkenhoff, and S. Mongey (2019) "Labor Market Power," Working Paper 25719, National Bureau of Economic Research.
- Blundell, R. and S. Bond (2000) "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, Vol. 19, No. 3, pp. 321–340.
- Boal, W. M. and M. R. Ransom (1997) "Monopsony in the Labor Market," *Journal of Economic Literature*, Vol. 35, No. 1, pp. 86–112.
- Bornstein, G. (2018) "Entry and Profits in an Aging Economy: The Role of Consumer Intertia," Manuscript.
- Brandt, L., J. van Biesebroeck, L. Wang, and Y. Zhang (2017) "WTO Accession and Performance of Chinese Manufacturing Firms," *American Economic Review*, Vol. 107, No. 9, pp. 2784–2820.
- Caselli, F. and J. Feyrer (2007) "The Marginal Product of Capital," *Quarterly Journal of Economics*, Vol. 122, No. 2, pp. 535–568.
- Cette, G., L. Koehl, and T. Philippon (2019) "Labor Shares in Some Advanced Economies," Working Paper 26136, National Bureau of Economic Research.
- Corrado, C., C. Hulten, and D. Sichel (2009) "Intangible Capital and US Economic Growth," *Review of Income and Wealth*, Vol. 55, No. 3, p. 661–685.
- David, J. M. and V. Venkateswaran (2019) "The Sources of Capital Misallocation," *American Economic Review*, Vol. 109, No. 7, p. 2531–2567.
- Davis, S. J., J. Haltiwanger, R. Jarmin, and J. Miranda (2006) "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms," *NBER Macroeconomics Annual*, Vol. 21, pp. 107–180.
- De Loecker, J. (2011) "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, Vol. 79, No. 5, p. 1407–1451.
- De Loecker, J., J. Eeckhout, and G. Unger (2018) "The Rise of Market Power and the Macroeconomic Implications," Manuscript.
- De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016) "Prices, Markups, and Trade Reform," *Econometrica*, Vol. 84, No. 2, p. 445–510.

- De Loecker, J. and F. Warzynski (2012) "Markups and Firm-level Export Status," *American Economic Review*, Vol. 102, No. 6, p. 2437–2471.
- Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014) "The Role of Entrepreneurship in US Job Creation and Economic Dynamism," *Journal of Economic Perspectives*, Vol. 28, No. 3, pp. 3–24.
- Demirer, M. (2019) "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups," Manuscript.
- Doraszelski, U. and J. Jaumandreu (2018) "Measuring the Bias of Technological Change," *Journal of Political Economy*, Vol. 126, No. 3, p. 1027–1084.
- Dube, A., J. Jacobs, S. Naidu, and S. Suri (2020) "Monopsony in Online Labor Markets," *American Economic Review: Insights*, Vol. 2, No. 1, pp. 33–46.
- Edmond, C., V. Midrigan, and D. Y. Xu (2018) "How Costly are Markups," Working Paper 24800, National Bureau of Economic Research.
- Eggertsson, G. B., J. A. Robbins, and E. G. Wold (2018) "Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States," Manuscript.
- Eisfeldt, A. L. and D. Papanikolaou (2013) "Organization Capital and the Cross-Section of Expected Returns," *Journal of Finance*, Vol. 68, No. 4, pp. 1365–1406.
- Elsby, M. W. L., B. Hobijn, and A. Sahin (2013) "The Decline of the U.S. Labor Share," *Brookings Papers on Economic Activity*, Fall.
- Farhi, E. and F. Gourio (2018) "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia," *Brookings Papers on Economic Activity*, Fall.
- Field, L. C. and J. Karpoff (2002) "Takeover Defenses of IPO Firms," *Journal of Finance*, Vol. 57, No. 5, pp. 1857–1889.
- Gandhi, A., S. Navarro, and D. Rivers (2016) "On the Identification of Production Functions: How Heterogeneous is Productivity?," Manuscript.
- Gordon, R. J. (2016) *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*: Princeton University Press.
- Gouin-Bonenfant, E. (2018) "Productivity Dispersion, Between-Firm Competition, and the Labor Share," Manuscript.
- Griliches, Z. and J. Mairesse (1984) "Productivity and R&D at the Firm Level," in *R&D, Patents, and Productivity*: National Bureau of Economic Research, Inc, pp. 339–374.

- Grossman, G., E. Helpman, E. Oberfield, and T. Sampson (2017) "The Productivity Slowdown and the Declining Labor Share: A Neoclassical Exploration," Working Paper 23853, National Bureau of Economic Research.
- Grullon, G., Y. Larkin, and R. Michaely (2019) "Are US Industries Becoming More Concentrated?" *Review of Finance*, Vol. 23, No. 4, pp. 697–743.
- Gutiérrez, G. and T. Philippon (2017) "Declining Competition and Investment in the U.S.," Working Paper 23583, National Bureau of Economic Research.
- Gutiérrez, G. and T. Philippon (2019) "Fading Stars," *AEA Papers and Proceedings*, Vol. 109, pp. 312–16.
- Gutiérrez, G. and S. Piton (2019) "Revisiting the Global Decline of the (Non-Housing) Labor Share," Manuscript.
- Hall, R. E. (1988) "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, Vol. 96, No. 5, pp. 921–947.
- (2001) "The Stock Market and Capital Accumulation," *American Economic Review*, Vol. 91, No. 5, p. 1185–1202.
- (2018) "New Evidence on the Markup of Prices over Marginal Costs and the Role of Mega-Firms in the US Economy," Working Paper 24574, National Bureau of Economic Research.
- Hall, R. E. and D. W. Jorgensen (1967) "Tax Policy and Investment Behavior," *American Economic Review*, Vol. 57, No. 3, p. 391–414.
- Haltiwanger, J. C. (1997) "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence," *Federal Reserve Bank St. Louis Review*, Vol. 79, No. 3, p. 55–77.
- Hsieh, C.-T. and P. J. Klenow (2009) "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1403–1448.
- Hubmer, J. (2018) "The Race Between Preferences and Technology," Manuscript.
- Jones, C. I. (2016) "The Facts of Economic Growth," in *Handbook of Macroeconomics*, Vol. 2, pp. 3–69: Elsevier.
- Jorgenson, D. W. and M. A. Sullivan (1981) "Inflation and Corporate Capital Recovery," in *Depreciation, Inflation, and the Taxation of Income from Capital* ed. by C. R. Hulten, pp. 171–238, 311–313: Washington: The Urban Institute Press.

- Jorgenson, D. W. and K.-Y. Yun (1991) *Tax Reform and the Cost of Capital*: Oxford University Press.
- Karabarbounis, L. and B. Neiman (2014) "The Global Decline of the Labor Share," *Quarterly Journal of Economics*, Vol. 129, pp. 61–103.
- (2018) "Accounting for Factorless Income," *NBER Macroeconomics Annual*, Vol. 33, pp. 167–228.
- Katz, L. F. and K. M. Murphy (1992) "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *Quarterly Journal of Economics*, Vol. 107, No. 1, pp. 35–78.
- Kehrig, M. and N. Vincent (2018) "The Micro-Level Anatomy of the Labor Share Decline," Working Paper 25275, National Bureau of Economic Research.
- Klenow, P. J. and A. Rodríguez-Clare (1997) "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual*, Vol. 12, pp. 73–103.
- Klette, T. J. and Z. Griliches (1996) "The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics*, Vol. 11, No. 4, pp. 343–361.
- Koh, D., R. Santaaulalia-Llopis, and Y. Zheng (2016) "Labor Share Decline and Intellectual Property Products Capital," Manuscript.
- Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000) "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, Vol. 68, No. 5, pp. 1029–1053.
- Levinsohn, J. and A. Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, Vol. 70, No. 2, pp. 317–341.
- Li, W. C. Y. and B. H. Hall (2018) "Depreciation of Business R&D Capital," *Review of Income and Wealth*.
- Liu, E., A. Mian, and A. Sufi (2019) "Low Interest Rates, Market Power, and Productivity Growth," Manuscript.
- Loughran, T. and J. R. Ritter (2004) "Why Has IPO Underpricing Changed Over Time?" *Financial Management*, Vol. 33, No. 3, pp. 5–37.

- Lucas, R. E., Jr. (1978) "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, Vol. 9, No. 2, pp. 508–523.
- Martinez, J. (2019) "Automation, Growth and Factor Shares," Manuscript.
- McGrattan, E. R. and E. C. Prescott (2005) "Taxes, Regulations, and the Value of U.S. and U.K. Corporations," *Review of Economic Studies*, Vol. 72, No. 3, pp. 767–796.
- Moll, B., L. Rachel, and P. Restrepo (2019) "Uneven Growth: Automation's Impact on Income and Wealth Inequality," Manuscript.
- Oberfield, E. and D. Raval (2014) "Micro Data and Macro Technology," Manuscript.
- Olley, G. S. and A. Pakes (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, Vol. 64, No. 6, pp. 1263–1297.
- Raval, D. R. (2019a) "The Micro Elasticity of Substitution and Non-Neutral Technology," *RAND Journal of Economics*, Vol. 50, No. 1, pp. 147–167.
- (2019b) "Testing the Production Approach to Markup Estimation," Manuscript.
- Rognlie, M. (2015) "Deciphering the Fall and Rise in the Net Capital Share: Accumulation or Scarcity," *Brookings Papers on Economic Activity*, Spring.
- (2018) "Comment on Accounting for Factorless Income," *NBER Macroeconomics Annual*, Vol. 33, pp. 235–248.
- Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2019) "Diverging Trends in National and Local Concentration," Manuscript.
- Solon, G., S. J. Haider, and J. M. Wooldridge (2015) "What are We Weighting For?" *Journal of Human Resources*, Vol. 50, No. 2, pp. 301–316.
- Solow, R. M. (1957) "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, Vol. 39, No. 3, pp. 312–320.
- Syverson, C. (2004) "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, Vol. 112, No. 6, pp. 1181–1222.
- (2019) "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, Vol. 33, No. 3, pp. 23–43.
- Traina, J. (2018) "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements," Manuscript.

Wooldridge, J. M. (1997) "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model," *Economics Letters*, Vol. 56, No. 2, pp. 129–133.

——— (2003) "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model," *Economics Letters*, Vol. 79, No. 2, pp. 185–191.

Appendix A Dynamic Model

The main text uses a static optimization problem to derive the estimator for the cost of capital. Here I derive the estimator using a dynamic model.

Suppose that firm i at time t produces real output Y_{it} using as inputs capital K_{it} and M variable inputs X_{it}^m according to the production function $Y_{it} = F_{it}(K_{it}, X_{it}^1, \dots, X_{it}^M)$. The firm minimizes costs subject to output at time τ being equal to some scalar $\bar{Y}_{i\tau}$ and the capital stock is chosen one-period ahead,

$$\begin{aligned} \min_{\{K_{i\tau+1}, X_{i\tau}^1, \dots, X_{i\tau}^M\}_{\tau=t}^{\infty}} \mathfrak{C}_{it}(\vec{K}_{it}, \vec{X}_{it}^1, \dots, \vec{X}_{it}^M) \\ \text{s.t. } F_{i\tau}(K_{i\tau}, X_{i\tau}^1, \dots, X_{i\tau}^M) = \bar{Y}_{i\tau}, \quad \forall \tau \geq t, \end{aligned} \quad (27)$$

where $\mathfrak{C}_{it}(\vec{K}_t, \vec{X}_t^1, \dots, \vec{X}_t^M) = \mathbb{E}_t \sum_{\tau=t}^{\infty} \beta_{i\tau} C_{i\tau}(K_{i\tau}, K_{i\tau+1}, X_{i\tau}^1, \dots, X_{i\tau}^M)$ denotes the expected sum of discounted future costs at time t where $C_{i\tau}(\cdot)$ are the costs in period τ . $\vec{\cdot}$ indicates a time-vector and $\beta_{i\tau}$ is the $(\tau - t)$ -period stochastic discount factor. The time- τ cost function $C_{i\tau}$ depends on capital at time τ and $\tau + 1$ because the costs include investment costs. Firm-specific input prices are suppressed in the $i\tau$ subscript of the cost function.

The following Lagrangian is associated with the cost minimization problem:

$$\begin{aligned} \mathcal{L}_{it}(\vec{K}_{it}, \vec{X}_{it}^1, \dots, \vec{X}_{it}^M) = & \mathfrak{C}_{it}(\vec{K}_{it}, \vec{X}_{it}^1, \dots, \vec{X}_{it}^M) + \\ & \mathbb{E}_t \sum_{\tau=t}^{\infty} \beta_{i\tau} \lambda_{i\tau} (\bar{Y}_{i\tau} - F_{i\tau}(K_{i\tau}, X_{i\tau}^1, \dots, X_{i\tau}^M)), \end{aligned} \quad (28)$$

where $\beta_{i\tau} \lambda_{i\tau}$ is the Lagrange multiplier. This means that $\beta_{i\tau} \lambda_{i\tau}$ is the increase in the (with respect to time- t) discounted costs of increasing output at time τ by one. Thus, $\lambda_{i\tau}$ is the marginal cost at time τ . Define the markup as the output price, $P_{i\tau}$, divided by marginal costs: $\mu_{i\tau} = \frac{P_{i\tau}}{\lambda_{i\tau}}$.

Taking the derivative of the Lagrangian with respect to input X_{it}^m gives after rewriting that the valued marginal product equals the markup times the input price,

$$VMPX_{it}^m = \mu_{it} P_{it}^{X^m}, \quad m \in \{1, \dots, M\}. \quad (29)$$

Now take the derivative of the Lagrangian with respect to capital K_{it+1} ,

$$\mathbb{E}_t \lambda_{it+1} \frac{\partial F_{it+1}(\cdot)}{\partial K_{it+1}} = \mathbb{E}_t R_{it+1} P_{it+1}^K,$$

where I have used that the derivative of the cost function with respect to capital equals $R_{it+1} P_{it+1}^K$ by the assumption that costs are linear in capital. Multiplying and dividing the left-hand side, within the expectation, by the output price gives that the expected cost of capital equals the expected valued marginal product of capital divided by the markup and the price of one unit of capital,

$$\mathbb{E}_t \frac{VMPK_{it+1}}{\mu_{it+1} P_{it+1}^K} = \mathbb{E}_t R_{it+1}, \quad (30)$$

where I have used that the price of capital is known one period ahead (i.e., capital at time $t + 1$ is purchased at time t). The difference with the first-order condition in the main text is that now expectations are taken.

For concreteness, the following is an example of a typical cost function, where I_t denotes investment and P_t^I the price of the investment good

$$\begin{aligned} \min_{\{K_{\tau+1}, X_{\tau}^1, \dots, X_{\tau}^M, I_{\tau}\}_{\tau=t}^{\infty}} \quad & \mathbb{E}_t \sum_{\tau=t}^{\infty} \beta^{\tau-t} \left[\sum_{m=1}^M P_{\tau}^{X^m} X_{\tau}^m + P_{\tau}^I I_{\tau} \right] \\ \text{s.t. } \quad & F_{\tau}(K_{\tau}, X_{\tau}^1, \dots, X_{\tau}^M) = \bar{Y}_{\tau} \quad \& \quad K_{\tau+1} = (1 - \delta)K_{\tau} + I_{\tau}, \quad \forall \tau \geq t. \end{aligned}$$

Substituting out investment I_{τ} from the cost function using the second constraint, setting up the Lagrangian and taking the derivative with respect to K_{t+1} gives after rewriting

$$\mathbb{E}_t \frac{VMPK_{t+1}}{\mu_{t+1} P_{t+1}^K} = \left[\frac{1}{\beta} - (1 - \delta) \mathbb{E}_t \frac{P_{t+1}^I}{P_t^I} \right] \equiv \mathbb{E}_t R_{t+1}.$$

With $\beta = \frac{1}{1+r}$ and π expected inflation of the investment good, this gives the usual condition that $\frac{VMPK}{\mu} = P^K(r + \delta - \pi + \delta\pi)$ where usually the last term is omitted because it is small (and which does not show up when the model is cast in continuous time).

Appendix B Relation to Production Function Estimation

Equation (9) obtained from Euler's theorem looks similar to a production function, which is a typical object of study in the industrial organization (IO) literature.⁷⁰ However, there are some notable differences which I discuss here. These dif-

⁷⁰See, e.g., Olley and Pakes (1996), Blundell and Bond (2000), Levinsohn and Petrin (2003), Akerberg et al. (2015) and Gandhi et al. (2016).

ferences are important since the recently popularized production approach to estimate markups relies on the estimation of a production function (see, e.g., De Loecker and Warzynski, 2012 and De Loecker et al., 2018). It is true that markups are a different object of interest than the profit share, since the profit share is a measure of the price relative to the average cost, while the markup is the price relative to the marginal cost, but both are clearly related. Another reason for why it is relevant to discuss the differences between my approach and the production function estimation literature is that estimating a production function would provide an alternative way to estimate the cost of capital. This would work as follows. Cost minimization still implies that the cost of capital is the marginal product of capital divided by the markup. Estimating a production function provides indirectly also an estimate of the marginal product of capital. Dividing this estimate with the markup found using the production approach provides an alternative estimate of the cost of capital. As I discuss here, the method I propose requires fewer assumptions than that are needed to estimate a production function. The differences in assumptions are the following.

First, equation (9) suggests that a production function linear in inputs is assumed. The opposite is true, I have assumed a general homogeneous production function and linearity follows from Euler's theorem. Instead, a specific functional form needs to be assumed when estimating production functions whereas here the functional form is not needed to be known.

Second, estimators employed in the IO literature typically require data on real output and real inputs. However, most data sets only comprise data on nominal inputs and output. See Klette and Griliches (1996) for a discussion of the bias that arises when no firm-specific output prices are available, and De Loecker et al. (2016) for the bias that arises when no input prices are available. To estimate the cost of capital I only need data on nominal quantities.

Third, when having only data on total inputs and not inputs by type one implicitly assumes that these different types of inputs are perfect substitutes when estimating a production function. For example, in the Compustat data that I use in this paper, expenditures on labor and materials are lumped together for around 90% of the observations. Hence, estimating a production function using these data requires assuming that materials and labor are perfect substitutes. This is not the case with the method I propose here. No assumption about the elasticity of substitution between different inputs X_i^m is made. A general production function, where each input enters separately, is assumed. This also means that different types of labor (such as workers with different skill levels) or different types of material or capital are not required to be perfect substitutes with each other.

Fourth, unobserved productivity leads typically to an omitted variable bias when estimating production functions. Most of the literature on production function estimation is about how to deal with unobserved (Hicks-neutral) productivity. The most common approach is to proxy for productivity using inputs (see, e.g., Olley and Pakes, 1996, Levinsohn and Petrin, 2003 and Akerberg et al., 2015.). Productivity is not an omitted variable in my approach since the productivity term does not show up in Euler's theorem. This comes at the cost that unobserved heterogeneity in the cost of capital leads to a bias.

Fifth, firms are usually assumed to have identical production functions up to a Hicks-neutral productivity term. If firms would differ in their factor-augmenting productivities this would lead to a bias.⁷¹ Doraszelski and Jaumandreu (2018) and David and Venkateswaran (2019) provide evidence that there is substantial heterogeneity in factor-augmenting productivities across firms. My method allows for production functions to be different across firms.

Sixth, the usual methods to estimate a production function assume that firms produce a single product. However, many firms produce several products. This is problematic because most data sets do not report output and inputs at the product level. To deal with multi-product firms authors have assumed identical production functions across products combined with input proportionality (De Loecker, 2011) or have estimated the production function using only single-product firms (De Loecker et al., 2016). My method does not require firms to produce a single product or that the production function is identical across different products. Suppose that firm i produces several products indexed by l according to the product-specific production function $Y_{il} = F_{il}(K_{il}, X_{il}^1, \dots, X_{il}^M)$. Then Euler's theorem for each individual product yields, after multiplying by the product specific output price and plugging in the first-order condition on the variable inputs (for notational simplicity I set the returns to scale equal to one)

$$P_{il}Y_{il} = \frac{VMPK_{il}}{P_{il}^K} P_{il}^K K_{il} + \mu_{il} P_{il}^X X_{il}, \quad (31)$$

where $P_{il}^X X_{il}$ is the total expenditure on variable inputs by firm i used to produce product l , and K_{il} is the capital allocated to the production of product l . I only observe total output and inputs, and not sales and inputs per product. Taking the sum of equation (31) over all products yields

$$P_i Y_i = \sum_l P_{il} Y_{il} = \frac{VMPK_i}{P_i^K} \sum_l P_{il}^K K_{il} + \mu_i \sum_l P_{il}^X X_{il}, \quad (32)$$

⁷¹Or, in the case of a Cobb-Douglas production function, if firms have a different output elasticity this would lead to a bias as well.

where $VMPK_i/P_i^K$ is the capital-weighted average marginal product of capital across different product lines ($\frac{VMPK_i}{P_i^K} = \frac{\sum_l P_{il}^K K_{il} VMPK_{il}/P_{il}^K}{\sum_l P_{il}^K K_{il}}$), and μ_i is the expenditure-weighted average markup across different product lines ($\mu_i = \frac{\sum_l P_{il}^X X_{il} \mu_{il}}{\sum_l P_{il}^X X_{il}}$). Note that only data on total output, total intermediate input expenditure and the total capital stock are needed.

Appendix C Simulations

To get a better sense of the magnitude of the bias due to heterogeneous capital costs, I apply the cost of capital estimator to simulated data. In order to simulate data I assume the following functional form of the production function,

$$Y_i = A_i \min\left\{\left(\alpha_i K_i^{\frac{\nu-1}{\nu}} + (1 - \alpha_i) L_i^{\frac{\nu-1}{\nu}}\right)^{\frac{\nu}{\nu-1}}, \alpha_i^M M_i\right\}. \quad (33)$$

There is a constant elasticity of substitution, ν , between capital and labor, and materials are perfect complements with capital and labor. A_i is a Hicks-neutral productivity term, α_i represents capital augmenting technology, and α_i^M represents materials augmenting technology.

I choose parameter values to match moments in the data and vary the dispersion in the cost of capital. Firms are still assumed to minimize cost and output is assumed to be log-normally distributed such that the dispersion in output equals the observed coefficient of variation of 2.1. All statistics reported here are within industry-year, and means are in thousands of dollars per employee. In the data I use, the wage is approximated reasonably well by a log-normal distribution with parameters 3.7 and 0.36. I assume $\frac{\alpha}{1-\alpha}$ is log-normally distributed with, conditional on the dispersion in the cost of capital, mean and standard deviation such as to match the observed average capital-labor ratio and coefficient of variation, which are 74 and 1.54, respectively. Furthermore, I normalize the price of materials to 1 and assume that α^M is log-normally distributed with parameters such as to match the average and coefficient of variation of the material-labor ratio (197 and 1.28, respectively). Finally, A is also log-normally distributed such that the average markup equals 1.1 and the standard deviation equals 0.05, which is roughly similar to what I find in the data.

I set the depreciation rate to 0.1, which is the average depreciation rate over the period I am studying. I further assume the interest rate, r , is log-normally distributed with mean 0.1, to match a cost of capital of 0.2 as I find it to be in the data, and I let the dispersion vary. I set the elasticity of substitution between capital and labor equal to 0.5 (Raval, 2019a). Setting a higher elasticity of substitution

leads to a similar bias (see Table 6 in Appendix G). I simulate data for 5,000 firms 100 times, and for each simulation I estimate the cost of capital and compare it with the true cost of capital. Table 3 shows how much larger the true cost of capital is relative to the estimated cost of capital; both the average and standard deviation across simulations are reported. Suppose that the value of this relative bias is 5%. This means that the true cost of capital and therefore the capital share are 5% larger than what I estimate them to be. In the data, I estimate the capital share to be around 8%. Thus, a 5% relative bias means that the true capital share would be 8.4%, and hence the bias in the capital (profit) share is minus (plus) 0.4 percentage points.

The second column of Table 3 shows the relative bias when the squared capital-input ratio is not included as a control, for different assumptions on the coefficient of variation of the interest rate (column 1). For each cell, technology parameters are re-calibrated to match the data moments. The bigger the dispersion in interest rates, the bigger the bias, but the bias is limited. When the coefficient of variation is 0.5, the relative bias is 11%. Thus, in this case, with an estimated capital share of 8% the absolute bias is less than a percentage point and can therefore not explain the four percentage point rise in the profit share I find. Of course, the coefficient of variation of the interest rate could be bigger than 0.5 and therefore the bias could be bigger. However, a coefficient of variation of 0.5 implies that the interquartile range is 0.06 and that eleven percent of firms face an interest rate lower than 0.05 while fourteen percent of firms face an interest rate higher than 0.15. Given my sample of publicly listed firms, this is already a substantial variation in interest rates. Moreover, when going to the data, I will control for financial data such as the leverage ratio. Thus, to get a substantial bias, a large variation in interest rates, after including these controls, is needed.

Note that the dispersion in the bias across simulations is sizable. For instance, when the coefficient of variation of the interest rate is 0.2, the standard deviation of the relative bias across simulations is 5.1%. This should not be a big concern since I estimate the cost of capital at the industry level. There are around 30 industries, so this standard deviation partially averages out when aggregating across industries. Figure 25 in Appendix H shows that the bias across simulations is symmetrically distributed around the mean.

The third column shows the bias when the squared capital-input ratio is included as a control. The resulting bias gets somewhat lower, but does not go to zero because the relationship between the capital-input ratio and the cost of capital is not linear. One reason for the low bias, both without and with control, is that there is a large dispersion in the capital-input ratio in the data, which yields a

Table 3: Monte carlo results relative bias

Coefficient of variation r	Var. capital-labor ratio I		Var. capital-labor ratio II	
	No control	With control	No control	With control
0	0.2% (2.5%)	0.3% (3.8%)	0.2% (2.8%)	-0.1% (4.3%)
0.1	0.8% (3.4%)	0.6% (4.2%)	0.8% (3.1%)	0.1% (4.5%)
0.2	2.4% (5.1%)	1.8% (5.5%)	2.6% (4.2%)	1.1% (5.1%)
0.3	4.7% (6.9%)	3.8% (7.1%)	5.4% (5.5%)	2.8% (6.1%)
0.4	7.7% (8.7%)	6.2% (8.8%)	8.9% (6.8%)	5.2% (7.2%)
0.5	11.1% (10.3%)	9.0% (10.3%)	12.8% (7.9%)	8.0% (8.4%)

Notes: The relative bias is $\left(\frac{\hat{R}}{R} - 1\right) 100\%$. Each cell shows the average and the standard deviation (within parentheses) of this statistic across simulations. The control refers to the squared capital-input ratio. For each cell, technology parameters are re-calibrated to match the moments of the data. In the columns referring to var. capital-labor ratio I, the simulations match the observed variation in the capital-input ratio, while in the columns referring to var. capital-labor ratio II, the simulations match half of the observed variation.

large dispersion in technology across firms in the simulations. The variation in the capital-input ratio could be high, partially because firms cannot adjust the capital stock within the period or because of measurement error. Therefore, the last two columns of Table 3 show the bias when the model only captures half of the observed variation in the capital-input ratio. This leads to only a small increase in the relative bias, and when the squared capital-input ratio is included as a control, the bias decreases in fact compared to the simulations that match all the variation in the capital-input ratio. The reason is that, in this case, with less variation in technology the relationship between the cost of capital and the capital-input ratio becomes closer to being linear.

Appendix D Relationship Between Profit Share and Market Capitalization

As a check of my estimate of the profit share, I study to what extent the profit share is related to market capitalization. All else equal, it would be expected that the more profitable an industry is the higher market capitalization. However, owning a stock provides claims on the future cash flow and this cash flow includes payments to capital. Therefore, to relate the profit share with market capitalization, we would want to exclude the value coming from future payments to capital from market capitalization. Therefore, I subtract net assets (i.e., total assets minus total liabilities) from market capitalization. The underlying rationale is that net assets equal the liquidation value of the firm, and that the value that exceeds the

liquidation value corresponds to the value derived from future economic profits.

Table 4 shows the results of regressing the profit share, at the industry level, on the asset-corrected market capitalization as a share of sales. The coefficient is statistically significant and is around 0.03. Thus when profits increase by 1 dollar, market capitalization increases by around 30 dollar. This is consistent with a stationary equilibrium in which the asset-corrected market capitalization equals the discounted sum of future profits when the discount rate is 0.97. The coefficient is insensitive to including industry fixed effects, and the coefficient is also similar when market capitalization not corrected for net assets is used.

Table 4: Relationship profit share and market capitalization across industries

	(1)	(2)	(3)	(4)
	π	π	π	π
<u>market capitalization - net assets</u> sales	0.0347*** (0.00367)	0.0223*** (0.00475)		
<u>market capitalization</u> sales			0.0328*** (0.00277)	0.0223*** (0.00410)
Year fixed effects	X	X	X	X
Industry fixed effects		X		X
Observations	1427	1427	1427	1427
R^2	0.117	0.435	0.148	0.439

Notes: Robust standard errors in parentheses. To remove outliers I trim the top and bottom fifth percentile of the dependent and independent variable.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix E Calculating the Depreciation Rate

Compustat reports the flow value of depreciation and accumulated depreciation. One way to calculate the depreciation rate is to divide the flow value of depreciation by the net capital stock in the previous year. However, this gives unreasonably large values for the depreciation rate. The reason for this can be seen as follows. Suppose a firm buys a machine for \$1000 at the end of period 0 and depreciates it linearly in 5 time periods. Table 5 shows the resulting gross and net capital stock, and the flow and accumulated depreciation as it would be reported in Compustat. The sixth row shows the resulting depreciation rate when dividing the flow value of depreciation by the net capital stock in the previous period. The depreciation rate increases quickly toward the end because the net capital stock is declining over time. The average depreciation rate in this example is 46%, which is high for a machine that depreciates in 5 years. The reason is that the company depreciates

Table 5: Example depreciation rates

time	0	1	2	3	4	5
Gross capital	100	100	100	100	100	100
Net capital	100	80	60	40	20	0
Flow depreciation	0	20	20	20	20	20
Accumulated depreciation	0	20	40	60	80	100
$\frac{\text{Flow depreciation}_t}{\text{Net capital}_{t-1}}$		20%	25%	33%	50%	100%
$1 - \left(\frac{\text{Net capital}_t}{\text{Gross capital}_t} \right)^{1/\text{age}}$		20%	23%	26%	33%	100%

linearly whereas economists typically use geometric depreciation. To avoid these high depreciation rates that are also volatile over time, I will calculate the depreciation rate in the following way. Start with the following equation, relating net capital to gross capital at time t , assuming capital has depreciated with a constant depreciation rate,

$$\text{Net capital}_t = (1 - \delta_t)^{\text{age}} \text{Gross capital}_t .$$

Rewriting this equation gives $\delta_t = 1 - \left(\frac{\text{Net capital}_t}{\text{Gross capital}_t} \right)^{1/\text{age}}$, where the age of the capital stock is measured as accumulated depreciation divided by the flow value of depreciation. This leads to values of the depreciation rate as reported in the seventh row. The average depreciation rate is at 40% still high, but not as high as for the naively calculated depreciation rate. The main reason why it is so high is because the net capital stock is 0 in the last period leading to a depreciation rate of 100%. If the company continuously replenishes capital this would not occur.

Appendix F Profit Share from Industry Data

I here estimate the profit share using industry level data from the BEA. This has as benefit that all firms are included to construct the industry values and not only firms that are present in Compustat, but has as downside that there are only few observations.⁷²

I use data on value added and its components (from the industry economic accounts), and the capital stock at current costs (table 3.1ESI). The components of value added are compensation of employees and taxes on production and imports. These data is only available from 1987 onward at the industry level. As for the Compustat data I exclude the utilities, mining, finance, insurance, real estate and

⁷²Hall (1988, 2018) estimates the markup using industry level data. His approach requires knowing the cost of capital.

government sectors. This leaves me with 50 industries. To estimate the cost of capital and the profit share, I construct the left-hand side as value added minus taxes, divided by labor compensation and the right-hand side as capital divided by labor compensation. Because there is not enough power to estimate the profit share year-by-year I include a cubic time trend for the marginal product of capital and the markup, and industry fixed effects.

Figure 23 shows the resulting profit share, and for reference the profit share as a share of value added in Compustat. The profit share I estimate using industry data is very similar in both level and trend to what I estimate using Compustat data. It increases during the late 1980s and 1990s and flattens out afterwards.

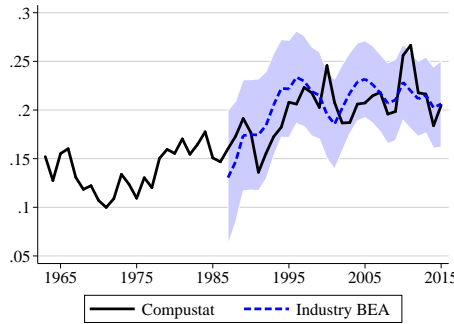


Figure 23: Profit share (of value added) using BEA industry data

Appendix G Additional Tables

Table 6: Monte carlo results relative bias when elasticity of substitution between capital and labor is 1.2

Coefficient of variation r	No control	Control
0	0.1% (2.4%)	0.2% (3.7%)
0.1	1.0% (3.1%)	0.7% (3.9%)
0.2	3.8% (4.6%)	2.5% (5.6%)
0.3	8.0% (6.2%)	5.6% (7.6%)
0.4	13.1% (7.6%)	9.5% (9.3%)
0.5	18.7% (8.8%)	13.9% (10.9%)

Notes: Average of $\left(\frac{R}{\hat{R}} - 1\right)$ 100%. Within parentheses is the standard deviation of this statistic across simulations.

Table 7: Relationship cost of capital and depreciation rate across industries

	R	R
Depreciation rate	0.812*** (0.122)	0.379* (0.179)
Year fixed effects	X	X
Industry fixed effects		X
Observations	1425	1425
R^2	0.106	0.377

Notes: Robust standard errors in parentheses. The depreciation rate is the average depreciation rate within an industry-year, weighted by capital. To remove outliers I trim the top and bottom fifth percentile of the dependent and independent variable.

* $p < .05$, ** $p < .01$, *** $p < .001$

Appendix H Additional Figures

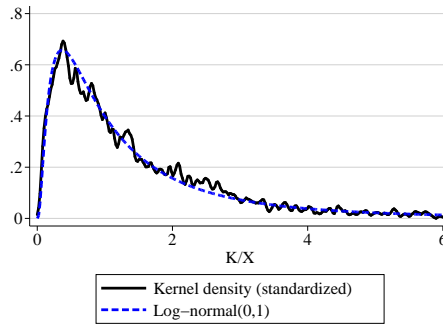


Figure 24: Density Capital-Input Ratio

Notes: The plot shows that the kernel density of the capital-input ratio after log standardizing at the industry-level (i.e., take the log, subtract the industry-mean and divide by the industry-standard deviation (of the logged variable) and then take the exponential) coincides with the log-normal distribution with parameters $\mu = 0$ and $\sigma = 1$. The plot uses data of the year 2000.

Estimating the Cost of Capital and the Profit Share

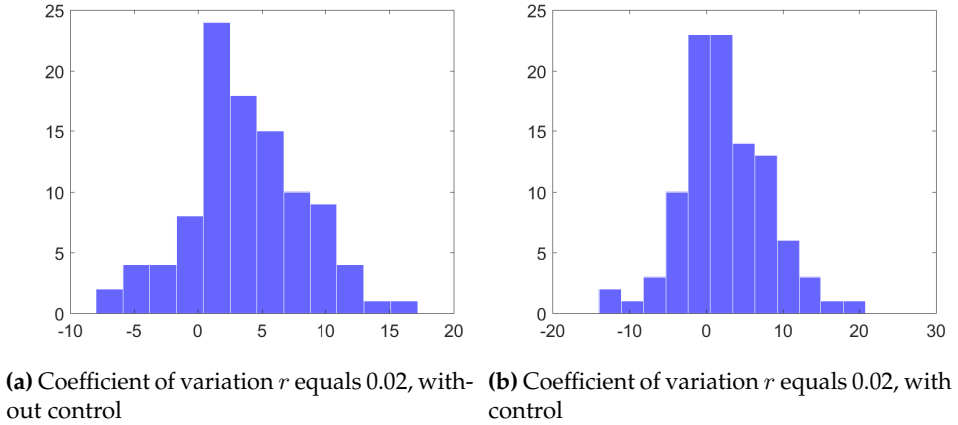


Figure 25: Histogram of the relative bias across simulations

Notes: The histogram shows the number of simulations with a certain relative bias. The total number of simulations for each specification is 100.

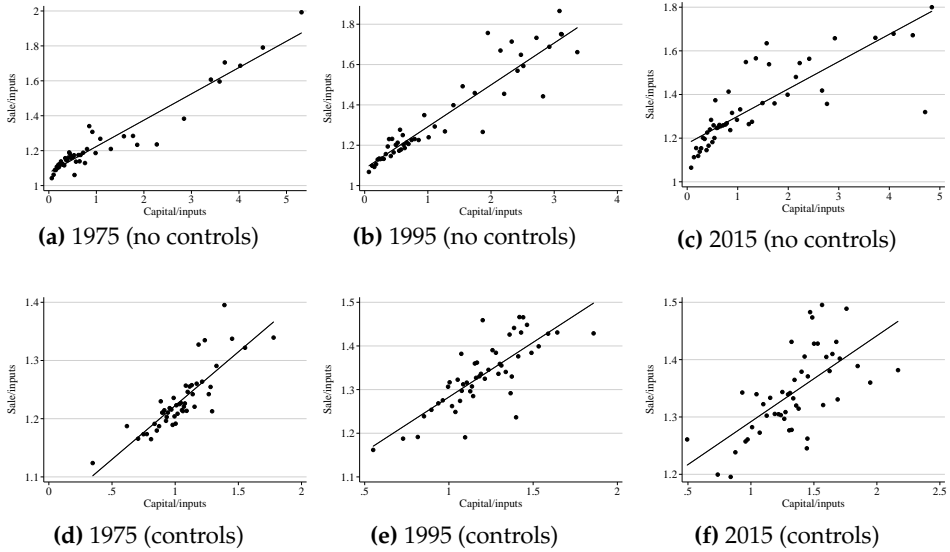


Figure 26: Binned scatter plot

Notes: The plots show the relationship between the capital-input and sales-input ratios for three different years, both with and without controls. All firms across industries within a year are included. The bins refer to equally sized groups and are weighted by capital, as the main regressions are.

Estimating the Cost of Capital and the Profit Share

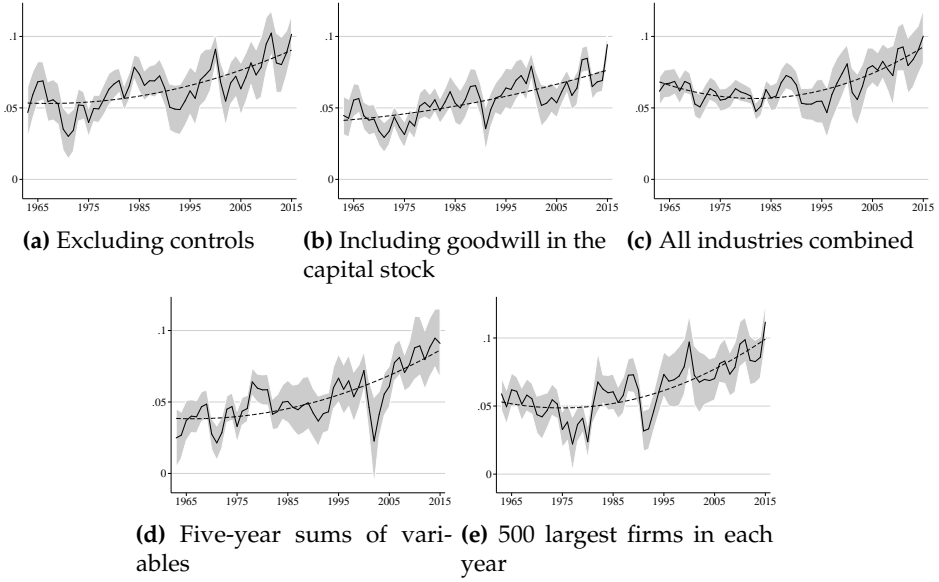


Figure 27: Profit share - robustness

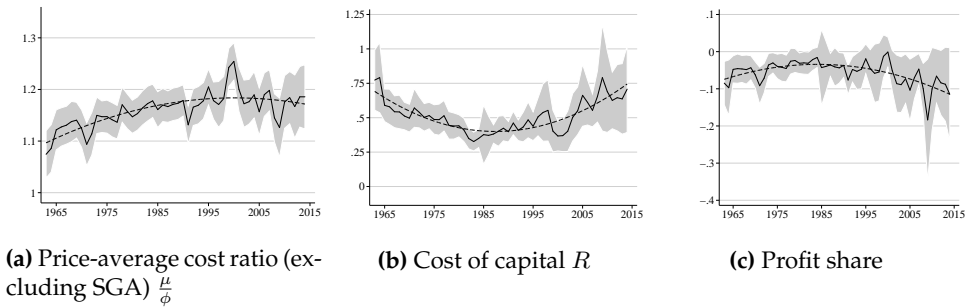


Figure 28: Results when SGA is considered to be a fixed cost (i.e., SGA is not included in the regression but still subtracted, together with all other costs, from sales to obtain profits)

Estimating the Cost of Capital and the Profit Share

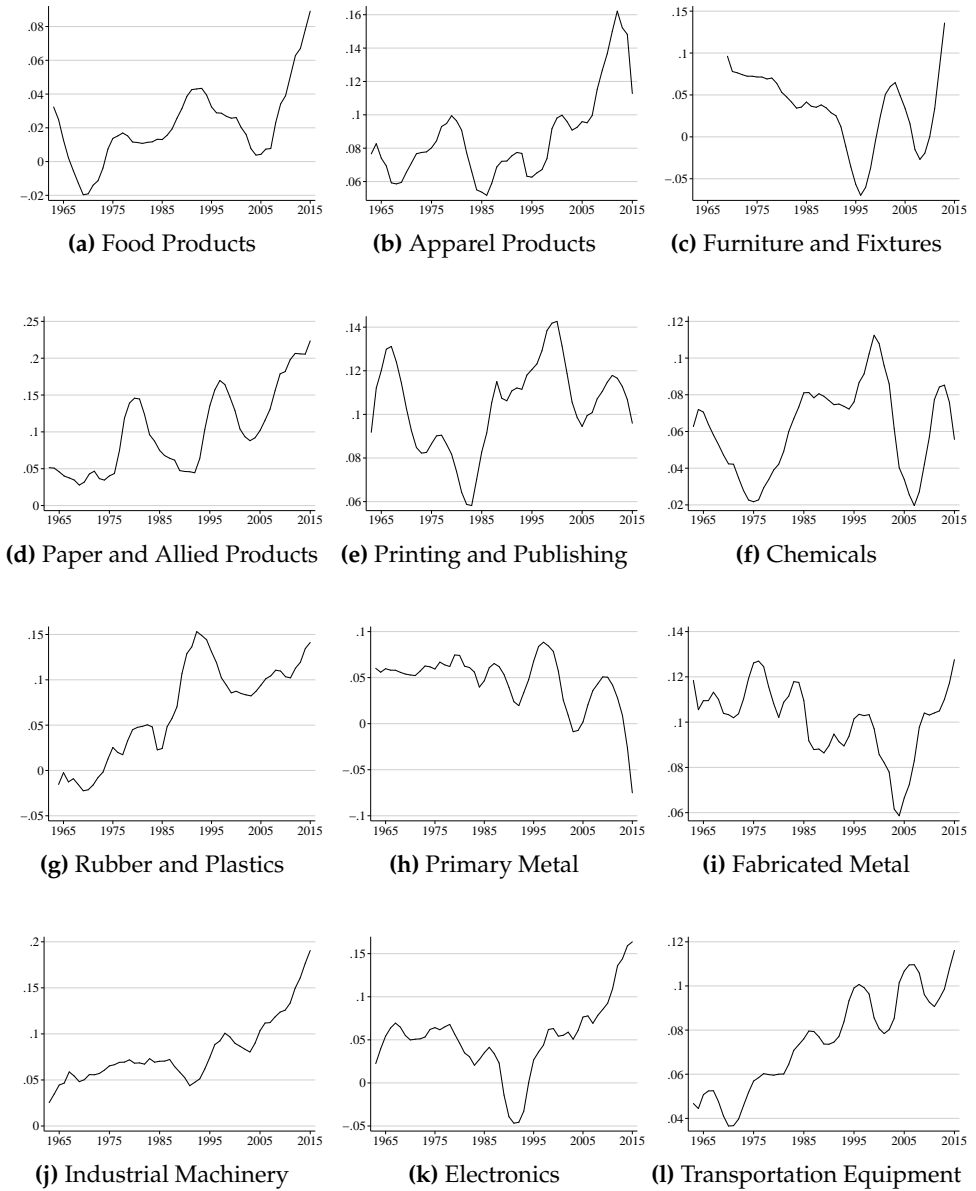


Figure 29: Profit share across industries (5-year moving average)

Estimating the Cost of Capital and the Profit Share

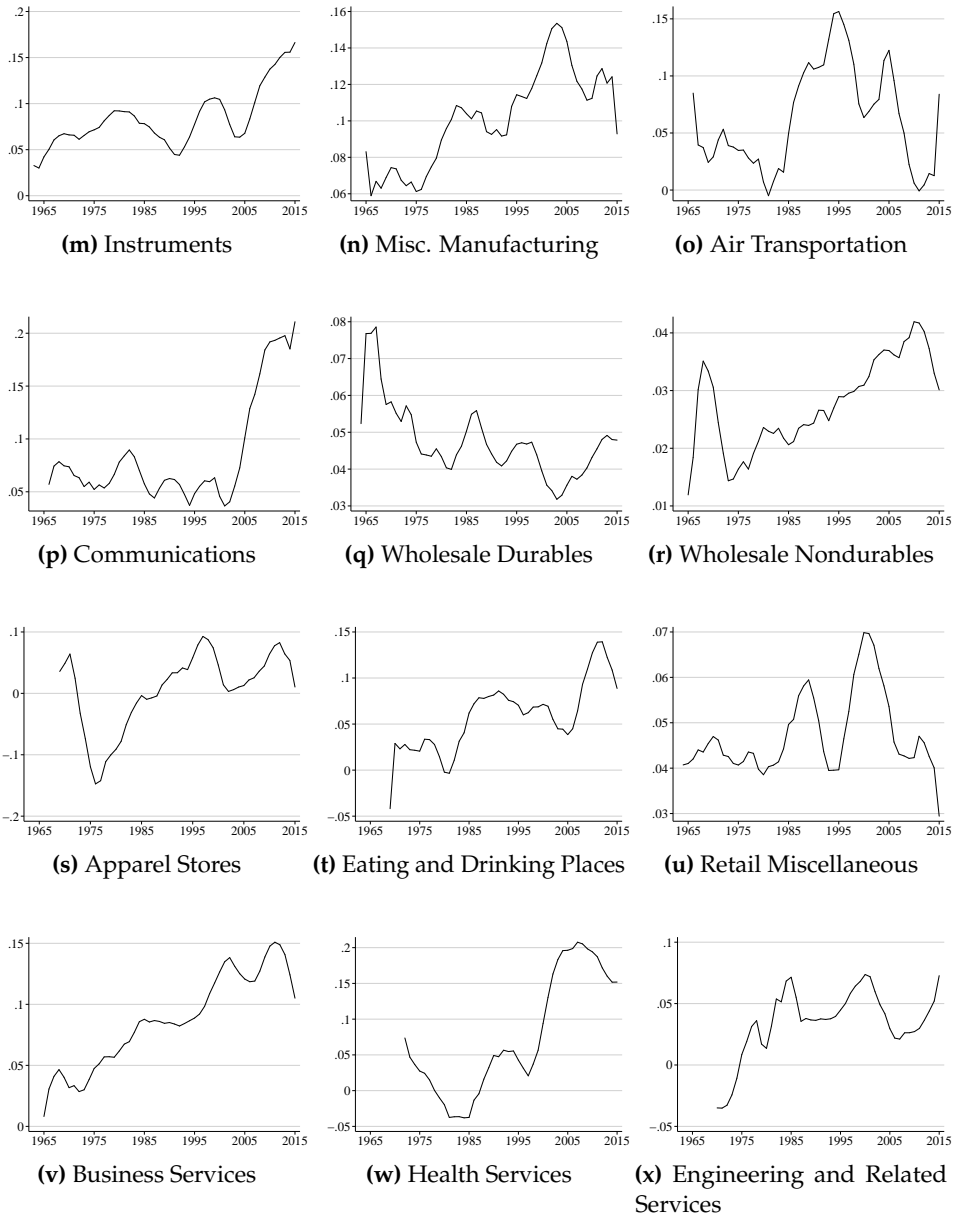


Figure 29: Profit share across industries - continued

Estimating the Cost of Capital and the Profit Share

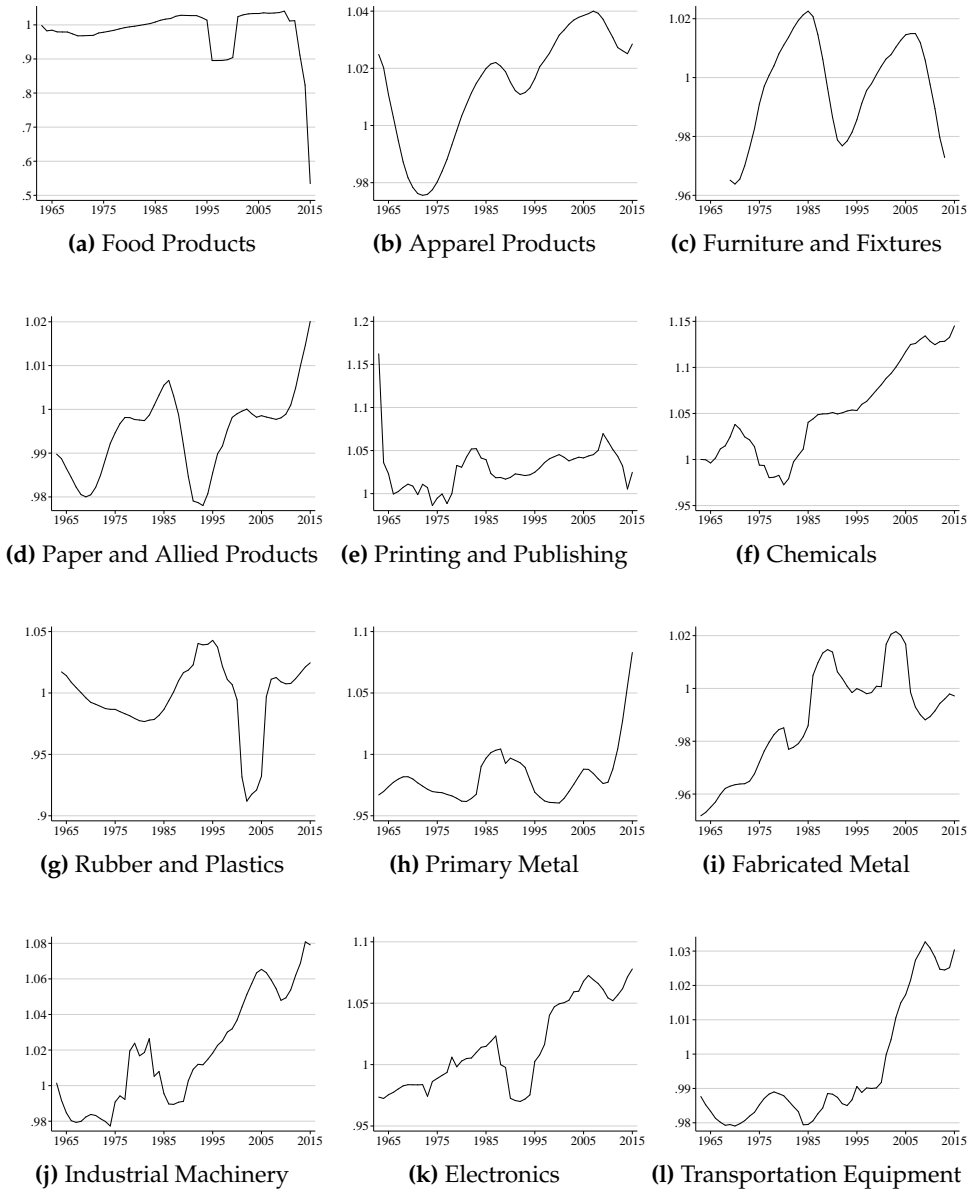


Figure 30: Returns to scale across industries (5-year moving average)

Estimating the Cost of Capital and the Profit Share

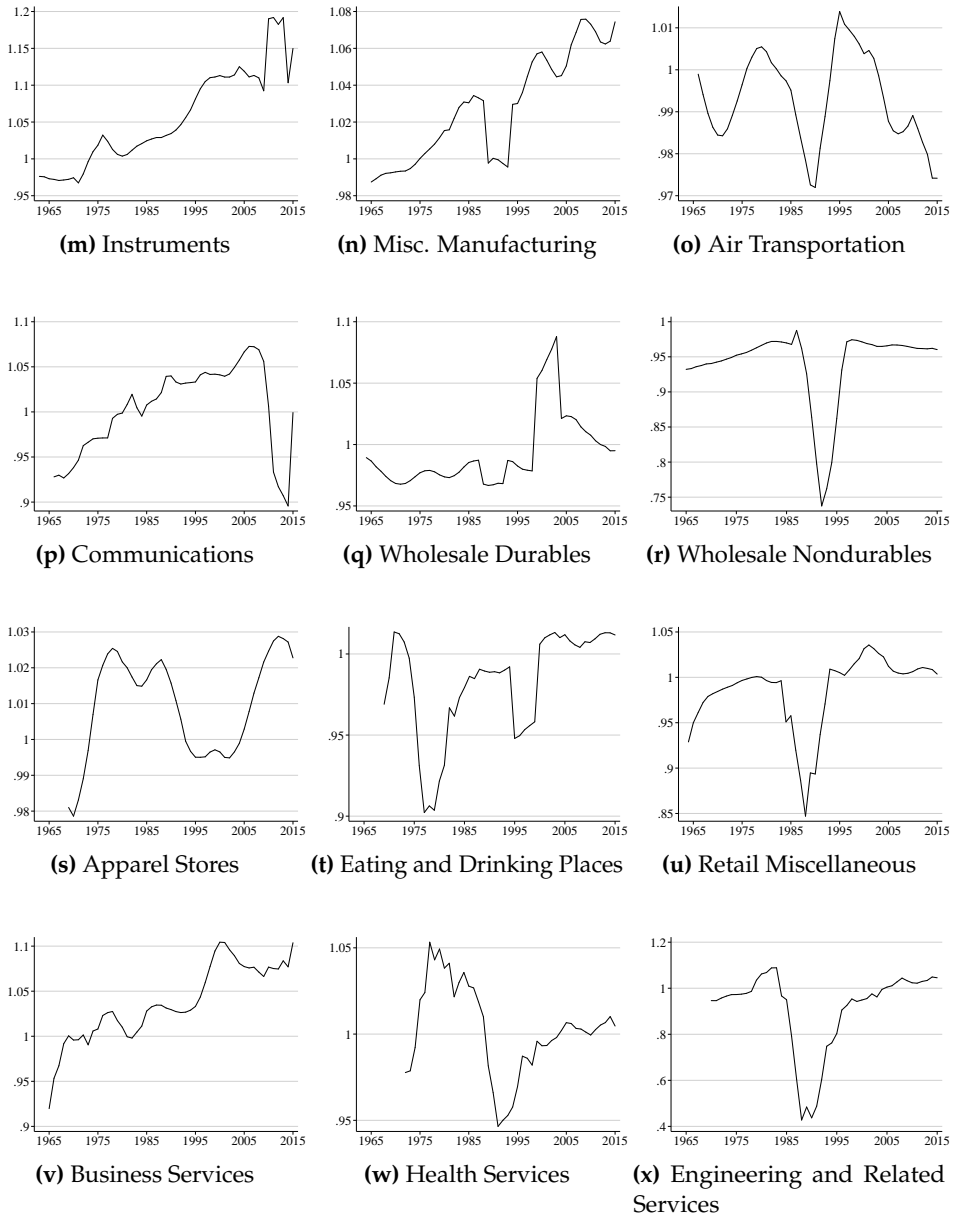


Figure 30: Returns to scale across industries - continued

Estimating the Cost of Capital and the Profit Share

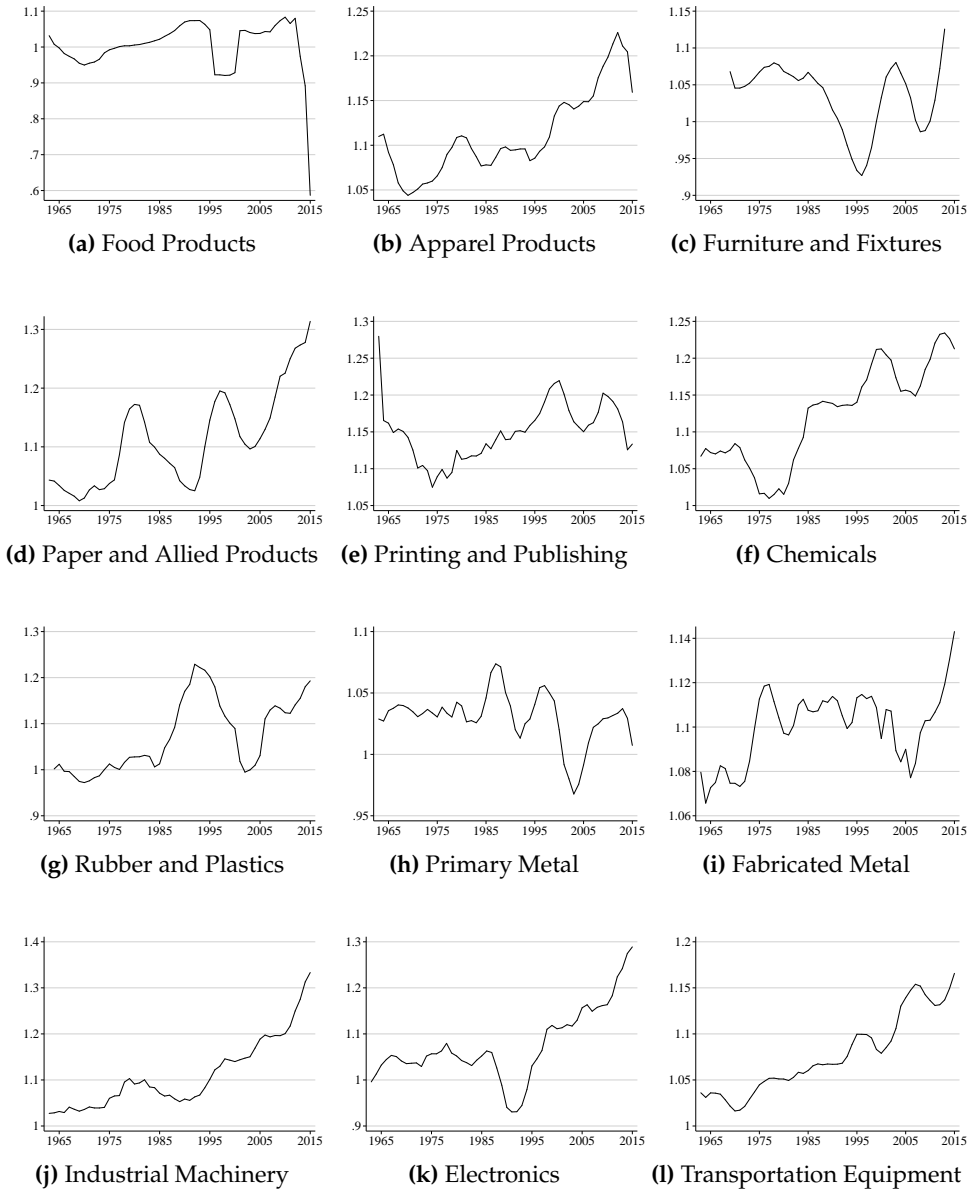


Figure 31: Markup across industries (5-year moving average)

Estimating the Cost of Capital and the Profit Share

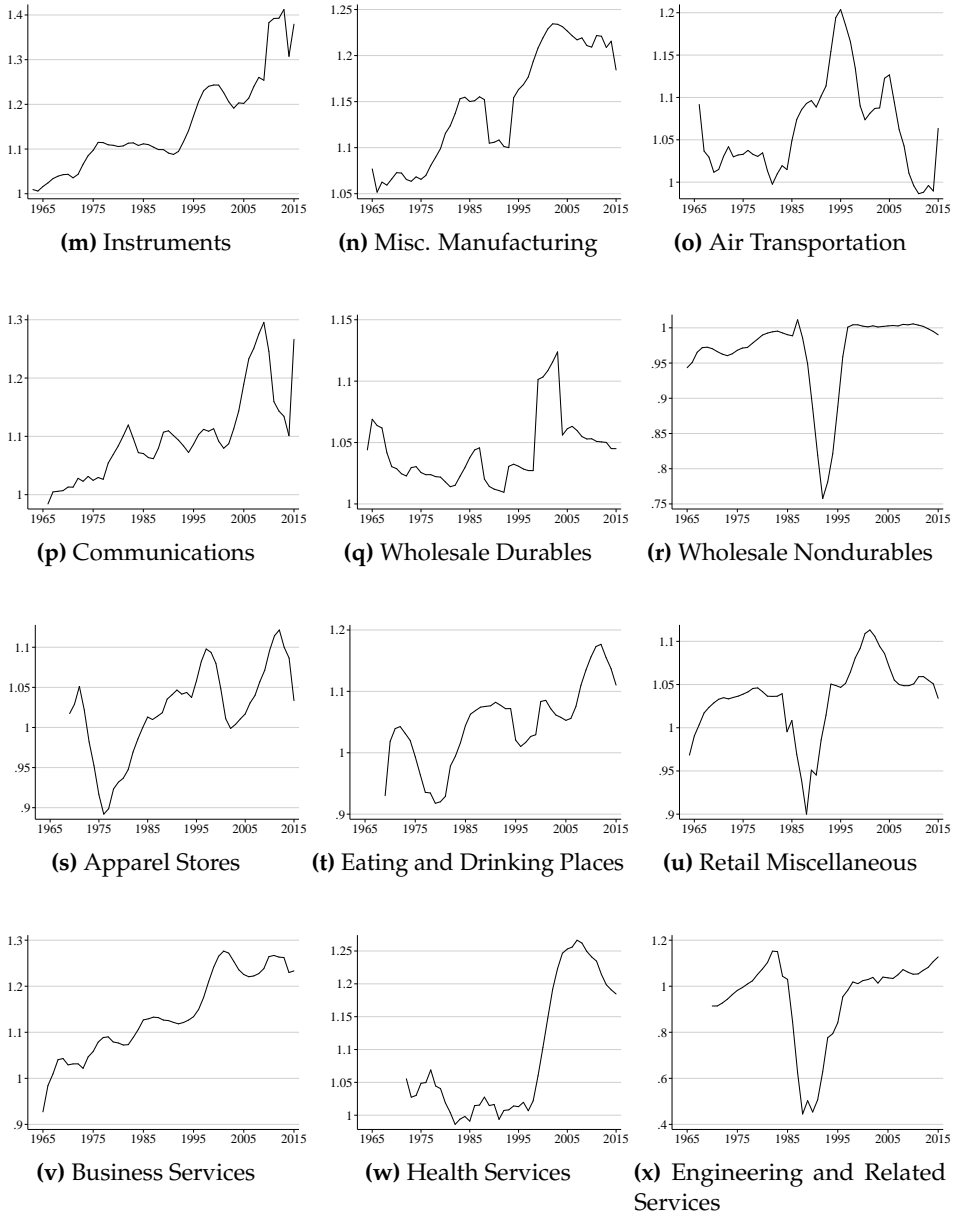


Figure 31: Markup across industries - continued

Appendix I Lag as Instrument

When there are productivity shocks, and capital responds with a lag while other inputs do not, then this leads to a correlation between the capital-input ratio and the marginal product of capital, and therefore to a bias. In the main text I deal with the bias arising from shocks, by using the lagged capital-input ratio as instrument. Here I show, using simulations, that this instrument is indeed appropriate.

I solve the following model. For simplicity, only consider two inputs: capital K and one other input X . The firm solves the following maximization problem,

$$\begin{aligned} \max_{\{P_t, Y_t, X_t, I_t, K_{t+1}\}_{t=0}^{\infty}} \quad & \mathbb{E} \sum_{t=0}^{\infty} \beta^t (P_t Y_t - P_t^x X_t - I_t) \\ \text{s.t.} \quad & Y_t = A_t K_t^\alpha X_t^{1-\alpha} \\ & P_t = Y_t^{-1/\sigma} \\ & K_{t+1} = (1 - \delta) K_t + I_t. \end{aligned}$$

It maximizes the discounted sum of its profits choosing output, inputs and the price where the firm takes into account an iso-elastic demand curve with price elasticity σ . Production is assumed to be Cobb-Douglas, as this allows for an analytic solution of the optimization problem. The advantages of an analytic solution are that it improves the reliability of the results and that it becomes easier to allow for rich heterogeneity.

Substituting the constraints into the objective and taking the derivative with respect to X_t and rewriting gives the following expression for X_t as a function of K_t and A_t ,

$$X_t = \left(\frac{\mu P_t^X}{1 - \alpha} A_t^{\frac{1-\sigma}{\sigma}} K_t^{\frac{1-\sigma}{\sigma} \alpha} \right)^{\frac{\sigma}{\alpha(1-\sigma)-1}}, \quad (34)$$

where $\mu = \frac{\sigma}{\sigma-1}$ is the markup. Taking the derivative with respect to K_{t+1} gives

$$\alpha K_{t+1}^{\frac{\alpha(\sigma-1)-\sigma}{\sigma}} \mathbb{E}_t \left(A_{t+1}^{\frac{\sigma-1}{\sigma}} X_{t+1}^{(1-\alpha)\frac{\sigma-1}{\sigma}} \right) = \mu \left(\frac{1}{\beta} + \delta - 1 \right). \quad (35)$$

Bringing the expression for X_t from equation (34) one period forward, plugging it into (35) and solving for capital gives

$$K_{t+1} = \left(\frac{\alpha}{1/\beta + \delta - 1} \right)^{1+\alpha(\sigma-1)} \left(\frac{w}{1-\alpha} \right)^{(\alpha-1)(\sigma-1)} \mu^{-\sigma} \left(\mathbb{E}_t A_{t+1}^{\frac{\sigma-1}{1+\alpha(\sigma-1)}} \right)^{1+\alpha(\sigma-1)}. \quad (36)$$

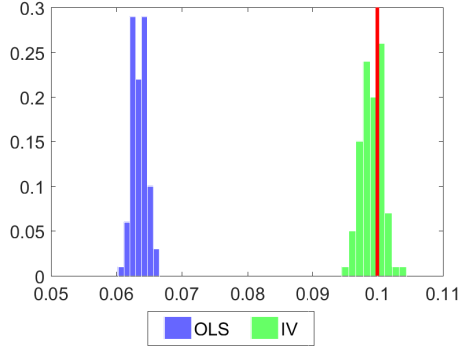


Figure 32: Histogram of the estimated cost of capital across simulations

Notes: The height of each bar is the relative number of observations in each bin. The red vertical line shows the true cost of capital, which is 0.1.

Thus, capital is a function of parameters and the expectation of (transformed) productivity, which we know conditional on the stochastic process of productivity. When the log of productivity follows an AR(1) according to $\log A_{t+1} = \rho_0 + \rho_1 \log A_t + \xi_{t+1}$ where the innovations ξ_t are mean-zero normally distributed with standard deviation σ_ξ then $\mathbb{E}_t A_{t+1}^{\frac{\sigma-1}{1+\alpha(\sigma-1)}} = e^{\rho_0 \frac{\sigma-1}{1+\alpha(\sigma-1)}} A_t^{\rho_1 \frac{\sigma-1}{1+\alpha(\sigma-1)}} e^{\frac{1}{2} \left(\frac{\sigma-1}{1+\alpha(\sigma-1)} \right)^2 \sigma_\xi^2}$.

I simulate 100 times 1000 firms for two periods. For each sample I run the OLS using the data from the second period (i.e., the baseline) and second I instrument the capital-input ratio with the lagged capital-input ratio. As parameters I take as discount factor $\beta = \frac{1}{1+0.04}$ and as depreciation rate $\delta = 0.06$ to obtain a cost of capital of 0.1. I normalize the input price P_t^X to 1. The demand elasticity σ is set such that the markup is 1.2. The output elasticity of capital α varies uniformly between 0.1 and 0.4. I set the persistence of the AR(1) process for productivity ρ_1 to 0.95, and set ρ_0 to 0. To obtain a constant standard deviation of productivity over time I set the standard deviation of the innovation to $\sigma_\xi = \sigma_\omega \sqrt{1 - \rho_1^2}$ where σ_ω is the standard deviation of log productivity that is used to draw the initial distribution of productivity.

Figure 32 shows the resulting distribution of the cost of capital for both the OLS and IV estimator. The true cost of capital is 0.1 in each simulation which is indicated by the red vertical line. The OLS estimate is severely biased with the average estimate being 0.0635 (note that in these simulations I do not use the squared capital-input ratio as a control). The IV estimate deals fully with the bias coming from productivity shocks. The average IV estimate is with 0.099 very close to the true value of 0.1. Moreover, the dispersion of the IV estimate across simulations is quite small. The IV estimate ranges between 0.095 and 0.105.

That there is such a big difference between the OLS and IV estimate in the simulated data but not in the Compustat data suggests that productivity shocks are not important for my estimator. This could either be because productivity shocks in the data are not sizable, or because it takes time for all inputs to respond to productivity shocks and not only for capital.

Appendix J Correcting for First-Order Approximation

$E(\hat{R})$

In the main text I used a first-order approximation for $\hat{R} = \frac{\widehat{VMPK}/\phi}{\mu/\phi}$ such that $E(\hat{R}) \approx R$. Here I study the error induced by this first-order approximation and show that it is small. For notational simplicity I assume constant returns to scale. This does not affect the results.

Denote \widehat{VMPK} by z and $\hat{\mu}$ by y , such that $\hat{R} = \frac{z}{y}$. Taking the second-order Taylor series expansion around the mean $E(z)$ and $E(y)$ gives

$$\begin{aligned} \hat{R} \approx & \frac{E(z)}{E(y)} + \frac{1}{E(y)} (z - E(z)) - \frac{E(z)}{E(y)^2} (y - E(y)) - \\ & \frac{1}{E(y)^2} (z - E(z)) (y - E(y)) + \frac{E(z)}{E(y)^3} (y - E(y))^2, \end{aligned}$$

and subsequently taking the expectation gives

$$E(\hat{R}) \approx \frac{E(z)}{E(y)} - \frac{1}{E(y)^2} \text{cov}(z, y) + \frac{E(z)}{E(y)^3} \text{var}(y).$$

The last two terms are second order. Plugging $z = \widehat{VMPK}$ and $y = \hat{\mu}$ back into the equation, and using that both are unbiased estimates of the marginal product and markup, respectively, gives

$$E(\hat{R}) \approx \frac{\overline{VMPK}}{\bar{\mu}} - \frac{1}{\bar{\mu}^2} \text{cov}(\widehat{VMPK}, \hat{\mu}) + \frac{\overline{VMPK}}{\bar{\mu}^3} \text{var}(\hat{\mu}),$$

where $\bar{\cdot}$ represent averages. With R being common across firms, we have that $\overline{VMPK} = \bar{\mu}R$, and hence $E(\hat{R}) = R$ up to a first-order approximation (Proposition 2). Using the regression output, it is straightforward to correct for the error induced by the first-order approximation compared to the second-order approximation. The resulting estimate of the profit share is shown in Figure 33, and it is identical to the baseline first-order approximation. Therefore, it is unlikely that

including higher order terms would lead to a substantially different estimate of the profit share.

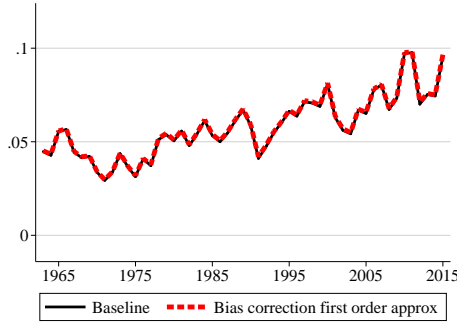


Figure 33: Profit share second-order approximation

Appendix K Growth Accounting

One frequently used way to estimate productivity growth is to calculate the Solow residual (Solow, 1957). Doing so, productivity growth is found to be lower after 1970 than before 1970. This is the so-called productivity slowdown (Gordon, 2016). However, in order to calculate the Solow residual, perfect competition and constant returns to scale are assumed. Here, I explore the effects on the Solow residual when diverting from perfect competition and constant returns to scale, using my estimates of the capital share and returns to scale.

Suppose there is an aggregate production function $Y_t = F(A_t, K_t, H_t)$, where Y is real value added, A is productivity, K is real physical capital and $H = hL$ is human capital where L is hours of labor.⁷³ Taking the derivative with respect to time and dividing by Y gives

$$\frac{\dot{Y}_t}{Y_t} = \theta_t^K \frac{\dot{K}_t}{K_t} + \theta_t^H \left(\frac{\dot{h}_t}{h_t} + \frac{\dot{L}_t}{L_t} \right) + \frac{\partial F_t}{\partial A_t} \frac{A_t}{Y_t} \frac{\dot{A}_t}{A_t},$$

where $\dot{X}_t = \frac{dX_t}{dt}$ refers to the time derivative, and θ_t^K and θ_t^H to the output elasticity with respect to physical and human capital respectively. Some of the growth in the capital stock is caused by growth in productivity. To credit such growth to productivity growth I follow Klenow and Rodríguez-Clare (1997) and subtract $\theta_t^K \frac{\dot{Y}_t}{Y_t}$ from both sides in the above equation, subsequently divide by

⁷³See Baqaee and Farhi (2020) for growth accounting in a network economy with imperfect competition.

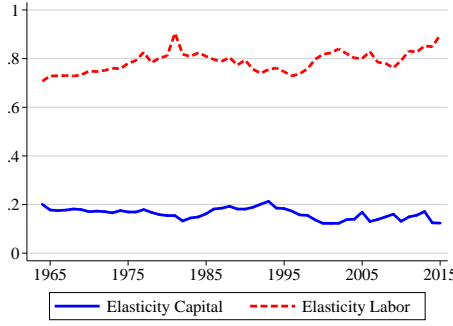


Figure 34: Output elasticities (value added production function)

$1 - \theta_t^K$ and then subtract labor growth $\frac{\dot{L}_t}{L_t}$ from both sides to obtain a growth decomposition of output per hour

$$\underbrace{\frac{\dot{Y}_t}{Y_t} - \frac{\dot{L}_t}{L_t}}_{\text{Growth output per hour}} = \underbrace{\frac{\theta_t^K}{1 - \theta_t^K} \left(\frac{\dot{K}_t}{K_t} - \frac{\dot{Y}_t}{Y_t} \right)}_{\text{Contribution physical capital}} + \underbrace{\frac{\theta_t^H}{1 - \theta_t^K} \frac{\dot{h}_t}{h_t}}_{\text{Contribution human capital}} + \underbrace{\frac{\theta_t^H + \theta_t^K - 1}{1 - \theta_t^K} \frac{\dot{L}_t}{L_t}}_{\text{Scale effect}} + \underbrace{\frac{1}{1 - \theta_t^K} \frac{\partial F_t}{\partial A_t} \frac{A_t}{Y_t} \frac{\dot{A}_t}{A_t}}_{\text{Contribution productivity}}.$$

Note that $\frac{\dot{Y}_t}{Y_t} - \frac{\dot{L}_t}{L_t}$ is the growth in output per hour $\frac{(\dot{Y}_t/L_t)}{Y_t/L_t}$. Growth in output per hour comes from capital deepening, improvements in human capital, changes in labor supply in case of no constant returns to scale and changes to productivity which is the Solow residual. The scale effect is non-standard as it is equal to zero in the case of constant returns to scale (i.e., when $\theta_t^H + \theta_t^K = 1$). Suppose there are decreasing returns to scale and that growth in the number of hours is positive. Then the contribution of the returns to scale being different from one is negative. This is because at the left-hand side is growth of *average* labor productivity. In case of decreasing returns to scale the marginal product is lower than the average product and therefore increasing the scale (i.e., increasing the total number of hours) lowers average productivity.

The contribution of productivity is productivity growth, $g_A = \frac{\dot{A}_t}{A_t}$, multiplied by $\frac{1}{1 - \theta_t^K} \frac{\partial F_t}{\partial A_t} \frac{A_t}{Y_t}$. In case of productivity being Hicks-neutral the contribution of productivity equals $\frac{1}{1 - \theta_t^K} g_A$, and when productivity is labor augmenting the contribution of productivity equals $\frac{\theta_t^H}{1 - \theta_t^K} g_A$.

Typically, perfect competition and constant returns are assumed such that θ_t^H is approximated by the payments to labor as a share of value added, and $\theta_t^K = 1 - \theta_t^H$. Instead, I estimate the output elasticity using cost shares. Recall that the output elasticity for capital is $\theta^K = \frac{RK}{RK + wH} \phi$ and likewise for θ^H . Note that I cannot use the estimate of the returns to scale, ϕ , obtained in the main text as

Table 8: Growth accounting

Period	Output per hour	Contributions from			
		Capital	Labor comp	Scale effect	Solow res
A. Perfect competition and constant returns					
1948-2015	2.4	0.1	0.3	0.0	1.9
1948-1973	3.3	-0.2	0.3	0.0	3.2
1973-1990	1.5	0.5	0.3	0.0	0.8
1990-1995	1.6	0.2	0.6	0.0	0.8
1995-2007	2.9	0.3	0.4	0.0	2.2
2007-2015	1.5	0.3	0.4	0.0	0.8
B. Imperfect competition and varying returns					
1948-2015	2.4	0.1	0.3	-0.1	2.1
1948-1973	3.3	-0.1	0.2	-0.1	3.2
1973-1990	1.5	0.2	0.3	0.0	1.1
1990-1995	1.6	0.1	0.6	-0.1	1.0
1995-2007	2.9	0.1	0.3	-0.1	2.5
2007-2015	1.5	0.1	0.4	0.1	0.9
C. Imperfect competition and constant returns					
1948-2015	2.4	0.1	0.3	0.0	2.0
1948-1973	3.3	-0.1	0.3	0.0	3.1
1973-1990	1.5	0.2	0.3	0.0	1.0
1990-1995	1.6	0.1	0.6	0.0	0.9
1995-2007	2.9	0.1	0.4	0.0	2.4
2007-2015	1.5	0.1	0.4	0.0	0.9

Notes: Average annual growth rates (in percent), the numbers in a row might not add up due to rounding errors. Source: BLS, Compustat and own calculations.

here a value added production function is used while in the main text I estimate a gross output production function. Therefore, I re-estimate the returns to scale in a similar way as in equation (26), but replacing sales by value added and having the input bundle only consist of labor and capital. This leads to similar returns to scale as found before, and Figure 34 displays the output elasticities. Due to the presence of markups the output elasticity with respect to labor is (much) higher than what is found using the labor share (i.e., 0.8 vs 0.65). When doing the growth decomposition I will take θ^K and θ^H to be the average across two subsequent periods.

In order to do the above growth accounting decomposition I use the historical multifactor productivity measures from the Bureau of Labor Statistics. This comprises data on output, hours worked, capital services and labor composition to

approximate h , all for the private business sector from 1948 onwards. In order to compare my growth decomposition with the literature, panel A of Table 8 shows the decomposition when perfect competition and constant returns to scale are assumed. I split the sample period in the same sub-periods as Jones (2016) does to highlight the productivity slowdown. As is a stylized fact, the majority of growth in output per hour for the period 1948-2015 comes from growth in the Solow residual. In the period 1948-1973 the Solow residual grows at 3.2% while it grows only at 0.8% during 1973-1995. This is the productivity slowdown. After 1995 the Solow residual increases rapidly again at a rate of 2.2% although not as fast as pre-1973. After the financial crisis, productivity growth slows down again.

Panel B of Table 8 shows the decomposition of growth when allowing for imperfect competition and the returns to scale being different from one, using my estimates for the output elasticities. I only estimate the output elasticity from 1964 onward. I impute the output elasticities for the period 1948-1963 using the averages for the period 1964-1973. The decomposition over the period 1948-2015 is very similar as in panel A although the Solow residual grows slightly faster in panel B. The contribution of the returns to scale being different from 1 is small, but negative because the returns to scale are less than 1. Furthermore, looking at the difference between the periods 1948-1973 and 1973-1990 the drop in the growth rate of the Solow residual between these two periods is in panel B not as big as in panel A, although still substantial. The fall in the growth rate of the Solow residual in panel B is with 2.1 percentage points 0.3 percentage points lower than in panel A. Thus, allowing for imperfect competition and non-constant returns to scale attenuates the productivity slowdown, but by no means makes the productivity slowdown disappear.

To obtain the estimate of the output elasticity I estimate the returns to scale at the firm level whereas for the growth accounting exercise the returns to scale at the aggregate is needed. Both might not coincide with each other. Therefore, panel C of Table 8 shows the decomposition when constant returns to scale are assumed such that the output elasticities are simply the cost shares. The results in panel C are very similar to the results in panel B.

Chapter 2

Profits and the Marginal Product of Capital Around the World

When international capital markets are frictionless the marginal product of capital will be equalized across countries.¹ When, on the other hand, the marginal product of capital varies substantially across countries, this implies that there are severe frictions. In the latter case, global output could be increased by reallocating capital from countries with a low marginal product of capital to countries with a high marginal product of capital.

It is well-known that the capital-labor ratio in rich countries is substantially larger than the capital-labor ratio in poor countries.² All else equal, this implies a larger marginal product of capital in poor countries. However, countries with a low capital-input ratio might differ with respect to other factors as well, such as the level of human capital and total factor productivity, which would affect the marginal product of capital (Lucas, 1990). Caselli and Feyrer (2007) take into account cross-country differences in endowments of natural resources and differences in the price of capital relative to the output good, and find that the marginal product of capital is not correlated with development. To estimate the marginal product of capital, Caselli and Feyrer (2007) multiply the capital share with the nominal output-capital ratio. The difficulty is in obtaining the capital share. By assuming perfect competition and constant returns to scale, Caselli and Feyrer (2007) estimate this as 1 minus the labor and natural resource shares. However, when firms make profits, this would overestimate the capital share and therefore overestimate the marginal product of capital.

Using the method developed in Chapter 1, I estimate the marginal product of capital across countries while allowing for differences in the degree of competition and in the returns to scale across countries. I find that the marginal product of capital is lower in poor countries than in rich countries. *Prima facie*, this implies

¹To be precise, the return on capital is what is expected to be equalized across countries. The return on capital is the marginal product of capital minus the depreciation rate, plus capital gains. I will take into account heterogeneity in the depreciation rate.

²An increase in GDP per worker of \$1 is associated with an increase in capital per worker of \$3.8 in 2013 (Penn World Tables 9.1).

that global output could be increased by reallocating capital from poor to rich countries.

One potential reason for getting this slightly surprising result compared to the literature, is that poor countries have a higher profit share. However, when I correlate development with profitability I find no relationship. Another potential reason for why I estimate the marginal product of capital to be lower in poor countries is lower data quality in poor countries as measurement error in the capital stock leads to a downward bias of the marginal product of capital with my method. Correcting for measurement error, I find that the relationship between the cost of capital and development gets attenuated but is still weakly positive.

What explains the lower marginal product of capital in poor countries is that the depreciation rate in poor countries is lower than in rich countries. This could, for instance, be due to rich countries having more equipment relative to structures than poor countries. When there is heterogeneity in the depreciation rate, the marginal product of capital is no longer equalized across countries under efficient markets. This is because investors take into account that if the depreciation rate is higher, they will be left with less capital in the future, and therefore, demand a higher marginal product of capital. That is, the return on capital—the marginal product of capital net of depreciation—is equalized across countries when markets are efficient.³ I find that there is no relationship between development and the net marginal product of capital. Thus, there are no severe frictions limiting the flow of capital between rich and poor countries, and global output net of depreciation—net domestic product (NDP)—cannot be increased by reallocating capital from poor to rich countries or vice versa.

The second contribution of this paper is to estimate how the profit share has evolved across different regions over time. I find that the profit share shows an inverted U-shape in Europe between 1990 and 2015, with an overall increase of around 2 percentage points. Profits have also been rising in Asia, Latin America and North America. This does not mean that profits in all countries have been increasing. For instance, the profit share has not increased in Canada. The global profit share has been rising by around 2 percentage points from 1990 to 2015, which is somewhat less than the increase in the United States. Finally, richer countries have experienced a somewhat faster increase in profitability.

³For simplicity, and due to data constraints, I do not take into account differences in capital gains across countries.

I The Marginal Product of Capital

Caselli and Feyrer (2007) estimate the marginal product of capital by multiplying the capital share with the inverse of the capital-output ratio. The difficulty is in obtaining the capital share. They assume perfect competition and constant returns to scale such that the capital share is 1 minus the labor and natural resource shares. However, in the presence of profits, the profit share also needs to be subtracted in order to obtain the capital share. This means that when the profit share is positive, Caselli and Feyrer (2007) overestimate the marginal product of capital.

To estimate the marginal product of capital in the presence of profits, consider a firm i that maximizes profits at time t ,

$$\begin{aligned} \max_{\{P_{it}, Y_{it}, X_{it}^1, \dots, X_{it}^M, K_{it}\}} & P_{it}(Y_{it})Y_{it} - \sum_{m=1}^M P_{it}^{X^m} X_{it}^m - R_{it}P_{it}^K K_{it} \\ \text{s.t. } & Y_{it} = F_{it}(X_{it}^1, \dots, X_{it}^M, K_{it}). \end{aligned} \quad (1)$$

The firm uses as inputs capital, K_{it} , and M other inputs, X_{it}^m , according to the production function $Y_{it} = F_{it}(X_{it}^1, \dots, X_{it}^M, K_{it})$. These inputs could, for instance, be (different types of) labor and materials. Different types of capital are also allowed, but for notational simplicity I only consider one capital good here. The firm internalizes that production affects the output price, P_{it} , while (domestic) input markets are assumed to be competitive. That is, input prices $P_{it}^{X^m}$, P_{it}^K and R_{it} do not depend on the input quantity demanded.

The first-order condition with respect to capital gives that the marginal revenue product of *real* capital, $MRPK_{it}$, divided by the price of capital, P_{it}^K , equals the cost of capital, R_{it} ,

$$\frac{MRPK_{it}}{P_{it}^K} = R_{it}.$$

The marginal revenue product of *real* capital is the increase in revenue when the *real* capital stock, K_{it} , increases by one. This marginal product divided by the price of capital, $\frac{MRPK_{it}}{P_{it}^K}$, is the increase in revenue when the *nominal* capital stock, $P_{it}^K K_{it}$, increases by one. The latter is predicted to be equalized across countries when capital is fully mobile, and when the depreciation rate and the price growth of the capital good is identical across countries. That is, an investor should get the same compensation independent of whether it invests capital worth one US dollar in the United States or in Indonesia. In what follows, I will refer to the marginal product of nominal capital as the marginal product of capital.

I here assume profit maximization while in Chapter 1 I assumed cost minimization. The difference is that cost minimization gives a link between the valued marginal product of capital and the cost of capital while profit maximization gives a link between the marginal revenue product of capital and the cost of capital. In this chapter I am interested in the capital allocation and therefore, I am interested in the marginal revenue product. Nevertheless, when marginal revenue equals marginal costs, the valued marginal product of capital and the marginal revenue product of capital are closely related to each other, namely, $\frac{MRPK_{it}}{P_{it}^K} = \frac{VMPK_{it}}{\mu_{it}P_{it}^K}$ where μ_{it} is the markup.⁴ Therefore, the first-order conditions implied by profit maximization are the same as under cost minimization, and the machinery developed in Chapter 1 can be used here as well. Estimating the marginal revenue product of capital comes down to estimating the cost of capital R .

Thus, assuming in addition to profit maximization that the production function is homogeneous of a constant degree gives a linear relationship between the sales-input ratio and the capital-input ratio. The intercept equals the price-average cost ratio while the slope equals the marginal revenue product of capital times the price-average cost ratio (or, equivalently, the slope equals the valued marginal product of capital divided by the returns to scale). Under the assumption that the cost of capital is equalized across producers within a country, the slope and intercept are identified by running a regression. Dividing the slope coefficient by the intercept coefficient gives the cost of capital or the marginal revenue product of capital, which is the object of interest. I refer the reader to Chapter 1 for details on the estimation.

Note that to estimate the marginal revenue product of capital, I do not need to make any assumptions about the returns to scale. This is different from Caselli and Feyrer (2007) who assume constant returns to scale.⁵ Furthermore, as I do not have to specify (or estimate) the production function, I allow for differences in technology across countries. This implies that countries can also vary in their

⁴This equality is easily obtained from the condition that marginal costs equal marginal revenue. The marginal cost is by definition of the markup equal to the price divided by the markup, and marginal revenue is $\frac{\partial Rev}{\partial Y} = \frac{\frac{\partial Rev}{\partial Y} \frac{\partial Y}{\partial K}}{\frac{\partial Y}{\partial K}} = \frac{MRPK}{MPK}$. Equalizing this with the price divided by the markup gives that $MRPK = \frac{VMPK}{\mu}$, where I have used that $VMPK = P \cdot MPK$. Alternatively, one can use the following derivation. Note that taking the derivative of revenue with respect to capital gives that $MRPK$ equals $P'(Y)Y \frac{\partial F}{\partial K} + P \frac{\partial F}{\partial K} = \left(P'(Y) \frac{Y}{P} + 1 \right) VMPK$. To get an expression for $P'(Y) \frac{Y}{P} + 1$, rewrite the profit maximization problem as $\max_{Y_{it}} P_{it}(Y_{it})Y_{it} - C_{it}(Y_{it})$, where $C_{it}(Y_{it})$ is the cost function. The associated first-order condition is $P'(Y)Y + P = C'_{it}(Y_{it}) \equiv mc_{it}$. Using that the marginal cost, mc_{it} , equals the price divided by the markup by the definition of the markup gives after rewriting that $P'(Y) \frac{Y}{P} + 1 = \frac{1}{\mu}$. Plugging this into the earlier equation gives that $MRPK = \frac{VMPK}{\mu}$.

⁵Note that Caselli and Feyrer (2007) estimate the valued marginal product of capital but this concept is identical to the marginal revenue product of capital when markets are competitive, which is the assumption in their paper.

human capital. The reason is that the factors of production, X^m , could refer to different types of workers in terms of their education level. And these different types are allowed to enter the production function in an arbitrary way as long as homogeneity is preserved. Nonetheless, for the estimation, only data on total operating expenses is needed and no information on how many employees with a certain education level each firm has is required.

An alternative method to estimate the marginal product of capital using micro data would be to specify and estimate a production function using standard IO techniques, of which then the analytic derivative could be taken. However, this has as its downside that if one wants to allow for differences in human capital, as is shown to be important by Lucas (1990), one has to know the educational attainment of the employees of each firm. I am not aware of any such dataset covering a wide range of countries and therefore, such an approach would be infeasible.

Data

The data used in this paper comes from Compustat. Compustat collects data on the balance sheet and income statement of a large number of firms across the world from 1987 onwards. These are mainly publicly listed firms. I estimate the profit share for each country-year pair separately, where I identify the country of a firm by the location of its headquarters. As there are relatively few firms for each country, I run the regression at the economy level and not at the industry level. See Chapter 1 that, for the United States, doing the estimation at the economy level gives similar results as when the estimation is done at the industry level. As in Chapter 1, I control for heterogeneity in the cost of capital. I use as controls the depreciation rate and the capital-input ratio. These controls control for heterogeneity across firms within a country, and do not control for heterogeneity in, let us say, depreciation rates across countries. Since there are relatively few firms for some countries, I decided to not include all possible controls, such as firm size and financial leverage, as these controls did not seem to impact my estimate of the cost of capital for the United States. Combining firms across all industries has as its downside that the heterogeneity in the cost of capital might be larger due to differences in risks, and therefore risk premia, across industries. To control for this, I include as an additional control the standard deviation of the growth rate across firms within a 2-digit industry. I trim each variable that shows up in the regression by the bottom and top percentile for each country-year pair.⁶

⁶I do not trim the 'risk' variable as this only differs across industries.

Profits and the Marginal Product of Capital Around the World

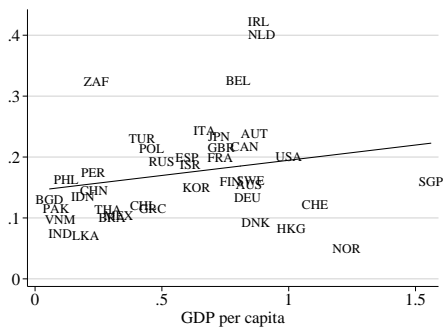


Figure 1: Cost of capital R , 2013

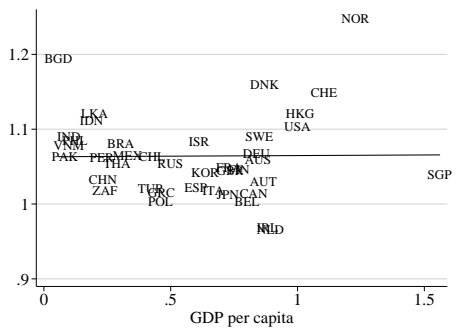


Figure 2: Price-average cost ratio, 2013

The variables I use are identical to the variables I use in Chapter 1 for the United States. The capital stock includes externally purchased intangibles, but not internally developed intangibles. The costs for developing intangibles are fully expensed and are therefore part of operating expenses. The depreciation rate is constructed using data on the reported value of depreciation and the capital stock.

Results

Figure 1 shows how the cost of capital is correlated with GDP per capita in 2013, where the cost of capital (as do the other variables in this section) refers to the 5-year average around 2013, and GDP per capita is normalized to the US level.⁷ There is a positive relationship between income and the cost of capital. The coefficient of regressing the cost of capital on normalized GDP is around 0.05, which is substantial, although not significant at the 5% level. Figure 10 in Appendix A shows that the pattern is very similar in 1998 for this figure and all coming figures.

That there is a positive relationship between income and the marginal revenue product of capital is somewhat surprising as the usual result is a negative relationship (Lucas, 1990) or no relationship (Caselli and Feyrer, 2007). One potential reason for this discrepancy is that Caselli and Feyrer (2007) assume that profits are zero. If poor countries have a higher profit share than rich countries, Caselli and Feyrer (2007) would overestimate the marginal product of capital to a larger extent in poor countries. However, Figure 2 shows that there was no relationship between development and the price-average cost ratio.⁸ And Figure 3 shows that the reason why richer countries have a higher marginal revenue product of

⁷GDP per capita is obtained from the World Bank World Development Indicators.
⁸In 1998, poorer countries had a weakly larger price-average cost ratio than rich countries, but this is not quantitatively large enough to explain the differences in the cost of capital.

Profits and the Marginal Product of Capital Around the World

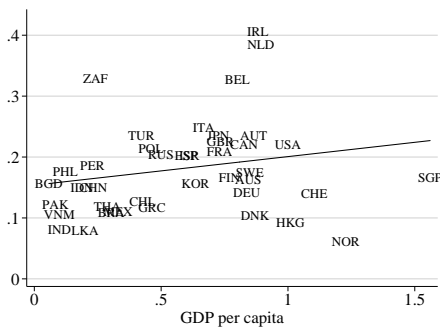


Figure 3: Valued marginal product of capital divided by the returns to scale, 2013

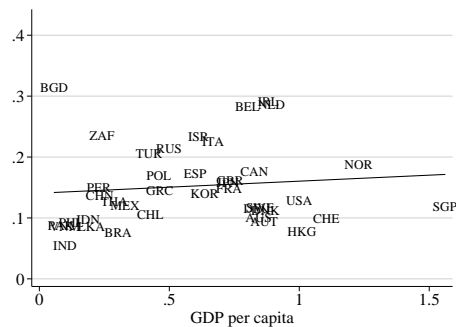


Figure 4: Cost of capital R (lag as instrument), 2013

capital is that they have a higher valued marginal product of capital divided by the returns to scale $\frac{VMPC}{PK\phi}$ (i.e. this is the slope coefficient in the regression), where ϕ are the returns to scale.

Thus, these results suggest *prima facie* that in poor countries production is too capital intensive and that global output would increase when capital is reallocated from poor to rich countries.

Measurement Error

One possible reason for finding a lower cost of capital in poor countries is that data quality is lower in poor countries. Classical measurement error of the capital stock would lead to attenuation bias meaning that the cost of capital is downward biased while the price-average cost ratio is upward biased. Thus, if there is more measurement error in the capital stock in poor countries than in rich countries, this could explain why I find the cost of capital to be smaller in poor countries.

To test this hypotheses, I instrument the capital-input ratio with the lagged capital-input ratio. This instrument deals with measurement error of the capital stock if the measurement error is not correlated across two periods. Figure 4 shows that, after taking measurement error into account, there is still a positive relationship between the cost of capital and development. However, the strength of this relationship is attenuated compared to the baseline. Increasing normalized GDP per capita by 1 increases the cost of capital by 0.02, while in the baseline this increase was 0.05.

Profits and the Marginal Product of Capital Around the World

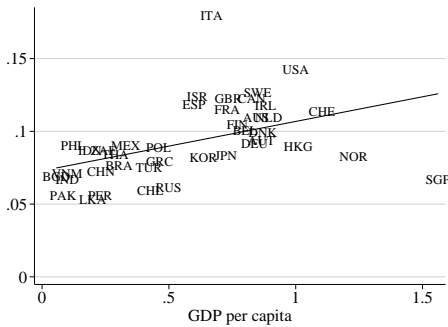


Figure 5: Depreciation rate, 2013

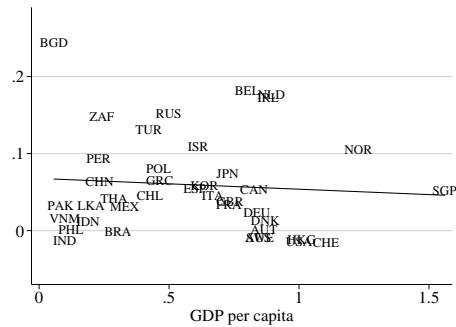


Figure 6: Cost of capital minus depreciation rate, 2013

Depreciation Rate

Another reason for there being a positive relationship between development and the marginal product of capital is that the composition of capital might differ across countries. Firms in rich countries might use more equipment relative to structures than firms in poor countries. Equipment has a higher depreciation rate than structures and therefore, the depreciation rate might be larger in rich countries. Figure 5 shows that rich countries indeed have a higher depreciation rate.

If there is heterogeneity in the depreciation rate, then the marginal product of capital would not be equalized across countries under efficient markets. The depreciation rate is a cost as this leads to less capital in the future. Therefore, an investor wants to be compensated for the higher depreciation rate by a higher marginal product. Thus, under efficient markets, the marginal product of capital net of depreciation would be equalized across countries.

Figure 6 shows that there is no relationship between development and the cost of capital after subtracting the depreciation rate. This figure uses the cost of capital corrected for measurement error. Thus, this implies that there are no severe frictions limiting the flow of capital between rich and poor countries and that NDP would not increase if capital is reallocated from rich to poor countries or vice versa.

II Profits over Time and across Continents

I now turn to how profitability has changed over time across different continents. For the interpretation of the results at the country level, it is problematic that

Profits and the Marginal Product of Capital Around the World

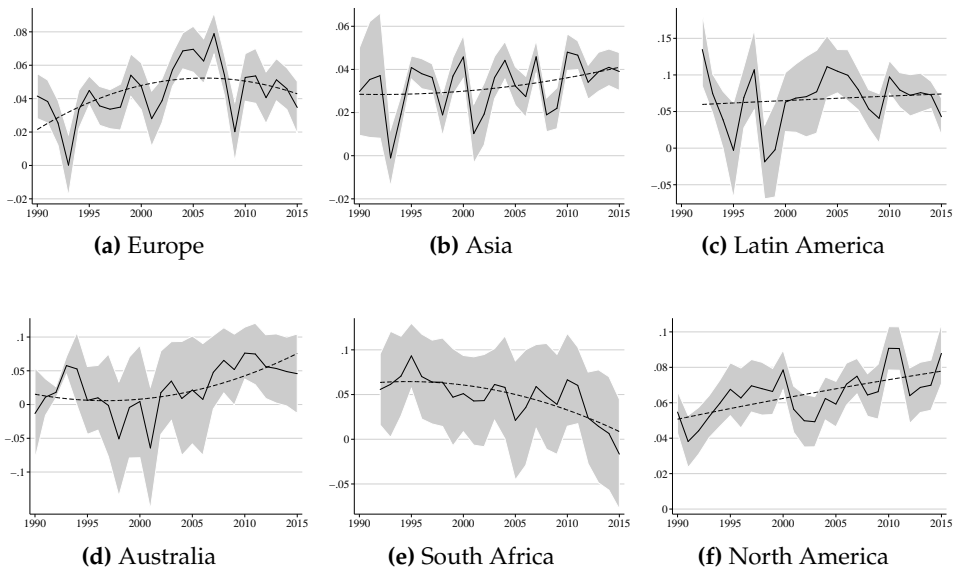


Figure 7: Profit share across continents

the data mainly consists of publicly listed firms. These are mainly larger firms that tend to be exporters and multinationals. For instance, Volvo is a Swedish firm while only a fraction of its sales occur in Sweden. If Volvo charges a high markup on its sales in other countries, this will lead to a high measured profit share in Sweden while the actual profit share in Sweden might, in fact, be much smaller. Hence, the results for small open economies in this paper are not reliable. Therefore, I aggregate the profit share up to the continent level and this will be the main focus here. This is under the assumption that the profitability of a firm to a larger extent reflects the market conditions on its own continent than elsewhere. Finally, I aggregate all countries and obtain the global profit share. As weights I use the GDP of each country.⁹

Figure 7 shows the profit share for each continent. In Europe the profit share displays an inverted U-shaped pattern over time with a peak in the mid-2000s. Overall, profits as a share of sales in Europe seem to have increased from being around 3% in the 1990s to being around 5% in the 2010s. There seems to be a weakly increasing profit share in Asia and Latin America, although the series are too erratic to tell with any certainty. For Oceania, there is only data for Australia and profits have been increasing in Australia. South Africa is the only African country with sufficient data to estimate the profit share, and profits have been declining in South Africa. Finally, in North America the profit share has been

⁹The results are similar when using sales in Compustat as weights, see Figure 15.

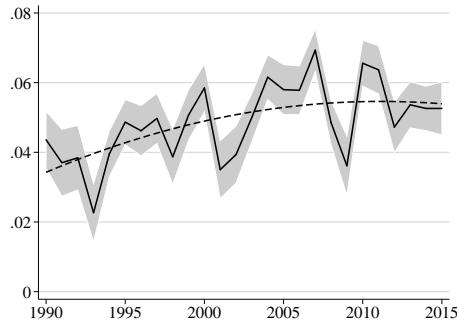


Figure 8: Global profit share

increasing by around 3 percentage points from 1990 until 2015.

These numbers mask important heterogeneity across countries. Figure 11, 12, 13 and 14 in Appendix A show the profit share for the European, Latin American, North American and Asian countries, respectively. Although profits have been increasing at the European level, there are several European countries for which the profit share has not been increasing or has even been declining, such as Belgium, Ireland and Greece. That there is heterogeneity across European countries is not surprising as there is also heterogeneity in the decline of the labor share across Europe (Gutiérrez and Piton, forthcoming; Cetto et al., 2019). Several Asian countries show a decline in profits over time. The increase in profits in Latin America is driven by Chile and Mexico while the profit share in Brazil and Peru has been declining. Finally, profits in Canada have not been increasing.

Aggregating to the global level, Figure 8 shows the global profit share. The global profit share has been increasing from around 4% of sales in 1990 to around 5–6% of sales in 2015. The increase in global profits is about half of what is observed for the United States where the profit share has increased by about 4 percentage points during this period. This differs from De Loecker and Eeckhout (2019) who estimate the global markup and find that it has been increasing at almost the same rate as the markup in the United States.¹⁰ One potential explanation for this difference in results between the two approaches is that De Loecker and Eeckhout (2019) do not estimate the output elasticity (which is a crucial part of their estimator of the markup) for each country separately but assume that it is the same as in the United States. Instead, I estimate the cost of capital for each country separately and do not assume that technology is the same across countries. If, for

¹⁰The baseline estimate in De Loecker and Eeckhout (2019) is the sales-weighted average markup across countries, which has increased at a slightly slower pace than the US markup. When they weight countries by GDP, the global markup has increased faster than the US markup.

Profits and the Marginal Product of Capital Around the World

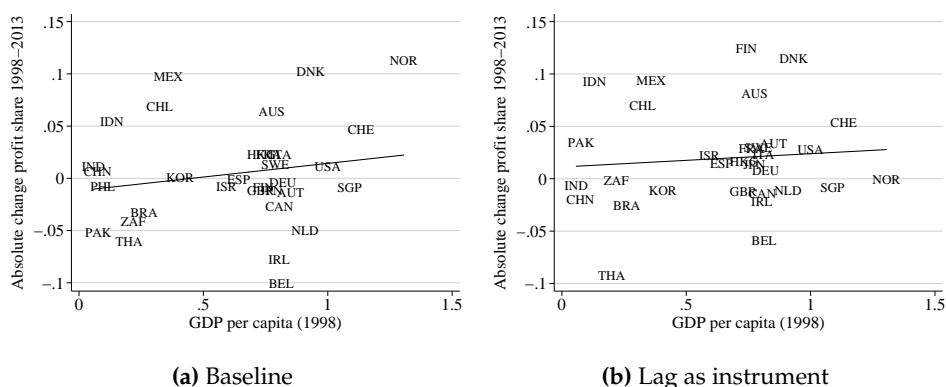


Figure 9: Change profit share

instance, labor costs differ across countries, this might lead to differences in factor augmenting technologies, and hence output elasticities, due to directed technical change.

Finally, Figure 9a shows that richer countries have experienced a somewhat faster increase in profits than low-income countries between 1998 and 2013. However, there are several low-income countries that experienced a substantial increase in profits as well. One reason for finding that low-income countries have experienced a slower growth in profitability is measurement error. If measurement error has been declining over time for low-income countries then I underestimate the rise in profits for these countries. Figure 9b shows that this is indeed the case to some extent since after using the lag as an instrument, the relationship between income and the change in the profit share is attenuated.

Figure 16 in Appendix A shows that also the global profit share increases by 2-3 percentage points when using the instrument compared to 1-2 percentage points when not using the instrument. But also the US estimate of the rise in the profit share is larger when using the lag as instrument. Thus, it is still the case that the global profit share has increased at a slower rate than the US profit share. Finally, for most continents, the profit share looks roughly similar when comparing the results with and without instruments, except for Asia, for which there is a more clearly marked increase in the profit share when using the instrument compared to the baseline (Figure 16).

III Conclusions

This paper studies whether the marginal product of capital is related to development and how the profit share has evolved globally over time. I find that richer countries have a higher marginal product of capital, but that this is mainly driven by richer countries having a higher depreciation rate. After subtracting the depreciation rate from the marginal product of capital, there is no relationship with development. Thus, global NDP would not increase after reallocating capital from rich to poor countries or vice versa.

Compared to the literature (Caselli and Feyrer, 2007), I allow for imperfect competition and non-constant returns to scale. Thus, both the degree of competition and the returns to scale are allowed to differ across countries in my framework. (Caselli and Feyrer, 2007) find that the marginal product of capital is not related to development and do not take into account differences in depreciation rates. Instead, I find that richer countries have a higher marginal product of capital, but that the marginal product of capital net of depreciation is not related to development.

Furthermore, I find that the rise in profits is not only a US phenomenon but that the profit share has also been increasing on other continents. This means that the global profit share has been increasing, although at a lower rate than the increase in the US. Finally, I find that richer countries have experienced a somewhat faster increase in profitability than poorer countries.

References

- Caselli, F. and J. Feyrer (2007) "The Marginal Product of Capital," *Quarterly Journal of Economics*, Vol. 122, No. 2, pp. 535–568.
- Cette, G., L. Koehl, and T. Philippon (2019) "Labor Shares in Some Advanced Economies," Working Paper 26136, National Bureau of Economic Research.
- De Loecker, J. and J. Eeckhout (2019) "Global Market Power," Manuscript.
- Gutiérrez, G. and S. Piton (forthcoming) "Revisiting the Global Decline of the (Non-Housing) Labor Share," *American Economic Review: Insights*.
- Lucas, R. E., Jr. (1990) "Why Doesn't Capital Flow from Rich to Poor Countries?" *AEA Papers and Proceedings*, Vol. 80, No. 2, pp. 92–96.

Appendix A Additional Figures

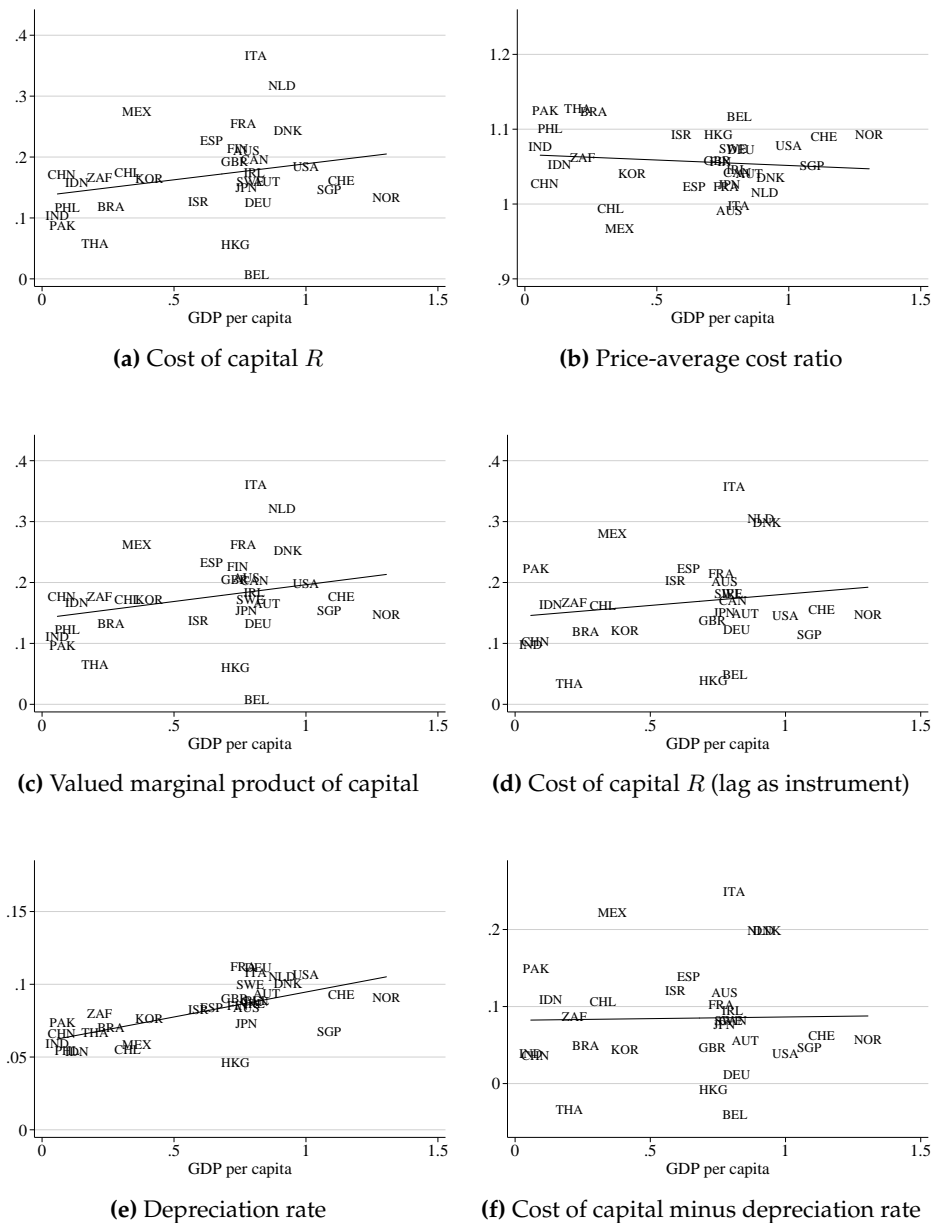


Figure 10: Results for 1998

Profits and the Marginal Product of Capital Around the World

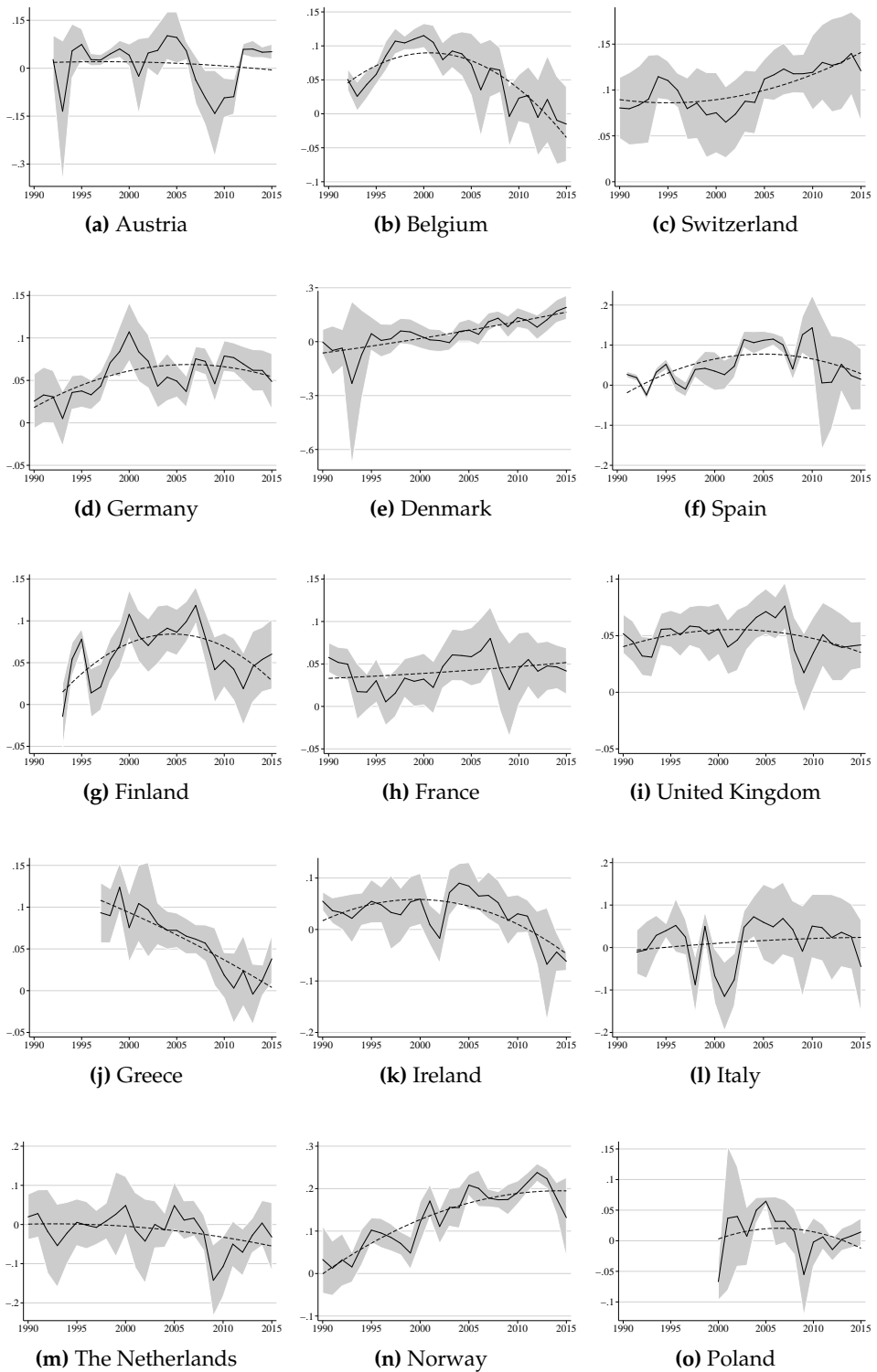


Figure 11: Profit share across European countries

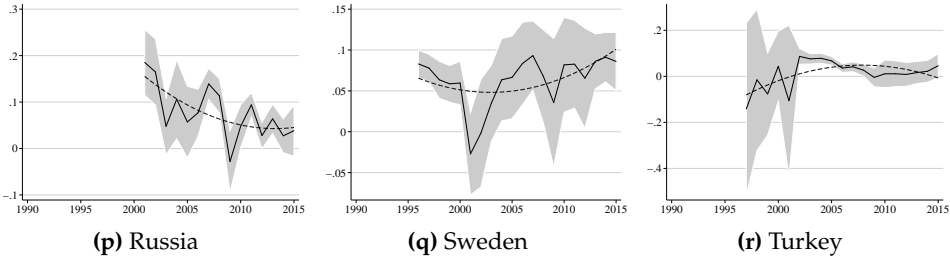


Figure 11: Profit share across European countries - continued

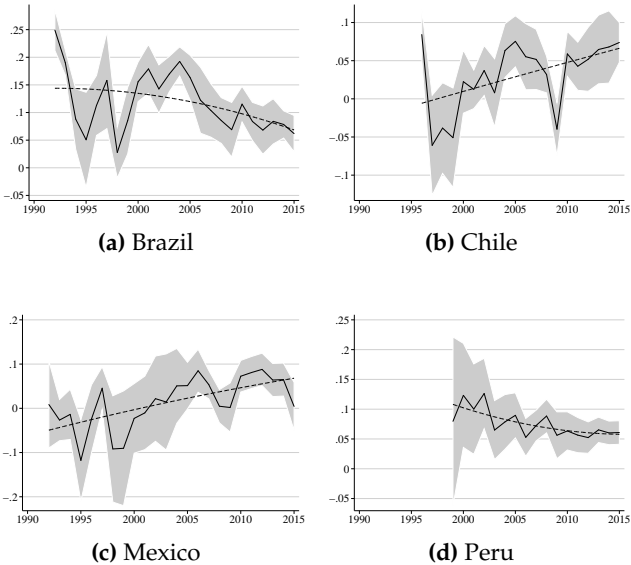


Figure 12: Profit share across Latin American countries

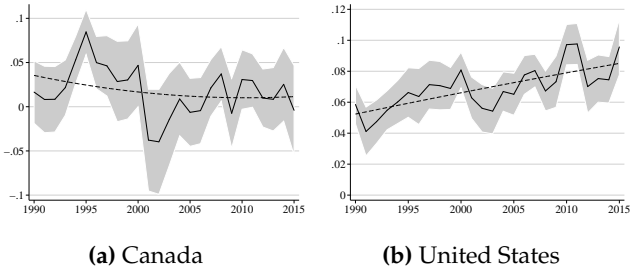


Figure 13: Profit share across North American countries

Profits and the Marginal Product of Capital Around the World

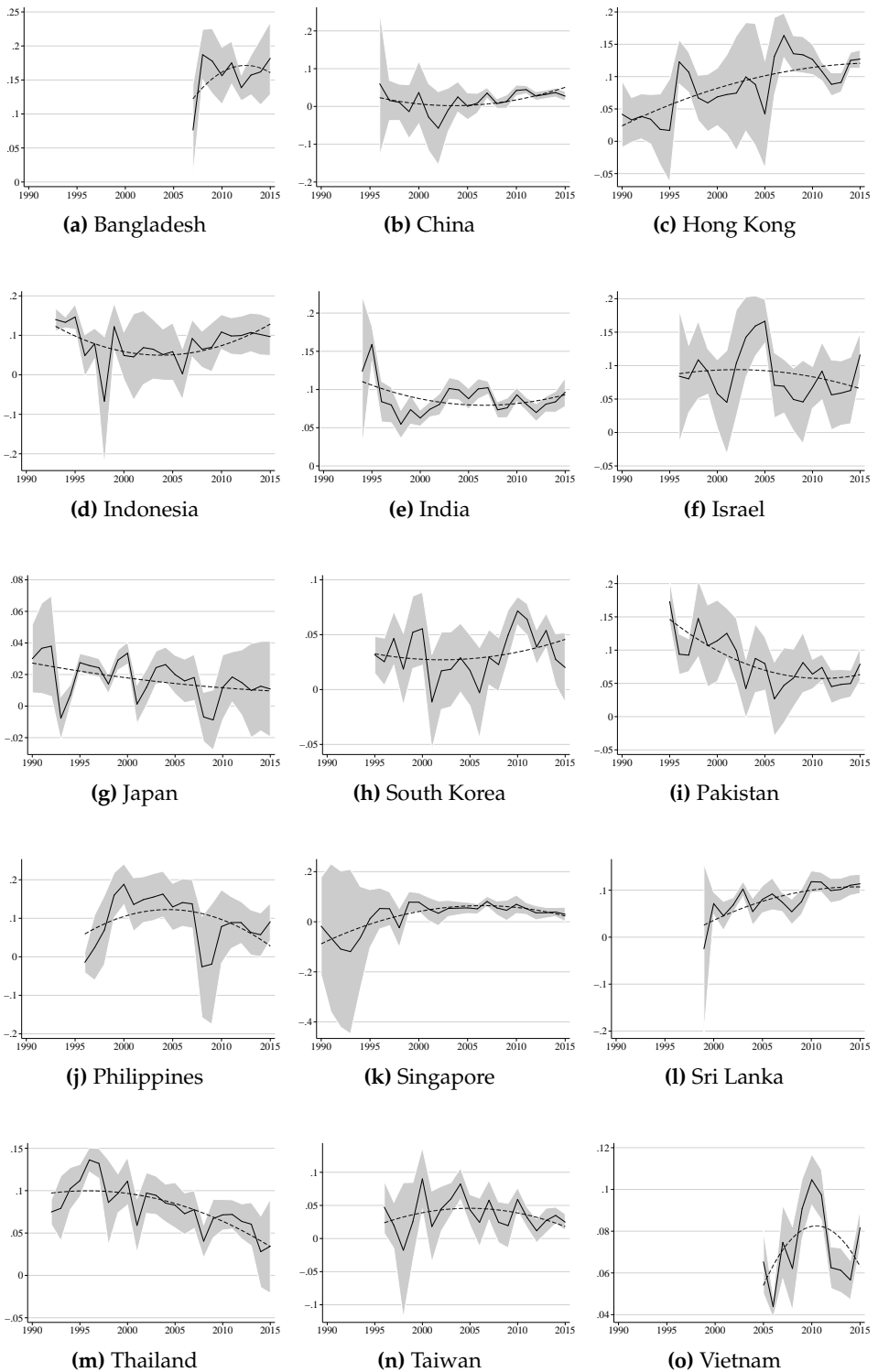


Figure 14: Profit share across Asian countries

Profits and the Marginal Product of Capital Around the World

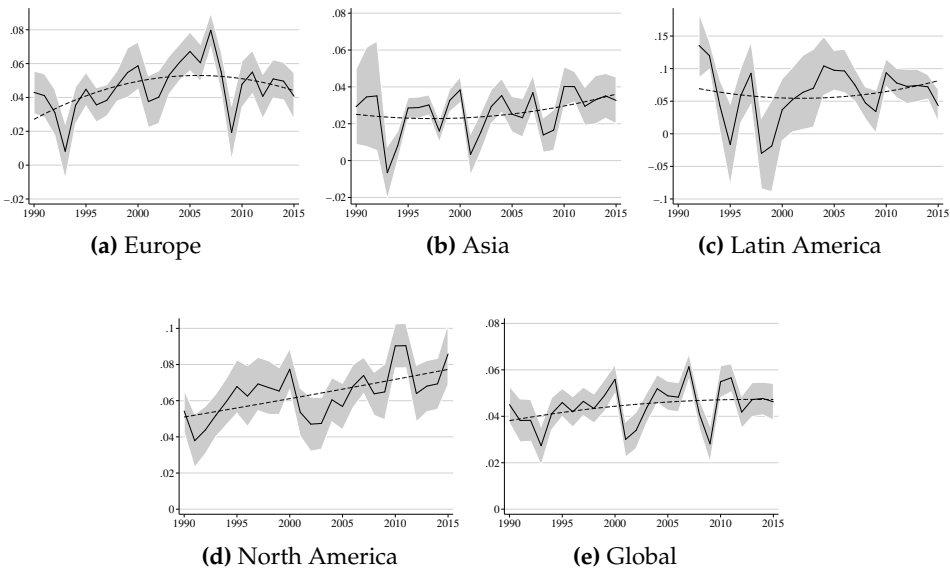


Figure 15: Profit share across continents (sales weights)

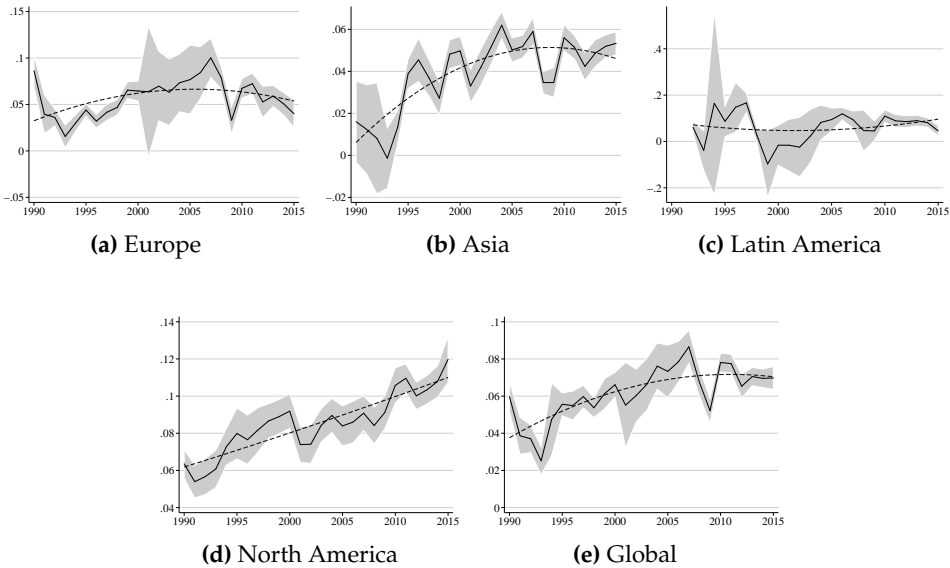


Figure 16: Profit share across continents (lag as instrument)

Chapter 3

The Life Cycle of Profits

Chapters 1 and 2 show that profits have been increasing during the last decades.¹ At the same time firm entry has been declining (Decker et al., 2014). In this chapter, I provide quantitative evidence for a new hypothesis that can explain both the rise in profits and the fall in firm entry, namely, that profits have become more back-loaded over the life cycle of the firm. Or, put differently, that the profits-firm age relationship has become steeper over time.

An entrepreneur enters the market when the value of having a firm is larger than the cost of entry. Thus, in equilibrium, entry costs equal the value of the firm,

$$c_e = V_0 = \sum_a \beta^a \Pi_a, \quad (1)$$

where c_e denotes the entry costs. The value of the firm upon entry is denoted by V_0 , which equals the discounted sum of profits Π_a over the life cycle of the firm, where firm age is denoted by a and the discount factor is denoted by $\beta < 1$. Naturally, profits might differ over the life cycle, and that is why profits are indexed by a . Due to discounting, total profits earned over the life cycle generally do not equal entry costs. And when profits are back-loaded, total profits exceed the discounted sum of profits and therefore exceed the entry costs (Atkeson and Kehoe, 2005).

Consider the following thought experiment. Suppose that the economy is initially in equilibrium, but that, over time, profits shift from a young firm age to an old age, in such a way that total profits over the life-cycle remain constant. That profits become more back-loaded lowers the value of entering the market as profits that appear later are more heavily discounted than profits that arrive early. Therefore, the entry condition (1) no longer holds as entry costs now exceed the value of the firm. Thus, as a response, firm entry goes down and hence competition goes down. This will, in turn, increase markups and the profits firms are making, until the entry condition holds again. Figure 1 illustrates this graphically. The blue solid line is an example of what profits over the life cycle might look like initially. Older firms make more profits than younger firms but

¹That profits have been increasing is also found by others in the literature (e.g., Barkai (2017) and De Loecker et al. (2018)).

The Life Cycle of Profits

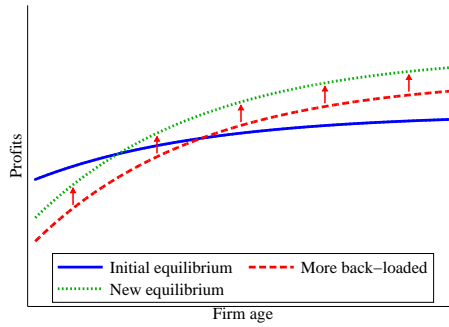


Figure 1: Illustration of profits over the life cycle

the profits-age relationship is quite flat. The red dashed line shows what the profits-age relationship might look like when profits become more back-loaded but before the equilibrium response; young firms make less profits and old firms make more profits, but the (undiscounted) sum of profits over the life cycle is the same as in the initial equilibrium. The green dotted line shows what profits might look like after the economy has moved to a new equilibrium. The discounted sum of profits is the same in the new as in the initial equilibrium, but total profits over the life cycle are larger in the new equilibrium than in the initial equilibrium. Therefore, an econometrician that observes the cross-section, and therefore weights all ages the same, concludes that aggregate profits have gone up. In Figure 1, it is assumed that the equilibrium response takes the form of a parallel shift in profits, but this does not necessarily need to be the case. Young and old firms might be differentially affected by the decrease in competition.

Section I shows that the profits-age relationship has indeed become steeper over time. Before 2000, older firms were only making moderately more profits than young firms, while after 2000 older firms were making much more profits than younger firms. Moreover, young firms after 2000 are making about as much profits as they did during the 1980s and the 1990s. These results hold for different estimates of the cost of capital, are not driven by changes in the industry composition over time and are not driven by the great recession. The results are also not sensitive to outliers as the life cycle of profits has also changed for the median firm. Thus, the hypothesis put forward in this chapter does not only apply to the so-called superstar firms.

Next, I decompose the change in the life-cycle pattern of profits into three components: the life-cycle patterns of i) profits as a share of sales, ii) firm size measured by sales and iii) the covariance between the profit share and firm size. I find that after 2000, firms younger than 15 years had a lower profit share, were

about equally large and had a larger covariance term compared to firms younger than 15 years before 2000. On the other hand, older firms have not experienced any change in their (unweighted average) profit share, while they have become much bigger, and only experienced a weak increase in the covariance between the profit share and firm size.

To study the equilibrium response to a change in the profits-age relationship, I build a quantitative model in Section II. The model features oligopolistic competition à la Atkeson and Burstein (2008) built into an overlapping generations model with an occupational choice between being an entrepreneur or a worker. Agents make the occupational choice at the beginning of their life, comparing the present value of earning a wage in the labor market with the present value of future profits. The present value of earning a wage is reminiscent of the entry cost in (1). It is the opportunity cost of choosing to be an entrepreneur. Furthermore, oligopolistic competition ensures that there is a relationship between the number of firms and profits. A result of the model is that a firm's level of profits depends on its productivity. I allow productivity to depend on age. And in order to generate a changing profits-age relationship I vary how productivity changes with age. I set the productivity-age relationship such as to match the observed profits-age relationship at different points in time. When calculating the profits-age relationship, I normalize profits such that average profits do not vary. I then let the model find the *level* of profits for which agents are indifferent between being an entrepreneur and a worker. Thus, in terms of Figure 1, I match the blue and red line exogenously, but the green line (the new equilibrium) follows endogenously.

Section III shows that the changing profits-age relationship explains about two-thirds of the rise in profits and more than fully accounts for the fall in firm entry. These results are robust to changing parameter values.

Related literature This paper is foremost related to the recently emerging literature that studies why markups and profits have been increasing over the last decades. Explanations include consumer inertia (Bornstein, 2018), an increase in common ownership (Azar and Vives, 2019), IT improvements leading to a fall in the firm-level costs of spanning multiple markets (Aghion et al., 2019), falling interest rates (Liu et al., 2019) and a decline in knowledge diffusion between frontier and laggard firms (Akcigit and Ates, 2019). I contribute to this literature by quantitatively analyzing to what extent a changing life cycle pattern of profits can explain the rise in profits.

I Evidence

This section provides evidence on how the profits-age relationship has changed over time among Compustat firms in the United States. To estimate profits, I use the estimated capital costs from Chapter 1. In addition, I need firm age which is not directly available in Compustat. I obtain the year at which the firm was founded from Field and Karpoff (2002) and Loughran and Ritter (2004). They have compiled a dataset of the founding year of all firms that went public between 1975 and 2018. Then, the age of the firm in a given year is simply the founding year subtracted from the reporting year. The founding year refers to the year of incorporation and it is not straightforward to measure the founding year due to companies changing their names and there being mergers and acquisitions. In case of a name change, these authors take the original year of cooperation and in case of mergers and acquisitions, they take the founding age of the oldest entity, or if there is a substantial difference in size, they take the founding year of the largest entity. Companies for which they do not have a reliable founding year are not included.

As the founding year is only available for a subset of firms, there is not enough data for the years before 1980 to calculate the life-cycle pattern of profits.² I bin firms into age bins of five years and I collapse the data into decades starting with the 1980s and ending with the period 2010–2015. As an example, suppose that I observe a firm that was founded in 1986 from 1990 onward. Then, in order to calculate profits by age bin in the 1990s, that firm is used twice to calculate profits of firms in the 0–5 years age bin (namely the observations for 1990 and 1991). The observations of that firm for the years 1992 until 1996 are used to construct profits among firms six to ten years old in the 1990s. And likewise for the years 1997 until 1999. The observations of this firm for the years 2000 and 2001 are used to construct profits among firms 11–15 years old in the 2000s etc.

Figure 2 shows the resulting average profits by firm age for the four different time periods. Profits are deflated by the GDP deflator and the vertical lines denote 95% confidence intervals in this and all subsequent figures. The profits-age relationship has become steeper over time. During the 1980s and the 1990s, old firms were only making moderately more profits than young firms. But after 2000, old firms started to make much more profits while the profits of young firms hardly changed. In 2005, firms younger than ten years old made essentially no profits, while firms that were more than twenty years old made more than 150 million dollar of profits on an annual basis. Figure 10a in Appendix A shows that

²The firm age is observed for around 20% of the observations for which I have estimated profits after 1980. This leaves me with around 25,000 observations.

The Life Cycle of Profits

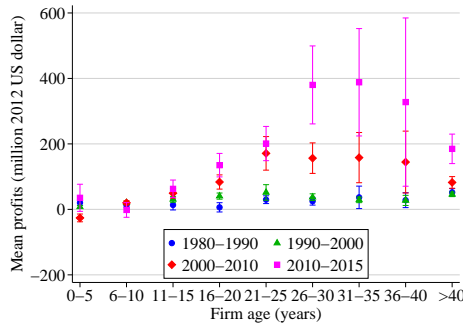


Figure 2: Profits over the life cycle by decade

the same pattern holds when considering median profits within an age-decade bin, and therefore this changing life-cycle pattern of profits is not driven by only a few firms becoming extremely profitable.

One possibility for this changing pattern is that the industry composition has changed over time. For instance, nowadays there might be more tech firms than there used to be, which might need to invest a great deal when young, which leads to a steeper profits-age relationship. To see to what extent a changing industry composition is driving the results, I weight 2-digit industries such as to keep the industry composition equal to that of the 1990s. Figure 10b shows that the same pattern emerges after correcting for the industry composition. Furthermore, Figure 10c shows that the patterns do not change after correcting for the industry composition within each age bin either. The latter industry correction deals with the concern that there might, for instance, nowadays be relatively more younger tech firms than in the earlier decades.

Moreover, the results are robust to using other estimates of the cost of capital. Figure 10d shows the results when the cost of capital varies across firms within an industry according to the controls I use in the regression. Figure 10e shows that the result is also robust to estimating the cost of capital with the required rate of return approach. Finally, the results are not driven by the great recession affecting younger firms to a larger extent than older firms since there is also a steeper life-cycle profile of profits for the period 2000–2005 (see Figure 10f).

One concern is that these results are for Compustat firms that tend to be older than the typical firm. However, Figure 11 shows the distribution of firm age in my sample (i.e., the firms after 1980 for which I observe firm age) and the majority of firms are younger than 25 years, and the mode is at around ten years. Figure 2 shows that the slope of profits with respect to age has increased over time for the age group six to thirty years. Therefore, for these results, it is not so much

The Life Cycle of Profits

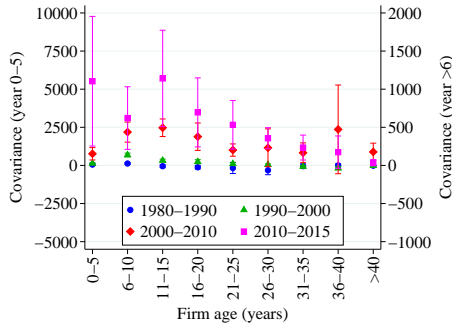


Figure 3: Covariance profit share and sales

of a concern that the typical Compustat firm is older than the typical firm in the economy. Nonetheless, the results could, of course, be driven by sample selection into Compustat and therefore, care should be taken into extrapolating these results to the rest of the economy. However, for these results to be driven by a changing sample selection, one would have to argue that firms that became public after 2000 (i.e., the relatively young firms that appear in the data for the later periods) were less profitable than they used to be. This does not seem likely since after 2000 IPOs have declined and one would expect this to have led to selection into more profitable firms among the young firms that decide to go public. However, it could, of course, be that the most profitable firms decided to stay private for a longer time. It is important to note here that Compustat also includes some firms before they go public. For instance, the first year Facebook appears in the data is 2008 while it went public in 2012.

Finally, Figure 12 shows the age distribution by decade. Over time, the average firm age has gone up. For the estimate of the oldest age bin in the 1980s there are no firms older than 125, but during the 2010s there are several firms that are around 150 years old.

Decomposition

Is the changing life-cycle pattern of profits due to a changing pattern of the profit share or of firm size? Profits of a firm equal its profit share times its size in terms of sales. Thus, average profits can be decomposed as

$$\overline{\text{Profits}} = \overline{\text{Profit Share}} \cdot \overline{\text{Sales}} + \text{cov}(\text{Profit Share}, \text{Sales}) . \quad (2)$$

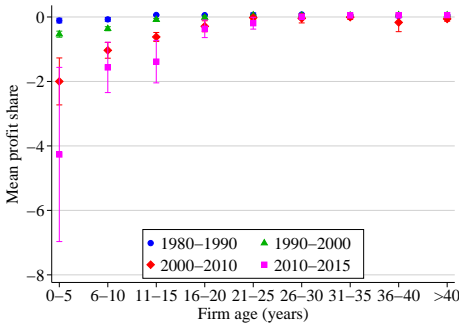


Figure 4: Profit share over the life cycle

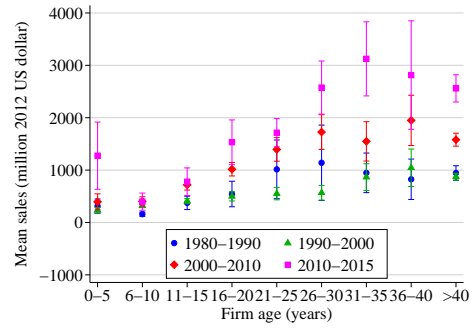


Figure 5: Sales over the life cycle

Figure 3 shows that the covariance between the profit share and sales has grown over time for all age bins, especially for the younger ages (note the different scale for the youngest age bin). Sales are here deflated by the GDP deflator. Based on equation (2) and that profits have not increased for the young firms, this means that for the young firms, the average profit share and/or average sales must be lower today compared to the 1980s and 1990s.

Figure 4 shows that the (unweighted) average profit share is much lower for the youngest firms after 2000 compared to before 2000, while the average profit share did not change much for the oldest firms. Figure 5 shows that the change in the life cycle pattern of firm size is very different. Young firms today are about equally large as young firms thirty years ago, but old firms are much larger than they used to be. Also these results are not driven by outliers as the same patterns hold when considering the median (see Figure 13).

That the covariance between the profit share and sales within an age bin has increased over time could be due to an increase in the variance of the profit share and/or sales, and/or due to an increase in the correlation between the profit share and sales within an age bin. In Chapter 1, I found that the relationship between firm size and the profit share has become larger over time, but this was without controlling for firm age. Thus, this result could potentially been fully driven by the life cycle profile of both firm size and the profit share. However, this is not the case. Figure 14 shows that the relationship between profits and firm size within an age bin has become stronger over time. However, this is only the case for the youngest firms, while for firms older than twenty years, the relationship between profits and firm size within an age bin has not changed. Thus, the increase in the covariance for firms older than twenty years is driven by an increase in the variance of the profit share and/or firm size. Moreover, these results are not driven by any outliers since a quantile regression gives similar results as the OLS.

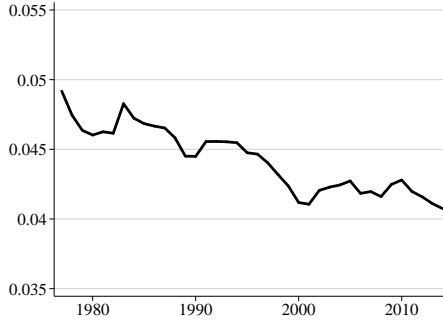


Figure 6: The number of firms relative to the population. Source: Longitudinal Business Database (US Census Bureau).

Firm Entry

It is well-known that firm entry has been declining over time. One way of illustrating this is by looking at the number of firms relative to the population, where the population is defined as the number of employees in the private sector plus the number of entrepreneurs.³ The reason for using this statistic is that it gives a clear mapping between the data and the model as will become clear momentarily. Figure 6 shows that the share of entrepreneurs has been declining by around half a percentage point from somewhat above 4.5% in the early 1980s to just above 4% nowadays.

II Model

The model is an overlapping generations model with an occupational choice. The economy is populated with N agents that live for T periods and, at birth, they have the option to become either an entrepreneur or to become a worker that supplies one unit of labor to the firms in a competitive market. For simplicity, it is assumed that agents do not save and thus consume all their income within each period, and that the agent's utility is linear in consumption. Firms (or, equivalently, entrepreneurs) only require labor for production. Firms engage in imperfect competition in the product market à la Atkeson and Burstein (2008). The amount of profits that firms make depends on productivity. Productivity is taken to be exogenous, but varies over the life cycle. Thus, also profits depend on age. Instead, for a worker, the wage w , and therefore consumption, is constant over time, leading to $V^{worker} = \sum_{a=0}^{T-1} \beta^a w = \frac{1-\beta^T}{1-\beta} w$. In equilibrium the value of

³I take the number of entrepreneurs as the number of firms.

being an entrepreneur and a worker is equal to each other,

$$V^{ent} = V^{worker}.$$

I will specify the value of being an entrepreneur, V^{ent} , in more detail later.

Firms

Suppose that there is a continuum of sectors, indexed by $i \in [0, 1]$, that produce a good y_i . These sectoral outputs are used by a competitive firm to produce final output, c , according to a CES production function with elasticity of substitution η :

$$c = \left(\int_0^1 y_i^{\frac{\eta-1}{\eta}} di \right)^{\frac{\eta}{\eta-1}}. \quad (3)$$

The first-order conditions of the competitive firm give the inverse demand functions

$$\frac{P_i}{P} = \left(\frac{y_i}{c} \right)^{-1/\eta}, \quad (4)$$

which combined with the zero profit condition give that the price index, P , for final consumption is given by

$$P = \left(\int_0^1 P_i^{1-\eta} di \right)^{\frac{1}{1-\eta}}. \quad (5)$$

Each sector consists of J firms, each producing a distinct variety in quantity q_{ij} . The output of a sector is given by the CES aggregate of these J varieties with elasticity of substitution ρ :

$$y_i = \left(\sum_{j=1}^J q_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}. \quad (6)$$

As before, this gives as inverse demand function

$$\frac{P_{ij}}{P_i} = \left(\frac{q_{ij}}{y_i} \right)^{-1/\rho}, \quad (7)$$

and as sectoral price index

$$P_i = \left(\sum_{j=1}^J P_{ij}^{1-\rho} \right)^{\frac{1}{1-\rho}}. \quad (8)$$

In a given period each firm produces according to a constant returns to scale production function with labor L as its only input:

$$q_{ij} = z_{ij} L_{ij}. \quad (9)$$

Productivity, z_{ij} , might vary by firm age.

As there is only a limited number of firms in a sector, firms do have market power. I assume that firms engage in Cournot competition, where firms choose their quantities, q_{ij} , taken as given the quantities chosen by the other firms. Firms do internalize that their quantity supplied affects the sectoral price, P_i , and quantity, y_i . However, as each sector is infinitesimally small, firms take the wage, w , and the price P and quantity c of final consumption as given. This is a static game as there are no adjustment frictions.

The above implies that firm ij solves the following maximization problem within a period

$$\max_{P_{ij}, q_{ij}} P_{ij} q_{ij} - w \frac{q_{ij}}{z_{ij}} \quad (10)$$

subject to the inverse demand function derived from (4) and (7):

$$\frac{P_{ij}}{P} = \left(\frac{q_{ij}}{y_i} \right)^{-\frac{1}{\rho}} \left(\frac{y_i}{c} \right)^{-\frac{1}{\eta}}, \quad (11)$$

where the firm takes into account that its quantity, q_{ij} , affects sectoral output, y_i , by (6).⁴ Substituting out P_{ij} from the objective, and taking the derivative with

⁴An alternative to Cournot competition would be Bertrand competition in which the firm takes the prices, instead of the quantities, of other firms as given. This leads to the following constraint for the maximization problem (where I have now combined (4) and (7) to substitute out y_i , instead of P_i as in the case of Cournot):

$$\frac{P_{ij}}{P} = \left(\frac{q_{ij}}{c} \right)^{-\frac{1}{\rho}} \left(\frac{P_i}{P} \right)^{1-\eta/\rho}. \quad (12)$$

Using this equation to substitute out the quantity, q_{ij} , in the objective, and taking the derivative with respect to P_{ij} , where the firm takes into account that changing its price affects P_i , gives

$$\varepsilon(s) = \rho(1-s) + \eta s. \quad (13)$$

respect to q_{ij} implies

$$P_{ij} = \frac{\varepsilon(s_{ij})}{\varepsilon(s_{ij}) - 1} \frac{w}{z_{ij}}, \text{ where} \quad (14)$$

$$\varepsilon(s) = \left[\frac{1}{\rho}(1-s) + \frac{1}{\eta}s \right]^{-1}, \quad (15)$$

and s_{ij} is the market share of firm j in its sector i .⁵ Furthermore, using (7) and (8) gives the following expression of the market share in terms of prices:

$$s_{ij} = \frac{P_{ij}^{1-\rho}}{\sum_{l=1}^J P_{il}^{1-\rho}}. \quad (16)$$

Thus, (14) is a system of J non-linear equations in the J equilibrium prices P_{ij} for each sector.

Discussion

The profit share at the firm level equals

$$\pi_{ij} = \frac{P_{ij}q_{ij} - wl_{ij}}{P_{ij}q_{ij}} = \frac{1}{\rho}(1-s_{ij}) + \frac{1}{\eta}s_{ij}. \quad (17)$$

Suppose that a firm has a monopoly in a sector, such that $s_{ij} = 1$, then the profit share equals $\frac{1}{\eta}$, the inverse elasticity of substitution between industries. This is the same result as when one would have monopolistic competition with elasticity of substitution η across goods. The opposite extreme is there being a continuum of small firms within an industry such that $s_{ij} = 0$. Then, the profit share equals $\frac{1}{\rho}$, the inverse elasticity of substitution between varieties within an industry. For intermediate cases, the profit share lies in between these extremes.

Equation (17) shows that in this model there is a linear relationship between the profit share and firm size. This is consistent with the data, as larger firms tend to have a higher profit share. However, this model is not able to capture the fact that for younger firms there is a stronger relationship between firm size and the profit share. Moreover, as I take ρ and η to be constant over time, this model will not be able to capture that this relationship has become stronger over time.

⁵The derivation uses that $s_{ij} = \frac{P_{ij}q_{ij}}{\sum_l P_{il}q_{il}} = \frac{q_{ij}^{1-1/\rho}}{y_i^{1-1/\rho}}$, where I have used (7) to obtain the second equality.

Equilibrium

To solve for the equilibrium in which agents are indifferent between being an entrepreneur and being a worker, we have to solve for the number of firms J . For this purpose, I set the number of sectors I to 10,000 and assume that when an agent decides to become an entrepreneur, it does not know in which sector it will be active such that the number of firms in each sector is the same. Each firm draws an age uniformly at random from 1 to T and subsequently draws a productivity from a log-normal distribution with age-dependent mean z_a and a standard deviation that does not depend on age.⁶ As an example, suppose that the number of firms J per sector equals T . Then, there are some sectors for which there is one firm of each age, but there are also sectors for which all firms have the same age etc.

I normalize the wage to 1 and, given the age and productivity draws, I solve for each sector the system of equations given by (14) to obtain P_{ij} . Then (8) gives the sectoral price P_i , and the price P of the final good is obtained by (5), where the integral is replaced by a sum as there is now a finite number of sectors. Now once we have solved for prices, we have to solve for quantities. For this, we need to obtain final consumption c . Labor market clearing gives that

$$\underbrace{N - \overbrace{I \cdot J}^{\text{Entrepreneurs}}}_{\text{Labor supply}} = \underbrace{\sum_{i=1}^I \sum_{j=1}^J \frac{q_{ij}}{z_{ij}}}_{\text{Labor demand}}. \quad (18)$$

Substituting out the quantity produced, q_{ij} , as a function of c and prices, using the inverse demand functions (4) and (7) gives, after rewriting, that

$$c = \frac{N - I \cdot J}{\sum_{i=1}^I \sum_{j=1}^J \frac{1}{z_{ij}} \left(\frac{P_{ij}}{P_i} \right)^{-\rho} \left(\frac{P_i}{P} \right)^{-\eta}}.$$

Having c and prices in hand, sectoral output y_i is given by (8), and the quantity produced by each firm q_{ij} is given by (7). The product market clears by Walras law. Knowing quantity and prices, profits are given. As an agent does not know in which sector it will be active when it decides to become an entrepreneur, the value of being an entrepreneur equals

$$V^{ent} = \sum_{a=0}^{T-1} \beta^a \Pi_a, \quad (19)$$

⁶To be precise, the standard deviation of the log of productivity does not depend on age.

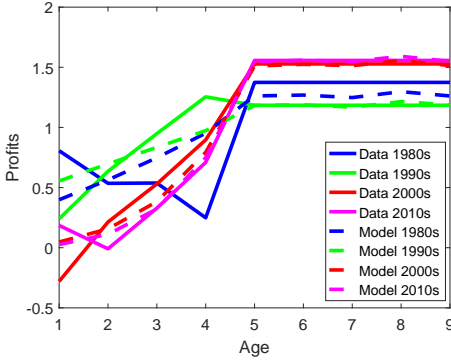


Figure 7: Normalized profits according to data and model

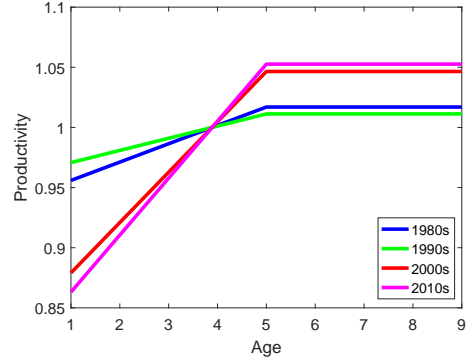


Figure 8: Average productivity according to model

where Π_a corresponds to the average profits across all sectors of all firms with an age a .

When for a guess of J the value of being an entrepreneur exceeds the value of being a worker, I increase the number of firms and vice versa until I have found an equilibrium.

Parametrization

I set the number of periods T equal to 9. Each period encompasses five years and corresponds to an age bin in the above graphs. In order to study to what extent the changing life-cycle pattern of profits explains the rise in profits and the fall in entrepreneurship, I vary how productivity evolves over the life cycle. For simplicity, I assume that average productivity increases linearly during the first five periods and is constant thereafter. I set average productivity over the life cycle equal to 1 and vary the slope of productivity during the first 5 periods in order to match the observed life-cycle profile of profits after normalizing. I normalize the observed profits to 1 on average over the life cycle and I set profits from age bin 5 onward equal to average profits during the last 5 age bins. The solid lines in Figure 7 show the normalized profits for the four periods.

When solving the model, I take the following parameter values. I set the discount factor β between two periods equal to 0.935^5 to reflect an annual discount factor of 0.935. At first, this might seem as agents being rather impatient but one should note that this refers to the stochastic discount factor. Moreover, the model does not take into account firm exit. Having a lower discount factor can be seen as taking exit into account in a reduced form. Nonetheless, I will vary the discount rate and the results are robust to taking a larger β . I set the elasticity of substitution

Table 1: Profit share and share being an entrepreneur according to data and model

	Data	Model I $\beta = 0.935^5$ $\eta = 2$	Model II $\beta = 0.96^5$ $\eta = 2$	Model III $\beta = 0.935^5$ $\eta = 1.5$
Profit share 1980s	0.056	0.084	0.082	0.093
Profit share 1990s	0.060	0.082	0.081	0.090
Profit share 2000s	0.068	0.094	0.088	0.10
Profit share 2010s	0.085	0.095	0.091	0.10
Δ profit share	0.017	0.011	0.0073	0.011
Entrepreneur share 1980s	0.046	0.065	0.070	0.065
Entrepreneur share 1990s	0.045	0.068	0.073	0.070
Entrepreneur share 2000s	0.042	0.053	0.065	0.058
Entrepreneur share 2010s	0.042	0.053	0.063	0.055
Δ entrepreneurs	-0.0036	-0.014	-0.0072	-0.011

Notes: Δ profit share and Δ entrepreneurs refer to the differences in the average profit share and the average entrepreneur share between 1980–1999 and 2000–2015.

ρ between different varieties equal to 50 and the elasticity of substitution η between different sectors equal to 2. A rather high elasticity of substitution of 50 is taken as the profit share will otherwise be (much) higher than the 6%, as it is observed in the data on average during this time period (see (17)). I set the standard deviation of log productivity (within an age bin) equal to 0.09 to roughly match the observed variance of the log of sales within an age bin.

III Results

The dashed lines in Figure 7 show normalized profits according to the model when the productivity profile is chosen to minimize the squared distance between normalized profits in the model and the data. The model matches the data closely for the last three decades. The model has a hard time matching the data for the 1980s as profits were then downward sloping during the first 4 periods. Moreover, the model cannot generate negative profits. Figure 8 shows the life-cycle profiles of productivity for the four different decades. It is inferred that productivity growth over the life cycle of the firm has become much larger over time. In the 1980s and the 1990s productivity was increasing by around 5% in total during the first five periods, while after 2000 it was increasing by around 20%.

The profits according to the model as displayed in Figure 7 are equal to 1 by normalization. Instead, the first five rows in Table 1 show how the profit share has changed over time. According to the baseline model (model I) the profit share

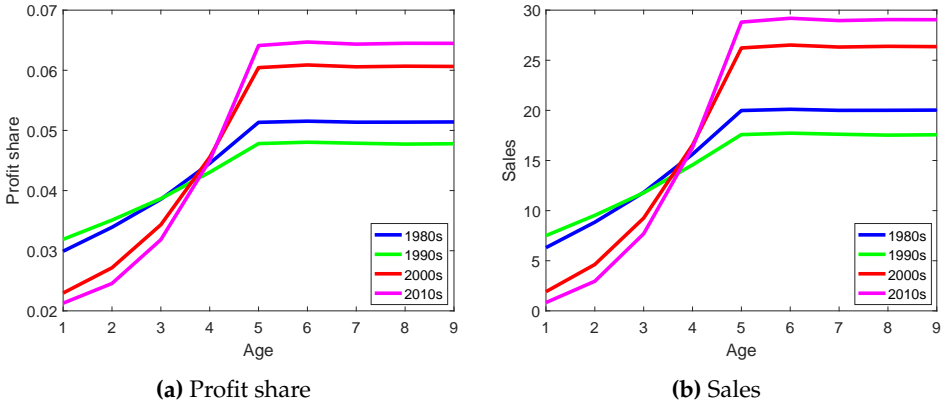


Figure 9: The profit share and firm size according to the model

has increased from 8.4% in the 1980s to 9.5% in the 2010s. The first column (data) shows that observed profits have increased from 5.6% in the 1980s to 8.5% in the 2010s. Observed profits refer to the average profit share for each period as estimated in Chapter 1. The row ‘ Δ profit share’ displays the difference in the profit share between 1980–1999 and 2000–2015. The profit share has increased by 1.7 percentage points while according to the model, profits have increased by 1.1 percentage points. Thus, the changing life-cycle pattern of profits can account for about two-thirds of the rise in profits.

The bottom part of Table 1 shows how the share of entrepreneurs has changed over time. The share of entrepreneurs in the model is defined in the same way as it is calculated in the data. Namely, it is calculated as the number of entrepreneurs $I \cdot J$ divided by the total population N . The share of agents that decide to become an entrepreneur falls more quickly according to the model than according to the data. According to the model, entrepreneurship would have fallen, all else equal, by 1.4 percentage points while in reality it fell by 0.4 percentage points.

The last two columns show that the results are robust to changing parameter values. Model II assumes that agents are more patient than in the baseline with an annual discount factor of 0.96. The rise in profits is attenuated as compared to the baseline but still substantial with 0.7 percentage points. Now the share of entrepreneurs falls by 0.7 percentage points. The last column shows that when the elasticity of substitution between sectors is 1.5, the results are very similar to when this elasticity is 2.

The share of agents being an entrepreneur is larger in the model than in the data. One reason is that there is no exit in the model while, in reality, not everyone that starts a firm continues to be an entrepreneur.

Figure 9 shows what the life-cycle profiles of profits and sales look like according to the model. Both the firm size and the profit share relationship with respect to age have become steeper over time. This is consistent with the data. However, the model cannot replicate that the profit share has become negative for young firms.

IV Conclusions

This chapter shows that the change in the life-cycle pattern of profits can account for two-thirds of the rise in profits and more than fully explains the decline in entrepreneurship. Nowadays, profits appear much later in the life cycle of the firm compared to thirty years ago. As agents discount, this lowers the value of the firm and therefore leads to less firm entry. This lowers competition and therefore increases profits.

The model does not take into account firm exit and this might be one reason why I overestimate the decline in entrepreneurship. Another reason why I find that entrepreneurship declines faster in the model than in the data is that the model only matches two-thirds of the observed rise in profits. This means there has also been another force that has led to an additional increase in profits. This could, for instance, be laxer antitrust regulation. An additional rise in profits would make it more attractive to become an entrepreneur and would therefore, lead to a slower decline in entrepreneurship.

I assume that parameters do not change over time. The risk free rate has been declining during the last decades. If this implies that the discount rate has been increasing over time, this would attenuate the rise in profits due to the changing life-cycle pattern of profits. However, the risk premium might have increased over time (Farhi and Gourio, 2018), and therefore, the stochastic discount factor might not have increased.

The model cannot generate negative profits, whereas young firms do make negative profits in the data. One reason is that the model does not take into account overhead costs. Overhead costs might be relatively large for a young small firm. In the data, total costs can be split up into three components: costs of goods sold (cogs), selling, general and administrative expenses (sga), and capital costs. Figure 15 shows these three separate costs as a share of sales over the firm life cycle. Cogs as a share of sales used to be uncorrelated with age, but are nowadays negatively correlated. The costs of goods sold do, on average, exceed the sales for young firms. However, this could partially be driven by outliers as the median cogs share of sales does not seem to be correlated with age. Sga as a share of

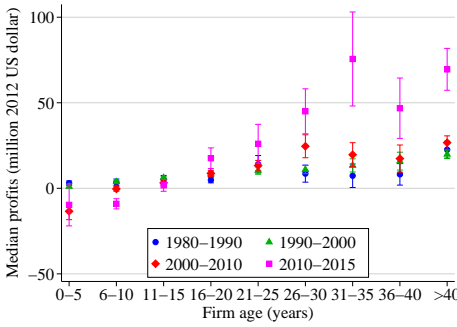
sales has always been negatively related to age, and this relationship has become stronger over time. This holds for both the average and the median. Capital costs as a share of sales are not correlated with age and this has not changed over time.

One explanation for why sga as a share of sales is now larger than it used to be for young firms is that intangibles have become more important over time. A young firm has to invest heavily into sga to build up intangible stock, while old firms already have a large intangible stock. This could explain why profits are negative for young firms. This could potentially also explain why productivity is nowadays steeper over the firm life cycle. I leave investigating what explains the changing life-cycle pattern of productivity to future research.

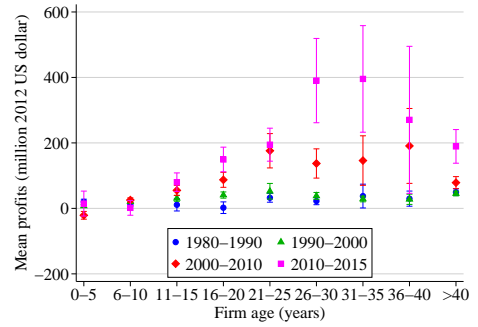
References

- Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019) "A Theory of Falling Growth and Rising Rents," Manuscript.
- Akcigit, U. and S. T. Ates (2019) "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory," CEPR Discussion Paper DP13668.
- Atkeson, A. and A. Burstein (2008) "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, Vol. 98, No. 5, pp. 1998–2031.
- Atkeson, A. and P. J. Kehoe (2005) "Modeling and Measuring Organization Capital," *Journal of Political Economy*, Vol. 113, No. 5, pp. 1026–1053.
- Azar, J. and X. Vives (2019) "Common Ownership and the Secular Stagnation Hypothesis," *AEA Papers and Proceedings*, Vol. 109, pp. 322–326.
- Barkai, S. (2017) "Declining Labor and Capital Shares," Manuscript.
- Bornstein, G. (2018) "Entry and Profits in an Aging Economy: The Role of Consumer Intertia," Manuscript.
- De Loecker, J., J. Eeckhout, and G. Unger (2018) "The Rise of Market Power and the Macroeconomic Implications," Manuscript.
- Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014) "The Role of Entrepreneurship in US Job Creation and Economic Dynamism," *Journal of Economic Perspectives*, Vol. 28, No. 3, pp. 3–24.
- Farhi, E. and F. Gourio (2018) "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia," *Brookings Papers on Economic Activity*, Fall.
- Field, L. C. and J. Karpoff (2002) "Takeover Defenses of IPO Firms," *Journal of Finance*, Vol. 57, No. 5, pp. 1857–1889.
- Liu, E., A. Mian, and A. Sufi (2019) "Low Interest Rates, Market Power, and Productivity Growth," Manuscript.
- Loughran, T. and J. R. Ritter (2004) "Why Has IPO Underpricing Changed Over Time?" *Financial Management*, Vol. 33, No. 3, pp. 5–37.

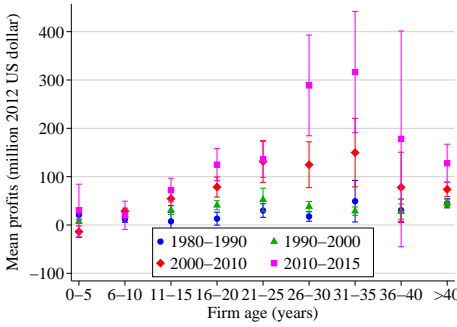
Appendix A Additional Figures



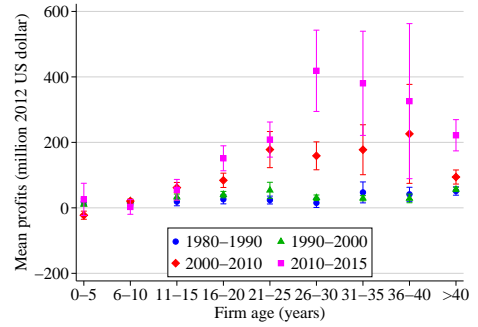
(a) Median



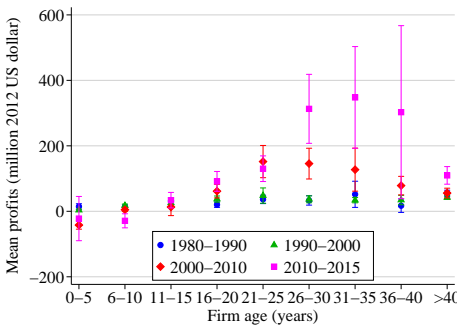
(b) Industry weights equal to 1990s



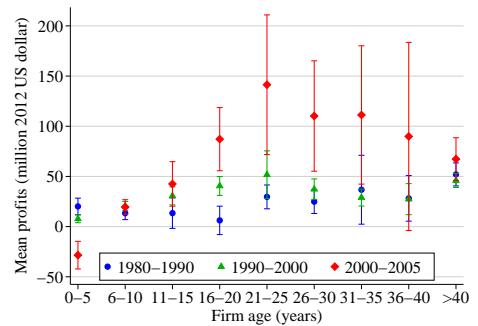
(c) Industry weights equal to 1990s by age



(d) Heterogeneous R



(e) R required rate of return approach



(f) Excluding the great recession

Figure 10: Life cycle of profits - robustness

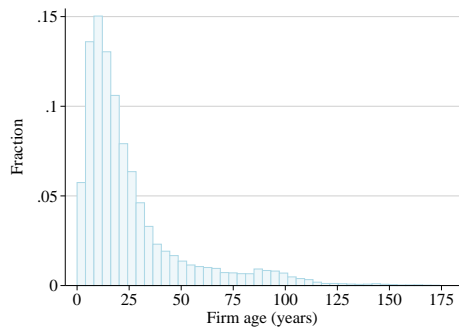
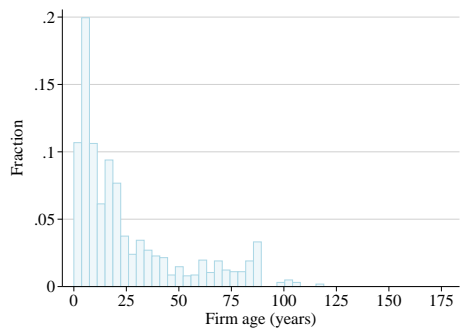
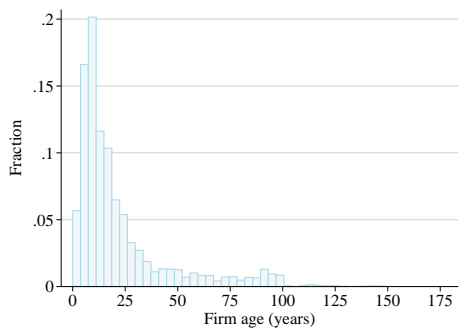


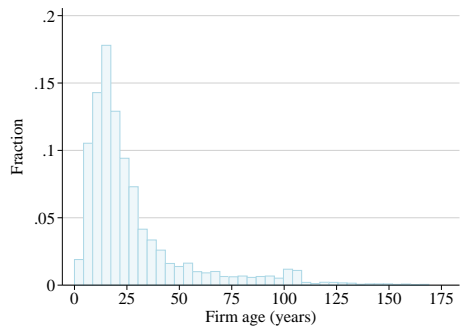
Figure 11: Histogram age distribution



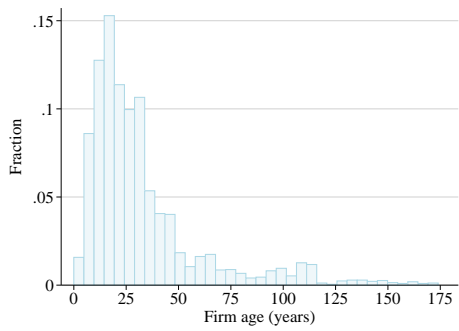
(a) 1980s



(b) 1990s

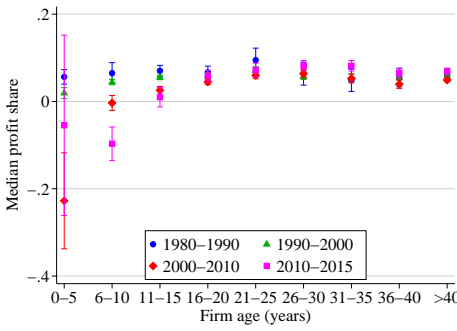


(c) 2000s

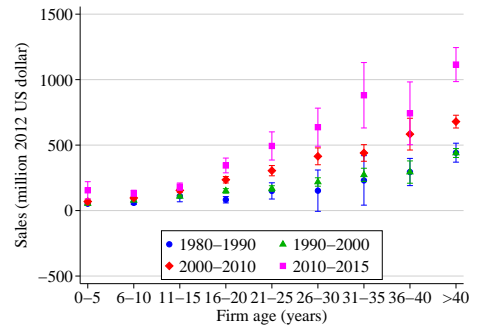


(d) 2010s

Figure 12: Histogram age distribution by decade

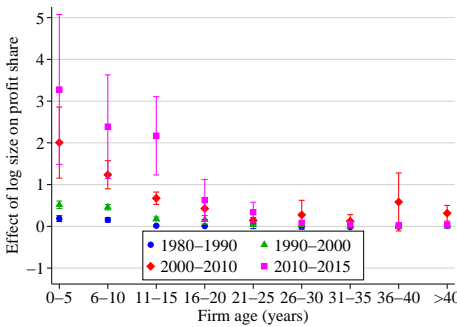


(a) Median profit share

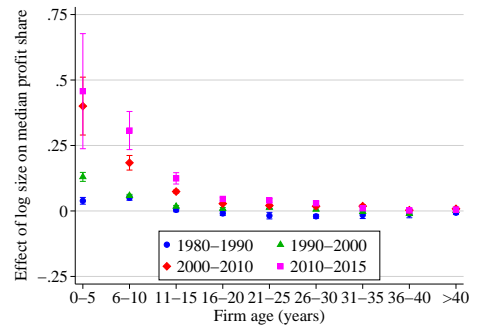


(b) Median sales

Figure 13



(a) OLS



(b) Quantile regression

Figure 14: Relationship between the profit share and log firm size over the life cycle

The Life Cycle of Profits

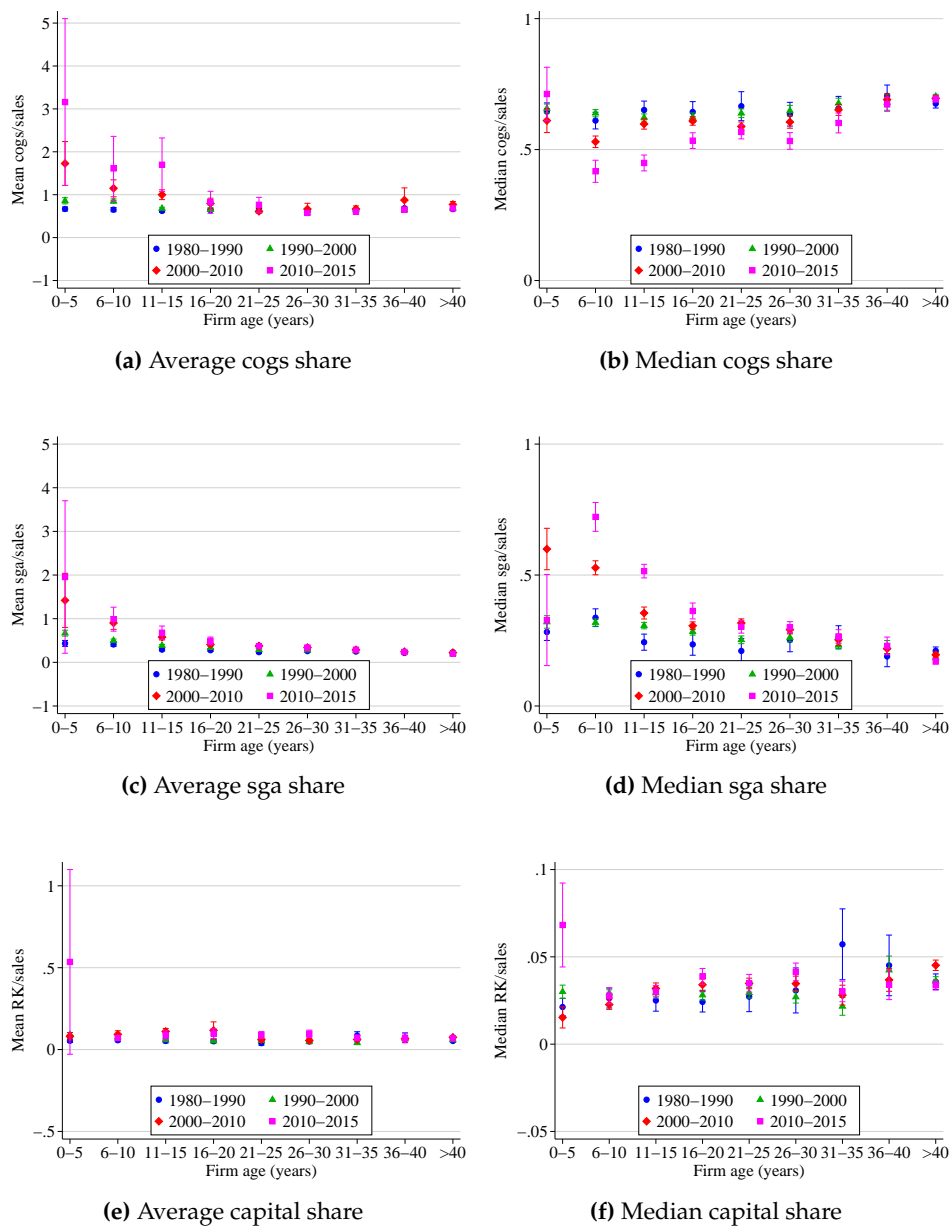


Figure 15: Decomposition of costs over the life cycle

Chapter 4

Diffusion of Ideas in Networks and Endogenous Search*

New ideas tend to spread gradually (Griliches, 1957) and agents that are directly connected to early adopters are more likely to adopt these ideas (Coleman et al., 1957).¹ These observations have led to an active literature studying the effects of the network on diffusion. Furthermore, whether and at what rate ideas diffuse depend on the effort agents put into searching for productivity-enhancing technologies, while this effort depends on the distribution of productivity across agents which, in turn, is a result of diffusion (Lucas and Moll, 2014; Perla and Tonetti, 2014). Despite the importance of search effort for diffusion, much of the literature has ignored how the network affects search effort.

In this chapter, I study which network properties are beneficial for diffusion when the effort put into search (or, equivalently, learning) is endogenous and depends on the network. In order to answer this question, I build a model in which agents differ in terms of their productivity level. Agents have the option to engage in costly learning. When an agent decides to pay the search (or learning) cost, it is matched with one of its first-degree connections. If the productivity of this connection is higher than the productivity of the searching agent, then the searching agent adopts the corresponding technology, meaning that its own productivity level will increase to the productivity level of the agent it is matched with.² The decision to learn depends on the network and interacts with the productivity distribution. The more productive an agent's connections, the higher the expected gains from learning and therefore, the higher its learning reservation productivity,

*This paper supersedes an earlier draft circulated under the title "Diffusion of Ideas and Network Linkages." I thank Mohammad Akbarpour, Timo Boppert, Ben Golub, Axel Gottfries, Matt Jackson, Karin Kinnerud, Pete Klenow, Per Krusell, Hannes Malmberg, Kurt Mitman, Ezra Oberfield, Chris Tonetti, Vincent Sterk, Marc Witte (discussant) and seminar participants at IIES Stockholm University, SSE PhD workshop, Stanford, UCL, the European Meeting of the Econometric Society 2018 and Warwick Economics PhD Conference 2018 for helpful comments. I thank Handelsbanken for providing financial support.

¹For more evidence that agents are more likely to learn from agents to which they are closely connected, see, for example, Conley and Udry (2010) for farmers, Carvalho and Voigtländer (2015) and Chaney (2014) for firms, and Coe and Helpman (1995) for countries.

²I focus on the decision of potential adopters. Hardy and McCasland (2018) study the decision of incumbent adopters to share technology.

which is the productivity at which an agent is indifferent between engaging in costly learning and not learning. Agents with a productivity below the reservation productivity find it optimal to search for productivity-enhancing technologies. Therefore, an increased reservation productivity compresses the support of the productivity distribution, and leads to a higher average productivity.

The main contribution of this paper is that I reveal a novel mechanism through which network density affects diffusion and therefore total factor productivity (TFP), namely that network density affects learning effort. The denser the network, the more links there are and the more agents are connected to the highly productive agents. This increases their gains from learning which will increase their reservation productivities (or their effort) and therefore their productivity on average. This leads to their connections, in turn, being connected to more productive agents, increasing their reservation productivities and so on. Hence, TFP will be higher and because it are the least productive that adopt, inequality will be lower with a dense network than with a sparse network.

Although productivity is higher in a dense network, the effect of network density on the share of agents that learn in *equilibrium* is ambiguous. The network affects the reservation productivity but there are two opposing effects on the share of agents that learn. The first effect is that when agents get more connected to more productive agents they become more willing to learn, thereby increasing the mass of learners. However, as the reservation productivity increases, the opportunity cost of learning also increases (as learning is disruptive to production) which has a negative effect on the mass of learners. When the model is calibrated to match moments of the firm-level distribution, these effects are approximately equally large and therefore the effect of the network on the share of agents that learn is negligible. This result has important implications for how one can empirically test which network properties are beneficial for learning. Suppose that one would, for different locations, have data on the network and the share of agents that are adopting new technologies in equilibrium in each location. Then, regressing the share of agents that is adopting on different network properties will not be informative. Instead, one should study how the share of agents that adopts changes during the transition to a new equilibrium after the network has changed. This result is the second contribution of the paper.

The third contribution is on the modeling side. For there to be an opportunity to learn, agents need to differ in their productivity levels. Most papers that study diffusion on a network assume that there is a binary productivity state. Instead, in the model presented here, the productivity state is continuous. To the best of my knowledge, this is the first paper that builds a model with both a continuous

productivity state and endogenous diffusion on a network. The advantages of a continuous productivity state are that it makes it possible to relate networks to aggregate productivity and other moments of the productivity distribution, and that the model can be disciplined using data on the distribution of productivity or income. Moreover, a continuous productivity state allows the gains from learning to increase in the productivity gap. There is recent evidence that this is indeed the case. Akcigit et al. (2018) find that the more productive the inventors are with whom an inventor interacts, the higher is subsequent productivity. Nix (2016), Jarosch et al. (2018) and Herkenhoff et al. (2018) find that learning from coworkers increases with knowledge gaps.

In order to achieve a continuous productivity distribution, agents face an exogenous probability of improving their productivity (i.e., creating new ideas) and this innovation intensity is allowed to differ across agents. Some agents are good at creating new ideas while others are not. Agents that have a high probability of innovating are more likely to have a high productivity than agents with a low probability of innovating. I take this heterogeneity in innovation intensities as given (and estimate it using firm level data) and study how this heterogeneity combined with the network structure affects the learning decision and the resulting productivity distribution. The presence of innovation leads to sustained long-run growth without needing to impose an unbounded distribution of productivity as initial condition. I analyze balanced growth paths in which the productivity distribution is a traveling wave, meaning that all quantiles of the distribution grow at the same rate. The model is an exogenous growth model in which the growth rate is given by the exogenous innovation process and is not affected by the network. However, the network affects the *level* of aggregate productivity. A network in which agents tend to be connected to more-productive agents leads to higher expected gains from learning, which increases the reservation productivities below which one decides to learn. Thus, the network affects aggregate productivity by affecting the learning decision of low productive agents.

Using the model, I investigate how different properties of the network affect the productivity distribution. As already discussed above, I find that adding more links increases TFP and reduces inequality (i.e., the variance of log productivity) by affecting reservation productivities. The effect of the number of links on TFP and inequality is especially strong for sparse networks while for already dense networks the effect of adding links is negligible. It could be the case that the increase in the number of links itself does not affect the equilibrium outcomes but that the effect goes through another network property. For instance, increasing the number of links also lowers the average path length, which is the average of

the lowest number of links needed to pass to ‘travel’ between any pair of nodes. Therefore, I also study networks that keep the number of links constant but vary the average path length. I find that decreasing the average path length increases TFP, while it has a negative effect on inequality. When the average path length is low, many nodes are closely connected to the highly innovative agents (i.e., only a few steps away), which increases reservation productivities for the same reason as why a high network density leads to higher reservation productivities. Agents that are directly connected to the most innovative agents exert more effort, thereby increasing their average productivity and, in turn, increasing the effort of their connections. The further an agent is away from the most innovative agents, the smaller is the effect on effort and therefore a low average path length is associated with high productivity. The positive effect of network density on TFP remains when controlling for the average path length, but the effect on inequality disappears.

It is not a novel result that denser networks lead to faster diffusion, but the mechanism through which it occurs here is novel. For instance, in the SIS model (a widely used model in the epidemiological literature), a larger number of links speeds up diffusion (see, e.g., Bailey, 1975; Jackson and Rogers, 2007) as is the case for the model in Fogli and Veldkamp (2016). However, in these models, the probability that an agent learns from each neighboring agent is exogenous and it is hard-wired that the more neighboring agents one has, the more likely it is to learn from a neighbor.³ In my model, the decision to learn is instead endogenous, and the reason that denser networks lead to higher productivity is that denser networks increase the learning reservation productivities. Having more connections does not affect the *probability* of learning since when agents decide to learn, they will only learn from exactly one of their neighboring nodes. Shutting down the probability effect highlights that a denser network also increases productivity by increasing the *gains* from learning. If, in addition, the density were to also affect the number of agents one learns from, the effect of network density on TFP would be higher.

Jackson and Yariv (2005) and López-Pintado (2008) have models where the decision to adopt is endogenous and the gains from adopting a given technology depend on the number of neighbors that have already adopted. However, this dependency is given exogenously, while here it is endogenous. Moreover, in these papers, the gains of a given technology that depend on the number of connections that are using it (think of the decision to use the QWERTY keyboard). These gains are known before the decision to adopt is made, while here there are

³Denote the number of neighbors of an agent by d and suppose that all of them have a better technology. Denote the probability of learning from each neighbor by p . Then, the probability of learning from a neighbor in these models is $1 - (1 - p)^d$ which is increasing in d .

multiple technologies (i.e., productivity is a continuous variable) and the gains are uncertain. A learning agent can be matched to a low productive agent, thus leading to productivity only improving slightly which, in turn, leads to a negative net gain (including the payment of the learning cost).

Related literature. My paper contributes to the following three strands of the literature.

First, my paper contributes to the literature on the effect of network properties on diffusion. Theory tends to originate from the epidemiological literature (Bailey, 1975) and therefore usually focuses on a binary state, i.e., either being sick or healthy, or either adopting or not adopting a given technology (see, e.g., Ellison, 1993; Young, 2003; Montanari and Saberi, 2010; Acemoglu et al., 2011; Akbarpour and Jackson, 2018). In these papers, the research question usually relates to how the network affects the time it takes until most agents have adopted, or what the probability is that everyone eventually adopts. I study diffusion when there is a continuous productivity state and study the effect of the network on aggregate TFP. In practice, there is a wide range of different technologies and, therefore, of different productivities. Some agents might have adopted multiple technologies and some technologies might have a larger effect on productivity than others. My model captures that the returns to learning are higher if one is connected to more-productive agents (Akcigit et al., 2018; Herkenhoff et al., 2018; Jarosch et al., 2018; Nix, 2016). Furthermore, I use moments of the firm-level productivity distribution to estimate parameters of my model. Most of the models in this literature take the adoption rate to be exogenous (see, e.g., Bailey, 1975; Granovetter, 1978; Jackson and Rogers, 2007; Acemoglu et al., 2011; Akbarpour and Jackson, 2018).⁴ I study the effect of the network on learning effort and adoption. Ellison (1993), Young (2003) and Montanari and Saberi (2010) are examples of models in which adoption is endogenous, but with a binary state. In these papers, the decision to take an action follows from a coordination game with one's neighbors. Fogli and Veldkamp (2016) is a notable exception that has a continuous productivity state, but it has an exogenous adoption rate.

Second, my paper contributes to the recently burgeoning literature on how the decision to learn interacts with the distribution of productivity across firms.⁵ Two prominent examples are Lucas and Moll (2014) and Perla and Tonetti (2014).⁶ In

⁴Kremer (1996) is a notable exception and studies the spread of HIV in a setting where sexual activity is endogenous.

⁵There is a related literature where learning depends on the productivity distribution as well, but the rate at which learning occurs is exogenous. An example is Luttmer (2012a). Furthermore, in Luttmer (2007, 2012b), Sampson (2016) and Lashkari (2018) entrants learn from incumbent firms. Alvarez et al. (2013), Buera and Oberfield (2020) and Cai et al. (2017) study the diffusion of ideas across countries.

⁶Other examples are Perla et al. (2015) and König et al. (2016). Hopenhayn and Shi (2017) incorporate

these papers, agents pay a learning or search cost and are subsequently randomly matched to another agent in the economy from which they will learn. The initial productivity distribution is assumed to have a Pareto tail which ensures the possibility of balanced growth. In a recent paper, Benhabib et al. (2017) study learning and innovation jointly and, as a consequence, no longer need the initial distribution to have infinite support to sustain long-run growth. In equilibrium, productivity is Pareto distributed where the thickness of the tail depends on the innovation intensity. I extend the model in Benhabib et al. (2017) to include a network.⁷ The existing models (e.g., Lucas and Moll, 2014; Perla and Tonetti, 2014; Benhabib et al., 2017) can be interpreted in two ways, either as models of the entire economy in which each agent is equally likely to learn from all other agents, no matter whether that agent is located next door or at the other side of the country. Or as models in which agents can only learn from agents that are similar (i.e., drawn from the same distribution) and have no possibility to learn from other types of agents. My model seeks a middle ground between these two extremes. Agents can learn from other types of agents as well as from agents with similar characteristics. The extent to which this happens is governed by an exogenous network.⁸

In this paper, search technology is modeled in a similar way as in Perla and Tonetti (2014) and Benhabib et al. (2017). However, the implications for the productivity distribution are somewhat different. First, in contrast to these papers, my model does not imply that only the least productive agents in the economy adopt. This is because the reservation productivities differ across nodes. Second, in my model, the productivity distribution is hump-shaped, whereas in Perla and Tonetti (2014) and Benhabib et al. (2017) the productivity distribution is downward sloping.

Third, my paper contributes to the renewed interest in the effect of networks on macroeconomic outcomes.⁹ Based on the seminal works by Long and Plosser (1983) and Hulten (1978), Carvalho (2010), Acemoglu et al. (2012), Acemoglu et al. (2016b), Acemoglu et al. (2017) and Baqaee (2018) study the network origins of aggregate fluctuations. In another strand of the literature, Jones (2011, 2013), Bigio and La'O (2017) and Baqaee and Farhi (2020) study how sectoral distortions spill over to the aggregate economy through the production network and affect aggregate TFP. Liu (2019) studies the effect of industrial policy in a production network with

search frictions into a model of learning.

⁷Appendix G discusses why including a network in Lucas and Moll (2014) and Perla and Tonetti (2014) does not admit any interesting dynamics.

⁸For an example of how to endogenize the network, see Oberfield (2018).

⁹See Carvalho and Tahbaz-Salehi (2019) for a survey.

distortions. In these papers, shocks or distortions affect the prices of intermediate goods which, in turn, affect the input choices of downstream producers.¹⁰ In other words, these papers study how distortions and shocks propagate through the production network. I study how the network affects TFP through the decision of firms to adopt existing technologies. Moreover, the above mentioned papers study networks based on trade flows of intermediate goods while, in my model, the network represents the network of idea flows which could indeed be based on trade flows, but could also be based on geography or movements of workers, for instance.¹¹ Acemoglu et al. (2016a) find that the historical network based on patent citations has a strong predictive power on innovation and Jaffe et al. (1993) find that patent citations are correlated with geographic distance. As these papers study patents, they focus on technology improvements at the frontier and not on those that are far behind in the productivity distribution which constitutes the focus of the present paper.

I Model

Time is continuous and there is a continuum of infinitely-lived agents which differ in their productivity level Z . The focus of this paper is on how productivity evolves over time. An agent's productivity can increase in two ways: i) through innovation and ii) through learning an existing technology. The decision to learn will be endogenously determined whereas innovation will be exogenous.¹² The learning and innovation processes are governed by some parameters, which I specify momentarily. I allow these parameters to vary across agents. Some agents are good at learning and others are good at innovating. However, the variation in these learning and innovating parameters across agents is assumed to be limited such that there are N different types of agents and all agents of the same type share the same parameter values. N is finite and types are indexed by n . Each type is assumed to consist of a continuum of agents. Then, the network is specified at the type level. Thus, the network says to which other types a type is connected. The network is important for the learning decision. Agents of a type have the

¹⁰Or the shocks and distortions affect input demand such that shocks travel upstream.

¹¹For instance, Ford adopted the assembly line in 1913 after an employee, William Klann, visited a slaughterhouse in Chicago where he saw one person removing the same piece over and over while the carcass moved along a moving line. Thus, the idea of the assembly line flowed from the meatpacking industry to Ford based on personal connections and not based on trade flows.

¹²I do not endogenize innovation because the network will only have a limited effect on innovation. If innovation is endogenous, it are mainly the most productive agents that are willing to innovate because the value of learning is close to zero for them (see Benhabib et al., 2017). The network has an effect on the learning value only and does therefore not affect the innovation decision of the most productive agents which is what is relevant for aggregate outcomes.

option to learn from the (types of) agents they are directly connected to (i.e., from their first-degree connections).

Suppose that income equals productivity and that agents discount expected future streams of income minus payments of the learning costs $\Theta_n(t)$ by ρ such that utility of an agent of type n at time t with productivity level Z equals

$$U_n(t, Z) = \mathbb{E}_t \left\{ \int_t^\infty e^{-\rho(\tau-t)} [Z(\tau) - \Theta_n(\tau) \mathbb{1}_{learning}(\tau)] d\tau \mid Z(t) = Z \right\}.$$

With abuse of notation, the learning costs $\Theta_n(t)$ only need to be paid when one decides to learn, which is indicated by the indicator function and is such that $\int_t^\infty \Theta_n(\tau) \mathbb{1}_{learning}(\tau) d\tau = \sum_k \Theta_n(t_k)$ when one decides to learn at times t_k .

Innovation

Innovation is modeled as follows. There are two innovation states $i \in \{l, h\}$; a low innovation state l and a high innovation state h . Agents switch from one state to the other following a continuous-time Markov process with transition intensities $\lambda_n^l > 0$ for transitioning from state l to state h and $\lambda_n^h > 0$ for the opposite transition. Productivity of agents in state h grows at the rate γ_n whereas those in state l have a constant productivity level.¹³ Types with a high λ_n^l and a low λ_n^h are ‘innovative’ types since a relatively large number of agents of these types will be in the high innovation state.

$V_n^i(t, Z)$ denotes the continuation value function, which is the value of continuing to produce at the current productivity level (i.e., the value of not learning), of an agent of type n in innovation state i at time t with current productivity Z . The Bellman equations are as follows for being in states l and h , respectively (see Appendix A for the derivation):

$$\rho V_n^l(t, Z) = Z + \lambda_n^l [V_n^h(t, Z) - V_n^l(t, Z)] + \partial_t V_n^l(t, Z), \quad (1)$$

$$\underbrace{\rho V_n^h(t, Z)}_{\text{Flow value}} = \underbrace{Z}_{\text{Profits}} + \underbrace{\lambda_n^h [V_n^l(t, Z) - V_n^h(t, Z)]}_{\text{Gains/losses from transitioning to other innovation state}} + \underbrace{\partial_t V_n^h(t, Z)}_{\text{Capital gains}} + \underbrace{\gamma_n Z \partial_Z V_n^h(t, Z)}_{\text{Gains from innovating}}, \quad (2)$$

where ∂_x denotes the partial derivative with respect to variable x which will later on also be indicated by a prime when a function is univariate. The Bellman equation equates the flow value to the instantaneous profits, plus the gain/loss

¹³Appendix H discusses geometric Brownian motions as an alternative stochastic process for modeling innovation. The results are similar to the Markov process. However, one difference is that the density turns out to be discontinuous in the case of a GBM whereas it is continuous in the case of the Markov process.

from transitioning to the other innovation state, plus the increase in the value function over time, plus the value from productivity growing at rate γ_n when in the high innovation state. Note that these equations do not depend on the other types and therefore, for each type, this is a system of two equations and can be solved for each type separately.

Denote the mass of type- n agents with productivity less than Z and in innovation state i at time t by $\Phi_n^i(t, Z)$. The technology frontier of a type is the highest productivity available for that type: $\bar{Z}_n(t) \equiv \sup\{\text{support}\{\Phi_n^l(t, \cdot)\}, \text{support}\{\Phi_n^h(t, \cdot)\}\}$. Furthermore, the maximum of support is equal across both innovation states since, due to the assumption of a continuum of agents, a continuum of agents will transition from one state to the other in each period of finite length, including agents at the frontier. For simplicity, suppose that the mass of agents of a type equals 1 such that $\Phi_n^l(t, \bar{Z}_n(t)) + \Phi_n^h(t, \bar{Z}_n(t)) = 1$. Denote the minimum of support by $M_n^i(t)$, which is the largest productivity such that $\Phi_n^i(t, M_n^i(t)) = 0$. Furthermore, denote the distribution unconditional on state i as $\Phi_n(t, Z) \equiv \Phi_n^l(t, Z) + \Phi_n^h(t, Z)$.

Learning

Besides innovating, an agent can also increase its productivity by learning and adopting an existing technology from another agent. When an agent decides to adopt, it has to pay a learning cost $\Theta_n(t)$ and will draw one of its first-degree connections at random.¹⁴ Learning is assumed to be perfect, meaning that if this connection has a higher productivity than its own, it adopts the technology and its productivity will instantaneously jump to this productivity level. To obtain the results in this paper, it is not important that the learning technology is perfect but it is important that the gains from learning are increasing in the productivity gap, which is consistent with what is found in Nix (2016), Akcigit et al. (2018), Jarosch et al. (2018) and Herkenhoff et al. (2018). Furthermore, random draws are consistent with the assumption that agents know the distribution of productivity across their connections but do not exactly know the productivity of each of their connections individually.¹⁵

Given the complexity of the system, I use a mean-field approximation which is a standard technique in the networks literature. That is, I assume that all agents of a certain type are connected to the same types of agents and that they are matched with the actual productivity distribution. A type- n agent that decides to learn will draw a productivity \hat{Z} from the learning distribution $\hat{\Phi}_n(t, \hat{Z})$, which is the

¹⁴I also sometimes refer to the learning cost as an adoption or search cost.

¹⁵This would, for instance, be the case if a connection represents a weak tie. There is evidence that a substantial amount of learning goes through weak ties (Granovetter, 1973).

probability that the productivity drawn is less than \hat{Z} . This learning distribution is a mixture of the productivity distributions of the neighboring types and is the same for all agents within a type.¹⁶ I will define the learning distribution more formally momentarily.

An agent decides to learn once the continuation value falls below the value of learning, where the value of learning equals the expected value of the continuation value associated with the productivity level drawn minus the learning cost. The continuation value function is increasing in productivity while the value of learning is independent of the productivity level of the learning agent. This implies that only those below a reservation productivity $M_n^a(t)$ will decide to learn. Furthermore, because learning is instantaneous, the minimum of support will be equal to the learning threshold (i.e., $M_n^a(t) = M_n(t)$). Hence, only the least productive of a type at a point in time will be learning since they have the most to gain from learning. At the optimal reservation productivity (i.e., $M_n(t)$), an agent must be indifferent between continuing to produce at its current productivity level and learning. This is summarized by the value matching condition,

$$\underbrace{V_n^i(t, M_n(t))}_{\substack{\text{Value (of not learning) \\ at threshold}}} = \underbrace{\int_{M_n(t)}^{\bar{Z}(t)} V_n^l(t, \hat{Z}) d\hat{\Phi}_n(t, \hat{Z})}_{\text{Gross learning value}} - \underbrace{\Theta_n(t)}_{\substack{\text{Learning} \\ \text{cost}}}, \quad \forall n. \quad (3)$$

Here, I have, for simplicity, assumed that a learning agent will be in the low innovation state after learning. In what follows, I take the learning costs to equal $\xi_n M_n(t)$. This represents that learning is disruptive to the production process and hence leads to a temporary loss of income. Moreover, this ensures that the learning cost is growing at the same rate as M_n , which turns out to equal the growth rate of the economy. That the learning costs grow at the same rate as the economy does is needed to ensure a balanced growth path.

Furthermore, note that I have omitted the innovation state superscript i on the minimum of support term (i.e., M_n and not M_n^i) because the minimum of support is equalized across innovation states. To see this, first note that the net learning value (i.e., the gross learning value minus the learning cost) does not depend on the innovation state of the learning agent. Furthermore, it turns out that the value functions at $M_n(t)$ are equalized across innovation states (see the proof of Proposition 1 in Appendix C). Therefore, the learning thresholds are equalized across innovation states.

¹⁶The latter is not restrictive since if, within a type, different agents are connected to different types of agents, this type can be split up into multiple types such that all agents of a type have the same connections. This works as long as there is a continuum of agents within a type.

The Network The decision to learn depends on the productivity of an agent's connections and therefore on the network. The network is specified at the type level and I will therefore also refer to types as nodes. Denote the set of types by $\mathcal{N} = \{1, \dots, N\}$ and let H denote an $N \times N$ matrix representing the links between agents. The graph (or network) is denoted by (\mathcal{N}, H) . As I will not vary \mathcal{N} , I will usually refer to H as being the network. An entry $h_{ij} = 1$ when type i is connected to j and zero otherwise. The network can be directed, meaning that h_{ji} does not necessarily need to equal h_{ij} . Define the degree, d_i , of type i as the number of types to which type i is connected, including a possible connection to agents of the same type (i.e., $d_i = \sum_j h_{ij}$). Define the entries of the $N \times N$ learning matrix A as $a_{ij} = \frac{h_{ij}}{d_i}$. By the definition of d_i , the rows of A sum to one which means that a learning agent will meet exactly one other agent (A is row stochastic).

An alternative modeling choice would have been that the number of draws increases in the degree, and that the agent adopts the highest productivity drawn. There are two reasons for not pursuing this approach. The first reason is that this is a growth model and in order to obtain a stationary system of equations, the economy needs to be normalized with a factor that is growing at the same rate as the economy. This normalization factor should not depend on the network in order to ensure that productivity distributions can be compared across networks, which is the aim of this paper. As I will show in Section II, this is ensured when the number of draws does not depend on the network but is not ensured when the number of draws does depend on the network. The second reason is that the main result of the paper is that more dense networks (with a higher average degree) lead to a higher TFP by increasing learning effort. Showing that this is the case even when the number of draws is independent of the degree highlights the role of learning effort. If the number of draws were to depend on the number of links, the relationship between network density and TFP would be even stronger.

Furthermore, I assume that there is directed search in the sense that learning agents will only draw agents with a productivity that is at least as high as the minimum of support of their own type $M_n(t)$. This is consistent with Jarosch et al. (2018) who find that having less-productive connections does not hinder learning from others through congestion. This leads to the following truncated distribution $\hat{\Phi}_n(\hat{Z})$ from which a learning agent draws,

$$\hat{\Phi}_n(\hat{Z}) = \frac{\sum_j a_{nj} (\Phi_j(\hat{Z}) - \Phi_j(M_n))}{\sum_j a_{nj} (1 - \Phi_j(M_n))},$$

where I have omitted the time subscripts for notational simplicity. It turns out to

be convenient to write the model in matrix-vector notation. Define the CDF $\Phi(Z)$ as a column vector where the n -th element represents $\Phi_n(t, Z)$ and similarly for the learning distribution $\hat{\Phi}(\hat{Z})$. This leads to,

$$\mathbf{1} - \hat{\Phi}(\hat{Z}) = X^{-1} A \left(\mathbf{1} - \Phi(\hat{Z}) \right), \quad (4)$$

where $\mathbf{1}$ is a column vector of which each entry equals 1. The probability that an agent draws a productivity larger than \hat{Z} (i.e., $\mathbf{1} - \hat{\Phi}(\hat{Z})$) equals the mass of agents with a productivity larger than \hat{Z} (i.e., $\mathbf{1} - \Phi(\hat{Z})$) times the learning matrix A times X^{-1} where X is an $N \times N$ diagonal matrix taking care of the truncation. Thus, the n -th diagonal entry of X equals the probability that a learning agent of type n draws a productivity larger than M_n in case the learning distribution would not be truncated (i.e., $X_{nn} = a_{n-} (\mathbf{1} - \Phi(M_n))$ where a_{n-} represents the n -th row of A).

Equation (4) does not take into account that this is only well defined for type n if $\hat{Z} \geq M_n$. Therefore, define $I_{\hat{Z} \geq M_n}$ as the matrix that has a 1 on the n -th diagonal if $\hat{Z} \geq M_n$. Rewriting equation (4), using that $A\mathbf{1} = \mathbf{1}$ and multiplying by $I_{\hat{Z} \geq M_n}$ gives

$$\hat{\Phi}(\hat{Z}) = I_{\hat{Z} \geq M_n} \left(\mathbf{1} + X^{-1} \left(A\Phi(\hat{Z}) - \mathbf{1} \right) \right), \quad (5)$$

which ensures that the vector $\hat{\Phi}(\hat{Z})$ has a zero on the n -th element if $\hat{Z} < M_n$.

The assumption that all draws have a productivity at least as high as the minimum of support of the type of the learning agent can easily be relaxed. However, also this assumption is needed for the specific application of the model in this paper, namely comparing aggregate productivity across networks. Relaxing this assumption would mean that learning agents could draw a lower productivity than their own. If this occurs they decide to keep their own productivity and it will be optimal for the agent to search again the next instant. This continues until the agent draws a higher productivity. Due to adoption being instantaneous, it will still be the case that all agents at the reservation productivity adopt. However, some agents have to search multiple times and hence have to pay the search cost multiple times. Therefore, the effective search cost is the expected number of draws times the search cost per draw, that is $X^{-1}\Theta$, where Θ is a column vector with Θ_n as the entries. This could easily be incorporated in the model, but the matrix X will depend on the network, and therefore, the effective search cost will differ across networks. This will imply that the normalization factor also depends on the network and will therefore not allow me to compare the productivity distribution across networks.

Having directed search is innocuous. The normalization factor is related to the productivity distribution of the most productive type and it turns out that the productivity distribution of the most productive type does not depend on the network in the case of directed search and therefore, neither does the normalization factor. A more dense network means that agents (except those of the most productive type) are more productive and therefore, in case of no directed search, it would become less likely for an agent of the most productive type to draw an agent with a lower productivity, thus lowering the expected number of times one needs to search. Hence, this would lower the effective search cost for this type and therefore lead to a higher reservation productivity and a higher TFP. Thus, incorporating undirected search would strengthen the result that more dense networks lead to a higher TFP.

In this paper, I will only consider connected networks. A network is connected when all pairs of nodes are connected by some path in the network where a path between nodes i_1 and i_K is a sequence of links i_1, i_2, \dots, i_K such that $h_{i_k i_{k+1}} = 1$ for each $k \in \{1, 2, \dots, K - 1\}$. Having a connected network means that each innovation can eventually reach all other nodes.

Lemma 1. *Suppose that the network is connected, that the maximum of support is finite and that a positive mass of agents of each type is learning, then the maximum of support is equal across types ($\bar{Z}_n(t) = \bar{Z}(t) \forall n$) and grows at $\gamma \equiv \max_n \gamma_n$.*

Proof. First show that $\bar{Z}_n(t) = \bar{Z}(t) \forall n$. Since for each type, there is a continuum of agents, there will also be a continuum of agents learning, drawing their productivity from $\hat{\Phi}_n(t, \hat{Z})$ which has a maximum of support equal to the maximum maximum of support of the types to which it is connected. As learning is instantaneous, the maximum of support of a type will be equal to the maximum of support of its connections. The connections will, in turn, have the same maximum of support as their connections etc. That the network is connected gives the result.

To show that the growth rate of the maximum of support equals $\max_n \gamma_n$, first suppose that there is no learning. Due to the continuum assumption, there are some agents that are lucky and are always in the high innovation state while being at the frontier. Therefore, if there were no learning, the maximum of support would grow at γ_n for each type. However, if there is learning, by the previous argument, the maximum of support is equal across all types at a given point in time and hence, the growth rate of the maximum of support for each type will be equalized to each other and be equal to $\gamma \equiv \max_n \gamma_n$. \square

Evolution of the Distribution

To see how the productivity distribution evolves over time, consider an agent with a productivity level just above its reservation productivity, and which is in the low innovation state, such that its productivity is constant over time. As long as this agent stays in the low innovation state, the gains from learning will increase over time. This occurs for two reasons. First, the productivity of those in the high innovation state grows over time and second, those that are at the reservation productivity will learn and increase their productivity. Both forces increase the productivity level of agents from which the original agent can learn, thus increasing the gains from learning, and once these gains exceed the learning cost, this agent will decide to learn. In other words, the learning threshold, and therefore the minimum of support, is growing over time. Denote its growth rate by $g_{M_n}(t)$. Now suppose that this agent would have been in the high innovation state such that its productivity grows at rate γ_n and suppose that $g_{M_n}(t) \leq \gamma_n$, then the reservation productivity is not catching up with the productivity level of this agent and hence, it will not learn as long as it is in the high innovation state. Only once it transitions to the low innovation state and falls back far enough relative to the other agents will it decide to learn. Only when $g_{M_n}(t) > \gamma_n$ will agents in the high innovation state learn.

Denote by $S_n^i(t)$ the flow of type- n agents in state i that decide to learn and by $S_n(t)$ the share of type- n agents that learn unconditional on the innovation state (i.e., $S_n(t) = S_n^l(t) + S_n^h(t)$). Since only the least productive of a type that are learning, the mass of learning agents of a type is equal to the mass of agents at the minimum of support times the rate at which the minimum of support increases faster than the growth of productivity in innovation state i :

$$\begin{aligned} S_n^l(t) &= M'_n(t) \partial_Z \Phi_n^l(t, M_n(t)), & \text{if } M'_n(t) > 0, \\ S_n^h(t) &= (M'_n(t) - \gamma_n M_n(t)) \partial_Z \Phi_n^h(t, M_n(t)), & \text{if } M'_n(t) - \gamma_n M_n(t) > 0. \end{aligned}$$

The Kolmogorov forward equations describe how the productivity distributions evolve over time for each n (see Appendix B for the derivation):

$$\begin{aligned} \partial_t \Phi_n^l(t, Z) &= \overbrace{\lambda_n^h \Phi_n^h(t, Z) - \lambda_n^l \Phi_n^l(t, Z)}^{\text{Transitioning to/from other innovation state}} - \overbrace{S_n^l(t)}^{\text{Learners of type l}} + \overbrace{\hat{\Phi}_n(t, Z) [S_n^l(t) + S_n^h(t)]}^{\text{Learners that draw productivity below } Z}, \quad (6) \\ \partial_t \Phi_n^h(t, Z) &= \overbrace{\lambda_n^l \Phi_n^l(t, Z) - \lambda_n^h \Phi_n^h(t, Z)}^{\text{Transitioning to/from other innovation state}} - \overbrace{S_n^h(t)}^{\text{Learners of type h}} - \underbrace{\gamma_n Z \partial_Z \Phi_n^h(t, Z)}_{\text{Innovation}}. \quad (7) \end{aligned}$$

The change in the mass of agents with a productivity below Z in state l of type n

equals the flow of agents transitioning from the high innovation state to the low innovation state (at the rate λ_n^h) and having a productivity below Z , minus the flow of agents with a productivity below Z that transition from the low to the high innovation state (at the rate λ_n^l), minus those that learn (recall that it are only the least productive that decide to learn), plus those that learn and draw a productivity below Z (according to $\hat{\Phi}_n(t, Z)$). The equation for the evolution of the CDF of the high innovation state is similar except that the last term is omitted because I have assumed that learning agents will be in the low innovation state after learning. Furthermore, there is an extra term that takes into account that productivity grows at the rate γ_n . Different from the Bellman equations, the Kolmogorov forward equations for a type n depend on the other types through $\hat{\Phi}_n(t, Z)$. Hence, this is a system of $2N$ coupled differential equations.

Finally, the learning decision problem is equivalent to an optimal stopping time problem and therefore, the following smooth pasting conditions must hold as well:

$$\partial_Z V_n^l(t, M_n(t)) = 0 \quad \text{if } M_n'(t) > 0, \quad \forall n, \quad (8)$$

$$\partial_Z V_n^h(t, M_n(t)) = 0 \quad \text{if } M_n'(t) - \gamma_n M_n(t) > 0, \quad \forall n. \quad (9)$$

II Balanced Growth Path

The previous section has outlined the model. In this section, I will solve for the (competitive) balanced growth path equilibrium defined as follows.¹⁷

Definition 1. Recursive Competitive Equilibrium

A recursive competitive equilibrium consists of distribution functions $\Phi_n^i(t, Z)$, value functions $V_n^i(t, Z)$, learning reservation productivity functions $M_n(t)$ and initial distributions $\Phi_n^i(0, Z)$ such that

1. Given $\Phi_n^i(t, Z)$, $M_n(t)$ are the optimal learning reservation productivity functions and $V_n^i(t, Z)$ are the associated value functions;
2. Given $M_n(t)$, $\Phi_n^i(t, Z)$ fulfill the Kolmogorov forward equations subject to the initial condition $\Phi_n^i(0, Z)$;

Definition 2. Balanced Growth Path Equilibrium

A balanced growth path equilibrium is a recursive competitive equilibrium such that aggregate output and the reservation productivities grow at constant rates. The productivity

¹⁷Note that agents, when deciding whether to learn, do not internalize that other agents might learn from them in the future, thereby providing an externality.

distribution functions are traveling waves: $\Phi_n^i(t, Z) = \Phi_n^i(0, Ze^{-g_Z t})$ where g_Z is the growth rate of each quantile of Φ_n and equal across types.

Lemma 2. *On a balanced growth path the growth rate of the reservation productivities equals the growth rate of average productivity.*

Proof. This follows directly from the definition of a balanced growth path, namely that the productivity distribution is a traveling wave. \square

Denote the growth rate of aggregate productivity by g , which equals g_Z on a balanced growth path by Lemma 2. Because each quantile of the productivity distribution grows at the rate g , also the minima of support $M_n(t)$ grow at the rate g . Therefore, to determine the growth rate it is only needed to determine the growth rate of $M_n(t)$. First suppose that the support $[M_n(t), \bar{Z}(t)]$ is of finite length in equilibrium. Then, since all quantiles grow at the same rate and the maximum of support grows at the rate γ by Lemma 1, the minimum of support will also grow at the rate γ . Hence, there is a unique constant growth rate consistent with balanced growth if there is finite support. However, if the support of the productivity distribution is infinite, i.e. $\bar{Z}(t) = \infty$, then this is no longer the case and g can differ from γ . The productivity frontier $\bar{Z}(t)$ can be infinite due to initial conditions or as an equilibrium outcome. The case with unbounded support as initial condition is studied by Lucas and Moll (2014) and Perla and Tonetti (2014), who show that in such a case, even without innovation (so $\gamma = 0$), there can still be growth, thus essentially $g > \gamma$. In what follows, I will assume that the support is of finite length as initial condition, so this case will not occur. However, even with finite initial support, the economy can still converge to an equilibrium with infinite support. It turns out that this will indeed be the case here. Moreover, Benhabib et al. (2017) show that in this case there are multiple equilibria with $g \leq \gamma$. However, this result is knife-edge in the sense that it will only appear if the length of the support goes to infinity. Adding Schumpeterian forces to the model will, in fact, ensure that the support will be bounded, leading to a unique equilibrium with $g = \gamma$.

Finally, when $g < \gamma$, the network does not affect the growth rate. This can be seen as follows. In equilibrium, the productivity distribution has a Pareto tail that is the same for each type and the growth rate is increasing in the thickness of the Pareto tail (Benhabib et al., 2017). Suppose that the economy is in equilibrium and that the network is changing. This will not affect the learning distribution of the most productive type due to directed search and will therefore not change the learning decision of this type. Hence, this type will still be in equilibrium and the

thickness of the Pareto tail is not affected by the network, thus making that the growth rate is not affected either.

The focus of this paper is on how the network affects aggregate productivity. As the network does not affect the growth rate, I will only study the equilibrium with $g = \gamma$ in order to keep the model as simple as possible.

Normalization

A stationary system of equations is needed in order to solve the model. In order to achieve this, I normalize productivity using a normalization factor that grows at the same rate as the economy. Because the goal of this paper is to compare the effect of the network on aggregate productivity and because I will only be able to solve for the stationary system of equations and not for the normalization factor, it is necessary that the normalization factor does not depend on the network.

These conditions are fulfilled when the normalization factor is the reservation productivity of the type with the largest reservation productivity, which is denoted by $\overline{M}(t) = \max_n M_n(t)$. By Lemma 2, this grows at the same rate as the economy and, as I will show momentarily, it does not depend on the network. All other reservation productivities turn out to depend on the network and would therefore not be adequate. Normalized productivity z becomes

$$z = \log \left(\frac{Z}{\overline{M}(t)} \right).$$

z can be negative for types that do not have the largest reservation productivity. Define $\alpha_n = \log \left(\frac{M_n(t)}{\overline{M}(t)} \right)$ as the minimum of support of the normalized distribution for each type. On a balanced growth path this will be independent of time since on a balanced growth path $M_n(t)$ and $\overline{M}(t)$ grow at the same rate. Define $\bar{z} = \log \left(\frac{\overline{Z}(t)}{\overline{M}(t)} \right)$, the normalized maximum of support. Furthermore, define the normalized productivity functions F_n^i and value functions v_n^i as

$$\begin{aligned} F_n^i(t, z) &= F_n^i \left(t, \log \left(\frac{Z}{\overline{M}(t)} \right) \right) \equiv \Phi_n^i(t, Z), \\ v_n^i(t, z) &= v_n^i \left(t, \log \left(\frac{Z}{\overline{M}(t)} \right) \right) \equiv \frac{V_n^i(t, Z)}{\overline{M}(t)}, \end{aligned}$$

such that these do not depend on time on a balanced growth path. The productivity distribution unconditional on the innovation state is denoted by $F_n = F_n^l + F_n^h$. When the index n is omitted, $F(t, z)$ denotes a column vector where the n -th entry is $F_n(t, z)$. Thus, $F(t, \alpha_n)$ is a vector of which the k -th entry is the CDF of type k

evaluated at α_n : $F_k(t, \alpha_n)$. On a balanced growth path, all time derivatives will be set to 0 and we can omit all time indices. The normalized system of equations looks as follows (for each type n):¹⁸

$$0 = g\partial_z F_n^l(z) - \lambda_n^l F_n^l(z) + \lambda_n^h F_n^h(z) + (S_n^l + S_n^h)\hat{F}_n(z) - S_n^l, \quad (10)$$

$$0 = (g - \gamma_n)\partial_z F_n^h(z) - \lambda_n^h F_n^h(z) + \lambda_n^l F_n^l(z) - S_n^h, \quad (11)$$

$$0 = F_n^l(\alpha_n) = F_n^h(\alpha_n), \quad (12)$$

$$1 = F_n^l(\bar{z}) + F_n^h(\bar{z}), \quad (13)$$

$$S_n^l = g\partial_z F_n^l(\alpha_n), \quad \text{if } g > 0, \quad (14)$$

$$S_n^h = (g - \gamma_n)\partial_z F_n^h(\alpha_n), \quad \text{if } g > \gamma_n, \quad (15)$$

$$(\rho - g)v_n^l(z) = e^z - g\partial_z v_n^l(z) + \lambda_n^l (v_n^h(z) - v_n^l(z)), \quad (16)$$

$$(\rho - g)v_n^h(z) = e^z - (g - \gamma_n)\partial_z v_n^h(z) + \lambda_n^h (v_n^l(z) - v_n^h(z)), \quad (17)$$

$$\partial_z v_n^l(\alpha_n) = 0, \quad \text{if } g > 0, \quad (18)$$

$$\partial_z v_n^h(\alpha_n) = 0, \quad \text{if } g > \gamma_n, \quad (19)$$

$$v_n(\alpha_n) = \int_{\alpha_n}^{\bar{z}} v_n^l(\hat{z}) d\hat{F}_n(\hat{z}) - \xi_n e^{\alpha_n}, \quad (20)$$

$$\hat{F}(z) = I_{z \geq \alpha_n} (\mathbf{1} + X^{-1} (AF(z) - \mathbf{1})). \quad (21)$$

Equations (10) and (11) are the Kolmogorov forward equations. Equations (12) and (13) say that the CDF is 0 at the minimum of support and 1 at the maximum of support for each type. The share of agents learning equals the mass of agents at the reservation productivity times the growth rate of the economy relative to productivity growth when innovating (equations (14) and (15)). Equations (16) and (17) are the Bellman equations. Equations (18) and (19) are the smooth pasting conditions. Equation (20) is the value matching condition and equation (21) gives the expression for the normalized learning distribution. Equations (10)-(20) hold for each type while equation (21) is in matrix-vector notation (and therefore comprises N equations).

As explained above, I will solve for the equilibrium in which $g = \gamma$. Further-

¹⁸The following equations are useful when normalizing:

$$\frac{\partial \Phi_n^i(t, Z)}{\partial t} = \frac{dF_n^i\left(t, \log\left(\frac{Z}{\bar{M}(t)}\right)\right)}{dt} = \frac{\partial F_n^i(t, z)}{\partial t} - \frac{\partial F_n^i(t, z)}{\partial z} \frac{\bar{M}'(t)}{\bar{M}(t)},$$

$$\frac{\partial \Phi_n^i(t, Z)}{\partial Z} = \frac{dF_n^i(t, z)}{dz} \frac{dz}{dZ} = \frac{dF_n^i(t, z)}{dz} \frac{1}{Z},$$

$$\frac{\partial V_n^i(t, Z)}{\partial t} = \frac{d\bar{M}(t)v_n^i\left(t, \log\left(\frac{Z}{\bar{M}(t)}\right)\right)}{dt} = \left(\bar{M}'(t)\right)v_n^i(t, z) + \bar{M}(t)\frac{\partial v_n^i(t, z)}{\partial t} - \bar{M}(t)\frac{\partial v_n^i(t, z)}{\partial z} \frac{\bar{M}'(t)}{\bar{M}(t)}.$$

more, for simplicity, I set γ_n equal to γ for each type. This means that agents in the high innovation state will not learn in equilibrium as those agents are not falling back relative to the reservation productivity. Hence, $S_n^h = 0$ and $S_n = S_n^l$. Furthermore, the smooth pasting condition for the high innovation state will not be needed. In what follows I assume, without loss of generality, that the types are ordered based on their minimum of support (i.e., $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_N = 0$).

Proposition 1. *Suppose that assumption 1 holds and $\rho > g$ then the balanced growth path equilibrium with $g = \gamma$ has the following productivity distribution functions for the intervals indexed by $k = 1, \dots, N - 1$,*

$$F(z) = \mathbf{1} - e^{-\frac{1}{g}SX^{-1}I_k\Lambda A(z-\alpha_k)} \underbrace{\prod_{j=k-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_j\Lambda A(\alpha_{j+1}-\alpha_j)} \right) \mathbf{1}}_{\mathbf{1}-F(\alpha_k)}, \quad \text{if } \alpha_k \leq z < \alpha_{k+1}, \quad (22)$$

where Λ is a diagonal matrix with $1 + \frac{\lambda_n^l}{\lambda_n^h}$ forming the diagonal entries, S is a diagonal matrix with S_n forming the diagonal entries and I_k are matrices where the first k diagonal elements are 1 and the other elements are zeroes. For the final interval

$$F(z) = \mathbf{1} - e^{-\frac{\mu}{g}z} \underbrace{\prod_{j=N-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_j\Lambda A(\alpha_{j+1}-\alpha_j)} \right) \mathbf{1}}_{\mathbf{1}-F(\alpha_N)}, \quad \text{if } 0 = \alpha_N \leq z \leq \bar{z}, \quad (23)$$

with

$$\begin{aligned} \mu &= g \frac{1 - \nu_N + \sqrt{(1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho-g}}{\xi_N} \nu_N}}{2}, \\ \nu_n &= \frac{(\rho - g)\bar{\lambda}_n}{g}, \\ \bar{\lambda}_n &= \frac{\lambda_n^l}{\rho - g + \lambda_n^h} + 1, \end{aligned} \quad (24)$$

and $\bar{z} \rightarrow \infty$. The value functions are

$$\begin{aligned} v_n^l(z) &= \frac{\bar{\lambda}_n}{g + (\rho - g)\bar{\lambda}_n} e^z + \frac{1}{(\rho - g)(\nu_n + 1)} e^{-z\nu_n + \alpha_n(\nu_n + 1)}, \\ v_n^h(z) &= \frac{e^z + \lambda_n^h v_n^l(z)}{\rho - g + \lambda_n^h}. \end{aligned}$$

S_n and α_n are such that

$$SX^{-1}\Lambda A(\mathbf{1} - F(0)) = \mu(\mathbf{1} - F(0)), \quad (25)$$

$$v_n^l(\alpha_n) = v_n^h(\alpha_n) = \frac{e^{\alpha_n}}{\rho - g} = \int_{\alpha_n}^{\infty} v_n^l(\hat{z}) \hat{f}_n(\hat{z}) d\hat{z} - \xi_n e^{\alpha_n}, \quad \forall n, \quad (26)$$

which implies $S_N = \frac{\mu}{1 + \lambda_N}$.

Proof. See Appendix C. □

The products in this proposition are products of matrices and therefore the order of taking the product matters. The order is indicated by the subscript. For example, $\prod_{j=k-1}^1 D_j = D_{k-1} D_{k-2} \dots D_1$. Furthermore, $\prod_{j=0}^1 (\cdot) = I$.

Proposition 1 shows that in a balanced growth path equilibrium, the normalized frontier converges to infinity as time goes to infinity. Note that in the expressions for the CDF (equations (22) and (23)), the matrix exponential shows up whenever there is a matrix (i.e., a capital letter) in the exponent of the exponential function. The matrix exponential is the exponential function applied to a matrix, of which the result will be a matrix.¹⁹ In contrast to the scalar exponential, the matrix exponential does not necessarily only have positive elements. Taking the derivative of equation (22) with respect to z gives that the equilibrium density function is a matrix exponential multiplied by a positive vector and therefore, the pdf could be negative for a general matrix exponential.²⁰ To ensure that the pdf will be non-negative on the entire domain, an extra condition has to be fulfilled. This condition is that the vector $\prod_{j=N-1}^1 \left(e^{-\frac{1}{g} SX^{-1} I_j \Lambda A(\alpha_{j+1} - \alpha_j)} \right) \mathbf{1}$ is an eigenvector of the matrix $SX^{-1}\Lambda A$ (see Lemma 3 in Appendix C). Equation (25) in Proposition 1 ensures that this condition is met and μ is the associated eigenvalue. Then, as a result, by the properties of the matrix exponential, the matrix exponential can be replaced by the scalar exponential in the expression for the CDF on the last interval (i.e., $z \geq \alpha_N = 0$).²¹ This is done to obtain equation (23) and leads to the following result.

Corollary 1. *The distribution of (unnormalized) productivity has a (right) Pareto tail for each type and the shape parameter is the same for each type and does not depend on the network.*

¹⁹Suppose that D is a matrix, then the matrix exponential is defined as $e^D = I + D + \frac{D^2}{2!} + \frac{D^3}{3!} + \dots$

²⁰To see that this vector is positive, note that all entries of the matrices S , X , Λ and A are non-negative as are the entries of $\mathbf{1} - F(\alpha_k)$.

²¹Suppose that v is the eigenvector of matrix D with the associated eigenvalue μ . Then,

$$e^D v = \left[I + D + \frac{D^2}{2!} + \frac{D^3}{3!} + \dots \right] v = \left[1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots \right] v = e^\mu v.$$

Proof. This follows directly from equation (23). The shape parameter is $\frac{\mu}{g}$. \square

A Pareto tail is consistent with the observed distributions of firm size and income (see, e.g., Luttmer (2007) and Gabaix et al. (2016), respectively). The shape parameter of the Pareto tail does not depend on the network. It depends on γ , ρ , ξ , λ^l and λ^h of the type with the highest reservation productivity.²² From equation (24) it follows that the shape parameter is decreasing in the learning cost ξ_N . An increase in the learning cost leads to a thicker Pareto tail because, in equilibrium, there need to be more highly productive agents to ensure that agents of type N are willing to pay the increased learning cost ξ_N .

In the right tail (i.e., $z \geq 0$), the pdf is the same for each type up to a multiplication factor. This multiplication factor is the mass of agents of a type with a productivity above 0, such that

$$f_n(z) = \frac{\mu}{g} e^{-\frac{\mu}{g}z} (1 - F_n(0)) , \quad \text{if } z \geq 0.$$

Recall that the n -th entry in X equals $a_{n-} (1 - F(\alpha_n))$, then the learning pdf for type N becomes

$$\hat{f}_N(\hat{z}) = \frac{\mu}{g} e^{-\frac{\mu}{g}\hat{z}} , \quad \text{if } \hat{z} \geq 0.$$

The learning distribution and hence the learning decision of type- N agents is independent of the network. This is because for a type- N agent, it does not matter from which type it learns, since from the perspective of that agent (i.e., only considering agents with a productivity above 0), all types are statistically the same since productivities follow the same Pareto tail.

That the network does not affect the type- N agents means that it neither does affect the normalization factor, $\max M_n(t)$. This feature allows me to compare the productivity distributions associated with different networks. As discussed above, when changes to the network would change the effective learning cost or the number of draws from the learning distribution this condition is violated. As an example of when the network affects the normalization factor, suppose that the number of draws equals the number of links d an agent has, such that the learning

²²The network does not affect which type has the highest reservation productivity. To see this, suppose that the two types with the largest reservation productivity are the same except that one has a lower λ^h and/or a higher λ^l such that this type has more agents in the innovation state, call this type-I and the other type-II. It can be verified that this makes that $v_n^l(z)/e^{\alpha_n}$ is larger for type-I, and hence by equation (26) type-I must have a larger reservation productivity than type-II. The only way in which it is possible for type-II to get a higher reservation productivity is that its position in the network is 'better' than for type-I (in the sense of being more likely to learn from a high productive type). But due to all types having the same Pareto tail, this is not possible and therefore, no matter what the network is, type-I will always have the highest reservation productivity.

CDF becomes $\left(\hat{F}(\hat{z})\right)^d$. Then, more links would make it more likely to meet a high productive agent, and would therefore increase the reservation productivity of the most productive type, and thus the normalization factor.

Finally, α_n and S_n need to be solved numerically such that equations (25) and (26) hold and once these variables are known, the distribution can be calculated directly using equation (22) (see Appendix F for details on the numerical algorithm).

Discussion

Having solved the model, I now discuss how the network affects the equilibrium distribution.²³ Here, I study networks with only two nodes in order to highlight the mechanisms present in the model and hence, changes to the network will only be on the intensive margin. This is an illustration of the working of the model. In the next section, I study networks with more nodes and do a serious calibration.

For now, suppose that the growth rate is $\gamma = 0.02$ and the discount rate is $\rho = 0.03$. Furthermore, take the learning cost and the transition rate from the high to the low innovation state to be the same for both types ($\xi_1 = \xi_2 = 25$ and $\lambda_1^h = \lambda_2^h = 2$) and that type-II agents are more likely to transition from the low innovation state to the high innovation state such that more type-II agents are in the high innovation state as compared to type-I agents ($\lambda_1^l = 0.5$ and $\lambda_2^l = 0.8$). This ensures that type-II has a higher reservation productivity and hence that the second row of A will have no effect on equilibrium outcomes. I vary the intensity at which type-I agents learn from type-I and type-II agents. Figure 1 shows the pdf of normalized productivity z for four different cases of the learning matrix. In the first panel, a type-I agent that decides to learn has a 95% probability of meeting a type-I agent and a 5% probability of meeting a type-II agent. In the second panel, the probability of meeting a type-II agent increases to 80%, in the third panel the probability is 50/50 and in panel (d) there is only a 5% probability that a type-I agent meets an agent of its own type. The change of the learning matrix has three effects.

The main effect is on the learning reservation productivities. When type-I agents have a high probability of meeting an agent of their own type, the value of learning is low because type-I agents are relatively unproductive (due to being less likely to be in the high innovation state). Therefore, agents wait with paying the learning costs until they have fallen back far enough relative to the other agents such that, in expected value, their productivity will increase enough to make it

²³See Appendix D for how the learning cost and the transition intensities between the low and the high innovation states affect the productivity distribution.

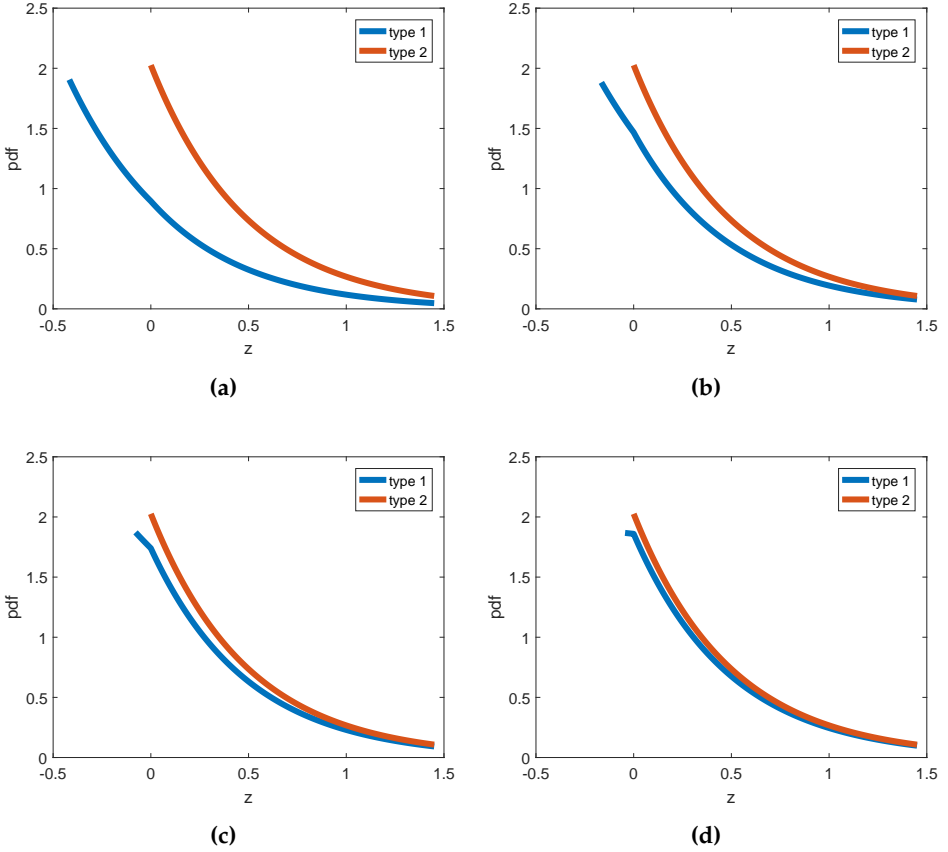


Figure 1: Pdf of normalized productivity z when having two types and the probability that a type-I agent learns from a type-II agent is (a) 5%, (b) 20%, (c) 50% and (d) 95%.

worth paying the learning costs. That is why in panel (a) the least productive of type I are much less productive than the least productive of type II. As it becomes more likely to meet a relatively high productive type-II agent (panels (b) through (d)), the gains from learning become larger and agents of type I do not need to fall back as far to be willing to learn. Through its effect on the reservation productivities, the learning matrix affects inequality and aggregate productivity. In panel (d), low productive agents have decided to learn, leading to higher aggregate productivity and less inequality than in panels (a) through (c).

The second effect is on the share of agents that decides to learn in equilibrium. Recall that the mass of learners of a type equals the growth rate of the reservation productivity times the mass of agents at the reservation productivity. First note that this means that not only the least productive in the economy are learning, which is an important difference from Perla and Tonetti (2014) and Benhabib et al.

(2017). Here, also the least productive type-II agents are learning, and depending on the network, these can have quite a high productivity compared to the least productive agents in the economy. Hence, this model shows that it is possible to have search at the extensive margin without having only the least productive deciding to learn. Going back to Figure 1, the growth rate is constant across the four panels and hence, the change in the mass that learns is proportional to the change in the densities at the reservation productivities. For type II, the density does not change and hence the mass of type-II learners is independent of the learning matrix. For type I, the density at the minimum of support goes down moving from panels (a) to (d). This gives the slightly surprising result that the more connected type-I agents are with the high productive type of agents, the lower is the mass of learners in equilibrium, although the effect is quantitatively small. Moreover, this is not a general result as it does not necessarily hold for other parameter values. The reason for this is that there are two opposing effects. The first effect is that if agents get more connected to more productive agents, they become more willing to learn, thereby increasing the mass of learners. However, this also increases the reservation productivity and therefore increases the opportunity cost of learning which has a negative effect on the mass of learners. That the share that learns goes down in equilibrium is still consistent with a higher aggregate productivity. This is because in the transition to an equilibrium with a higher reservation productivity, learning goes up temporarily.

The third and final effect of the learning matrix on the distribution is that when type-I agents are less likely to meet an agent of their own type, the more flat the pdf becomes in the region $[\alpha_0, \alpha_1]$. The reason for this is rather technical. For the distribution to be in equilibrium, the inflow of agents at the normalized productivity level z must be equal to the outflow at that productivity level. The outflow consists of the agents falling back relative to the minimum of support at the rate g times the slope of the pdf.²⁴ The inflow is those that are learning and drawing a productivity z . If the learning matrix becomes such that agents of type I are less likely to meet an agent of type I, the inflow will fall on $[\alpha_0, \alpha_1]$. Hence, in equilibrium, the outflow must also fall, meaning that the slope of the pdf has to become smaller.

Finally, with the current model, it is easy to generate a hump-shaped productivity distribution with a Pareto tail. One easy way of achieving this is to have the learning cost vary across types while holding all other parameters fixed. The first panel of Figure 2 shows the hump-shaped density across all nodes when there are

²⁴It is the slope of the pdf that is relevant because those just above z will flow to z whereas those at z will flow to just below z . Hence the difference in the pdf between z and $z + \varepsilon$ (i.e., the slope) defines the net outflow.

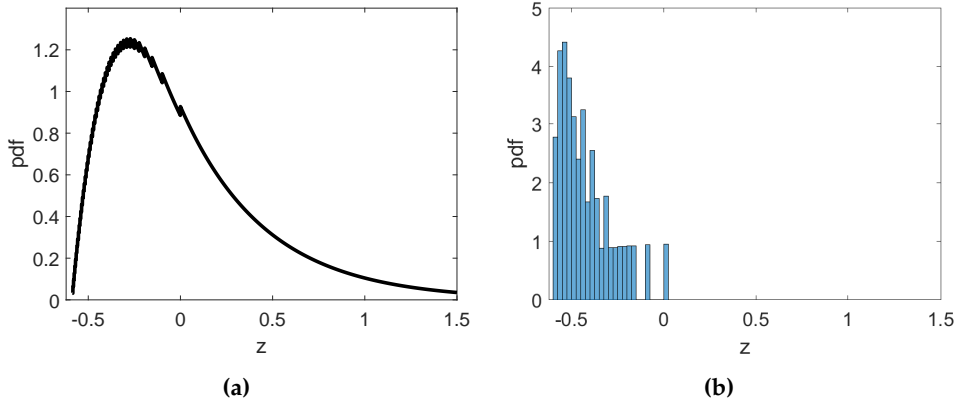


Figure 2: (a) Pdf of normalized productivity z of the entire economy and (b) histogram of those that are adopting when there is only variation in learning costs across types.

50 nodes with a complete network and the learning costs are uniformly distributed between 20 and 30. This resulting distribution is different from the models in Perla and Tonetti (2014) and Benhabib et al. (2017), which have the feature that the productivity distribution is downward sloping. The second panel shows the distribution of productivity of those that are adopting. As already noted above, it are not only the least productive in the economy that are adopting.

III Results

The previous section has highlighted the mechanisms of the model. However, since there were only two nodes, all nodes were connected to each other and therefore changes to the network did only reflect changes in the intensity at which agents met agents of the other type. In this section, I study economies with more than two nodes and test how different network properties affect the distribution of productivity and hence aggregate TFP.

To study the effect of the network, nodes need to differ ex ante, which is achieved by having heterogeneity in the innovation intensity. In order to get realistic outcomes, I estimate the innovation transition intensities using firm-level data. For this purpose, I interpret each type as an industry.²⁵ The dataset I use is the Amadeus dataset for France which covers private firms. France is chosen to maximize the number of observations. My analysis uses manufacturing firms for the period 2004-2016. I estimate the λ parameters for each industry as follows. The λ parameters govern at which intensity firms transition between

²⁵See Bernstein (1989) about there being knowledge diffusion across industries.

the two innovation states. Suppose that a firm is in the high innovation state if its detrended number of employees grows at a rate higher than 5% between two years, where the detrending is done at the industry level,²⁶ and that it is in the low innovation state if the growth rate is less than 5%. In this way, I construct a time series of the innovation state for each firm. Using this time series, I can determine how often firms transition from one state to the other. Aggregating, I calculate the transition probabilities at the industry level. For each industry n , I estimate λ_n^l and λ_n^h using these transition probabilities as follows

$$\begin{bmatrix} -\lambda_n^l & \lambda_n^l \\ \lambda_n^h & -\lambda_n^h \end{bmatrix} = \log \left(\begin{bmatrix} P_n^{ll} & P_n^{lh} \\ P_n^{hl} & P_n^{hh} \end{bmatrix} \right),$$

where P_n^{ll} is the probability of staying in the low innovation state etc. In the data, I do not want to mistake an adoption for an innovation. As it are only the low productive firms that adopt, I calculate the transition probabilities excluding data on firms below the 25th percentile of the firm size distribution within an industry. There is quite some heterogeneity in the estimates. The values of λ^h are in general quite high (ranging between 1.4 and 3.4), indicating that firms only spent a short amount of time in the high innovation state, and λ^l ranges between 0.5 and 1.1 (see Figure 10 in Appendix E).

The remaining calibration is as follows. The productivity growth rate of those in the high innovation state, γ , is set to 0.02 to target an annual growth rate of 2%. The discount rate, ρ , is set to 0.03. Finally, the learning costs are set such that the thickness of the Pareto tail matches the estimated Pareto tail of the firm size distribution. Luttmer (2007) estimates a Pareto shape parameter of 1.06 for the size distribution.²⁷ However, this is not directly comparable to the shape parameter in the model (i.e., $\frac{\mu}{g}$) as this is the shape parameter of the productivity distribution and there is not necessarily a linear mapping between productivity and firm size. Benhabib et al. (2017) show that in a monopolistic competition model, the tail parameter of the size distribution needs to be multiplied by the elasticity of substitution between varieties minus one to obtain the tail parameter of the productivity distribution. Therefore, an elasticity of substitution of 3 makes that the shape parameter of the tail of the productivity distribution is 2.12. Targeting $\frac{\mu}{g} = 2.12$ gives a learning cost of 23.2503 which is taken to be the same for all types.

²⁶Employment is used as a measure of productivity for the following reason. Suppose that marginal costs are proportional to the inverse of productivity. Then a 1% increase in productivity will lead to a 1% drop in marginal costs. When markups are constant the price will drop by 1% as well. Then, if the elasticity of demand is larger than 1 (which is the empirically relevant case for most products), the number of items sold will grow by more than 1% and hence, there is also an increase in labor demand.

²⁷The Amadeus dataset cannot be used to estimate the tail parameter because it omits the largest (public) firms.

For the remainder of this paper, this calibration will be used. I replicate each of the sixteen industries ten times such that each simulated network consists of 160 nodes.

Dense Networks Lead to High Aggregate Productivity and Low Inequality

To study the role of different network properties, I first focus on the role of network density, which is the proportion of potential links that are actual links. For this purpose, I draw networks randomly and the probability of a link forming between two nodes is independent of each other (i.e., an Erdős-Rényi random network). I assume that the network is undirected which means that learning is symmetric; if there is a link between two nodes, both nodes can learn from each other. In addition, all agents are assumed to also be able to learn from agents of their own type. To summarize, the network matrix h has the following properties: $h_{ii} = 1$ and $h_{ij} = h_{ji}$ where the probability that $h_{ij} = 1$ is p .

To create variation in the density I vary p . If p is high, many links will be formed such that the network density is high while a low p implies a sparse network. I also refer to dense networks as networks with a high average degree. I only consider networks that turn out to be connected. For each simulated network, I solve for the equilibrium distribution of productivity using the numerical algorithm outlined in Appendix F. The first panel of Figure 3 shows the relationship between the network density and TFP across networks. When calculating network density, the links between agents of the same node are not included. TFP is the average productivity across agents: $\frac{1}{N} \sum_n \int e^z dF_n(z)$. In all graphs, I normalize TFP such that the average TFP within a graph equals 1.

Each dot in Figure 3 represents a different network. There is a positive relationship between network density and TFP.²⁸ The return to adding more links is high for sparse networks but decreases sharply as the average degree increases.²⁹ One thing that is apparent is that the variation of TFP across simulations is larger for sparse networks than for dense networks. This reason for this is that when there are few links, it is important how the types are distributed across nodes. When a highly innovative type is centrally located, TFP will be large because all nodes connected to this node will have a high reservation productivity. These

²⁸The relationship between network density and income after subtracting the learning costs from productivity is identical.

²⁹The effect of increasing the number of links for societies with a sparse network is, in fact, larger than what is suggested by Figure 3. To construct Figure 3 I required the network to be connected for each simulation. When there are only a few links, it is likely that the network is not connected and, in that case, if an innovative type were isolated, TFP would be lower.

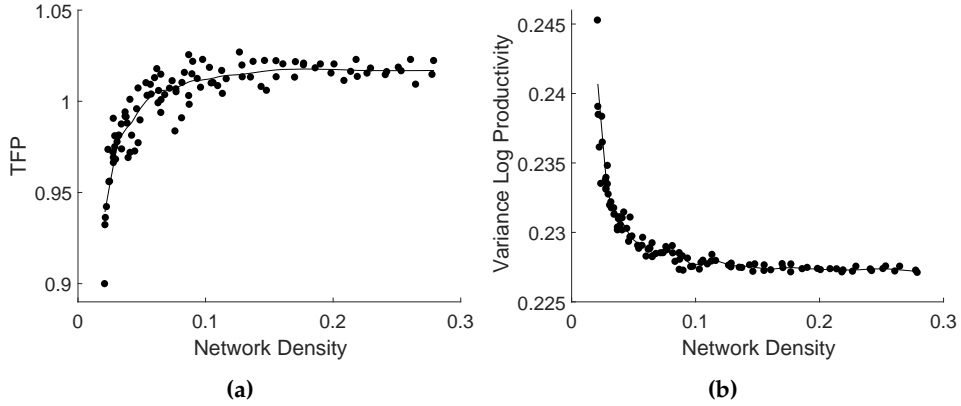


Figure 3: Relationship between network density and (a) TFP and (b) the variance of log productivity for simulated Erdős-Rényi random networks. Network density is the proportion of potential connections that are actual connections. Each data point represents a simulation and the line represents the best fit according to a local linear regression.

are the networks located at the left top of the figure. At the left bottom are the sparse networks in which the low innovative types happen to be centrally located, leading to low reservation productivities and a low TFP. When the network is dense, the variation across simulations is lower because there is less variation in centrality measures across nodes and therefore it matters less how the different types are distributed across nodes. The second panel of Figure 3 shows that there is a negative relationship between network density and the variance of log productivity (within a simulation), which is a measure of inequality. The reason for this negative relationship is the same as why there is a positive relationship between network density and TFP. When there are more links, agents are more likely to be connected to the high innovative types which tend to have a higher productivity. This increases the gains from learning for these nodes and therefore increases their reservation productivity (or learning effort), which, in turn, increases TFP and because it are the less productive types that increase their reservation productivity, this also lowers inequality (recall that the network does not affect the most productive type). The nodes connected to these nodes with an increased reservation productivity will also increase their effort and so on.

An issue with simulating Erdős-Rényi random networks is that it does not capture that the empirical degree distribution (across nodes) tends to have fat tails (see Bernard et al. (2018) and Price (1965) for the degree distribution among firms and scientific citations, respectively). To test whether this matters for learning effort, I simulate networks using preferential attachment (Barabási and Albert, 1999) which generates more plausible degree distributions. The algorithm works

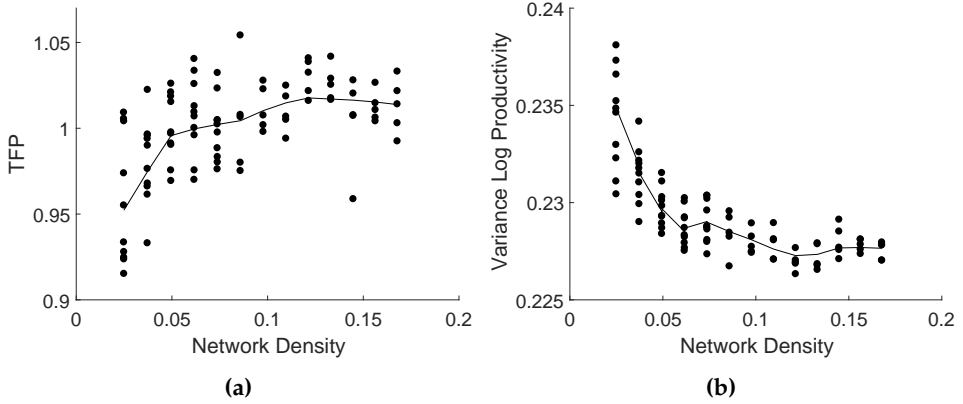


Figure 4: Relationship between network density and (a) TFP and the (b) variance of log productivity when networks are simulated using preferential attachment. Network density is the proportion of potential connections that are actual connections. Each data point represents a simulation.

as follows. The network starts with m_0 nodes and all links between these nodes are present. New nodes are added one-by-one and they form m links with existing nodes. The probability that a link is formed with an existing node is proportional to the number of links the existing node already has (this is why it is called preferential attachment). This ensures that the nodes that were there from the beginning have many links while nodes that were added at the end only have a few links. I vary m_0 between 3 and 15 and m between 2 and 14 to generate differences in network density. Figure 4 shows that the resulting relationship between network density and TFP and inequality is similar to what is found for Erdős-Rényi random networks, namely a positive effect of network density on TFP and a negative effect on inequality. When the degree distribution has fat tails, the dispersion across simulations increases. This is because when there are nodes with many connections, it matters whether these nodes are highly innovative since the productivity of these nodes directly affects the learning decision of many other nodes.

A Low Average Path Length Leads to High Aggregate Productivity and Low Inequality

Figures 3 and 4 show that higher degree networks are associated with a higher TFP, but it is still unclear what the importance of different network properties is, since changing the number of links also affects other network properties. For instance, increasing the number of links lowers the average path length.

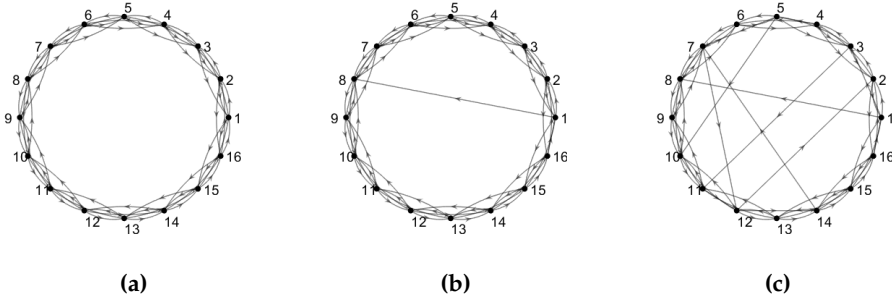


Figure 5: Example Watts-Strogatz Algorithm. (a) Initialization of the algorithm, (b) 1 link has been rewired and (c) 6 links have been rewired.

To study the effect of the average path length on the productivity distribution, I simulate networks in which I vary the average path length while keeping the number of links constant using the algorithm proposed by Watts and Strogatz (1998). The algorithm works as follows. Start with a circle network in which each node is connected to its four neighbors at both sides (so eight links per node in total). Draw a node randomly and with probability q break up the link with its first neighbor to the right and form a new link with a randomly drawn node. Do this for all nodes once. Then, do the same for all nodes again but now break up the link with its second neighbor to the right with probability q and do the same for the third and fourth neighbor to the right. The advantage of this algorithm is that it changes the average path length while keeping the total number of links constant.³⁰ Figure 5 shows an example of how the algorithm works in case of 16 nodes and four links for each node. In this graph, an arrow originating from a node means that that node is learning from the node to which the arrow is pointing. The first network shows the circle network from which the algorithm starts. The second network is the same as the first network except that the link between node 1 and node 16 is broken up and replaced by a link between node 1 and node 8. This lowers the average path length as node 1 (and its neighboring nodes) are now closely connected to node 8 and its neighboring nodes while the average degree is not affected. The third figure shows the network when multiple links have been broken up and this is the network with the lowest average path length of the three networks displayed. This is an example of what the network looks like once the algorithm has been finished. I differ the probability q across simulations to create dispersion in the average path length.

Figure 6 shows the resulting relationship between the average path length and

³⁰The clustering coefficient (i.e., the fraction of neighbors that are directly connected to each other) also changes.

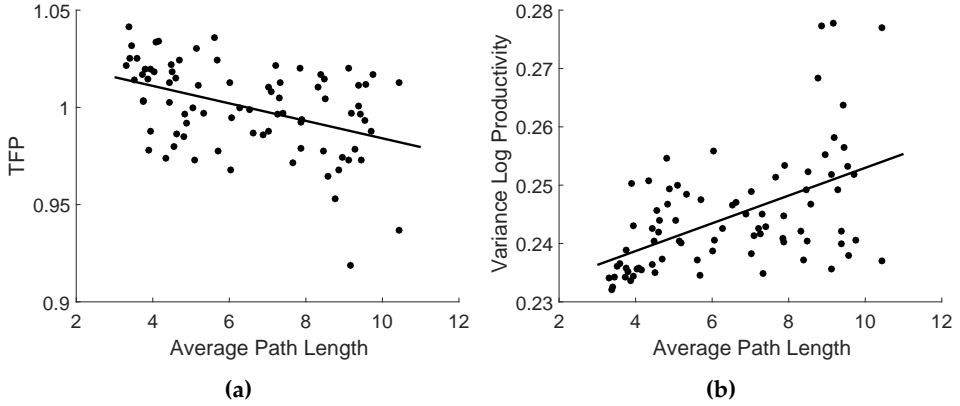


Figure 6: Relationship between average path length and (a) TFP and the (b) variance of log productivity for networks simulated using the Watts and Strogatz algorithm. Each data point represents a simulation and the line represents the linear best fit.

TFP and inequality. There is a decreasing relationship between the average path length and TFP, while increasing the average path length increases the variance of log productivity.³¹ When the average path length is low, nodes are closely connected to each other. One might not be directly connected to an innovative type but with a low average path length, it is likely that a neighboring node is connected to an innovative type, thus making this neighboring node more productive due to learning. This, in turn, increases the gains from learning for the original node. This process increases the reservation productivities and hence leads to higher productivity and lower inequality.

A low average path length is a property of what is called a ‘small world’ network.³² Duernecker and Vega-Redondo (2018) provide a theory for how ‘small worlds’ develop and argue that ‘small worlds’ are beneficial for economic growth as they increase the opportunities for collaboration. My model highlights another positive effect of ‘small worlds’, namely the enhancement of the diffusion of ideas by increasing the effort agents put into learning.

Finally, Table 1 projects TFP and inequality on network density and average path length, and confirms the earlier result that network density has a positive effect on TFP while the average path length has a negative effect on TFP. However, controlling for the average path length makes the effect of network density on inequality disappear.

³¹The relationship between the diameter, which is the maximum distance between any pair of nodes in a network, and TFP and the variance is similar.

³²‘Small worlds’ is a typical property found in real world networks (see Travers and Milgram (1969) for an example).

Table 1: Effect of network properties on TFP and inequality (variance of log productivity).

	(1) TFP	(2) Inequality
Network Density	0.24*** (0.048)	-0.0068 (0.0046)
Average path length	-0.013*** (0.0011)	0.0032*** (0.00031)
Mean dependent variable	1.54	0.23
Observations	271	271
R^2	0.53	0.68

Robust standard errors in parentheses. Based on simulations of Erdős-Rényi, Watts-Strogatz and preferential attachment networks.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Closeness Centrality is Important for Productivity

In the above I have discussed the effect of network properties on aggregate TFP. Here I will discuss how the location of a node within the network affects the productivity of that node. For instance, is it the case that having more connections leads to a higher productivity? To this end, I construct the following centrality measures that are standard in the literature. Degree centrality is the degree of a node divided by the maximum degree possible (159 in this case). Eigenvector centrality is the eigenvector of the network h associated with the largest eigenvalue.³³ The eigenvector is normalized such that the sum is 1. The eigenvector centrality takes into account the centrality of the connections of a node. Closeness centrality measures the relative distance to other nodes: $\text{Closeness}_i = \frac{N-1}{\sum_j \text{path length}_{i,j}}$. The shorter the distances the higher the closeness centrality. Bonacich centrality gives each node a base value equal to the degree of that node and then adds all nodes that are distance 1 away times b times the base value of those nodes. This continues subsequently with all nodes that are distance 2 away times b^2 times the base value of those nodes etc. (i.e., Bonacich centrality is $(I - bh)^{-1}h\mathbf{1}$). I use a value of $b = 0.05$.

In addition, I add a centrality measure based on the model. Recall from Proposition 1 that $\mathbf{1} - F(0)$ is an eigenvector of $SX^{-1}\Lambda\Lambda$. $\mathbf{1} - F(0)$ is a vector where the n -th entry denotes the share of type- n agents with a normalized productivity above 0. In other words, the larger an entry in this eigenvector, the more highly productive agents a type has. Ignoring the equilibrium matrices SX^{-1} , I propose the eigenvector of $\Lambda\Lambda$ as an alternative centrality measure. This centrality measure

³³The largest eigenvalue is chosen to ensure that the eigenvector only consists of positive entries (by Perron-Frobenius).

Table 2: Effect of centrality measures on TFP of a type

	(1) TFP	(2) TFP
Degree centrality	-0.0040*** (0.00068)	-0.0040*** (0.00065)
Closeness centrality	0.045*** (0.0010)	0.045*** (0.00096)
Eigenvector centrality	-0.00045 (0.00044)	-0.00036 (0.00043)
Bonacich centrality	0.00024 (0.00022)	0.00024 (0.00022)
Eigenvector centrality ΛA		0.041*** (0.00085)
Type fixed effects	X	X
Mean dependent variable	1.54	1.54
Observations	43360	43360
R^2	0.50	0.56

Robust standard errors in parentheses. Point estimates denote the effect of a one standard deviation increase. Based on simulations of Erdős-Rényi, Watts-Strogatz and preferential attachment networks.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

weights connections by their innovation intensity.

Table 2 shows the results from projecting TFP of each type on the centrality measures. The point estimates shown indicate the effect of a one standard deviation increase in the centrality measure on TFP. Since the eigenvector of ΛA needs more information than only the network, the first column omits this centrality measure while it is included in the second column. Furthermore, type fixed effects are included. The closer a type is to all other types, the higher is its productivity. The reason why closeness centrality is important for productivity is that a high closeness centrality means that a node is closely connected to many types and is therefore likely to be connected to some highly productive types, thereby increasing the learning value. There is a negative effect of degree centrality on productivity. The reason is that degree centrality is highly correlated with closeness and eigenvector centrality. Including the centrality measures separately, all of them have a significant positive effect on TFP. The innovation intensity weighted eigenvector has a strong positive effect, as should have been expected from the model.

The results from Table 2 are consistent with what is found in Duernecker et al. (2016). They construct the country-level network based on trade flows and

show that a higher closeness centrality is positively correlated with economic growth. They argue that to study the effect of openness, one should not only need to take into account the intensity of trade with a country's connections but should also take into account the higher-order connections (i.e., the full network). Here I provide a theoretical foundation for taking the network into account when studying the effect of openness on TFP. If only the first-degree connections had mattered, there would have been no effect of closeness centrality on TFP, and degree centrality would have been the only significant variable. Instead I find that having a higher closeness centrality has a large and positive effect on TFP. One does not only want to be connected to many nodes but ideally, these nodes are also connected to many other nodes themselves. Increasing closeness centrality by 1 standard deviation increases TFP for that node by 3%.

Share of Agents Learning Unaffected by Network

That denser networks have a larger TFP does not necessarily imply that these networks have more agents learning in equilibrium, as was already indicated in the discussion following Figure 1. Figure 7 plots the relationship between the network density and the share of agents that learn, and although there is a positive relationship the effect is negligible. In a network where 30% of the potential links are active, agents are less than 0.01 percentage points more likely to be learning during a year than in a network where 3% of the links are formed. At first this seems contradictory to the result that a dense network has a 5% higher TFP than a sparse network. The reason is that the density of the network affects the reservation productivity, and in a dense network, agents do not need to fall back as far to make it advantageous to learn. This leads to a rightward shift of the productivity distribution of a type and therefore, it hardly affects the mass of agents at the reservation productivity, which is what determines the share of agents that are learning.

This result has implications for empirical studies studying the effect of the network on learning. Suppose that one has a data set where, for each observation, the full network is known and it is known when agents adopt a new technology. By the above result, regressing the share of agents that decides to adopt in a given period on properties of the network will find a zero effect even though there might be a large effect of the network on productivity through diffusion. Instead, one can run a regression of productivity on network properties, but the problem with this approach is that it is contaminated by initial productivity differences across observations. Therefore, to empirically study how networks affect diffusion, one needs to focus on economies that are not in equilibrium and study whether a

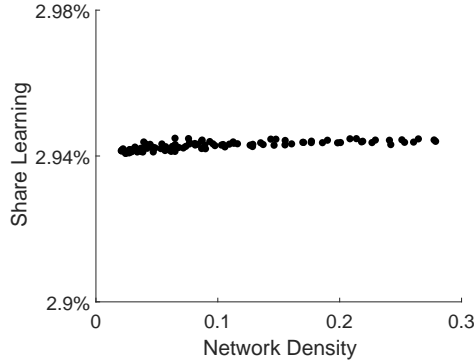


Figure 7: Relationship between average degree and the share of agents learning for simulated Erdős-Rényi random networks. Network density is the proportion of potential connections that are actual connections. Each data point represents a simulation.

change in network properties temporarily affects the share of adopters.

IV Conclusions

In this paper I study how the decision to learn depends on the network. I find that the more dense the network is, the more learning effort agents apply, thus leading to a higher TFP and a lower inequality. The effect of adding links is especially strong in networks with a small number of links. This suggests that in economies where the network is sparse, TFP and equality could be increased substantially by adding links to the network. For economies in which the network is already highly connected, the gains from adding more links are much lower. In addition to the number of links, a low average path length is beneficial for the diffusion of ideas. A low average path length means that agents are closely connected to all other agents. Then, the innovation of agents spill over more easily to the rest of the economy.

Empirically it is known that the dispersion of productivity across firms is large (Syverson, 2004) and tends to be larger in low-income countries (Hsieh and Klenow, 2009). One possibility for this difference in dispersion is that different network structures emerge in different countries. This paper shows that if low-income countries have a more sparse network, then this can explain part of the differences in the dispersion of productivity and in the level of TFP. However, the effect is modest and network differences therefore do not have the potential to explain all of the cross-country differences in TFP and dispersion. One important remark, though, is that the effect of network density on TFP and the distribution

of productivity I found here is a lower bound. If, for instance, having more links were to imply having multiple draws from the learning distribution, this would provide an additional force, making dense networks have a higher TFP and a lower dispersion. In addition, I have only considered connected networks. A sparse network makes it more likely that some nodes are isolated, such that their innovations would never spill over to the rest of the economy.

There are several reasons for why the network could be more sparse in low-income countries. One reason is having lower trust which would prevent the formation of links along which ideas diffuse. Consistent with this, Algan and Cahuc (2010) show that trust is positively associated with economic development and Knack and Keefer (1997) find that trust is negatively correlated with inequality. An alternative reason is differences in institutions. Boehm and Oberfield (2018) document that in Indian states with a low formal contract enforcement, there is more vertical integration which implies a sparser network.

In order to study whether network differences can indeed explain part of the cross country variation in income, one needs to have network data for a large set of countries, which is generally unavailable. However, Fogli and Veldkamp (2016) provide a promising first step. Based on data from immigrants to the US, they construct a network index for 69 countries and find that a higher index of network quality leads to faster diffusion. Their index of network quality consists of measures of the degree, average path length and clustering and their results are therefore consistent with the results of my model.

The model is also consistent with the observed negative relationship between aggregate productivity and inequality across countries. Here it is not the case that less inequality causes higher productivity or the reverse, but a third factor, namely the density of the network, causes both a higher TFP and a lower inequality. For this reason, in a cross-country analysis studying the relationship between inequality and development, one should control for the network to avoid omitted variable bias.

One final remark is that the networks studied here are formed exogenously. If agents tend to predominantly connect to highly productive agents, the effect of network density on TFP will be lower than what is suggested above since when agents are already connected to the most productive nodes, there is no gain from adding more links (unless this leads to multiple draws from the productivity distribution). However, if the decision to form links is orthogonal to productivity, then studying exogenous networks is adequate. Plausibly, ideas diffuse along all kinds of connections, being both formed exogenously and endogenously taking learning into account. The results from this paper mainly relate to diffusion over

links that are formed orthogonal to productivity. It would be interesting to study what types of networks are formed endogenously, in order to study how policies can affect the network and therefore development. However, in order to do this, a stance needs to be taken on along what type of links ideas flow (e.g., do agents learn from agents they trade with, that are geographically close or that have similar personal traits etc.?).

References

- Acemoglu, D., U. Akcigit, and W. Kerr (2016a) "Innovation network," *Proceedings of the National Academy of Sciences*, Vol. 113, No. 41, p. 11483–11488.
- (2016b) "Networks and the Macroeconomy: An Empirical Exploration," *NBER Macroeconomics Annual*, Vol. 30, pp. 273–335.
- Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012) "The Network Origins of Aggregate Fluctuations," *Econometrica*, Vol. 80, No. 5, pp. 1977–2016.
- Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2017) "Microeconomic Origins of Macroeconomic Tail Risks," *American Economic Review*, Vol. 107, No. 1, pp. 54–108.
- Acemoglu, D., A. Ozdaglar, and E. Yildiz (2011) "Diffusion of Innovations in Social Networks," in *Proceedings of the 50th IEEE Conference on Decision and Control*, pp. 2329–2334.
- Akbarpour, M. and M. O. Jackson (2018) "Diffusion in Networks and the Virtue of Burstiness," *Proceedings of the National Academy of Sciences*, Vol. 115, No. 30, p. E6996–E7004.
- Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi (2018) "Dancing with the Stars: Innovation Through Interactions," Working Paper 24466, National Bureau of Economic Research.
- Algan, Y. and P. Cahuc (2010) "Inherited Trust and Growth," *American Economic Review*, Vol. 100, No. 5, pp. 2060–2092.
- Alvarez, F. E., F. J. Buera, and R. E. Lucas, Jr. (2013) "Idea Flows, Economic Growth, and Trade," Working Paper 19667, National Bureau of Economic Research.
- Bailey, N. T. J. (1975) *The Mathematical Theory of Infectious Diseases and its Applications*: Oxford University Press, 2nd edition.
- Baqaei, D. R. (2018) "Cascading Failures in Production Networks," *Econometrica*, Vol. 86, No. 5, pp. 1819–1838.
- Baqaei, D. R. and E. Farhi (2020) "Productivity and Misallocation in General Equilibrium," *Quarterly Journal of Economics*, Vol. 135, No. 1, pp. 105–163.
- Barabási, A.-L. and R. Albert (1999) "Emergence of Scaling in Random Networks," *Science*, Vol. 286, No. 5439, pp. 509–512.

- Benhabib, J., J. Perla, and C. Tonetti (2017) "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier," Working Paper 23095, National Bureau of Economic Research.
- Bernard, A. B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes (2018) "The Origins of Firm Heterogeneity: A Production Network Approach," Manuscript.
- Bernstein, J. I. (1989) "The Structure of Canadian Inter-Industry R&D Spillovers, and the Rates of Return to R&D," *Journal of Industrial Economics*, Vol. 37, No. 3, pp. 315–328.
- Bigio, S. and J. La'O (2017) "Distortions in Production Networks," Manuscript.
- Boehm, J. and E. Oberfield (2018) "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," Working Paper 24937, National Bureau of Economic Research.
- Buera, F. J. and E. Oberfield (2020) "The Global Diffusion of Ideas," *Econometrica*, Vol. 88, No. 1, pp. 83–114.
- Cai, J., N. Li, and A. M. Santacreu (2017) "Knowledge Diffusion, Trade and Innovation Across Countries and Sectors," Working Paper 2017-029A, Federal Reserve Bank of St. Louis.
- Carvalho, V. M. (2010) "Aggregate Fluctuations and the Network Structure of Intersectoral Trade," Economics Working Papers 1206, Universitat Pompeu Fabra.
- Carvalho, V. M. and A. Tahbaz-Salehi (2019) "Production Networks: A Primer," *Annual Review of Economics*, Vol. 11, pp. 635–663.
- Carvalho, V. M. and N. Voigtländer (2015) "Input Diffusion and the Evolution of Production Networks," Working Paper 20025, National Bureau of Economic Research.
- Chaney, T. (2014) "The Network Structure of International Trade," *American Economic Review*, Vol. 104, No. 11.
- Coe, D. T. and E. Helpman (1995) "International R&D Spillovers," *European Economic Review*, Vol. 39, No. 5, pp. 859–887.
- Coleman, J., E. Katz, and H. Menzel (1957) "The Diffusion of an Innovation Among Physicians," *Sociometry*, Vol. 20, No. 4, pp. 253–270.

- Conley, T. G. and C. R. Udry (2010) "Learning About a New Technology: Pineapple in Ghana," *American Economic Review*, Vol. 100, No. 1, pp. 35–69.
- Duernecker, G., M. Meyer, and F. Vega-Redondo (2016) "The Network Origins of Economic Growth," Manuscript.
- Duernecker, G. and F. Vega-Redondo (2018) "Social Networks and the Process of Globalization," *Review of Economic Studies*, Vol. 85, No. 3, pp. 1716–1751.
- Ellison, G. (1993) "Learning, Local Interaction, and Coordination," *Econometrica*, Vol. 61, No. 5, pp. 1047–1071.
- Fogli, A. and L. Veldkamp (2016) "Germs, Social Networks and Growth," Working Paper 18470, National Bureau of Economic Research.
- Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016) "The Dynamics of Inequality," *Econometrica*, Vol. 84, No. 6, pp. 2071–2111.
- Granovetter, M. S. (1973) "The Strength of Weak Ties," *American Journal of Sociology*, Vol. 78, No. 6, pp. 1360–1380.
- (1978) "Threshold Models of Collective Behavior," *American Journal of Sociology*, Vol. 83, No. 6, pp. 1420–1443.
- Griliches, Z. (1957) "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, Vol. 25, No. 4, pp. 501–522.
- Hardy, M. and J. McCasland (2018) "It Takes Two: Experimental Evidence on the Determinants of Technology Diffusion," Manuscript.
- Herkenhoff, K., J. Lise, G. Menzio, and G. Phillips (2018) "Production and Learning in Teams," Manuscript.
- Hopenhayn, H. and L. Shi (2017) "Knowledge Creation and Diffusion with Limited Appropriation," Manuscript.
- Hsieh, C.-T. and P. J. Klenow (2009) "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1403–1448.
- Hulten, C. R. (1978) "Growth Accounting with Intermediate Inputs," *Review of Economic Studies*, Vol. 45, No. 3, pp. 511–518.
- Jackson, M. O. and B. W. Rogers (2007) "Relating Network Structure to Diffusion Properties through Stochastic Dominance," *The B.E. Journal of Theoretical Economics*, Vol. 7.

- Jackson, M. O. and L. Yariv (2005) "Diffusion on Social Networks," *Économie publique/Public economics*, No. 16, pp. 3–16.
- Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, Vol. 108, No. 3, pp. 577–598.
- Jarosch, G., E. Oberfield, and E. Rossi-Hansberg (2018) "Learning from Coworkers," Manuscript.
- Jones, C. I. (2011) "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macroeconomics*, Vol. 3, No. 2, pp. 1–28.
- (2013) "Misallocation, Economic Growth, and Input–Output Economics," in *Advances in Economics and Econometrics* eds. by D. Acemoglu, M. Arellano, and E. Dekel, pp. 419–456: New York: Cambridge University Press.
- Knack, S. and P. Keefer (1997) "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *Quarterly Journal of Economics*, Vol. 112, No. 4, pp. 1251–1288.
- König, M. D., J. Lorenz, and F. Zilibotti (2016) "Innovation vs. Imitation and the Evolution of Productivity Distributions," *Theoretical Economics*, Vol. 11, No. 3.
- Kremer, M. (1996) "Integrating Behavioral Choice into Epidemiological Models of AIDS," *Quarterly Journal of Economics*, Vol. 111, No. 2, pp. 549–573.
- Lashkari, D. (2018) "Innovation, Knowledge Diffusion, and Selection," Manuscript.
- Liu, E. (2019) "Industrial Policies in Production Networks," *Quarterly Journal of Economics*, Vol. 134, No. 4, pp. 1883–1948.
- Long, J. B., Jr. and C. I. Plosser (1983) "Real Business Cycles," *Journal of Political Economy*, Vol. 91, No. 1, pp. 39–69.
- López-Pintado, D. (2008) "Diffusion in Complex Social Networks," *Games and Economic Behavior*, Vol. 62, No. 2, pp. 573–590.
- Lucas, R. E., Jr. and B. Moll (2014) "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, Vol. 122, No. 1, pp. 1–51.
- Luttmer, E. G. (2007) "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, Vol. 122, No. 1, p. 1103–1144.

- (2012a) “Eventually, Noise and Imitation Implies Balanced Growth,” Working Paper 699, Federal Reserve Bank of Minneapolis.
- (2012b) “Technology Diffusion and Growth,” *Journal of Economic Theory*, Vol. 147, No. 2, pp. 602–622.
- Montanari, A. and A. Saberi (2010) “The Spread of Innovations in Social Networks,” *Proceedings of the National Academy of Sciences*, Vol. 107, No. 47, p. 20196–20201.
- Nix, E. (2016) “Learning Spillovers in the Firm,” Manuscript.
- Oberfield, E. (2018) “A Theory of Input-Output Architecture,” *Econometrica*, Vol. 86, No. 2, pp. 559–589.
- Perla, J. and C. Tonetti (2014) “Equilibrium Imitation and Growth,” *Journal of Political Economy*, Vol. 122, No. 1, pp. 52–76.
- Perla, J., C. Tonetti, and M. E. Waugh (2015) “Equilibrium Technology Diffusion, Trade, and Growth,” Working Paper 20881, National Bureau of Economic Research.
- Price, D. J. (1965) “Networks of Scientific Papers,” *Science*, Vol. 149, No. 3683, pp. 510–515.
- Sampson, T. (2016) “Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth,” *Quarterly Journal of Economics*, Vol. 131, No. 1, pp. 315–380.
- Syverson, C. (2004) “Product Substitutability and Productivity Dispersion,” *Review of Economics and Statistics*, Vol. 86, No. 2, pp. 534–550.
- Travers, J. and S. Milgram (1969) “An Experimental Study of the Small World Problem?” *Sociometry*, Vol. 32, No. 4, pp. 425–443.
- Watts, D. J. and S. H. Strogatz (1998) “Collective Dynamics of ‘Small-world’ Networks,” *Nature*, Vol. 393, pp. 440–442.
- Young, H. P. (2003) “The Diffusion of Innovations in Social Networks,” in *The Economy as an Evolving complex System, III* eds. by L. E. Blume and S. N. Durlauf, p. 267–282: Oxford: Oxford University Press.

Appendix A Derivation of the Bellman Equations

This appendix derives the value functions $V_n^i(t, Z)$ of continuing, i.e. not learning. To do so, approximate the continuous time limit by taking time steps of Δ and let Δ go to zero. Agents value their stream of current income plus the discounted expected future value. Agents discount at rate $\Delta\rho$. For agents in the low innovation state l , the value function looks as follows

$$V_n^l(t, Z) = \Delta Z + (1 - \Delta\rho) \left[(1 - \Delta\lambda_n^l) V_n^l(t + \Delta, Z) + \Delta\lambda_n^l V_n^h(t + \Delta, Z) \right],$$

where the expected future value takes into account transitioning from the low to the high innovation state occurring at the rate $\Delta\lambda_n^l$ during time interval Δ . Subtracting $(1 - \Delta\rho)V_n^l(t, Z)$ from both sides and dividing by Δ subsequently gives

$$\rho V_n^l(t, Z) = Z + (1 - \Delta\rho) \left(\lambda_n^l [V_n^h(t + \Delta, Z) - V_n^l(t + \Delta, Z)] + \frac{V_n^l(t + \Delta, Z) - V_n^l(t, Z)}{\Delta} \right).$$

Taking the limit of $\Delta \rightarrow 0$ gives the Bellman equation:

$$\rho V_n^l(t, Z) = Z + \lambda_n^l [V_n^h(t, Z) - V_n^l(t, Z)] + \partial_t V_n^l(t, Z).$$

Doing similarly for agents in the high innovation state h , the value function $V_n^h(t, Z)$ is

$$V_n^h(t, Z) = \Delta Z + (1 - \Delta\rho) \left[(1 - \Delta\lambda_n^h) V_n^h(t + \Delta, Z + \Delta\gamma_n Z) + \Delta\lambda_n^h V_n^l(t + \Delta, Z) \right].$$

This takes into account that agents that continue to be in the high innovation state during the time period Δ experience improvements in their productivity at the rate γ_n . Again, subtracting $(1 - \Delta\rho)V_n^h(t, Z)$ from both sides and dividing by Δ gives

$$\rho V_n^h(t, Z) = Z + (1 - \Delta\rho) \left[\frac{V_n^h(t + \Delta, Z + \Delta\gamma_n Z) - V_n^h(t, Z)}{\Delta} + \lambda_n^h (V_n^l(t + \Delta, Z) - V_n^h(t + \Delta, Z + \Delta\gamma_n Z)) \right].$$

Adding and subtracting $(1 - \Delta\rho) \frac{V_n^h(t, Z)}{\Delta}$ on the right-hand side gives

$$\rho V_n^h(t, Z) = Z + (1 - \Delta\rho) \left[\frac{V_n^h(t + \Delta, Z) - V_n^h(t, Z)}{\Delta} + \frac{V_n^h(t + \Delta, Z + \Delta\gamma_n Z) - V_n^h(t + \Delta, Z)}{\Delta} + \lambda_n^h (V_n^l(t + \Delta, Z) - V_n^h(t + \Delta, Z + \Delta\gamma_n Z)) \right].$$

Taking the limit of $\Delta \rightarrow 0$ gives the Bellman equation for type h :

$$\rho V_n^h(t, Z) = Z + \lambda_n^h [V_n^l(t, Z) - V_n^h(t, Z)] + \partial_t V_n^h(t, Z) + \gamma^n Z \partial_Z V_n^h(t, Z).$$

Appendix B Derivation of the Kolmogorov Forward Equations

This appendix derives the Kolmogorov forward equations. Similarly as for the Bellman equations, approximate the continuous time limit by taking time steps of Δ and let Δ go to zero. The evolution of the distribution over time looks as follows for the low innovation state l . Denote the CDF by $\Phi(\cdot)$ and the pdf by $\phi(\cdot)$.

$$\begin{aligned} \Phi_n^l(t + \Delta, Z) &= P(\text{productivity below } Z \text{ at } t \text{ and no adoption higher than } Z \text{ plus} \\ &\quad \text{remained in } l \text{ or transitioned from } h \text{ to } l \text{ during time period } \Delta) \\ &= \int_{M_n(t)}^Z \phi_n^l(t, z) (1 - \Delta \lambda_l^n) + \phi_n^h(t, z) \Delta \lambda_n^h - \Delta S_n^l(t) \left[1 - \hat{\phi}_n^l(t + \Delta, z) \right] + \\ &\quad \Delta S_n^h(t) \hat{\phi}_n^l(t + \Delta, z) dz \\ &= \Phi_n^l(t, Z) (1 - \Delta \lambda_l^n) + \Phi_n^h(t, Z) \Delta \lambda_n^h - \Delta S_n^l(t) + \Delta \hat{\Phi}_n^l(t + \Delta, Z) \left[S_n^l(t) + S_n^h(t) \right]. \end{aligned}$$

Subtracting $\Phi_n^l(t, Z)$ and dividing by Δ gives

$$\begin{aligned} \frac{\Phi_n^l(t + \Delta, Z) - \Phi_n^l(t, Z)}{\Delta} &= -\lambda_l^n \Phi_n^l(t, Z) + \lambda_n^h \Phi_n^h(t, Z) - S_n^l(t) + \hat{\Phi}_n^l(t + \Delta, Z) \left[S_n^l(t) + S_n^h(t) \right], \\ \partial_t \Phi_n^l(t, Z) &= -\lambda_l^n \Phi_n^l(t, Z) + \lambda_n^h \Phi_n^h(t, Z) - S_n^l(t) + \hat{\Phi}_n^l(t, Z) \left[S_n^l(t) + S_n^h(t) \right]. \end{aligned}$$

where I took the limit of $\Delta \rightarrow 0$ to obtain the last expression. For the high innovation state h , the evolution of the distribution looks as follows.

$$\begin{aligned} \Phi_n^h(t + \Delta, Z) &= P(\text{productivity below } \frac{Z}{1 + \Delta \gamma} \text{ at } t \text{ and no adoption higher than } Z \text{ plus} \\ &\quad \text{remained in } h \text{ or transitioned from } l \text{ to } h \text{ during time period } \Delta) \\ &= \int_{M_n(t)}^Z \phi_n^h(t, \frac{z}{1 + \Delta \gamma}) (1 - \Delta \lambda_n^h) + \phi_n^l(t, z) \Delta \lambda_n^l - \\ &\quad \Delta S_n^h(t) \left[1 - \hat{\phi}_n^h(t + \Delta, z) \right] + \Delta S_n^l(t) \hat{\phi}_n^h(t + \Delta, z) dz \\ &= \Phi_n^h(t, \frac{Z}{1 + \Delta \gamma}) (1 - \Delta \lambda_n^h) + \Phi_n^l(t, Z) \Delta \lambda_n^l - \Delta S_n^h(t) + \\ &\quad \Delta \hat{\Phi}_n^h(t + \Delta, Z) \left[S_n^l(t) + S_n^h(t) \right], \\ \frac{\Phi_n^h(t + \Delta, Z) - \Phi_n^h(t, Z)}{\Delta} &= \frac{\Phi_n^h(t, \frac{Z}{1 + \Delta \gamma}) - \Phi_n^h(t, Z)}{\Delta} - \Phi_n^h(t, \frac{Z}{1 + \Delta \gamma}) \lambda_n^h + \Phi_n^l(t, Z) \lambda_n^l - S_n^h(t) + \\ &\quad \hat{\Phi}_n^h(t + \Delta, Z) \left[S_n^l(t) + S_n^h(t) \right], \end{aligned}$$

$$\partial_t \Phi_n^h(t, Z) = -\gamma_n Z \partial_Z \Phi_n^h(t, Z) - \lambda_n^h \Phi_n^h(t, Z) + \lambda_n^l \Phi_n^l(t, Z) - S_n^h(t) + \hat{\Phi}_n^h(t, Z) \left[S_n^l(t) + S_n^h(t) \right],$$

where again the last expression follows from taking the limit of $\Delta \rightarrow 0$.

Appendix C Proof of Proposition 1

The following lemma will be needed for the proof of proposition 1.

Lemma 3. *Suppose that B is an $n \times n$ diagonalizable matrix of which all entries are positive and v is an $n \times 1$ vector with all entries positive. Then $e^{-Bt}v \geq 0 \ \forall t \geq 0$ if and only if v is an eigenvector of B .*

Proof. Since B is diagonalizable, we have $B = PDP^{-1}$ where D is a diagonal matrix containing the eigenvalues ordered from high to low and each column of P contains the associated eigenvectors. By the properties of the matrix exponential the following holds.

$$\begin{aligned} e^{-Bt}v &= e^{-PDP^{-1}t}v \\ &= Pe^{-Dt}P^{-1}v. \end{aligned}$$

Because B is a positive matrix, the Perron Frobenius theorem applies and hence, there is a unique largest eigenvalue (denoted by D_1) and the associated eigenvector has positive entries, call this eigenvector P_1 . All other eigenvectors have at least one negative entry.

If-statement: Suppose v is an eigenvector of B . Because v is assumed to have all entries positive we have $v = P_1$, then

$$\begin{aligned} Pe^{-Dt}P^{-1}v &= Pe^{-Dt}P^{-1}P_1 \\ &= Pe^{-Dt} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= e^{-D_1t}P_1 > 0. \end{aligned}$$

This proves that v being an eigenvector is a sufficient condition.

Only-if-statement: I prove this by contradiction, suppose that v is not an an

eigenvector. So $v \neq P_1$, then

$$\begin{aligned} P e^{-Dt} P^{-1} v &= P e^{-Dt} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= x_1 e^{-D_1 t} P_1 + x_2 e^{-D_2 t} P_2 + \dots + x_n e^{-D_n t} P_n, \end{aligned}$$

where at least two of the x_i are non-zero (since v would otherwise equal an eigenvector). By Perron Frobenius there is at least one positive and one negative element in P_i , $i \neq 1$. Suppose that $D_1 > D_2 > \dots > D_n$. Then the largest i for which $x_i \neq 0$ will dominate and taking t large will give that the pdf equals $x_i e^{-D_i t} P_i$ which has at least one negative element and hence $e^{-Bt} v \not\geq 0 \quad \forall t$. Therefore, we need $v = P_1$.

Here I have assumed that all eigenvalues are unique but Perron Frobenius only guarantees that the largest eigenvalue is unique (therefore, the above proof is complete for $n = 2$ but not necessarily for $n > 2$). So what happens if two eigenvalues are equal? Suppose that $D_{n-1} = D_n$ and suppose that v is such that $x_i = 0 \quad \forall i < n - 1$. Then $e^{-Bt} v = (x_{n-1} P_{n-1} + x_n P_n) e^{-D_n t}$. Then x_n and x_{n-1} could potentially be chosen such that $e^{-Bt} v$ has all entries positive. However, the choice of x_n and x_{n-1} is not free since $x = P^{-1} v$ where v is positive. If the eigenvalue with multiplicity 2 has a unique eigenvector, then the solution becomes: $((x_{n-1} + x_n t) P_n + \iota) e^{-D_n t}$ where ι solves $(B - D_n I) \iota = P_n$ and due to the presence of t , the $x_n t P_n$ term will dominate for large t and hence at least one element is negative making $e^{-Bt} v \not\geq 0 \quad \forall t$. \square

Now I turn to the proof of proposition 1.

Proof. I start by studying the Kolmogorov forward equations to get expressions for the distribution functions. These CDFs will depend on some variables, namely S_n and α_n , for which I will solve using the value matching condition.

Using that $g = \gamma$, $S_n^h = 0$ and $S_n^l = S_n$, equation (11) becomes

$$F_n^h(z) = \frac{\lambda_n^l}{\lambda_n^h} F_n^l(z) = \hat{\lambda}_n F_n^l(z), \quad (27)$$

where $\hat{\lambda}_n = \frac{\lambda_n^l}{\lambda_n^h}$. Plugging this into equation (10) gives

$$0 = g \partial_z F_n^l(z) + S_n \hat{F}_n^l(z) - S_n. \quad (28)$$

It will be easier to work with vectors and matrices. Therefore, stack all distribution functions $F_n^l(z)$ in the column vector $F_l(z)$. Similarly for $F_n^h(z)$ and $F_n(z)$. Then equation (27) becomes:

$$F_h(z) = \hat{\Lambda} F_l(z),$$

where $\hat{\Lambda}$ is a diagonal matrix with $\hat{\lambda}_n$ forming the diagonal entries. And

$$F(z) = F_h(z) + F_l(z) = (\hat{\Lambda} + I)F_l(z) = \Lambda F_l(z), \quad (29)$$

where Λ is a diagonal matrix with $1 + \hat{\lambda}_n$ on the diagonal entries. And equation (28) becomes:

$$\mathbf{0} = gF_l'(z) + S(\hat{F}(z) - \mathbf{1}),$$

where S is a diagonal matrix with S_n on the diagonal entries, and $\mathbf{0}$ and $\mathbf{1}$ are vectors where all elements are zero and one, respectively. Note that this differential equation is only well defined for $\alpha_n \leq z \leq \bar{z}$ as it is the Kolmogorov forward equation on the interior. For $z \leq \alpha_n$, we have $F_n(z) = 0$ and for $z \geq \bar{z}$, we have $F_n(z) = 1$. Incorporating this gives the following linear ordinary differential equation:

$$F_l'(z) = \frac{1}{g} I_{\alpha_n \leq z \leq \bar{z}} S (\mathbf{1} - \hat{F}(z)), \quad (30)$$

where $I_{\alpha_n \leq z \leq \bar{z}}$ is a diagonal matrix such that the n -th diagonal entry is 1 if $\alpha_n \leq z \leq \bar{z}$ and 0 otherwise.

Plug the expression for the learning distribution (21) into equation (30) and use equation (29) to get

$$F_l'(z) = \frac{1}{g} S I_{\alpha_n \leq z \leq \bar{z}} X^{-1} (\mathbf{1} - \Lambda F_l(z)). \quad (31)$$

Now I will solve this differential equation interval by interval. Define the matrix I_k as a matrix where the first k diagonal entries are a 1 and all other elements are a 0 (to represent $I_{\alpha_k \leq z \leq \bar{z}}$). Furthermore, define $\alpha_{N+1} = \bar{z}$. Then, the solution of this linear ODE, for each interval $[\alpha_k, \alpha_{k+1})$ ³⁴, is

$$F_l(z) = e^{-\frac{1}{g} S X^{-1} I_k \Lambda z} C_k + \Lambda^{-1} \mathbf{1}, \quad \text{if } \alpha_k \leq z < \alpha_{k+1},$$

³⁴ n will denote different types, whereas k will denote different intervals. However, they are related as only the first $n = k$ types are active on the k -th interval.

where C_k are constant vectors such that the CDF is continuous at the α_n (i.e., there is no point mass anywhere) and the initial condition holds. $F_l(\alpha_1) = 0$ gives $C_1 = -\Lambda^{-1}\mathbf{1}$ and hence

$$F_l(z) = \left(I - e^{-\frac{1}{g}SX^{-1}I_1A\Lambda z} \right) \Lambda^{-1}\mathbf{1}, \quad \text{if } 0 \leq z < \alpha_2. \quad (32)$$

In general:

$$F_l(z) = \left(I - e^{-\frac{1}{g}SX^{-1}I_kA\Lambda(z-\alpha_k)} \prod_{j=k-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_jA\Lambda(\alpha_{j+1}-\alpha_j)} \right) \right) \Lambda^{-1}\mathbf{1} \quad \text{if } \alpha_k \leq z < \alpha_{k+1}. \quad (33)$$

The product is denoted as $\prod_{j=k-1}^1$ to indicate that the matrix multiplication goes from high to low, i.e., $\prod_{j=k-1}^1 B_j = B_{k-1}B_{k-2} \dots B_1$. And if $k = 1$, $\prod_{j=k-1}^1 B_j = I$.

Evaluating equation (33) at the maximum of support \bar{z} and using equation (29) gives

$$\begin{aligned} F(\bar{z}) &= \Lambda F_l(\bar{z}) \\ &= \mathbf{1} - \Lambda e^{-\frac{1}{g}SX^{-1}A\Lambda(\bar{z}-\alpha_N)} \prod_{j=N-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_jA\Lambda(\alpha_{j+1}-\alpha_j)} \right) \Lambda^{-1}\mathbf{1}, \end{aligned}$$

which needs to equal $\mathbf{1}$ by the definition of \bar{z} . In other words, \bar{z} needs to be such that $\prod_{j=N-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_jA\Lambda(\alpha_{j+1}-\alpha_j)} \right) \Lambda^{-1}\mathbf{1}$ is in the null space of $e^{-\frac{1}{g}SX^{-1}A\Lambda(\bar{z}-\alpha_N)}$. Note that the matrix exponential is always invertible (for finite \bar{z}) and therefore, by the invertible matrix theorem, the null space of a matrix exponential only contains the zero vector. However, we have that $\prod_{j=N-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_jA\Lambda(\alpha_{j+1}-\alpha_j)} \right) \Lambda^{-1}\mathbf{1}$ equals $\Lambda^{-1}\mathbf{1} - F_l(\alpha_N)$ which does not equal the zero vector since the last element of $F_l(\alpha_N)$ is zero by equation (12). Therefore, \bar{z} cannot be finite and the only possibility for $F(\bar{z}) = \mathbf{1}$ to hold is if the matrix exponential $e^{-\frac{1}{g}SX^{-1}A\Lambda(\bar{z}-\alpha_N)}$ becomes the zero matrix as $\bar{z} \rightarrow \infty$. Define $\hat{A}_k = \frac{1}{g}SX^{-1}I_kA\Lambda$. Momentarily it will be shown that $\prod_{j=N-1}^1 \left(e^{-\hat{A}_j(\alpha_{j+1}-\alpha_j)} \right) \Lambda^{-1}\mathbf{1}$ is the eigenvector of $SX^{-1}A\Lambda$ and therefore only the eigenvalue associated with the eigenvector with only positive values needs to be positive to guarantee that $F(\bar{z}) = \mathbf{1}$. Since \hat{A}_N is a positive matrix, the Perron-Frobenius theorem guarantees that there is a unique eigenvector with positive values and that the associated eigenvalue is always positive.³⁵

All of the above has only used equations (10)-(13) and (21). Now continue with

³⁵Define $v = \prod_{j=N-1}^1 \left(e^{-\frac{1}{g}SX^{-1}I_jA\Lambda(\alpha_{j+1}-\alpha_j)} \right) \Lambda^{-1}\mathbf{1}$ and for the moment, assume that v is the eigenvector of \hat{A}_N , so $\hat{A}_N v = \mu v$. Since $v = \Lambda^{-1}\mathbf{1} - F_l(\alpha_N)$, all entries of v are positive as \hat{A}_N is positive. By Perron-Frobenius there is only one eigenvector with all entries positive and the associated

equation (14). First take the derivative of equation (33) with respect to z to get the pdf on $\alpha_k \leq z < \alpha_{k+1}$,

$$f_l(z) = \frac{1}{g} e^{-\frac{1}{g} SX^{-1} I_k A \Lambda (z - \alpha_k)} SX^{-1} I_k A \Lambda \prod_{j=k-1}^1 \left(e^{-\frac{1}{g} SX^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}. \quad (34)$$

Evaluating equation (14) for type n gives

$$S_n = e_n SX^{-1} I_n A \Lambda \prod_{j=n-1}^1 \left(e^{-\frac{1}{g} SX^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}, \quad \text{for } 1 \leq n \leq N, \quad (35)$$

where e_n is a row vector with a 1 on the n -th position and zeroes otherwise. By the definition of X , we have that equation (35) always holds. To see this, move X^{-1} to the right of I_n (which is allowed because both are diagonal), then by the definition of X we have that everything on the right-hand side of I_n is a vector with a 1 on the n -th position, because X is such that $X^{-1} A (1 - F(\alpha_n))$ has a 1 for the n -th element. I_n times this vector will still be a vector with a 1 on the n -th position. Multiplying by S and taking the n -th element gives S_n . Hence, equation (14) holds.

The pdf in equation (34) is only well-defined if its values are non-negative for all z . However, note that the elements of $e^{-\hat{A}_k \Lambda (z - \alpha_k)}$ are not necessarily all positive and therefore $f_l(z)$ is not necessarily non-negative. Therefore, I will now derive the necessary and sufficient conditions such that $f_l(z)$ is non-negative.

Define $v = \prod_{j=N-1}^1 \left(e^{-\hat{A}_j (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}$. Since $v = \Lambda^{-1} \mathbf{1} - F_l(\alpha_N)$, all entries of v are positive. Then, by Lemma 3, f is non-negative on the last interval if and only if $SX^{-1} A \Lambda v$ is an eigenvector of $SX^{-1} A \Lambda$. Because v is positive, Perron-Frobenius gives that the associated eigenvalue is the largest eigenvalue of $SX^{-1} A \Lambda$, which I call μ . So we have:

$$SX^{-1} A \Lambda SX^{-1} A \Lambda v = \mu SX^{-1} A \Lambda v, \quad (36)$$

$$SX^{-1} A \Lambda \prod_{j=N-1}^1 \left(e^{-\hat{A}_j (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1} = \mu \prod_{j=N-1}^1 \left(e^{-\hat{A}_j (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}, \quad (37)$$

eigenvalue, μ , is positive. By the properties of the matrix exponential we have that

$$e^{-\hat{A}_N (\bar{z} - \alpha_N)} v = e^{-\mu (\bar{z} - \alpha_N)} v,$$

which goes to zero as $\bar{z} \rightarrow \infty$.

$$\begin{bmatrix} \sim \\ \vdots \\ \sim \\ S_N \end{bmatrix} = \mu \begin{bmatrix} \sim \\ \vdots \\ \sim \\ \frac{1}{1+\lambda_N} \end{bmatrix}, \quad (38)$$

where the vector on the left-hand side follows from the property that $\mathbf{1} - F(\alpha_N) = \Lambda \prod_{j=N-1}^1 \left(e^{-\frac{1}{g} S X^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}$ and, by the definition of X , (i.e. the last element in X is such that $X^{-1} A (\mathbf{1} - F(\alpha_N))$ has a 1 as its last element), and the right-hand side follows because $F_N(\alpha_N) = 0$.³⁶ Thus, the condition that is required to have $f_l(z)$ positive for $z \geq \alpha_N$ implies that

$$S_N = \frac{\mu}{1 + \hat{\lambda}_N}. \quad (39)$$

An expression for μ will be given later.

Furthermore, the pdf for the final interval becomes:

$$f_l(z) = \frac{\mu}{g} e^{-\frac{\mu}{g}(z - \alpha_N)} v, \quad \text{if } z \geq \alpha_N,$$

where e is no longer the matrix exponential but the regular (scalar) exponential which is, indeed, positive everywhere.

Now that I have established that the pdf is positive on the final interval, I still have to show that it is positive on all other intervals. First, note by taking the derivative of equation (34) with respect to z that the slope of the pdf will be negative whenever the pdf is positive. Furthermore, note that the pdf is continuous at α_k .³⁷ Imagine that you are walking from \bar{z} towards α_1 . We know that for the last interval, the pdf is positive and hence the derivative will be negative. Once you are in a neighborhood around α_N , the pdf is positive and the derivative is negative, so the further you walk to the left, the larger becomes the pdf and hence the pdf remains positive. This continues until you reach α_1 . Hence, the pdf will be

³⁶Furthermore, note that $\mathbf{1} - F(\alpha_N)$ is the eigenvector of $\Lambda S X^{-1} A$, so the measure of agents with productivity in a sector higher than α_N is related to eigenvector centrality. Types with a higher eigenvector centrality will have more highly productive agents.

³⁷ $\lim_{z \uparrow \alpha_{k+1}} f_l(z) = f_l(\alpha_{k+1})$ (for the first k elements) because

$$\begin{aligned} \lim_{z \uparrow \alpha_{k+1}} f_l(z) &= \frac{1}{g} I_k S X^{-1} A \Lambda e^{-\frac{1}{g} S X^{-1} I_k A \Lambda (\alpha_{k+1} - \alpha_k)} \prod_{j=k-1}^1 \left(e^{-\frac{1}{g} S X^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1} \\ &= \frac{1}{g} I_k S X^{-1} A \Lambda \prod_{j=k}^1 \left(e^{-\frac{1}{g} S X^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}, \end{aligned}$$

while $f_l(\alpha_{k+1}) = \frac{1}{g} I_{k+1} S X^{-1} A \Lambda \prod_{j=k}^1 \left(e^{-\frac{1}{g} S X^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}$ is identical for the first k elements.

positive everywhere.

To finalize the expression for the distribution function, I need to solve for α_n and S_n . Those will be pinned down by the eigenvalue condition and the value matching condition. First get expressions for the value functions. Solve equation (17) for $v_n^h(z)$:

$$v_n^h(z) = \frac{e^z + \lambda_n^h v_n^l(z)}{\rho - g + \lambda_n^h}.$$

Plugging this into equation (16) and solving for $v_n^l(z)$ gives:

$$(\rho - g)v_n^l(z) = e^z - \frac{g}{\bar{\lambda}_n} \partial_z v_n^l(z),$$

where $\bar{\lambda}_n = \frac{\lambda_n^l}{\rho - g + \lambda_n^h} + 1$. The value function of one type does not depend on the value function of the other types and therefore this is a 1-dimensional ordinary differential equation. Solving it subject to the smooth pasting condition (18) gives:

$$v_n^l(z) = \frac{\bar{\lambda}_n}{g + (\rho - g)\bar{\lambda}_n} e^z + \frac{1}{(\rho - g)(\nu_n + 1)} e^{-z\nu_n + \alpha_n(\nu_n + 1)}, \quad (40)$$

where $\nu_n = \frac{(\rho - g)\bar{\lambda}_n}{g}$. Evaluating at α_n gives:

$$v_n^l(\alpha_n) = \frac{e^{\alpha_n}}{\rho - g} = v_n^h(\alpha_n).$$

Plugging this into the value matching condition (20) gives for each n

$$\frac{e^{\alpha_n}}{\rho - g} = \int_{\alpha_n}^{\infty} \left(\frac{\bar{\lambda}_n}{g + (\rho - g)\bar{\lambda}_n} e^z + \frac{1}{(\rho - g)(\nu_n + 1)} e^{-z\nu_n + \alpha_n(\nu_n + 1)} \right) \hat{f}_n(z) dz - \xi_n e^{\alpha_n}. \quad (41)$$

Now investigate the value matching condition for type N . For this purpose, we only need to consider the learning distribution on the last interval and hence we can leverage the result that due to the eigenvector, the matrix exponential disappears for this interval. The learning distribution for type- N becomes:

$$\hat{f}_N = e_N \frac{\mu}{g} X^{-1} A \Lambda e^{-\frac{\mu}{g}(z - \alpha_N)} v.$$

Hence, the value matching condition for type N becomes (recall $\alpha_N = 0$)

$$\xi_N + \frac{1}{\rho - g} = \frac{\mu}{g} \int_0^{\infty} \left(\frac{\bar{\lambda}_N}{g + (\rho - g)\bar{\lambda}_N} e^z + \frac{1}{(\rho - g)(\nu_N + 1)} e^{-z\nu_N} \right) e^{-\frac{\mu}{g}z} e_N X^{-1} A \Lambda v dz$$

$$= \frac{\mu}{g} \int_0^\infty \left(\frac{\bar{\lambda}_N}{g + (\rho - g)\bar{\lambda}_N} e^z + \frac{1}{(\rho - g)(\nu_N + 1)} e^{-z\nu_N} \right) e^{-\frac{\mu}{g}z} dz,$$

where the last equality holds because $e_N X^{-1} A \mathbf{1} v = 1$ by the definition of X (e_N is a row vector with its last element 1 and all others are 0). Solving the integral gives (note that $\frac{\mu}{g}$ needs to be larger than 1 for the integral to be well defined. See footnote 38 that this holds):

$$\xi_N + \frac{1}{(\rho - g)} = \frac{\mu}{g} \left[-\frac{\bar{\lambda}_N}{g + (\rho - g)\bar{\lambda}_N} \frac{1}{1 - \frac{\mu}{g}} + \frac{1}{(\rho - g)(\nu_N + 1)} \frac{1}{\nu_N + \frac{\mu}{g}} \right].$$

This equation is a quadratic equation in μ (the only unknown). To see that it is quadratic, rewrite it with $x = \frac{\mu}{g}$, $b = \frac{\bar{\lambda}_N}{g + (\rho - g)\bar{\lambda}_N}$ and $c = (\rho - g)(\nu_N + 1)$:

$$\xi_N + \frac{1}{\rho - g} = x \left[-b \frac{1}{1 - x} + \frac{1}{c} \frac{1}{\nu_N + x} \right], \quad (42)$$

$$\left(\xi_N + \frac{1}{\rho - g} \right) (1 - x) (\nu_N + x) = x \left[-b (\nu_N + x) + \frac{1}{c} (1 - x) \right]. \quad (43)$$

Collecting terms gives:

$$\left(\xi_N + \frac{1}{\rho - g} - b - \frac{1}{c} \right) x^2 + \left(\left(\xi_N + \frac{1}{\rho - g} \right) (\nu_N - 1) + \frac{1}{c} - b\nu_N \right) x - \left(\xi_N + \frac{1}{\rho - g} \right) \nu_N = 0.$$

This can be simplified by noting that $b + \frac{1}{c} = \frac{1}{\rho - g}$, $b - \frac{1}{c} \nu_N = 0$ and $\frac{1}{c} - b\nu_N = -\frac{\nu_N - 1}{\rho - g}$ (can be seen by using the definition of ν_N):

$$\xi_N x^2 + \xi_N (\nu_N - 1) x - \left(\xi_N + \frac{1}{\rho - g} \right) \nu_N = 0. \quad (44)$$

Solving gives:

$$\mu_{1,2} = g \frac{1 - \nu_N \pm \sqrt{(1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho - g}}{\xi_N} \nu_N}}{2}.$$

First of all note that the term in the square root is positive and therefore $\mu_{1,2}$ is real. Second, note that one of the solutions is positive and one is negative (because $4 \frac{\xi_N + \frac{1}{\rho - g}}{\xi_N} \nu_N > 0$). By Perron-Frobenius, the eigenvalue associated with the eigenvector with only positive entries is positive and therefore, the solution to

μ is unique:³⁸

$$\mu = g \frac{1 - \nu_N + \sqrt{(1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho-g}}{\xi_N} \nu_N}}{2}.$$

And we get:

$$S_N = g \frac{1 - \nu_N + \sqrt{(1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho-g}}{\xi_N} \nu_N}}{2(1 + \hat{\lambda}_N)}.$$

For the other types, it is not possible to get closed form expressions for S_n because the integral cannot be solved analytically due to the presence of the matrix exponentials.

The value matching condition (41) can also be written in matrix notation:

$$\begin{aligned} \text{diag}(e^{\alpha_n}) \left(\frac{1}{\rho - g} I + \text{diag}(\xi_n) \right) \mathbf{1} = \\ \frac{1}{g} \sum_{k=1}^N \int_{\alpha_k}^{\alpha_{k+1}} \left(\text{diag} \left(\frac{\bar{\lambda}_n}{g + (\rho - g)\bar{\lambda}_n} e^z \right) + \text{diag} \left(\frac{1}{(\rho - g)(\nu_n + 1)} e^{-z\nu_n + \alpha_n(\nu_n + 1)} \right) \right) \\ I_k X^{-1} A \Lambda e^{-\frac{1}{g} S X^{-1} I_k A \Lambda (z - \alpha_k)} S X^{-1} I_k A \Lambda \prod_{j=k-1}^1 \left(e^{-\frac{1}{g} S X^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1} dz. \end{aligned} \quad (45)$$

The first part of the integral can be explicitly solved for, using $\hat{A}_k = \frac{1}{g} S X^{-1} I_k A \Lambda$:

$$\begin{aligned} \frac{1}{g} \sum_{k=1}^N \int_{\alpha_k}^{\alpha_{k+1}} \text{diag} \left(\frac{\bar{\lambda}_n}{g + (\rho - g)\bar{\lambda}_n} e^z \right) I_k X^{-1} A \Lambda e^{-\frac{1}{g} S X^{-1} I_k A \Lambda (z - \alpha_k)} S X^{-1} I_k A \Lambda \\ \prod_{j=k-1}^1 \left(e^{-\frac{1}{g} S X^{-1} I_j A \Lambda (\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1} dz \end{aligned}$$

³⁸Furthermore note that $\mu > g$ (to ensure that the integral is well defined) is always fulfilled:

$$\begin{aligned} \frac{\mu}{g} > 1 &\Leftrightarrow \frac{1 - \nu_N + \sqrt{(1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho-g}}{\xi_N} \nu_N}}{2} > 1 \\ &\Leftrightarrow \sqrt{(1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho-g}}{\xi_N} \nu_N} > 1 + \nu_N \\ &\Leftrightarrow (1 - \nu_N)^2 + 4 \frac{\xi_N + \frac{1}{\rho-g}}{\xi_N} \nu_N > (1 + \nu_N)^2 \\ &\Leftrightarrow 4 \frac{1}{(\rho - g)\xi_N} \nu_N > 0. \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k=1}^N \text{diag} \left(\frac{\bar{\lambda}_n}{g + (\rho - g)\bar{\lambda}_n} \right) I_k X^{-1} A \Lambda (\hat{A}_k - I)^{-1} \left(e^{I\alpha_k} - e^{-\hat{A}_k(\alpha_{k+1} - \alpha_k) + I\alpha_{k+1}} \right) \\
 &\quad \hat{A}_k \prod_{j=k-1}^1 \left(e^{-\hat{A}_j(\alpha_{j+1} - \alpha_j)} \right) \Lambda^{-1} \mathbf{1}.
 \end{aligned}$$

The second part of the integral cannot be solved explicitly and has to be solved numerically. The unknowns in equation (45) are S and α_n . This means there are $2N - 2$ unknowns (because we know that $\alpha_N = 0$ and S_N). There are $N - 1$ remaining equations in equation (45) (because we have used the value matching condition for the last type to pin down μ). The remaining $N - 1$ conditions come from the eigenvector condition (where up until now I only used the last equation to pin down S_N). Thus, α_n and S_n are such that equations (37) and (45) hold. Knowing α_n and S_n , equation (33) gives the CDF. \square

Appendix D The Learning Cost and Innovation Intensity

This appendix discusses the effect of the learning cost and the transition intensities between the low and high innovation states on the productivity distribution.

Effect of the Learning Cost

Figure 8 shows how the distribution depends on the learning cost when both types are the same except for their learning cost.³⁹ In panel (a) both have the

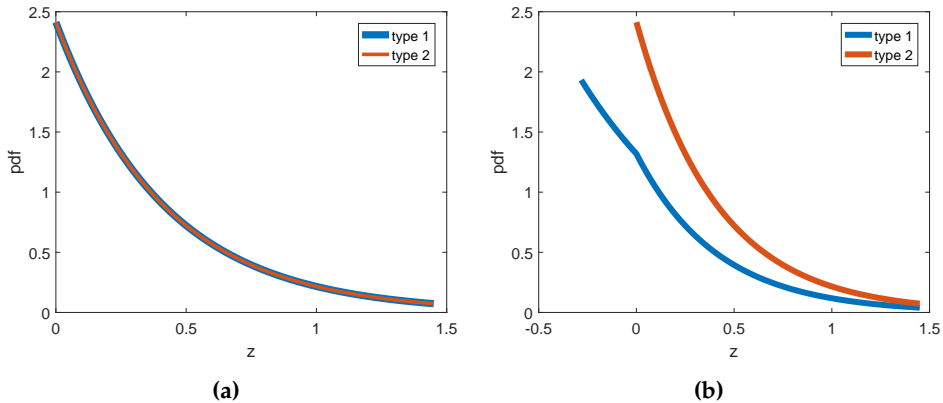


Figure 8: Pdf of normalized productivity z when (a) both types have a learning cost of 15 and (b) type I has a learning cost of 20 while type II has a learning cost of 15.

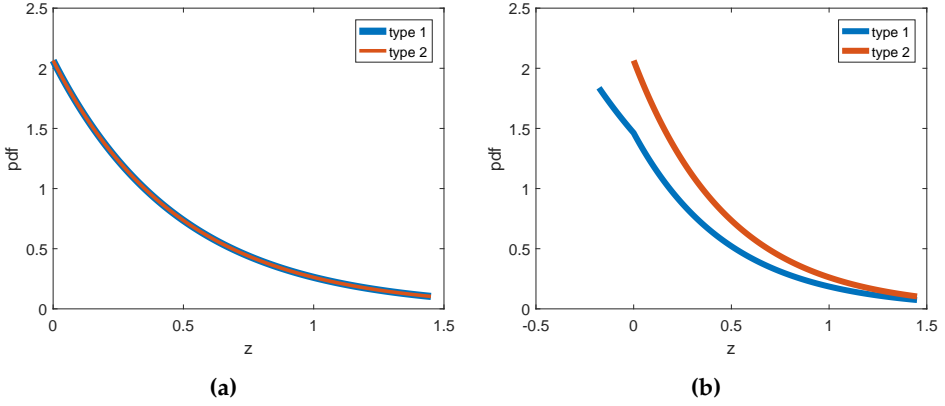


Figure 9: Pdf of normalized productivity z when (a) the intensity to go from the low to the high innovation state is 1 for both types and (b) when it is 0.5 for type I and 1 for type II.

same learning cost and hence the distributions are identical. In panel (b) the learning cost increases from 15 to 20 for type I while staying at 15 for type II. The increase in learning costs for type I makes that the gross learning value for type I has to increase in equilibrium to ensure that learners are indifferent between learning (and paying the learning costs) and continuing to produce at their current productivity level. This is achieved by type-I agents falling back relative to type-II agents. Furthermore, the increase in learning costs lowers the mass of agents learning in equilibrium.

Effect of the Innovation Intensity

Figure 9 shows the density for different parameter values governing the intensity at which type-I agents transition from the low innovation to the high innovation state. In panel (a), for both types this intensity is 1 and in panel (b), this intensity changes to 0.5 for type-I. For both graphs and both types, the intensity at which agents transition from the high to the low innovation state is 2, and the network consists for 70% of agents of their own type. The learning costs are 25 for both types. This implies that in panel (b) there are fewer type-I agents in the high innovation state than in panel (a). Due to this exogenous innovation process, average productivity, ignoring learning for the moment, will be lower for type-I agents. This means that the value of learning goes down for type-I agents (nothing changes for type-II agents because the shape parameter of the Pareto tail is not affected) and hence the minimum of support for type I goes down relative to the

³⁹ $\lambda^h = 2$, $\lambda^l = 0.6$ and, for both types, the network consists of 70% of agents of their own type. Furthermore, $\gamma = 0.02$ and $\rho = 0.03$.

minimum of support for type II. Also the mass of learners goes down.

Appendix E Additional Figures

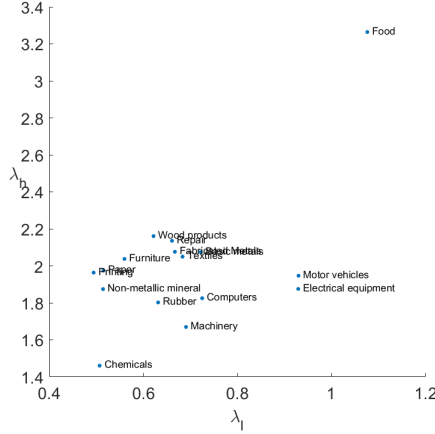


Figure 10: Estimated transition intensities between low and high innovation states by industry (Source: Amadeus). See Section III for details.

Appendix F Numerical Algorithm

α_n and S_n need to be solved numerically such that equations (25) and (26) hold, and once these variables are known, the distribution can be calculated directly using equation (22). This section describes how the numerical algorithm works (superscripts and subscripts denote the iteration steps).

Step 0. Make the initial guess for the vector α^1 and matrix S_0^1 . Then, for $t = 1, 2, \dots$, do the following.

Step 1. Update S_j^t for $j = 1, 2, \dots, J$ as follows.

Step 1a. Calculate the entries of X using the guesses α^t and S_{j-1}^t .

Step 1b. Calculate S_j^t using equation (25), iterate until convergence.

Step 2. Using equation (26), update α^{t+1} using the S and X found under step 1. If α^{t+1} is close to α^t stop, otherwise return to step 1 with $S_0^{t+1} = S_j^t$.

To calculate the entries of X under step 1a, note that $X_1 = 1$ by definition (because there are no agents with a productivity below α_1). Then X_2 will be

calculated as follows:

$$X_2 = a_{2-} \left(e^{-\frac{1}{g} S X^{-1} I_1 \Lambda A (\alpha_2 - \alpha_1)} \right) \mathbf{1}. \quad (46)$$

Due to the presence of I_1 , the only element of X that shows up in this equation is the first element of X which we already know to be 1. Hence, given the guess for S and α , X_2 can be directly calculated. Likewise, the expression for X_3 only depends on X_1 and X_2 . Hence, given X_2 , X_3 directly follows. This way, the entire matrix X can be calculated in $N - 1$ steps.

To update S under step 1b equation (25) is used. This is done as follows. Calculate $F(0)$ using the S from the previous iteration. Then $X^{-1} \Lambda A (\mathbf{1} - F(0))$ and $\mu (\mathbf{1} - F(0))$ are two vectors based on the guessed α and S . Because S is a diagonal matrix, the entries of the updated S are found by the elementwise division of these two vectors.

In step 2 α is updated using equation (26). First calculate the integral for each type using the α resulting from the previous iteration. Then update α element by element by dividing this integral by $\xi_n + \frac{1}{\rho - g}$ and taking the logarithm subsequently. In both step 1 and step 2, homotopy is used to update the guess for S and α .

Appendix G Modelling Choices

My model is based on the model proposed by Benhabib et al. (2017). The main differences of this model compared to the models in Lucas and Moll (2014) and Perla and Tonetti (2014) is that there is innovation besides learning and that the initial distribution does not need to be unbounded in order to make long-run growth possible. As I show here, incorporating a network into the models of Lucas and Moll (2014) and Perla and Tonetti (2014) would not lead to any interesting dynamics.

In Lucas and Moll (2014) and Perla and Tonetti (2014), the only dimension in which agents differ is their productivity level. The thickness of the Pareto tail of the productivity distribution determines the growth rate. Suppose that there would be a network present in Lucas and Moll (2014) or Perla and Tonetti (2014) similar to the main text here with each node reflecting a type containing a continuum of agents. If each type were the same (i.e., having the same tail parameter) then it does not matter for the learning decision to which types a certain node is connected since the learning distribution is the same. To make this clear, suppose that there are two types of agents and the distribution of productivity across agents is the

same for both types. Then, whether a learning agent has a 50% probability of meeting an agent from each type or a 100% probability of meeting one type and 0% for the other type does not matter because the distribution from which the learning agent is going to draw will be the same for both networks. Therefore, types need to differ ex ante for the network to be relevant.

Lucas and Moll (2014) assume that the initial support of the productivity distribution is unbounded and follows a Pareto tail. The natural way to extend this model to multiple types would be to assume that the initial distribution for each type has a Pareto tail with the thickness of the tail differing across types. It turns out that this extension admits no interesting dynamics in equilibrium. To see this, suppose for simplicity that each type is connected to each other type, potentially with the weights differing across types. Then, the tail of the learning distribution, for each type, will be a Pareto tail of which the thickness equals the maximum thickness of the original distributions. Hence, no matter what the network is, the tail of the distribution of each type will converge to a Pareto tail with the same thickness, namely equal to the thickness of the tail of the distribution of the type with the thickest tail. Since in both Lucas and Moll (2014) and Perla and Tonetti (2014) the Pareto tail is what affects the decision to learn, the network has no interesting role.

Appendix H Model the Innovation Process as a GBM

In this paper, I model innovation as following a two-state Markov Process. Here I discuss another possible stochastic process, namely a geometric Brownian motion. The technical appendix of Benhabib et al. (2017) contains a version of their model where the Markov innovation state is replaced by a geometric Brownian motion. Here, I solve the model with a geometric Brownian motion in case there are two types and I discuss why I chose the Markov Process as the model in the main text. Without loss of generality, suppose that the second type has the largest reservation productivity. The productivity of a single agent evolves according to a geometric Brownian motion: $dZ = \gamma_n Z + \sigma_n Z dW$ where dW is a standard Brownian motion. The drift γ_n and the variance σ_n are allowed to vary by type. The normalized Kolmogorov forward equation for type n in steady state is as follows

$$0 = (g - \gamma_n)F'_n(z) + \frac{\sigma_n^2}{2}F''_n(z) + S_n\hat{F}_n(z) - S_n, \quad (47)$$

where g is the economy-wide growth rate. The difference with respect to the main text is that there are no longer two innovation states and that the second derivative

shows up in the KFE. The expression for the learning distribution is the same as in the main text (using matrix-vector notation):

$$\hat{F}(z) = I_{z > \alpha_n} (\mathbf{1} + X^{-1} (AF(z) - \mathbf{1})) . \quad (48)$$

Start by considering the first interval $[\alpha_1, 0)$ at which only agents of the first type are active. Hence, the KFE on this interval is a single second-order differential equation. Solving this subject to the initial condition $F_1(\alpha_1) = 0$ gives

$$F_1(z) = \frac{1}{a_{11}} - \frac{1}{a_{11}} e^{\overbrace{\left(\sqrt{\frac{(g - \gamma_1)^2}{\sigma_1^4}} - 2S_1 \frac{a_{11}}{\sigma_1^2} - \frac{g - \gamma_1}{\sigma_1^2} \right)}^{T_1} (z - \alpha_1)} , \quad \text{for } \alpha_1 \leq z < 0 , \quad (49)$$

where a_{11} is the probability that a learning type-I agent learns from another type-I agent. Evaluating equation (49) at $z = 0$ gives the initial condition needed to solve the differential equation for the second interval:

$$F_1(0) = \frac{1}{a_{11}} - \frac{1}{a_{11}} e^{-T_1 \alpha_1} . \quad (50)$$

The solution to the KFE (47) for $z \geq 0$ is

$$F(z) = \mathbf{1} - e^{Bz} C , \quad z \geq 0 , \quad (51)$$

for some matrix B and where C is a vector such that $F(0) = [F_1(0) \ 0]'$ holds. Hence, $C = [1 - F_1(0) \ 1]'$. Taking the derivative gives that the pdf is $-e^{Bz} BC$. Suppose that all entries of B are non-negative, then by Lemma 3, it is needed that C is an eigenvector of B for the density to be positive for all $z > 0$. Denote the associated eigenvalue by μ . Then, the calculation rules of the matrix exponential imply

$$F(z) = \mathbf{1} - e^{\mu z} C , \quad z \geq 0 . \quad (52)$$

Taking the first- and second-order derivatives and plugging these back into the KFE (47) gives

$$\left(\Gamma \mu + \frac{1}{2} \Sigma^2 \mu^2 + S X^{-1} A \right) C = \mathbf{0} , \quad (53)$$

where Γ is a diagonal matrix with $g - \gamma_n$ as its entries and Σ is a diagonal matrix with σ on the diagonal. Equation (53) is a system of two equations in three unknowns (μ , S_1 and S_2). Suppose that S_1 is known, then the first equation pins down μ and the second equation pins down S_2 . Using the definition of X , rewrite the second equation

$$S_2 = -(g - \gamma_2)\mu - \frac{1}{2}\sigma_2^2\mu^2, \quad (54)$$

which is the same expression as could have been obtained by evaluating the KFE at $z = 0$ for type 2. The first equation of (53) can be rewritten as

$$\left((g - \gamma_1)\mu + \frac{1}{2}\sigma_1^2\mu^2 \right) (1 - F_1(0)) + S_1(1 - F_1(0)a_{11}) = 0, \quad (55)$$

which gives S_1 as a function of μ and α_1 (α_1 is hidden in $F_1(0)$).⁴⁰

In case of the geometric Brownian motion, the normalized Bellman equation becomes

$$(\rho - g)v_n(z) = e^z + (\gamma_n - g)v'_n(z) + \frac{\sigma_n^2}{2}v''_n(z), \quad (56)$$

with the smooth pasting condition

$$v'_n(\alpha_n) = 0. \quad (57)$$

Solving gives

$$v_n(z) = \frac{1}{\rho - \gamma_2 - \frac{\sigma_2^2}{2}} \left(e^z + \frac{1}{\nu_n} e^{-\nu_n z + \alpha_n(\nu_n + 1)} \right), \quad (58)$$

with

$$\nu_n = \frac{\gamma_n - g}{\sigma_n^2} + \sqrt{\left(\frac{g - \gamma_n}{\sigma_n^2} \right)^2 + 2 \frac{r - g}{\sigma_n^2}}. \quad (59)$$

The value matching condition for type II is

$$v(0) = \frac{1 + \frac{1}{\nu_2}}{\rho - \gamma_2 - \frac{\sigma_2^2}{2}} = - \int_0^\infty v_n(z) \mu e^{\mu z} dz - \xi_2 \quad (60)$$

$$= \frac{-\mu(\nu_n + 1)(\nu_2 - \mu - 1)}{(-\mu - 1)\nu_2(\nu_2 - \mu)(r - \gamma_2 - \frac{\sigma_2^2}{2})} - \xi_2. \quad (61)$$

⁴⁰Note, that evaluating equation (47) at $z = \alpha_1$ is not informative about S_1 given the first and second order derivatives of equation (49).

Solving this for ν_n , equating to equation (59) and rewriting gives an expression for the growth rate in terms of μ

$$g = \frac{-((\mu+1)\mu\xi_2(2\gamma_2 + \sigma_2^2) + 2)((-\mu-1)\xi_2(-\mu\sigma_2^2 + 2\gamma_2) + 2) - 4(\mu+1)^2\xi_2^2\rho^2}{2(\mu+1)^2\xi_2((-\mu-1)\xi_2(-2\gamma_2 + 2\rho - \sigma_2^2) - 2)} - \frac{2(\mu+1)\xi_2\rho(-(\mu+1)\xi_2((\mu^2+1)\sigma_2^2 + 2(1-\mu)\gamma_2) + 4)}{2(\mu+1)^2\xi_2((-\mu-1)\xi_2(-2\gamma_2 + 2\rho - \sigma_2^2) - 2)}. \quad (62)$$

The value matching condition for type I becomes

$$\begin{aligned} \frac{1 + \frac{1}{\nu_1}}{\rho - \gamma_1 - \frac{\sigma_1^2}{2}} e^{\alpha_1} = & -\frac{1}{\rho - \gamma_1 - \frac{\sigma_1^2}{2}} \frac{T_1}{a_{11}} e^{-T_1\alpha_1} \left[\frac{1}{T_1 + 1} (1 - e^{(T_1+1)\alpha_1}) + \right. \\ & \left. \frac{1}{\nu_1(T_1 - \nu_1)} (e^{\alpha_1(\nu_1+1)} - e^{\alpha_1(T_1+1)}) \right] + \\ & \frac{1}{\rho - \gamma_1 - \frac{\sigma_1^2}{2}} \mu(1 - F_1(0)a_{11}) \left[\frac{1}{\mu + 1} + \frac{1}{\nu_1(\mu - \nu_1)} e^{\alpha_1(\nu_1+1)} \right] - \xi_1 e^{\alpha_1}. \end{aligned} \quad (63)$$

There are 5 unknowns: $\alpha_1, g, \mu, S_2, S_1$ and 4 equations: (54), (55), (62) and (63) meaning that there will be multiple equilibria in terms of the growth rate. A fatter tail leads to a higher growth rate. Note that given μ the growth rate does not depend on the network. Suppose that the economy is in an equilibrium for a certain μ . Then, equation (62) determines the growth rate and equation (54) gives S_2 where both do not depend on the network (this is for similar reasons as in the two-state Markov model). Equations (55) and (63) determine S_1 and α_1 and these equations do depend on the network. Now, suppose that the network changes. Since this will not affect the second type, μ and g will not be affected. This is the reason why I focus on the effect of the network on TFP and not on the growth rate in the main text.

Figure 11 shows the distribution of productivity as an illustration. The parameter values are as follows: $\gamma_1 = 0.005$, $\gamma_2 = 0.01$, $\sigma_1 = \sigma_2 = 0.1$, $\xi_1 = \xi_2 = 25$, $\rho = 0.04$ and $g = 0.02$. The network differs across the two panels of the figure. The distribution looks very similar as for the model in the main text (see Figure 1). The more connected type-I agents are to the type-II agents, the higher is the reservation productivity of type I and the flatter is the density of type I on $[\alpha_1, 0)$. Therefore, it seems plausible that the effect of the network in the GBM model is similar to the effect in the model of the main text. However, one difference between the two models is that the pdf is not continuous here whereas it is continuous in the main text. At $z = 0$ there is a jump in the pdf for type 1. It turns out that this jump depends on the growth rate of the economy. If the growth rate of the economy, g , were 90% instead of 2% then the discontinuity would (almost) be gone. Both

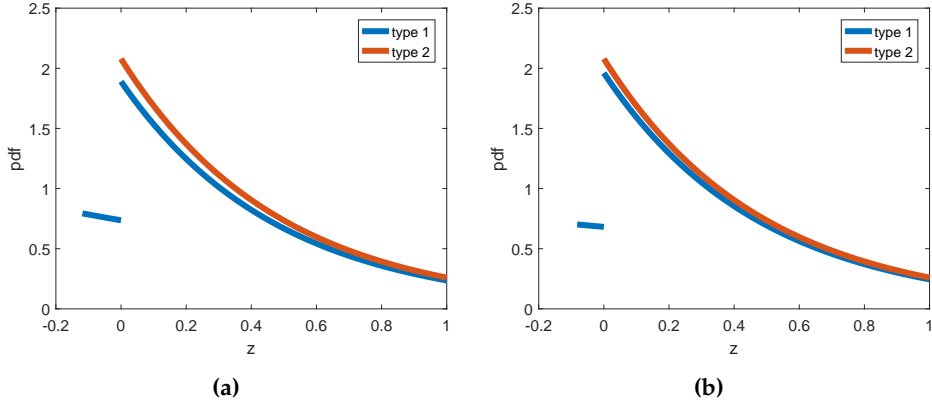


Figure 11: Pdf of normalized productivity z when innovation is modeled as a geometric Brownian motion when the probability that a type-I agent learns from a type-II agent is (a) 20% and (b) 50%.

a growth rate of 90% (while the largest drift is only 1%) and the discontinuities observed in Figure 11 seem implausible, providing a justification for using the two-state Markov chain as the main model.

Sammanfattning

Den här avhandlingen består av fyra kapitel med det gemensamma temat företagsdynamik. I de två första kapitlen dokumenterar jag att företag har blivit mer lönsamma över tiden och i det tredje kapitlet studerar jag vad som skapar denna lönsamhetsökning. I det fjärde kapitlet studerar jag hur innovationer sprider sig mellan företag.

Det första kapitlet, **Att beräkna kapitalkostnaden och vinstandelen (Estimating the Cost of Capital and the Profit Share)**, beräknar hur stora vinster företagen gör. Vinsterna är lika med produktion minus samtliga kostnader. Emellertid så observeras inte kapitalkostnaderna direkt då de inte rapporteras i företagens resultaträkning. Därför behöver vi först beräkna kapitalkostnaderna för att kunna beräkna de ekonomiska vinsterna. I det här kapitlet så utvecklar jag en ny metod för att kunna beräkna kapitalkostnaderna. Den här metoden använder det faktum att företags val av insatsvaror visar kapitalkostnaden när företagen minimerar kostnaderna och producerar i enlighet med en homogen produktionsfunktion. Genom att använda den här metoden så finner jag att kapitalkostnaderna som en del av produktionen har varit svagt avtagande i USA under de senaste 50 åren. Genom att dra av dessa beräknade kapitalkostnader, tillsammans med alla andra observerade kostnader, från produktionen finner jag att vinsterna ungefärligen har fördubblats från att vara ca 4% av produktionen på 1960-talet till att utgöra ca 8% av produktionen i dag.

Att veta hur stora vinster företagen gör är avgörande för att förstå ojämlikhet. Under de senaste årtiondena har arbetstagarnas andel av inkomsterna minskat i USA. Detta innebär att en lägre andel av inkomsterna tillfaller arbetstagarna och en större andel tillfaller de som äger företagen och aktiekapitalet. Denna minskning i arbetskraftens andel kan antingen bero på att kapitalet blivit viktigare i produktionen (exempelvis till följd av automatiseringen) eller till följd av en ökning av företagens marknadskraft. Att kapitalandelen har minskat medan vinstandelen har ökat innebär att minskningen i arbetskraftsandelen beror på en ökning i företagens marknadskraft och inte på att kapitalet blivit viktigare för produktionen.

När jag beaktar lönsamhetsfördelningen mellan företagen så finner jag att hela fördelningen har flyttats till höger. Sålunda är det inte enbart de mer lönsamma företagen som har blivit mer lönsamma, medianföretaget har också blivit mer

lönsamt över tiden. I vilket fall så beror vinstökningen på att den ekonomiska aktiviteten flyttas från företag med en låg vinstandel till företag med en stor vinstandel. Jag finner även att större företag har en högre vinstandel än små företag och att detta samband har blivit starkare över tiden.

Det andra kapitlet, **Vinster och marginalprodukten av kapital världen över (Profits and the Marginal Product of Capital around the World)**, använder den metod som utvecklats i det första kapitlet för att studera hur vinstandelen har utvecklats världen över. Jag finner att vinstandelen visar en inverterad U-kurva i Europa mellan 1990 och 2015, med en total ökning uppgående till cirka 2 procentenheter. Vinster har även ökat i Asien, Latinamerika och Nordamerika. Detta betyder inte att vinster i samtliga länder har ökat. I exempelvis Kanada har vinstandelen inte ökat. Den globala vinstandelen har ökat med ca 2 procentenheter från 1990 till 2015, vilket är något mindre än ökningen i USA. Slutligen har rikare länder upplevt en något snabbare lönsamhetsökning än fattiga länder.

Vidare så studerar det här kapitlet i vilken omfattning som marginalprodukterna av kapital är utjämnade länder emellan. Detta är viktigt för att förstå hur de internationella kapitalmarknaderna fungerar. När marginalprodukten av kapital skiljer sig mellan länder, fungerar inte de internationella kapitalmarknaderna väl, och den globala produktionen skulle kunna öka genom att omallokera kapital från länder med en låg marginalprodukt av kapital till länder med en hög marginalprodukt av kapital. Jag beräknar marginalprodukten av kapital mellan länder med hänsyn tagen till imperfekt konkurrens, och att stordriftsfördelarna kan avvika från ett. Jag finner att rikare länder har en högre marginalprodukt av kapital än fattigare länder men att detta enbart drivs av skillnader i avskrivningstakten. Sålunda verkar internationella kapitalmarknader fungera väl och det finns ingen vinst, i termer av produktionen efter avskrivningar, av att omallokera kapital från fattiga till rika länder och vice versa.

I det tredje kapitlet, **Vinsternas livscykel (The Life Cycle of Profits)**, dokumenterar jag att över tiden har vinster kommit att skjutas fram mer över ett företags livscykel. Ett företag som är yngre än tio år i dag gör lika stora vinster på årsbasis som ett ungt företag gjorde för trettio år sedan. Emellertid så gör ett gammalt företag i dag mycket större vinster än vad ett gammalt företag brukade göra för 30 år sedan. Det finns två skäl till detta ändrade livscykelmönster för vinster. Unga företag i dag är bara aningen större i termer av försäljning än vad de brukade vara, medan äldre företag har blivit mycket större jämfört med äldre företag för trettio år sedan. För det andra har unga företag börjat göra mindre vinster i förhållande till sin storlek medan gamla företag i dag har en vinstandel som är ungefär densamma som den brukade vara.

Jag bygger sedan en kvantitativ modell för att förstå i vilken utsträckning detta förändrade livscykelmönster för vinster förklarar vinstökningen. En företagare kommer att starta en företagsverksamhet när värdet av att ha denna företagsverksamhet (dvs den diskonterade summan av vinster) överstiger inträdeskostnaderna. Allt annat lika så uppstår vinsterna i dag i ett senare skede än vad de brukade göra, företagets värde är lägre till följd av diskontering. Detta gör det mindre tilltalande att starta en företagsverksamhet och kommer därför att leda till ett lägre företagsinträde. Detta minskar konkurrensen företag emellan och leder i sin tur till en vinstökning. Jag finner att den observerade skillnaden i vinsternas livscykelmönster förklarar ungefär två tredjedelar av de vinstökningar som jag fann i kapitel 1 och detta kan mer än helt förklara den minskning i företagsinträde som observeras.

Slutligen studerar det fjärde kapitlet, **Spridningen av idéer i nätverk och endogen sökning (Diffusion of Ideas in Networks and Endogenous Search)**, spridningen av teknologi. Nya idéer tenderar att spridas gradvis och aktörer som är direkt relaterade till de som tidigt antar tekniken är mer sannolika att själva anta tekniken. Detta innebär att hur nätverket av interaktioner mellan aktörer ser ut påverkar hur snabbt spridningen sker. Vidare så beror sökansträngningen på nätverket och produktivetsfördelningen. När man är knuten till högproduktiva aktörer så är man villig att anstränga sig mer eller mindre för att lära sig och anta de teknologier som dessa högproduktiva aktörer använder jämfört med när man enbart är knuten till lågproduktiva aktörer. Det här kapitlet studerar teoretiskt vilka nätverksegenskaper som är fördelaktiga för spridningen när beslutet att leta efter produktivitetshöjande teknologier beror på nätverket av interaktioner mellan aktörerna.

Aktörer har optionen att involvera sig i kostsamt lärande av sina förstagradskontakter. Ju mer produktiva en aktörs förbindelser är, desto villigare är denne att lära. Sålunda påverkar nätverket den reservationsproduktivitet vid viken aktörer väljer att lära sig och det påverkar därför den aggregerade produktiviteten. Jag finner att ju tätare nätverket är (dvs ju fler förbindelser som finns mellan företag), desto större är ansträngningen att lära sig och därav den högre totala faktorproduktiviteten och den lägre ojämlikheten. Emellertid så är nätverkets effekt på den andel aktörer som lär sig i jämvikt tvetydig. Vidare så finner jag att noder som är centrala i termer av deras närhet till andra noder tenderar att använda en större ansträngning vid inlärnin och ha en högre produktivitet.

IIES Monograph Series

1. Michaely, Michael *The Theory of Commercial Policy: Trade and Protection*, 1973
2. Söderström, Hans Tson *Studies in the Microdynamics of Production and Productivity Change*, 1974
3. Hamilton, Carl B. *Project Analysis in the Rural Sector with Special Reference to the Evaluation of Labour Cost*, 1974
4. Nyberg, Lars and Staffan Viotti *A Control Systems Approach to Macroeconomic Theory and Policy in the Open Economy*, 1975
5. Myhrman, Johan *Monetary Policy in Open Economies*, 1975
6. Krauss, Melvyn *International Trade and Economic Welfare*, 1975
7. Wihlborg, Clas *Capital Market Integration and Monetary Policy under Different Exchange Rate Regimes*, 1976
8. Svensson, Lars E.O. *On Competitive Markets and Intertemporal Resources Allocation*, 1976
9. Yeats, Alexander J. *Trade Barriers Facing Developing Countries*, 1978
10. Calmfors, Lars *Prices, Wages and Employment in the Open Economy*, 1978
11. Kornai, János *Economics of Shortage, Vols I and II*, 1979
12. Flam, Harry *Growth, Allocation and Trade in Sweden. An Empirical Application of the Heckscher-Ohlin Theory*, 1981
13. Persson, Torsten *Studies of Alternative Exchange Rate Systems. An Intertemporal General Equilibrium Approach*, 1982
14. Erzan, Refik *Turkey's Comparative Advantage, Production and Trade Patterns in Manufactures. An Application of the Factor Proportions Hypothesis with Some Qualifications*, 1983
15. Horn af Rantzien, Henrik *Imperfect Competition in Models of Wage Formation and International Trade*, 1983

16. Nandakumar, Parameswar *Macroeconomic Effects of Supply Side Policies and Disturbances in Open Economies*, 1985
17. Sellin, Peter *Asset Pricing and Portfolio Choice with International Investment Barriers*, 1990
18. Werner, Ingrid *International Capital Markets: Controls, Taxes and Resources Allocation*, 1990
19. Svedberg, Peter *Poverty and Undernutrition in Sub-Saharan Africa: Theory, Evidence, Policy*, 1991
20. Nordström, Håkan *Studies in Trade Policy and Economic Growth*, 1992
21. Hassler, John, Lundvik, Petter, Persson, Torsten and Söderlind, Paul *The Swedish Business Cycle: Stylized facts over 130 years*, 1992
22. Lundvik, Petter *Business Cycles and Growth*, 1992
23. Söderlind, Paul *Essays in Exchange Rates, Business Cycles and Growth*, 1993
24. Hassler, John A.A. *Effects of Variations in Risk on Demand and Measures of Business Cycle Comovements*, 1994
25. Daltung, Sonja *Risk, Efficiency, and Regulation of Banks*, 1994
26. Lindberg, Hans *Exchange Rates: Target Zones, Interventions and Regime Collapses*, 1994
27. Stennek, Johan *Essays on Information-Processing and Competition*, 1994
28. Jonsson, Gunnar *Institutions and Incentives in Monetary and Fiscal Policy*, 1995
29. Dahlquist, Magnus *Essays on the Term Structure of Interest Rates and Monetary Policy*, 1995
30. Svensson, Jakob *Political Economy and Macroeconomics: On Foreign Aid and Development*, 1996
31. Blix, Mårten *Rational Expectations and Regime Shifts in Macroeconometrics*, 1997
32. Lagerlöf, Nils-Petter *Intergenerational Transfers and Altruism*, 1997
33. Klein, Paul *Papers on the Macroeconomics of Fiscal Policy*, 1997
34. Jonsson, Magnus *Studies in Business Cycles*, 1997
35. Persson, Lars *Asset Ownership in Imperfectly Competitive Markets*, 1998

36. Persson, Joakim *Essays on Economic Growth*, 1998
37. Domeij, David *Essays on Optimal Taxation and Indeterminacy*, 1998
38. Flodén, Martin *Essays on Dynamic Macroeconomics*, 1999
39. Tangerås, Thomas *Essays in Economics and Politics: Regulation, Elections and International Conflict*, 2000
40. Lidbom, Per Pettersson *Elections, Party Politics and Economic Policy*, 2000
41. Vestin, David *Essays on Monetary Policy*, 2001
42. Olofsgård, Anders *Essays on Interregional and International Political Economics*, 2001
43. Johansson, Åsa *Essays on Macroeconomic Fluctuations and Nominal Wage Rigidity*, 2002
44. Groth, Charlotta *Topics on Monetary Policy*, 2002
45. Gancia, Gino A. *Essays on Growth, Trade and Inequality*, 2003
46. Harstad, Bård *Organizing Cooperation: Bargaining, Voting and Control*, 2003
47. Kohlscheen, Emanuel *Essays on Debts and Constitutions*, 2004
48. Olovsson, Conny *Essays on Dynamic Macroeconomics*, 2004
49. Stavlöt, Ulrika *Essays on Culture and Trade*, 2005
50. Herzing, Mathias *Essays on Uncertainty and Escape in Trade Agreements*, 2005
51. Bonfiglioli, Alessandra *Essays on Financial Markets and Macroeconomics*, 2005
52. Pienaar, Natalie *Economic Applications of Product Quality Regulation in WTO Trade Agreements*, 2005
53. Song, Zheng *Essays on Dynamic Political Economy*, 2005
54. Eisensee, Thomas *Essays on Public Finance: Retirement Behavior and Disaster Relief*, 2005
55. Favara, Giovanni *Credit and Finance in the Macroeconomy*, 2006
56. Björkman, Martina *Essays on Empirical Development Economics: Education, Health and Gender*, 2006
57. Larsson, Anna *Real Effects of Monetary Regimes*, 2007

58. Prado, Jr., Jose Mauricio *Essays on Public Macroeconomic Policy*, 2007
59. Tonin, Mirco *Essays on Labor Market Structures and Policies*, 2007
60. Queijo von Heideken, Virginia *Essays on Monetary Policy and Asset Markets*, 2007
61. Finocchiaro, Daria *Essays on Macroeconomics*, 2007
62. Waisman, Gisela *Essays on Discrimination and Corruption*, 2008
63. Holte, Martin Bech *Essays on Incentives and Leadership*, 2008
64. Damsgaard, Erika Färnstrand *Essays on Technological Choice and Spillovers*, 2008
65. Fredriksson, Anders *Bureaucracy, Informality and Taxation: Essays in Development Economics and Public Finance*, 2009
66. Folke, Olle, *Parties, Power and Patronage*, 2010
67. Drott, David Yanagizawa, *Information, Markets and Conflict*, 2010
68. Meyersson, Erik, *Religion, Politics, and Development*, 2010
69. Klingelhofer, Jan, *Models of Electoral Competition*, 2010
70. Perrotta, Maria Carmela, *Aid, Education and Development*, 2010
71. Caldara, Dario, *Essays on Empirical Macroeconomics*, 2011
72. Mueller, Andreas, *Business Cycles, Unemployment and Job Search*, 2011
73. von Below, David, *Essays in Climate and Labour Economics*, 2011
74. Gars, Johan, *Essays on the Microeconomics of Climate Change*, 2012
75. Spiro, Daniel, *Some Aspects of Resource and Behavioral Economics*, 2012
76. Ge, Jinfeng, *Essays on Macroeconomics and Political Economy*, 2012
77. Li, Yinan, *Institutions, Political Cycles and Corruption*, 2012
78. Håkanson, Christina, *Changes in Workplaces and Careers*, 2013
79. Qin, Bei, *Essays on Empirical Development and Political Economics*, 2013
80. Jia, Ruixue, *Essays on the Political Economics of China's Development*, 2013

81. Campa, Pamela, *Media Influence on Pollution and Gender Equality*, 2013
82. Seim, David, *Essays on Public, Political and Labor Economics*, 2013
83. Shifa, Abdulaziz B., *Essays on Growth, Political Economy and Development*, 2013
84. Panetti, Ettore, *Essays on the Economics of Banks and Markets*, 2013
85. Schmitt, Alex, *Beyond Pigou: Climate Change Mitigation, Policy Making and Decisions*, 2014
86. Rogall, Thorsten, *The Economics of Genocide and War*, 2015
87. Baltrunaite, Audinga, *Political Economics of Special Interests and Gender*, 2016
88. Harbo Hansen, Niels-Jakob, *Jobs, Unemployment, and Macroeconomic Transmission*, 2016
89. Stryjan, Miri, *Essays on Development Policy and the Political Economy of Conflict*, 2016
90. Karadja, Mounir, *On the Economics and Politics of Mobility*, 2016
91. Kitamura, Shuhei, *Land, Power and Technology, Essays on Political Economy and Historical Development*, 2016
92. Malmberg, Hannes, *Human Capital in Development Accounting and Other Essays in Economics*, 2017
93. Öberg, Erik, *On Money and Consumption*, 2017
94. Lane, Nathaniel, *States of Development: Essays on the Political Economy of Development in Asia*, 2017
95. Prawitz, Erik, *On the Move: Essays on the Economic and Political Development of Sweden*, 2017
96. Iwanowsky, Mathias, *Essays in Development and Political Economics*, 2018
97. Dehdari, Sirus Håfström, *Radical Right, Identity, and Retaliation*, 2018
98. Harmenberg, Karl, *Essays on Income Risk and Inequality*, 2018
99. Kilström, Matilda, *Household's Responses to Policy in Labor and Credit Markets*, 2018
100. Meriläinen, Jaakko, *Essays in Political Economics*, 2019

101. Sigurdsson, Jósef, *Essays on Labor Supply and Adjustment Frictions*, 2019
102. Cocciolo, Serena, *Participatory Governance and Public Service Provision*, 2019
103. Mitrunen, Matti, *Essays on the Political Economy of Development*, 2019
104. Olsson, Jonna, *Work, Wealth, and Well-being: Essays in Macroeconomics*, 2019
105. Darougheh, Saman, *Search and Mismatch*, 2019