UNIVERSITÀ COMMERCIALE LUIGI BOCCONI

PhD SCHOOL

PhD Program in Economics and Finance
Cycle: 36th
Disciplinary Field (code): SECS-P/01

Essays in Firm Dynamics and the Role of Capital Intensity

Advisor: **Basile Grassi**
Co-Advisor: **Dmitriy Sergeyev**

PhD Thesis by
Geraud Desazars de Montgailhard
ID number: 3103440

Year 2026

Contents

Capital Intensity and Firm Dynamics*

Geraud Desazars

March 13, 2026

Most recent version here

Abstract

Large firms exhibit systematically higher capital intensity relative to peers, a fact at odds with standard models where firms technological differences are factor-neutral. This paper develops a general equilibrium framework in which firms expand by adapting their technology in order to adopt capital goods when they become cheaper. Firms can pay a firm-specific switching cost to change their production technology for a more capital-intensive one. As the cost of capital falls due to capital-embodied technical change, the firms that face the smallest switching costs become relatively more capital intensive and gain a competitive edge. This allows them to build market shares and markups. Markups endogeneity generates strategic interactions which widens further the dispersion in capital intensity and market shares. The model provides a unified explanation for rising dispersion in markups, a falling aggregate labor share despite a rising median labor share, investment weakness, and the divergence between sales and employment concentration. *JEL Codes:* D21, D24, E22, E25, L11, L16.

*Bocconi University. `geraud.desazars@phd.unibocconi.it`. This paper previously circulated under the name "Growing through Capital". I am grateful to Basile Grassi and Dmitriy Sergeyev for their continuous advice and support. I also thank Nicolas Crouzet and Yueran Ma for their extensive feedback and valuable discussions. I am further grateful for discussions with: Philippe Aghion, Ufuk Akcigit, Vivek Bhattacharya, Giuseppe Berlingieri, Florin Bilbiie, Lawrence Christiano, Joel David, Gustavo de Souza, Piero de Dominicis, Jan Eeckhout, Martin Eichenbaum, Carlo Favero, Veronica Guerrieri, François Gourio, Benjamin Jones, Greg Kaplan, Martin Kornejew, Francis Kramarz, Thomas Le Barbanchon, Guido Lorenzoni, Clément Mazet-Sonilhac, Tommaso Monacelli, Marco Ottaviani, Christina Patterson, Nicola Pavoni, Marta Prato, Nick Pretnar, Xavier Ragot, Matthew Rognlie, Nicolas Serrano-Velarde, Vincent Sterk, Enrico Stivella, and Carolina Villegas-Sanchez. I also thank conference and seminar participants at: American Finance Association, Bocconi, Chicago Fed, European Economics and Financial Society, HEC Paris, Northern Finance Association, University of Cambridge, University of Chicago, Universita Catolica di Milano, Northwestern University, Politecnico di Milano, XXVIII Workshop on Dynamic Macroeconomics in Vigo, and Swiss Financial Institute.

1 Introduction

Capital intensity is at the center of longstanding economic debates. It can shape the distribution of income between workers and capitalists, as argued by Marx (1867, Vol. I, Ch. 25), and has recently been scrutinized in the context of the U.S. labor-share decline (Karabarbounis and Neiman, 2014; Oberfield and Raval, 2021). It also bears on growth according to Schumpeter's definition of development (Schumpeter, 1911, Ch. II) and to more recent studies of capital-embodied technological change (Jones and Liu, 2024). Traditionally, capital intensity is also associated with employment: from Keynes' predictions of rising "technological unemployment" due to mechanization (Keynes, 1930) to the revival of this concept in the face of AI (Acemoglu, Autor, Hazell, and Restrepo, 2022; Eloundou, Manning, Mishkin, and Rock, 2024).

This paper brings a new perspective to these debates and to the role of capital intensity by documenting three stylized facts: (i) across- and within-sector dispersion in firms' capital intensity has widened; (ii) capital-intensive firms are larger; and (iii) capital-intensive firms are associated with higher markups. These three stylized facts motivate two questions: why are firms heterogeneously capital-intensive, and what is the link between a firm's capital intensity and its size? What are the implications for aggregate trends?

I answer these questions by proposing that firms differ in their ability to adopt increasingly productive capital goods. I model this by allowing firms to update the output elasticity to capital of their production technology as a reaction to capital goods becoming more productive, or equivalently, cheaper (Gordon, 1990; Greenwood, Hercowitz, and Krusell, 1997; Cummins and Violante, 2002). I then embed this mechanism in a general-equilibrium model with oligopolistic competition, where heterogeneity in updating arises both from uneven adoption capability among firms and from strategic interactions. I use the model to derive the macroeconomic implications of this heterogeneity. Calibrated to U.S. public firms (1980-2019), the model accounts for four aggregate trends: the aggregate markup rises by 6.7%; the aggregate labor share falls by 5.6 percentage points (pp); the aggregate investment gap—i.e., the distance between the aggregate marginal product of capital and the marginal cost of capital—increases by 18%; and large firms' sales shares outpace their employment shares, yielding a 15pp difference in 2019.

The paper proceeds in three steps. First, I document dispersion in capital intensity and its link with size and markups using public firm data and Bureau of Economic Analysis (BEA) data. I show that this heterogeneity can be represented by firm-specific output elasticity to capital. A firm's output elasticity to capital in turn determines the sensitivity of its marginal cost to changes in the cost of capital goods. When capital goods become more productive—equivalently, cheaper in efficiency units—capital-intensive firms see their marginal costs decline compared to those of their labor-intensive competitors. In other words, because they adopt the cheaper input more intensively, capital-intensive firms are more efficient and therefore capture higher market shares. Under some general conditions, I show that a firm's output elasticity to capital nests its level of automation and its capital-biased productivity.

If capital-intensive technologies are more efficient than labor-intensive technologies, why don't all firms converge to a capital-intensive production function? I answer this question in the second step of the paper by building a general equilibrium model in which firms choose their output elasticity to capital and compete in oligopolistic markets. In the model, a capital-goods producer sees its productivity increase, resulting in a declining cost of capital goods relative to labor. Firms respond by updating their output elasticity to capital, but differ in their capacity to do so for two reasons: differing exogenous capabilities of updating their technology, and endogenous strategic interactions.

I model adoption capability with a switching cost: to update its technology, a firm pays a cost that increases with the distance between its current and target output elasticity to capital. I assume that firms differ in the steepness of their switching cost function to capture, in a stylized manner, drivers of differing capabilities such as entrepreneurial/managerial differences, organizational features or information asymmetries. Strategic interactions also shape heterogeneity in firms' technological updating because in oligopolistic models, a firm's markup is increasing and convex in its market share.[1] As a result, large capital-intensive firms stand in the steeper region of the markup curve while small labor-intensive firms stand in its flatter part implying heterogeneous marginal returns to firms' technological updating.

To understand the relative role of these two channels, I present the mechanism through comparative statics. At the beginning of the model, firms share the same technology and differ only in their switching cost, i.e, their capability to update their production function. When the relative cost of capital (RCK) declines as the capital-goods producer becomes more productive, all firms adopt the cheaper input, raising their capital-to-labor ratio. Firms also decide to update their technology—specifically, their output elasticity to capital—to make their output more elastic to the cheaper input. Firms with lower switching costs update more and see their capital-to-labor ratio rise faster than the aggregate. Capital-intensive firms now display a marginal cost that is more sensitive to the decline of the RCK than their labor-intensive competition. Their marginal cost is therefore lower, allowing them to set lower prices and attract demand. With only switching-cost heterogeneity, firm dynamics materialize and dispersion in capital intensity and market shares emerges.

With endogenous markups, the mechanism is similar to the one just detailed up to one extra sub-mechanism: strategic interactions. Now that firms have dispersed market shares, they stand at different points on the markup curve, which is increasing and convex in market share. As such, the marginal return to increasing one's market share is higher for firms that are already capital-intensive as they are on a steeper tangent to the markup curve. The difference in the amplitude of the update of firms' technologies is hence magnified and capital

[1]This assumption is common to other models of endogenous markups such as duopoly with limit pricing as in Aghion and Howitt (1992) or in monopolistic competition models with a Kimball (1995) aggregator.

intensity, market shares and markups disperse at a faster rate. As the model unfolds, and the RCK declines secularly, the fanning-out of the distribution of firms' capital intensity, market shares and markups accelerates and is driven almost solely by strategic interactions.

In the third step of the paper, I derive the macroeconomic implications of this dimension of firm heterogeneity and compare some model-implied moments to their data counterpart. To do so, I calibrate my model on American public firms between 1980 and 2019, targeting firms' capital intensity and the fall in the RCK. The model provides a unified quantitative account for four key aggregate trends that have characterized the American economy over the past decades.

First, the model reproduces a rise in firm-level markup dispersion together with a higher aggregate markup (de Loecker, Eeckhout, and Unger, 2020; de Ridder, Grassi, and Morzenti, 2022). In the model, when the RCK falls and firms update heterogeneously their output elasticity to capital, firms that end up more capital intensive capture higher market shares. The lower the RCK, the higher the competitive edge for capital-intensive firms and hence the dispersion in market shares. Because markups are increasing and convex in market shares, this translates in an even wider dispersion in markups. Overall, the model predicts that between 1980 and 2019, the variance of firms' markups is multiplied by four, which, together with reallocation towards high-markup firms, results in an increase of 6.7% in the aggregate markup (half of the increase estimated by de Loecker et al. (2020)).

The model also generates a 5.6 percentage point (pp) decline in the aggregate labor share and a 16.9pp decline in the manufacturing labor share (respectively 70% and 84% of their empirical counterpart) while micro-founding the empirical regularities documented by Autor, Dorn, Katz, Patterson, and Van Reenen (2020) and Kehrig and Vincent (2021). In the baseline model, following Antras (2004) and Oberfield and Raval (2021), I assume that firms' production technologies exhibit complementarity between capital and labor. Initially, a decline in the RCK therefore raises all firms' labor share. This is dampened by the fact that all firms raise their capital share when they increase their output elasticity to capital following a fall in the RCK. Crucially, heterogeneity in technological updating yields dispersion in market shares and labor shares. Capital-intensive firms (which update the most) see their capital share rise the most and attract more demand, boosting their markup and hence their profit share. This results in capital-intensive firms growing in size while their labor share plunges. On the contrary, labor-intensive firms lose market shares as their labor share grows (because their profit share plummets as they lose market shares and hence markups). Overall, reallocation from increasing labor share firms (labor-intensive) towards decreasing labor share firms (capital-intensive) generates an aggregate decline in the labor share through a mild increase in the capital share (+0.4pp) and a sharp rise in the profit share (+5.2pp). Importantly, the model implies that capital-labor substitution and markup dynamics are complementary drivers of the labor share decline, operating jointly rather than as competing explanations.

While firms grow through capital, capital misallocation and the investment gap increase. Capital misallocation is measured as dispersion in firms' marginal productivity to capital following Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). As large capital-intensive firms charge higher markups, they step further away from their efficient frontier because markups increase the distance between the marginal productivity and the marginal cost of capital, as in Baqaee and Farhi (2020) and Edmond, Midrigan, and Xu (2023). Given capital-intensive firms' disproportionately large share of capital, this deviation translates into pulling the entire economy's capital investment away from the aggregate efficient frontier, in line with what Gutiérrez and Philippon (2016) call a rising aggregate investment gap. Overall, the model accounts for roughly one-third of the investment gap estimated by Gomme, Ravikumar, and Rupert (2011) or Gutiérrez and Philippon (2016).

Finally, in the model, large firms concentrate different portions of sales and employment, consistently with Autor et al. (2020), Kwon, Ma, and Zimmermann (2024) and Ma, Zhang, and Zimmermann (2024). Specifically, as large firms grow by becoming more capital intensive, their demand for workers per unit of output declines. While in 1980 the sales and employment share of large firms were comparable, in 2019, the sales share of the largest firms is close to 15pp higher than their employment share.

Literature. This paper maps microeconomic heterogeneity into macroeconomic trends and contributes to the literature by documenting a novel dimension of firm heterogeneity and proposing a new mechanism of reallocation across heterogeneous firms. Closest to my mechanism is Hubmer and Restrepo (2025) who build on Acemoglu and Restrepo (2018, 2022) and Hubmer (2023) to propose a model of firm-specific automation. In their model, automation comes at a fixed cost but only firms that are ex-ante large (due to high TFP) can afford such a cost and automate when the RCK falls. On the contrary, in my paper, firms are ex-ante of the same size and differing capabilities to update their technology is the cause for heterogeneous market share dynamics. While we share the same implications regarding the labor share, Hubmer and Restrepo (2025) introduce an extra change to match the empirical regularities of Kehrig and Vincent (2021): rising competition intensity. This implies that the dispersion of markups compresses while reallocation towards efficient firms increases, yielding a stable aggregate markup. This yields two implications that are opposite to mine. First, in their model, the fall of the aggregate labor share is carried by a rise in the capital share while the profit share is stable. Second, increasing competition intensity diminishes misallocation and stabilizes the aggregate markup, implying that the investment gap does not increase. Despite those differences, both our papers point at the importance of capital-biased technological heterogeneity among firms to explain structural changes in the American economy.

My paper is also related to Oberfield and Raval (2021) who show that complementarity between capital and labor implies that the RCK and the labor share are negatively correlated.

I build on their result and show that introducing endogenous updating of output elasticity to capital allows to match the evolution of firms' capital-to-labor ratio. This addition and the presence of endogenous markups yields a positive relationship between RCK and the labor share. While my model focuses on trends and hence ignores nominal frictions, Martinez (2021) proposes a model of putty-clay automation where firms adjust their capital stock episodically and irreversibly. This also generates dispersion in the output elasticity to capital and offers rich short- vs. long- terms insights. Finally, my paper also builds on the results of: Grazzini and Rossi (2023) who show that entrant cohorts become more capital-intensive as the RCK falls; Guntin and Kochen (2025) who document that firms destined to become large start more capital-intensive than their peers; Lashkari, Bauer, and Boussard (2024) who show, using French administrative data, that demand for tangible and intangible IT capital increases when its relative cost falls; Cho, Manaresi, and Martin (2025) who link the fall in the aggregate labor share to capital-intensive superstar firms for 26 OECD countries.[2]

Naturally, this paper relates to alternative models that aim at understanding the same structural changes. Rising markups can be demand-driven markups (see Gourio and Rudanko (2014) for customer capital), the result of common ownership (Azar, Schmalz, and Tecu, 2018) or of decreased competition intensity (de Loecker, Eeckhout, and Mongey, 2021), or supply-driven and resulting from heterogeneous TFP (Aghion and Howitt, 1992; Atkeson and Burstein, 2008; Grassi, 2017; Burstein, Carvalho, and Grassi, 2020; Edmond, Midrigan, and Xu, 2015). Different causes for the falling aggregate labor share have been put forward: Kaymak and Schott (2023) propose a mechanism related to mine as they show that declining corporate tax rates favor capital-intensive firms whose labor shares are lower, Elsby, Hobijn, and Şahin (2013) propose offshoring, Koh, Santaeulàlia-Llopis, and Zheng (2020) look at changes in accounting methods, Garibaldi and Turri (2024) focus on monopsony. Some focus on the profit share: de Loecker et al. (2020), Barkai (2020), Hartman-Glaser, Lustig, and Xiaolan (2019), David, Ranciere, and Zeke (2024) while others focus on the capital share: Karabarbounis and Neiman (2014), Acemoglu and Restrepo (2023), or Berlingieri, Boeri,

[2]Papers that focus on the role of heterogeneous adoption of intangible capital are also related. Closest is de Ridder (2024), who attributes the slowdown in productivity growth, lower dynamism, and higher markups to the rise of intangibles that lower marginal but raise fixed costs, giving adopters a cost advantage that deters entry. Aghion, Bergeaud, Boppart, Klenow, and Li (2023) likewise connect rising concentration with falling growth and rising rents through the expansion of the most efficient firms across product lines due to a fall in the overhead cost of spanning multiple product lines. Finally, Crouzet, Eberly, Eisfeldt, and Papanikolaou (2024) formalize replicability of intangibles, showing that easier replication scales firm scope and raises concentration, valuation ratios, and the profit share while lowering long-run growth. All these results speak to specificities of intangible capital vs. tangible capital. I show that heterogeneity in tangible capital intensity is at least as wide as heterogeneity in intangible capital. Because, as shown by Gourio and Rognlie (2020), software are among the capital goods who have seen their cost decline the most, I include intangibles in my baseline measure of capital intensity. My results are hence complementary to those of papers focusing on intangible capital. Alternative driving forces of reallocation across heterogeneous firms comprise, among others, falling imitation rates (Akcigit and Ates, 2023), falling demographic dynamism (Peters and Walsh, 2022), falling interest rates (Liu, Mian, and Sufi, 2022), or rising returns-to-scale (Hubmer, Chan, Ozkan, Salgado, and Hong, 2024)

Lashkari, and Vogel (2024). The rising investment gap, first documented by Abel, Mankiw, Summers, and Zeckhauser (1989), is attributed to mismeasurement of the intangible capital stock (Crouzet and Eberly, 2021), rising markups (Gutiérrez and Philippon, 2016), frictions that prevent productive firms from entering new markets (Peters, 2020), or rising risk premia (Farhi and Gourio, 2018; David, Schmid, and Zeke, 2022). David and Venkateswaran (2019) propose to test empirically the relative importance of those possible drivers of misallocation and find that technology and markups are important drivers. Finally, diverging sales and employment concentration has been freshly documented by Autor et al. (2020), Kwon et al. (2024) and Ma et al. (2024) and to the best of my knowledge, no model exists that explains this pattern.

The rest of the paper is organized as follows. Section 3 first documents rising dispersion in the capital-to-labor ratio of firms and lays out the capital-intensity modeling at the firm-level. Second, I show that capital-intensive firms have seen their market shares rise. In Section 3, I build the general-equilibrium model, following the directions established in Section 3, I define the equilibrium and describe the calibration technique. Section 3 derives the aggregate implications of the model and presents the results obtained from the simulated baseline model in terms of firm, markups, income shares, investment, and concentration dynamics. Section 5 presents counterfactual exercises and extensions and Section 6 concludes.

2 Stylized Facts and Firms' Technologies

In this section, I start by presenting the data I use in section 2.1. I then document the stylized facts that motivate this paper and show how these can be modeled in firms' production technologies.

2.1 Data

This section presents data sources and measure constructions. Further information is detailed in Empirical Appendix A.1.

Compustat North American Fundamentals (Annual). To build firm- and year-specific measures I use the publicly available Annual North American Fundamentals cut of the Compustat database. It provides the financial statements (balance sheets, income statements and cash flow statements) of publicly traded firms in the United States. The dataset is retrieved from the Wharton Research Data Services. The sample starts in 1980, because data prior to that date is sparse. The sample ends in 2019 to ignore the COVID episode specificity. I focus mainly on firms' sales, physical and intangible capital and employment. I exclude observations with missing or negative sales, capital and employees. I exclude observations not nested within a sector (no information or "Other"). The sample counts 162,282 unique firm-year observations, covers 17,708 firms across 19 2-digit NAICS industries[3].

Bureau of Economic Analysis (BEA). I build three sets of sector-specific measures using data from the BEA: real gross output, deflators for output and capital and the constant-quality relative cost of capital.

Real gross output by sector is retrieved from the "GDP-by-industry" tables of the BEA which provides a series for real gross output (in 2017 billions of dollars) per 3-digit NAICS sector. Because some 3D sectors don't have data prior to 1997 (every 3D sectors in the 2D sector "Retail trade" for example), I focus on 2D sectors.

To deflate firms' output, I compute a deflator for each 2D sector by dividing real gross output by nominal gross output from the "GDP-by-industry" tables of the BEA. Similarly, for capital goods I use data from the "Fixed Asset Tables" of the BEA. I divide the Current-Cost Stock of Capital (Table 2.1) with the Historical-Cost Stock of Capital (Table 2.3) which gives me 3 separate deflators for "Structures", "Equipment" and "Intellectual Property Products" for each 2D sector. I then use a weighted sum (by relative stock) of the deflator for Structures and Equipment as the deflator for physical capital and the deflator for Intellectual Property Products as the deflator for intangible capital.

Finally, I use data from the BEA to construct the relative cost of *constant-quality* capital (RCK) following Gourio and Rognlie (2020). To build sector-specific cost of constant-quality capital, I take the ratio between current-cost investment in fixed assets (Table 2.7) and the quantity index for investment in fixed assets (Table 2.8). The quantity index is built in two steps. First, the historical-cost investment in fixed asset measure is deflated for each sector. Second, using hedonic methods pioneered by Gordon (1990) and Cummins and Violante (2002) the measure is adjusted for changes in the quality of capital goods.[4]. The cost of consumption is the price index for non-durable goods obtained from Table 1.2.4 from National Income and Product Accounts (NIPA) tables. The cost of labor per industry is "Wages and Salaries Per Full-Time Equivalent Employee by Industry" (Tables 6.6 B, C and D). The resulting dynamics for the RCK is plotted in Figure A.12 in Appendix A.2. In the baseline, I do not consider the fall in the interest rate as an additional driver for the decline in the relative cost of capital (in the form of a rental rate) because, as shown by Farhi and Gourio (2018) for example, risk and liquidity differences between the required return to capital and the risk-free interest rate make their correlation uncertain. Furthermore, as

[3]Public administration is not considered. Manufacturing codes 31, 32 and 33 are consolidated. Same for Retail Trade (44 and 45) and Transportation and Warehousing (48, 49)

[4]Importantly, as noted by Greenwood et al. (1997), the cost of capital goods and the deflator for capital goods don't coincide because the capital stock and the investment flow don't have the same weights in terms of capital goods types and vintages.

documented by Barkai and Benzell (2024), considering the Feds Fund rate as a measure for the required return makes the capital and profit shares over-sensitive to the rental rate of capital (the profit share collapses from 15% to 0% in the Volcker years and then reverts back to 15%). For completeness, I repeat my results including the fall of the interest rate in Appendix A.4 for the empirics and in Appendix B.10 for the model.

2.2 Rising Heterogeneity in Firms' Capital Intensity

2.2.1 Heterogeneous Capital Intensity in the Aggregate

There has been a substantial divergence in firms' capital intensity across time. To show this, I build a firm-year measure of capital intensity in the following way: capital is the sum of (i) physical capital (Property, Plant and Equipment in Compustat) deflated by the stock-weighted sum of the cost of structures and equipment from the BEA at the 2D-sector level and (ii) intangible capital (Intangible Assets in Compustat) deflated by the stock-weighted deflator from the BEA for intangible capital (Intellectual Property Products in the BEA) estimated at the 2D-sector level as well. I include tangible and intangible capital due to the rising importance of the latter documented in Corrado, Hulten, and Sichel (2009), McGrattan and Prescott (2010), Crouzet and Eberly (2019), Koh et al. (2020), Crouzet, Eberly, Eisfeldt, and Papanikolaou (2022), and Crouzet and Eberly (2023), and visible in Appendix A.2, Figure A.1.[5] Labor is simply the count of employees of each firm at each period. In Figure 1, I plot unweighted percentiles of the distribution of firms' capital intensity over time. It reveals three main facts. First, all firms seem to have increased their capital intensity. Second, the variance of the cross-sectional distribution of firms' capital intensity has increased. Third, the increase in the variance is driven by the upper decile: the 10% most capital-intensive firms in 2019 have a capital intensity five times higher than in 1980 while the median firm has seen its capital intensity be multiplied by a little less than 2.5.[6]

The common growth in firms' capital intensity is well understood. Over the same period, the cost of capital goods has increased at a lower pace than that of labor, yielding a decline

[5]Note, however, that none of the two types of capital drives exclusively any of the trends presented in this section. Exercises carried in the empirical section are revamped in appendix A.4.5 separating tangible and intangible capital.

[6]In Appendix A.4, Figure A.13, I plot moments of the distribution of capital intensity changing the denominator of my measure to cost of goods sold (COGS). COGS capture labor costs as well as other intermediate costs. This measure also suggests a fanning out although with a smaller takeoff of the top decile. Among the intermediate inputs captured by COGS figures outsourcing. In principle, firms that outsource labor would feature lower capital per employee but similar capital per COGS than firms who don't. While outsourcing has been affecting firms unequally, as documented by Boehm, Flaaen, and Pandalai-Nayar (2020) or Bergeaud, Malgouyres, Mazet-Sonilhac, and Signorelli (2025), the fact that the same fanning out appears in capital per employee and capital per COGS implies that outsourcing is not the main driver of capital-intensity dispersion.

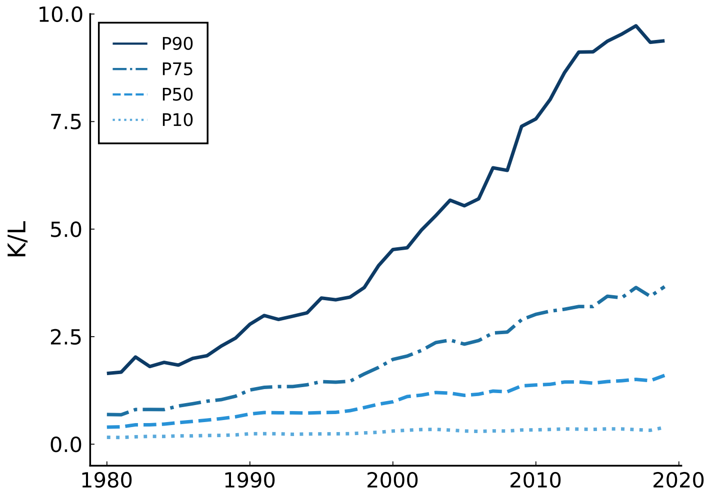


Figure 1: Percentiles of the distribution of firms' capital intensity.

Note: Source is Compustat. Sample is 1980-2019. Measure is (PPE+INT)/L. PPE stands for "Property, Plant and Capital (Gross)" and measures physical capital while INT stands for "Intangible Assets" and measures intangible capital. EMP is "Employees" and is the count of employees. P90 is the 90th percentile, P10, the 10th. Units are 100'000s of 2000 dollars per employee.

in the relative cost of capital (RCK) and a rise in aggregate capital intensity.[7] However, a falling RCK does not imply dispersion if firms face the same costs of inputs. To see this, consider a production function of firm i in time t: $Y_{i,t} = \mathcal{F}_{i,t}(K_{i,t}, L_{i,t})$ where $K_{i,t}$ is capital, and $L_{i,t}$ is labor. At time t, for a given pair of costs $\{w_t, r_t^K\}$ of labor and capital respectively, supposing that firms minimize their costs yields the following identity for a firm's capital intensity:

$$\frac{K_{i,t}}{L_{i,t}} = \frac{\zeta_{i,t}^K}{\zeta_{i,t}^L} \frac{w_t}{r_t^K} \tag{1}$$

Where $\zeta_{it}^{\mathcal{I}} = \frac{\partial \ln(Y_{it)}}{\partial \ln(\mathcal{I}_{it})}$ is the output elasticity to input $\mathcal{I} \in \{K, L\}$. Clearly, if a firm is price-taker on input markets, only differences in its output elasticities to inputs can drive heterogeneity in capital intensity.[8]

[7]See for example: Gordon (1990), Greenwood et al. (1997), Greenwood, Hercowitz, and Krusell (2000), Cummins and Violante (2002), Fisher (2006), Primiceri (2006), Gourio and Kashyap (2007), Gourio and Klier (2015), Karabarbounis and Neiman (2014), Jones and Liu (2024), Dechezleprêtre, Hémous, Olsen, and Zanella (2025), Hartley, Hassett, and Rauh (2025). See also Kaymak and Schott (2023) for the role of corporate tax changes.

[8]In Appendix B.1, I show that capital intensity is orthogonal to TFP, demand shifters and returns to scale in the case of a CES production function.

2.2.2 Heterogeneity Across and Within Sectors

If rising dispersion in capital intensity comes from input-biased technological differences, it can come from two levels of granularity. The first one is sectoral and is well established (Acemoglu and Guerrieri, 2008). Sectors with high structural capital intensity (agriculture, manufacturing, information) have been subject to automation and robotization (see Acemoglu and Restrepo (2020, 2022) among others) or offshoring of labour-intensive technologies (see for example Elsby et al. (2013)) while lowly capital-intensive sectors, such as services, haven't. Diverging capital intensity can also materialize within sector. It would imply that firms use differently capital-intensive technologies within sector. For example, the use of satellites by Walmart to centralize real-time inventory management makes its technology more capital-intensive than its competitors.[9] I find that this within-sector dispersion is of the same order of magnitude than cross-sector dispersion. In Appendix A.2.1, Figure A.2, I plot percentiles of the distribution of firms' capital intensity for manufacturing, trade and services; pictures similar to Figure 1 obtain. I also show this in a more systematic way by comparing dispersion measures of capital intensity within and across sectors in Table A.1 in the same Appendix.

Another useful way to see this is by looking at firm-level output elasticities to capital. To do so, notice that inverting equation (1) allows to estimate the ratio of output elasticities at the firm-year level. Indeed, supposing constant returns to scale allows, by Sheppard's lemma, to estimate output elasticities as cost shares:

$$\zeta_{i,t}^K = \frac{r_t^K K_{i,t}}{TC_{i,t}}$$

and $\zeta_{i,t}^L = \frac{w_t L_{i,t}}{TC_{i,t}}$ where $TC_{i,t}$ stands for total costs of firm i in time t[10]. To perform this estimation, I use the same measures for $K_{i,t}$ and $L_{i,t}$ as for capital intensity. For the relative cost of capital, r_t^K/w_t, I measure the rental rate of capital as the opportunity cost of capital times the cost of investment at the 2D sector level from the BEA following Gourio and Rognlie (2020). The opportunity cost of capital is $r+\delta$ where r is the real interest rate and δ is the depreciation rate of capital. In the baseline, the real interest rate is 0.05 and the depreciation rate is 0.1.[11] The wage is defined at the 2D level of granularity and taken from the BEA (see

[9] See this article: Satellites for details.

[10] I suppose constant returns to scale because, as shown in Appendix B.1, heterogeneity in returns to scale cannot account for heterogeneity in capital intensity under a CES technology. It could still be that firms' technology share a common homogeneity of degree γ. In such a case, the distribution of their output elasticity to capital is the same up to a constant, that is, we would have: $\frac{\zeta_{i,t}^K}{\gamma} = \frac{r_t^K K_{i,t}}{TC_{i,t}}$. It could also be that rising output elasticity to capital increases a firms' return to scale, as is the case for intermediate inputs in Hubmer et al. (2024). This possibility does not affect the four key results of the paper and is therefore left for further research.

[11] In Appendix A.4, I repeat this estimation with a falling real interest rate that follows the path of the policy rate. As is visible in Figure A.16, while this slightly dampens the growth rate of firms' ζ_{it}^K, those

the Data section above for details about these measures).[12] Total cost is then the sum of the cost of capital and the cost of labor. Some moments of the resulting distribution of output elasticity to capital are plotted in Figure 2 for the Aggregate, Manufacturing (NAICS codes 31, 32 and 33), Trade (NAICS codes 42, 44 and 45) and Services (NAICS codes 54, 55, 61, 62 and 71).[13]

Aggregate and Manufacturing feature different patterns than Trade and Services. For the last two, the distribution of output elasticity to capital is dispersed but close to stationary across time. This is enough to generate the widening dispersion in capital intensity in these two sectors (displayed in Figure A.2 in Appendix A.2) as the relative cost of capital falls. For Manufacturing, the output elasticity to capital of the median firm is rising. However, within this sector, the rise in the output elasticity to capital is unequal as rising dispersion in firms' capital intensity is too large to be explained only by the initial heterogeneity in firms' output elasticity to capital. The observed rising dispersion in the distribution of output elasticity to capital in the aggregate is therefore the result of the patterns across and within sectors.

Heterogeneity in firms' capital intensity, even within sectors, relates to the literature on absorptive capacity that puts forward heterogeneous capacity to assimilate external knowledge (see Cohen and Levinthal (1990) for the seminal paper). This is also consistent with some recent empirical findings; Baley and Veldkamp (2025) document that firms utilization of data is widely dispersed, Acemoglu, Anderson, Beede, Buffington, Childress, Dinlersoz, Foster, Goldschlag, Haltiwanger, Kroff, et al. (2022) show that firms differ in their adoption rates of robots, AI, cloud computing, specialized equipment and specialized software, even within narrowly defined sectors. Atkin, Chaudhry, Chaudry, Khandelwal, and Verhoogen (2017) show that when firms that produce soccer balls in Pakistan are proposed to adopt a technology that reduces material waste, only a handful of those adopt the technology.

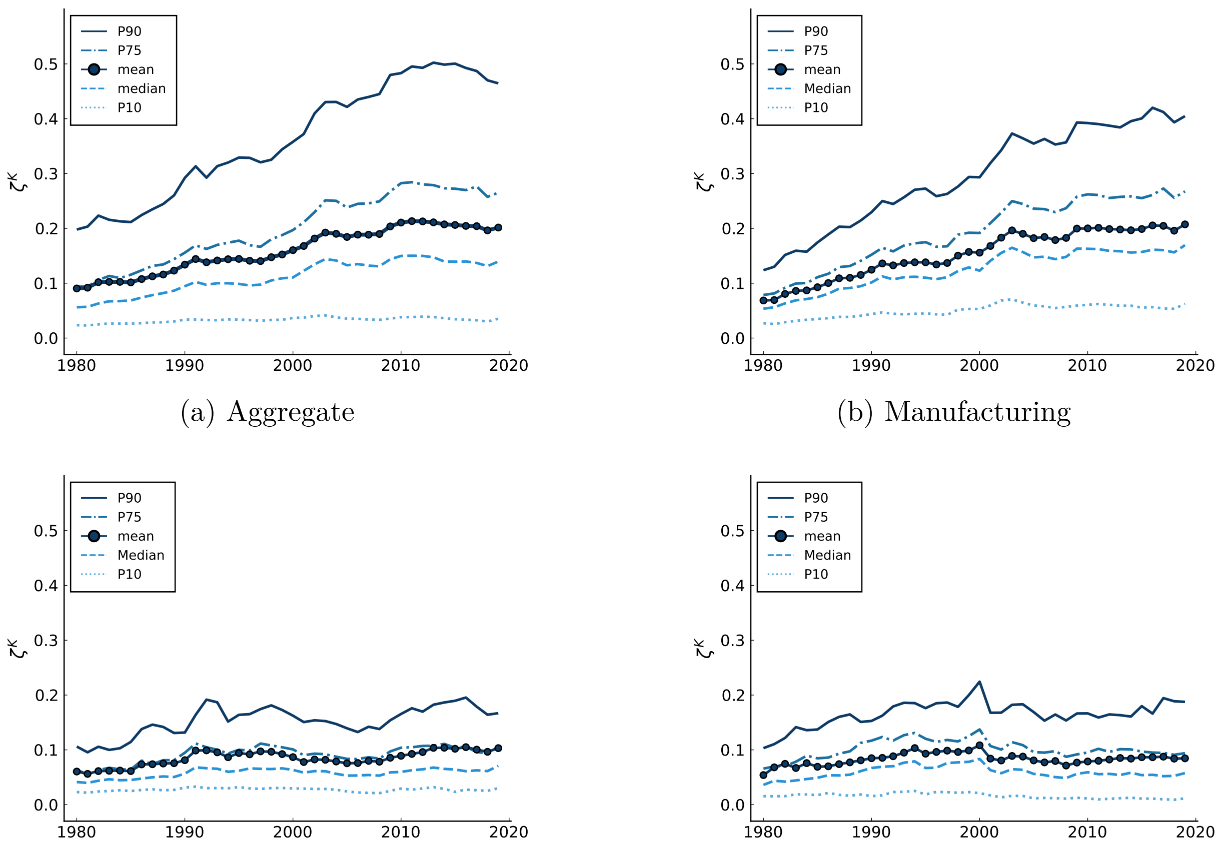
For completeness, in Appendix A.2, I plot in Figure A.3 the evolution of the output elasticity to capital for other 2-digit sectors. In Figure A.4, I plot the percentiles of the distribution of firms' output elasticity to capital within the 9 largest 4-digit sectors in Manufacturing to show that rising dispersion carries across all sectors of Manufacturing and at thinner levels of granularity.

Before moving to the next fact, note that there are three concerns with using Compustat data that might affect the robustness of these facts. First, the universe of publicly listed firms is not representative of the universe of all firms in the US along many dimensions.

remain significantly heterogeneous.

[12] To account for potential differences in firms' wages, in Figure A.15 in Appendix A.4, I repeat the estimation of $\zeta_{i,t}^K$ using Cost of Goods Sold as a proxy for $w_t L_{i,t}$ although this measure is polluted by other intermediate goods. The conclusion of such exercises is that the distribution of $\zeta_{i,t}^K$ is unaffected by the choice of measure for $w_t L_{i,t}$.

[13] In Appendix A.4, I look into the possibility that firms face different opportunity costs of capital. That is, I redefine $r_t + \delta$ into $r_{i,t} + \delta$ to allow for larger firms to have lower risk premia for example. I show that, supposing homogeneous ζ across firms, the implied increase in the dispersion of firms' opportunity costs of capital is counterfactually high.



(a) Aggregate (b) Manufacturing

(c) Trade (d) Services

Figure 2: Percentiles of the distribution of firms' output elasticity to capital per sector.

Note: Source is Compustat. Sample is 1980-2019. $\zeta_{i,t}^{K} = (r_t^K K_{i,t})/TC_{i,t}$ is the output elasticity to capital, or equivalently, the cost share of capital of firm i at time t. K_{it} is the sum of physical capital (PPE in Compustat) and intangible capital (INT in Compustat). $TC_{i,t}$ is total cost and is the sum of cost of capital and cost of labor defined as the wage bil. P90 is the 90th percentile, P10, the 10th. Manufacturing gathers 2-digit NAICS sectors 31, 32 and 33. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).

Second, capital might be mismeasured. Third, listing and de-listing patterns do no track entry and exit rates both in the aggregate and within sectors.[14] To make sure that my measures aren't biased by these defects, in Figure A.17 of Appendix A.4, I compute the aggregate output elasticity to capital for the American economy and for the manufacturing sector using BEA data. To do so, I use the same measure for the relative cost of capital as in my baseline exercise. For capital, I use the "quantity index for net stock of private fixed assets by industry" (Fixed Asset table 3.2), for labor I use "full-time equivalent employees by industry" (National Income and Product Accounts table 6.5). The output elasticity to

[14] See Fama and French (2001), Grullon and Michaely (2002), Fama and French (2004), Davis, Haltiwanger, Jarmin, Miranda, Foote, and Nagypal (2006), Ali, Klasa, and Yeung (2008), Keil (2017), Kahle and Stulz (2020) or Decker and Williams (2023) for a thorough documentation of these shortcomings.

capital of both the whole economy and manufacturing rise between 1980 and 2019 with an amplitude in line with that of the median or the mean of Compustat firms displayed in Figure 2.

We have seen that rising heterogeneity in firms' capital intensity is a pervasive phenomenon, across and within sectors. This heterogeneity can be expressed as differing output elasticity to inputs across firms. While for most sectors the distribution of firms' output elasticity to capital is stationary, it has been fanning out for manufacturing. I now turn to investigating the relationship between firms' size and capital intensity, focusing on manufacturing (results for the aggregate and other sectors in Appendix A.2).

2.3 Capital Intensity Correlates with Size

Figure 3 plots the correspondence between capital intensity deciles and their relative size within manufacturing for the beginning of my sample (1980) and its end (2019). Relative size is defined as the sales of the firms present in a decile to the total sales in the sector. It suggests a permanent positive correlation between capital intensity and size in manufacturing. While the relationship seems close to linear in 1980, in 2019 the 10% most capital-intensive firms seem to concentrate a higher share of sales. Rising concentration is even more visible for top percentiles: in Figure A.9 of Appendix A.2, I plot the time series of the relative size of the 90th, 95th, and 99th percentiles of the distribution of capital intensity in manufacturing over my whole sample. Concentration seems to have accelerated between 2000 and 2010 before cooling down and to be concentrated at the top of the distribution: in 2019, the 1% most capital-intensive firms are responsible for 14% of the sales of all public manufacturing firms and for 40% of the 10% most capital intensive firms, those numbers were respectively 3% and 12% in 1980. Capital-intensive firms also appear to be persistent. In Figure A.10 of Appendix A.2, I plot the persistence rate of firms in the top 10%, 5% and 1% of the distribution of capital intensity within manufacturing. The persistence rate is the frequency with which a firm stays in its percentile over a 5 year window, conditional on it being listed consistently. For the top decile, the persistence rate is centered around 70% over the whole sample. The pattern is similar for the 95th percentile, despite a sharp decline below 60% around the financial crisis. The top percentile is more volatile around a mean of 57%.

The correspondence between capital intensity and size for the aggregate, trade and services are reported in Figure A.8 of Appendix A.2. In the aggregate, the correlation is positive but seems to not change with time. For trade, the correlation is also positive but smaller and the correspondence is s-shaped. For services, the correlation is null.

To understand the correlation visible in Figure 3, consider that output elasticities impact the marginal cost which in turn determines prices and hence demand. To see this, suppose that a firm produces using the following constant elasticity of substitution production

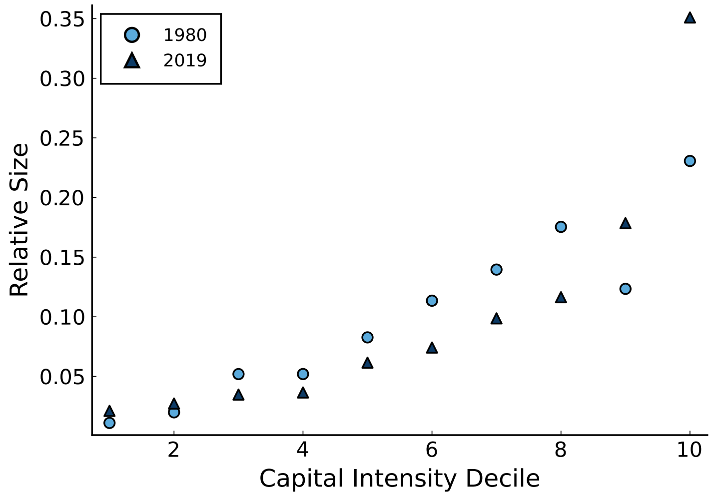


Figure 3: Relative size per capital intensity deciles in manufacturing.

Note: Source is Compustat. Sample is 1980-2019. Manufacturing is composed of 2D NAICS sectors 31, 32 and 33. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Relative size is the sales of a decile over total sales of the sector.

technology:[15]

$$Y_{i,t} = z_{i,t}\left(\alpha_{i,t}K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,t})L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}} \tag{2}$$

where $\alpha_{i,t}$ is the share of capital in production and ψ is the elasticity of substitution between capital and labor. In such a case, capital intensity is: $\frac{K_{i,t}}{L_{i,t}} = \left(\frac{\alpha_{i,t}}{1-\alpha_{i,t}}\frac{w_t}{r_t^K}\right)^{\psi}$ and the output elasticity to capital writes: $\zeta_{i,t}^K = \alpha_{i,t}\left(\frac{Y_{i,t}}{K_{i,t}}\right)^{\frac{1}{\psi}}$ so that α_{it} captures heterogeneity in ζ_{it}^K.[16] I plot estimates of percentiles of the distribution of $\alpha_{i,t}$ for the baseline $\psi = 0.5$ and alternative ψ in Appendix A.2.2, Figure A.20. Supposing that firms are cost-minimizing,

[15]In Appendix B.2, I allow for technology to feature heterogeneous labor-augmenting productivity instead of alpha, as advocated by Demirer (2020) for example. If workers are paid at their marginal productivity, this yields no heterogeneity in marginal costs and hence no correlation between capital intensity and size. In Appendix B.3, I show that if large firms have monopsonic power, my estimates for $\alpha_{i,t}$ are downward biased for the most capital-intensive firms, suggesting that dispersion might be even higher. Furthermore, Kehrig and Vincent (2021) show, using Census Data where firms' wages are observable (unlike in Compustat), that there is no significant difference between the wage paid by high- and low- labor share firms. In Appendix A.3.1, I show that firms' labor share and capital intensity correlate negatively, implying that differently capital-intensive firms don't charge differentiated wages.

[16]Note that the Cobb-Douglas case (equation (2) with $\psi \to 1$) is sufficient to capture heterogeneity in output elasticity to inputs. However, as shown in Karabarbounis and Neiman (2014) and Oberfield and Raval (2021), ψ plays a crucial role in the relationship between the dynamics of the RCK and the labor share. I therefore chose to root heterogeneity of ζ^K into α to be able to test my results under different values of ψ. In the baseline, I use the estimate of $\psi = 0.5$ from Oberfield and Raval (2021).

one obtains the following identity for the marginal cost:

$$mc_{i,t} = \frac{1}{z_{i,t}} \left({\alpha_{i,t}}^{\psi} {r_t^K}^{(1-\psi)} + (1-\alpha_{i,t})^{\psi} {w_t}^{(1-\psi)} \right)^{\frac{1}{1-\psi}} \tag{3}$$

From equation (3), a fall in the rental rate of capital diminishes the marginal cost of production and therefore prices and boosts demand. If firms feature heterogeneous α, the elasticity of their marginal cost to a fall in r^k is in turn heterogeneous; high-alpha firms (capital intensive) will see their marginal cost fall by more than low-alpha firms (labor intensive). In other words, in an environment where the RCK falls, heterogeneity in α implies heterogeneity in efficiency. Intuitively and anticipating on the fully-fledged model, if firms' marginal costs determine their prices which in turn affects their demand, most capital-intensive firms (higher α) are able to post lower prices than their labor-intensive competitors and hence to steal demand from them.

In order to understand what can microfound this heterogeneity, notice that α can be interpreted in two manners. First, α can be interpreted as normalized capital-biased productivity. It is capital-biased in that a rise in α increases the relative marginal product of K to L (Acemoglu, 2002; Jones, 2005; Peretto and Seater, 2013) and normalized in that capital-biased and labor-biased productivity are supposed to sum up to one.[17] Heterogeneity in capital intensity can then be interpreted as heterogeneous directed technical change (towards capital). Another way to interpret a firm's α comes from looking at task-based production functions pioneered by Zeira (1998). In such a case, $\alpha_{i,t}$ is the share of tasks that are performed by capital, as opposed to labor. It can rise if tasks are automated, as in Acemoglu and Restrepo (2022) or Hubmer and Restrepo (2025) or if new tasks performed by capital are added, as in Hémous and Olsen (2022).[18] In both cases (automation/new task and capital-biased productivity), α is a choice of the firm.

By equation (3) and Figure 3, it is advantageous for firms to increase the capital intensity

[17]Too see this, suppose that a firm produces with the following technology: $Y_{i,t} = A_{i,t}\left(A^K_{i,t} K_{i,t}^{\frac{\psi-1}{\psi}} + A^L_{i,t} L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$. One can rewrite this equation as equation (2) by defining $\alpha_{i,t} = \frac{A^K_{i,t}}{A^K_{i,t}+A^L_{i,t}}$ and factorizing out the non-homothetic terms: $z_{i,t} = A_{i,t}\left(A^K_{i,t} + A^L_{i,t}\right)^{\frac{\psi}{\psi-1}}$.

[18]Consider for example the following production technology: $Y_{i,t} = a_{i,t}\left(\int_0^\tau y_{i,t}(x)^{\frac{\psi-1}{\psi}} dx\right)^{\frac{\psi}{\psi-1}}$ where $y_{i,t}(x)$ is a task necessary to production. Each task can be produced using labor $L_{i,t}$ or capital $K_{i,t}$ such that a number $a^K_{i,t} \in [0,\tau]$ of tasks is performed by capital and a number $a^L_{i,t} = \tau - a^K_{i,t}$ is produced by labor. This allows to rewrite: $Y_{i,t} = a_{i,t}\left(a^K_{i,t} K_{i,t}^{\frac{\psi-1}{\psi}} + a^L_{i,t} L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$ Defining $\alpha_{i,t} = a^K_{i,t}/\tau$ yields: $Y_{i,t} = z_{i,t}\left(\alpha_{i,t} K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,t}) L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$ where $z_{i,t} = \tau^{\frac{\psi}{\psi-1}} a_{i,t}$. In such a case, automation appears if τ remains fixed but a^K increases, and the addition of more tasks performed by capital corresponds to an equal scalar increment in τ and in a^K.

of their technology (i.e, their alpha) as the relative cost of capital falls. The observed heterogeneity in α therefore suggests that firms differ in their capacity to become more capital intensive. To understand why, in the next section, I build a general equilibrium model where α is a choice of the firm.

To discipline the necessary "ingredients" of the model, in Appendix A.5.2, I document an extra stylized fact: capital-intensive firms also charge a significantly higher markup than their labor-intensive competitors. To the trained eye, this flows directly from the fact that capital-intensive firms are larger as the positive correlation between a firm's size and its markup is an already established relationship (see de Loecker et al. (2020) or de Ridder et al. (2022) for example).

3 General Equilibrium Model

Building on these empirical findings, I construct a General Equilibrium model to understand what drives the heterogeneity in firms' capital intensity and to derive the aggregate implications of firm-specific capital intensity. In this section, I focus on laying out the model, presenting the equilibrium, and detailing the calibration. In the result section 3, I present the aggregate implications of the model and compare simulated moments of the model to actual moments from the data.

3.1 Environment

Time is discrete. The supply side of the model has three layers. Final good producers aggregate from sectoral good producers who, in turn, aggregate intermediate goods from intermediate firms who compete in oligopolistic industries. Intermediate firms chose the capital intensity of their technology on top of their usual control variables (capital, labor, output, price). The household is representative and standard, he supplies labor, consumes and saves using both a bond and the capital good.

3.1.1 Final Good Producers

There are two final goods: one for consumption and one for investment in the capital stock. Each of the two goods is sold to the representative household and produced by two producers (one for each type of final good) representative of a continuum of perfectly competitive final good producers. Both producers aggregate sectoral goods using a Constant Elasticity of Substitution (CES) function and take sectoral goods' prices as given.

The final consumption good is produced using the technology:

$$C_t = \left(\int_0^1 C_{j,t}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} \tag{4}$$

While the final investment good is produced using the technology:

$$X_t = \frac{1}{\xi_t} \left(\int_0^1 X_{j,t}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} \tag{5}$$

Where σ is the elasticity of substitution of demand across sectoral goods. $C_{j,t}$ and $X_{j,t}$ denote, respectively, the quantity of inputs the final consumption-goods producer and the final investment-goods producer purchase to sector j. The scalar ξ denotes the relative productivity of the consumption producer with respect to the investment producer. If it decreases, the relative productivity of the investment good technology increases. It is supposed to vary with time and will capture capital good producers becoming more productive.

Both final good producers minimize their total cost and maximize their profits subject to their production function to choose their optimal input mix and output. This yields the following pricing function:

$$p_t^X = \xi_t \left(\int_0^1 p_{j,t}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} = \xi_t p_t \tag{6}$$

Where p_t^X is the price of the final investment good, p_t is the price of the final consumption good and $p_{j,t}$ is the price of the good produced by sector j. Final good producers take aggregate demand from households as given: $Y_t := C_t + \xi_t X_t$ and $Y_{j,t} = C_{j,t} + X_{j,t}$, so the inverse demand functions for the good of individual sectors is given by:

$$p_{j,t} = \left(\frac{Y_{j,t}}{Y_t} \right)^{-\frac{1}{\sigma}} p_t \tag{7}$$

3.1.2 Sectoral Good Producers

Sectoral good producers aggregate from intermediate firms and sell to final good producers. They are perfectly competitive and the variety they supply is lowly substitutable (low σ). The intermediate firms to which they buy their inputs are not atomistic. Output in each sector j writes:

$$Y_{j,t} = \left(\sum_{i=1}^{N} Y_{i,j,t}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{8}$$

Where $Y_{i,j,t}$ is output of firm i in sector j, N is the number of firms present in sector j and ϵ is the elasticity of substitution of demand across different varieties within the sectors. I suppose that varieties within a sector are more substitutable than across: $\epsilon > \sigma$. This yields the price aggregator for sector j in time t, $p_{j,t}$:

$$p_{j,t} = \left(\sum_{i=1}^{N} {p_{i,j,t}}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \tag{9}$$

And the inverse demand functions for varieties within sectors:

$$p_{i,j,t} = \left(\frac{Y_{i,j,t}}{Y_{j,t}} \right)^{-\frac{1}{\epsilon}} p_{j,t} \tag{10}$$

3.1.3 Intermediate Good Producers

In each sector, an integer number of oligopolistic firms use labor and capital to produce output using a CES production function. They compete in a static differentiated Bertrand or Cournot game to set their prices and quantities. Firms also chose their technology. They start with common capital intensity, α, but can increase it by paying a sunk switching cost. I first present the quantities and price problems of the firm, conditional on its technology, and then move to its technology choice.

Quantities and price. Consistently with the equation that guided the empirical exercise, firms use a CES production function with heterogeneous and time-dependent capital-share in production, α. Namely:

$$Y_{i,j,t} = \left(\alpha_{i,j,t} {K_{i,j,t}}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,j,t}) {L_{i,j,t}}^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}} \tag{11}$$

Where ψ is the elasticity of substitution between capital (K) and labor (L). Conditional on technology (8), and supposing that firms are price-taker on input markets where r_t^K is the rental rate of capital and w_t is the rental rate of labor, cost minimization yields the following marginal cost and implicit input demand functions:

$$mc_{i,j,t} = \left({\alpha_{i,j,t}}^{\psi} {r_t^K}^{(1-\psi)} + (1-\alpha_{i,j,t})^{\psi} {w_t}^{(1-\psi)} \right)^{\frac{1}{1-\psi}} \tag{12}$$

$$r_t^K = \alpha_{i,j,t} mc_{i,j,t} \left(\frac{Y_{i,j,t}}{K_{i,j,t}} \right)^{\frac{1}{\psi}} \tag{13}$$

$$w_t = (1-\alpha_{i,j,t}) mc_{i,j,t} \left(\frac{Y_{i,j,t}}{L_{i,j,t}} \right)^{\frac{1}{\psi}} \tag{14}$$

Taking the ratio of Equations (13) and (14) yields the capital intensity identity already seen in equation (1):

$$\frac{K_{i,j,t}}{L_{i,j,t}} = \left(\frac{\alpha_{i,j,t}}{1-\alpha_{i,j,t}}\frac{w_t}{r_t^K}\right)^{\psi}$$

capital intensity is increasing in $\alpha_{i,j,t}$ and decreasing in the cost of capital relative to that of labor $\frac{r_t^K}{w_t}$, i.e, in the inverse of the Relative Cost of Capital (RCK).

On top of quantities, firms also have freedom in setting their prices because they sell differentiated products and are non-atomistic. Specifically, the oligopolistic structure of Atkeson and Burstein (2008) implies that static profit maximization of intermediate firms subject to their individual inverse demand curve (equation (19)) yields:

$$p_{i,j,t} = \frac{\varepsilon(s_{i,j,t})}{\varepsilon(s_{i,j,t})-1} mc_{i,j,t} = \mu(s_{i,j,t}) mc_{i,j,t} \tag{15}$$

Where $s_{i,j,t} = \frac{p_{i,j,t}Y_{i,j,t}}{p_{j,t}Y_{j,t}}$ is the sales share of a firm in its sector (or, loosely speaking, its market share).[19] $\varepsilon(s_{i,j,t})$ is the elasticity of demand to the price posted by firm i in sector j at time t. Importantly, $\varepsilon(s_{i,j,t}) = -\frac{\partial \ln(Y_{i,j,t})}{\partial \ln(p_{i,j,t})} \neq \epsilon$ because, in an oligopolistic organization, a firm internalizes the impact of its individual pricing on the aggregate price and hence on the distribution of demand. Intuitively, an efficient firm faces a smaller limit price (equal to its marginal cost) than its competitors. It can chose to charge a markup over its marginal cost while still posting a lower price and hence attracting the highest chunk of demand. The key difference with a classic monopolistic competition setting *à la* Dixit and Stiglitz (1977) is that, as firms are non-atomistic, they recognize this possibility.[20] This game can be solved as a Bertrand or Cournot static differentiated game yielding the following formulae for the elasticities of demand to prices:

$$\varepsilon(s_{i,j,t}) = \begin{cases} \epsilon - (\epsilon - \sigma)s_{i,j,t} & \text{under Bertrand competition} \\ \left(\frac{1}{\epsilon} - (\frac{1}{\epsilon} - \frac{1}{\sigma})s_{i,j,t}\right)^{-1} & \text{under Cournot competition} \end{cases} \tag{16}$$

In both cases, a firm's markup is increasing and convex in its market share and depends on the elasticity of substitution of demand across varieties (ϵ and σ). Finally, individual accounting profits can be written:

$$\pi_{i,j,t} = \left(1 - \frac{1}{\mu(s_{i,j,t})}\right) p_{i,j,t}Y_{i,j,t} \tag{17}$$

[19] I also define $s_{i,t} = s_{i,j,t}s_{j,t}$ for $s_{i,t} = \frac{p_{i,j,t}Y_{i,j,t}}{p_tY_t}$ and $s_{j,t} = \frac{p_{j,t}Y_{j,t}}{p_tY_t}$.

[20] An alternative possibility would be to model imperfect competition with a Kimball aggregator for demand, as shown by Edmond et al. (2023). I chose the oligopolistic model for its clear microfoundations that allow me to analyze the role of strategic innteractions.

Technology. There are two drivers to firms' heterogeneous choice of $\alpha_{i,j,t}$: an exogenous component that reflects a firm's "capability" to update its technology and an endogenous component that comes from the strategic interactions induced by oligopolistic competition.

For the exogenous component, suppose that firms can pay a cost to update their production technology. This cost increases in the distance between the technology with which a firm starts the period (summarized by $\alpha_{i,j,t-1}$) and its targeted technology ($\alpha_{i,j,t}$). I hence refer to it as a switching cost. I also suppose that once a firm has paid to obtain a certain $\alpha_{i,j,t}$, it becomes available in subsequent periods. The switching cost is hence a sunk cost. Suppose further that firms differ in the slope of their switching cost function, i.e, that some firms need to pay a higher cost than others for a same change of α.[21] I denote this slope as $\varphi_{i,j,t}$.[22] The switching cost function faced by a firm i in sector j and at time t is hence defined:

$$sc_{i,j,t}(\Delta\alpha_{i,j,t}) = \varphi_{i,j,t}\left(\alpha_{i,j,t} - \alpha_{i,j,t-1}\right) p_t Y_t \tag{18}$$

Where $\Delta\alpha_{i,j,t} = \alpha_{i,j,t} - \alpha_{i,j,t-1}$, and $p_t Y_t$ scales the cost with nominal GDP to avoid making the switching cost negligible as the economy grows.[23] This exogenous heterogeneity could represent a variety of realities. Possible micro-foundations include: differing capacities to re-organize production when firms change their technology, in the spirit of Aghion and Tirole (1997), heterogeneous entrepreneurial or managerial capabilities, as in Bresnahan, Brynjolfsson, and Hitt (2002) or Bloom, Sadun, and Van Reenen (2012), mismatches between the skills necessary to implement new vintages of capital and those available to the firm, as in Chari and Hopenhayn (1991), or informational frictions due for example to managers or employees not learning about the availability of alternative technologies as in Lucas and Moll (2014). In this model, I do not distinguish between these possibilities but rather use $\varphi_{i,j,t}$ as a way to capture them in a stylized manner in order to distinguish them from the second determinant of firms' choice of α: strategic interactions.

To understand how this second determinant of firms' choice of α appears, consider that an intermediate firm decides on its $\alpha_{i,j,t}$ at the beginning of a period, after observing the productivity of the capital good producer, ξ_t, and given its $\alpha_{i,j,t-1}$ as well as the $\alpha_{-i,j,t-1}$ of its competitors. A firm chooses its optimal α by maximizing the present value of its current and future pure economic profits defined as its accounting profit minus the sunk cost of updating its technology.[24]:

[21] Alternatively, one could suppose that firms face the same cost but heterogeneous returns in terms of α.

[22] For now, I suppose that the slope of the switching cost function can be time-specific and hence index it with t. This is not strictly necessary as I discuss in the calibration section.

[23] Importantly, $p_t Y_t$ is considered as given by the firm (unlike $p_{j,t} Y_{j,t}$) because sectors are atomistic while firms aren't. This is essential to avoid the firm facing an Euler equation and the oligopolistic problem to become dynamic and hence impossible to solve without extra limiting assumptions (Mongey, 2022; Wang and Werning, 2022).

[24] There is no entry, which I justify in section 4.3, the number of firms is finite and goods are differentiated

$$\max_{\alpha_{i,j,t}} \sum_{\tau=0}^{+\infty} \beta^{\tau} \pi_{i,j,t+\tau} - sc_{i,j,t}(\Delta\alpha_{i,j,t})$$

This problem is subject to the implied marginal cost (equation (12)), the demand curve (equation (19)), the optimal pricing rule (equations (15) and (16)), the profit identity (equation (17)) as well as a the functional form of the switching cot function (equation (18)). To account for sectoral differences in the distribution of capital intensity, it is also useful to constrain the sectoral support of α by supposing a sector-specific technological frontier: $\alpha_{i,j,t} \leq \overline{\alpha}_{j,t}$. The impact on permanent profits of a change in α flows through the marginal cost first $\frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}$, which affects the market share $\frac{\partial s_{i,j,t}}{\partial mc_{i,j,t}}$ which in turn affects the markup $\frac{\partial \mu(s_{i,j,t})}{\partial s_{i,j,t}}$. Solving for this problem yields the following implicit optimality condition for α (complete algebra is refereed in Appendix B.6):

$$\underbrace{\frac{1}{1-\beta}\frac{\partial s_{i,t}}{\partial \alpha_{i,j,t}}\left(\overbrace{1-\frac{1}{\mu(s_{i,j,t})}}^{\text{Volume effect}}+\overbrace{\frac{\partial \mu(s_{i,j,t})}{\partial s_{i,j,t}}\frac{s_{i,j,t}}{\mu(s_{i,j,t})^2}}^{\text{Markup effect}}\right)}_{\Xi_{i,j,t}} = \varphi_{i,j,t} \tag{19}$$

Where the right-hand side is the exogenous determinant of a firm's optimal α while the left-hand side contains the endogenous determinant of a firm's α. The left-hand side of equation (19) is the marginal return to increasing α, that I will refer to as $\Xi_{i,j,t}$. This component can be broken down in different components. $\frac{1}{1-\beta}$ comes from the fact that firms' new alpha is available for all subsequent periods. $\frac{\partial s_{i,t}}{\partial \alpha_{i,j,t}}$ refers to how a firm's update of α impacts its market share. A change in a firms' market share has in turn two effects. First, a volume effect: absent a change in markups, higher market shares imply higher profits. Second, a markup effect, which is central: as a firm becomes larger, its markup increases. Because markups are increasing and convex in market shares, the size of the markup effect depends on the original market share of a firm.

This means that, already capital-intensive firms are large (because they are more efficient) and hence stand on the steeper tangent line to the markup function than a small/labor-intensive firm. This implies that for a same $\Delta\alpha$, a larger firm secures a higher marginal return than a smaller firm, driven by a higher markup increase. This markup effect, or strategic interaction effect, also drives firms' updating of α. Already capital-intensive firms are large and hence update their α more due to the implied high markup effect. On the contrary, labor-intensive firms are small and so is the marginal gain from increasing their α, leading to a more modest updating. Furthermore, small firms are discouraged to chose a large $\Delta\alpha$ to catch-up with capital-intensive firms because this would decrease the market share, and hence the markup, associated to high capital intensity. This can create inertia

so pure economic profits > 0 are feasible.

for the firms endowed with a high switching cost as they are discouraged to increase their capital intensity due to a marginal return too low to cover the sunk cost.

In the result section dedicated to firm dynamics, I show the relative quantitative role of these two drivers of firms' choice of α by performing a numerical comparative exercise. In the counterfactual section, I propose to relax the hypothesis of markup endogeneity to understand how differences in firms capabilities to update their technologies, alone, affect the predictions of the model.

3.1.4 The Household

An infinitely-lived representative household enjoys consumption and dislikes labor. He buys the final goods and rents out his stock of labor and capital to intermediate firms that he owns. He has access to a bond that pays out $1 + r_t$ in the next period. He discounts the future at a constant rate β. His life-time utility reads:

$$\sum_{t=0}^{+\infty} \beta^t U(C_t, L_t)$$

Where $U(C_t, L_t)$ is the utility function of the household that is increasing and concave in consumption and decreasing and concave in labor. He maximizes it subject to an initial stock of capital K_0 and bonds B_0, a capital accumulation equation that reads $K_{t+1} = (1-\delta)K_t + X_t$, where δ is the depreciation rate of capital, and a budget constraint that reads:

$$p_t(C_t + \xi_t X_t) + B_{t+1} = (1 + r_t)B_t + w_t L_t + r_t^K K_t + \pi_t \tag{20}$$

Where $K_t = \int_0^1 \sum_{i=0}^N K_{i,j,t} dj$, $L_t = \int_0^1 \sum_{i=0}^N L_{i,j,t} dj$ and $\pi_t = \int_0^1 \sum_{i=0}^N \pi^{i,j,t} dj$. His intra-temporal optimality condition is:

$$-\frac{U_{L_t}(C_t, L_t)}{U_{C_t}(C_t, L_t)} = \frac{w_t}{p_t} \tag{21}$$

Where $U_{L_t}(C_t, L_t) = \frac{\partial U(C_t, L_t)}{\partial L_t}$ and $U_{C_t}(C_t, L_t) = \frac{\partial U(C_t, L_t)}{\partial C_t}$. His Euler equation is:

$$\frac{U_{C_t}(C_t, L_t)}{U_{C_{t+1}}(C_t, L_t)} = \beta(1 + r_{t+1}) \frac{p_t}{p_{t+1}^C} \tag{22}$$

And a Hall and Jorgenson (1967) no-arbitrage condition that reads:

$$r_{t+1}^K = p_t^X (1 + r_{t+1}) - p_{t+1}^X (1 - \delta) \tag{23}$$

That is, the return on renting capital must be equal to that of buying a bond corrected by the rate of depreciation of the capital stock.

3.2 Equilibrium

An equilibrium in time t, is a steady-state that is defined, given the parameters β, δ, ϵ, σ, ξ_t, the wage (normalized to 1), a per-sector distribution of $\alpha_{i,j,t-1}$ and of frontiers $\underline{\alpha}_j, \overline{\alpha}_j$, a distribution of the switching cost parameter $\varphi_{i,j,t}$ per sector and time and a functional form for the utility function, by the solved system of equations (4) to (23) for which every variable is constant (conditional on ξ_t). Equations (4) to (23) imply that the household maximizes his utility, final and sectoral producers minimize their costs, intermediate producers maximize their pure economic profits, all markets clear and every constraint is satisfied.

As is well known, the non-linearity of the sub-system composed of equations (19), (15), (16) and (19) implies that the model must be solved numerically.

Computations and numerical implementations to solve for the equilibrium are provided in Appendix B.4 and B.5.

3.3 Calibration

There are two manners to calibrate the model: I can either target the distribution of firms' capital intensity for every period, or only for the first period and let the model determine the following ones. In the main text, I focus on the one with the highest degree of freedom, i.e, where I target the distribution of capital intensity for every period. In Appendix B.8.1, I detail the calibration where only the distribution of capital intensity at the beginning of the sample is targeted and comment on the differences in terms of results.

The model is solved at the annual frequency. There are two exercises to be undertaken. First, calibrating parameters necessary to solve equations (4) to (23) without the technological choice of the firm -equations (18) and (19). Once this is done, I solve numerically and simulate the model without the endogenous updating of alpha but rather feeding the model with the actual distribution of α estimated in Appendix section A.2.2. The results are then used to estimate the cost of updating α that prevails at each period to retrieve a distribution of $\varphi_{i,j,t}$ for each sector j and time t. The model can then be simulated again with equations (18) and (19) joining the system of equations to be solved numerically.

I choose the Greenwood, Hercowitz, and Huffman (1988) utility function to wash away wealth effects on labor supply:

$$U(C_t, L_t) = \frac{\left(C_t - \frac{L_t^{1+\eta}}{1+\eta}\right)^{1-\gamma}}{1-\gamma}$$

Where γ is the coefficient of relative risk aversion and η the inverse Frisch elasticity. Then, the following parameters need to be set: the discount rate β, the depreciation rate of capital δ, the sequence of relative productivity of the final investment good producer to the final consumption good producer for every period: $\{\xi_t, \forall t\}$, the micro elasticity of substitution between capital and labor ψ, the elasticity of substitution between goods within markets, ϵ and across markets, σ, and the distribution of firm-specific cost parameter for the updating of α for each sector and time: $\{\varphi_{i,j,t}, \forall i, j, t\}$ as well as sector specific technological frontiers: $\overline{\alpha}_{j,t}$.

Parameters to be calibrated are listed in Table 1 together with their value and the corresponding moment or source. I set $\beta = 0.95$ to obtain a steady-state interest rate of 0.05, similarly to the one assumed in the empirical section. I set δ to 0.1, its yearly average according to the estimation of Gutiérrez and Philippon (2016). I don't consider its variations over time because they are minimal and would only add noise to the model. The dynamics of ξ_t traces that of the relative cost of capital documented in Figure A.12 of Appendix section A.2. Its initial value is set so that the aggregate estimated ζ^K in section 3 is equal to the one found if one uses an Olley and Pakes (1996) or Ackerberg, Caves, and Frazer (2015) method to estimate it: 0.12. The micro elasticity of substitution between labor and capital, ψ, is set to 0.5, as suggested by Oberfield and Raval (2021).

I set $\sigma = 1.01$ to make sectors nor complements nor substitutes and avoid composition effects across sectors to influence the results (Like Atkeson and Burstein (2008), I deviate from $\sigma = 1$ because the model is not well behaved in this case). I set $\epsilon = 10$ to match the fact described in Antras (2004) or Oberfield and Raval (2021) that the aggregate elasticity of substitution is also below one but close to Cobb-Douglas. To do so, I build an aggregate CES-CRS production function that reads:

$$Y_t = \left(\alpha K_t^{\frac{\theta-1}{\theta}} + (1-\alpha) L_t^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}$$

Which yields the aggregate capital intensity equation in logarithm:

$$\ln\Big(\frac{K_t}{L_t}\Big) = \theta\Bigg(\ln\Big(\frac{\alpha}{(1-\alpha)}\Big) + \ln\Big(\frac{w_t}{r_t^K}\Big)\Bigg)$$

I simulate my model for different levels of ϵ over a fixed distribution (the one of 1980) of α to obtain a change in the log of aggregate capital intensity purified of changes in α, i.e: $\Delta \ln(\frac{K_t}{L_t}) = \theta \Delta \ln(\frac{w_t}{r_t^K})$. An $\epsilon = 10$ yields an aggregate elasticity of substitution of 0.86.[25]

To build the sectors of the nested CES, I follow the standard procedure (see Burstein et al. (2020) or de Loecker et al. (2021)). I suppose 100 sectors with 10 firms each. The number 10 is chosen to match the aggregate markup in 1980 estimated as the harmonic sales-weighted average of firm-level markups in Appendix section A.5.2.[26] Each sector j is endowed with a

[25] Any $\epsilon \in [6, 12]$ satisfies this condition but $\epsilon = 10$ is common in the literature: Atkeson and Burstein (2008), Grassi (2017), de Loecker et al. (2021), Atkeson, Heathcote, and Perri (2022).

[26] I prove that this is the correct equation for aggregation in equation (37) in Section 4.2. I take the freedom of using this equation without having defined it yet as it is already a well established result (Grassi,

Symbol	Name	Value	Source/Moment
β	Discount rate	0.95	
η	Inverse Frisch elas.	2	Chetty, Guren, Manoli, and Weber (2011)
γ	Coeff. relative risk aversion	0.5	
ξ_t	Relative productivity of K producer	[1:0.4]	Figure A.12
δ	Depreciation rate	0.1	Gutiérrez and Philippon (2016)
ψ	Elas. of substitution between K and L	0.5	Oberfield and Raval (2021)
$\varphi_{i,j,t}$	Firm-specific switching-cost slope		$K_{i,j,t}/L_{i,j,t}$ (Eq. (10))
$\underline{\alpha}_j, \overline{\alpha_j}$	Technological frontier		$\min(K_{i,j,t}/L_{i,j,t}), \max(K_{i,j,t}/L_{i,j,t}), \forall i, t$
ϵ	Elas. of substitution of demand within sectors	10	$\theta = 0.86$
σ	Elas. of substitution of demand across sectors	1.01	Atkeson and Burstein (2008)
N	Number of firms within sectors	10	$\mathcal{M}_{1980}$

Table 1: Model Calibration.

technological frontier: $\alpha_{i,j,t} \in [\underline{\alpha}_j, \overline{\alpha}_j]$ where the boundaries are respectively the ones implied by the minimum and maximum capital intensity achieved within a sector along the whole sample. In the baseline model, every firm has TFP $Z = 1$. In the counterfactual and extensions (section 5), I'll include heterogeneous TFP to analyse the relative impact of TFP and structural capital intensity on my results.

To calibrate $\varphi_{i,j,t}$, each firm in each sector draws an $\alpha_{i,j,t}$ from a sector-year specific distribution. Those distributions are built to match the distributions of α in the 2D NAICS sector from the data that count more than 10 firms using Kernel density estimators. Typically, "Construction" (code 23) counts 21 firms out of 1095 in 1980. 2 of the 100 sectors will hence have their firms draw their α from the Kernel density of sector 23.

Once firms' α are calibrated, the model can be simulated. From the result of the simu-

2017; Edmond et al., 2023)

lation, firms' switching cost functions can be estimated. To estimate $\varphi_{i,j,t}$, the firm-specific slope of the switching cost, I simply rely on the optimality condition for $\alpha_{i,j,t}$; equation (19). That is, I compute $\Xi_{i,j,t}$ and use it to calibrate each firms' $\varphi_{i,j,t}$.

With firms' switching cost functions calibrated, the model can be simulated again with the parameters $\varphi_{i,j,t}$ as ex-ante parameters. Now, $\alpha_{i,j,t}$ is a control of the firm. Supposing Cournot competition, this boils down to solving numerically for $s_{i,j,t}$, $\mu(s_{i,j,t})$ and $\alpha_{i,j,t}$ the system:

$$\begin{cases} \mu_{i,j,t} = \frac{\varepsilon(s_t^{i,j})}{\varepsilon(s_t^{i,j})-1} \\ \varepsilon(s_t^i) = \left(\frac{1}{\epsilon} - (\frac{1}{\epsilon} - \frac{1}{\sigma})\left(\frac{\mu_{i,j,t} mc_{i,j,t}}{p_t^j}\right)^{1-\epsilon}\right)^{-1} \\ \Xi_{i,j,t} = \varphi_{i,j,t} \end{cases}$$

While in the previous step, only the two first equations where solved for, as $\alpha_{i,j,t}$ was a state variable to the firm. This modeling is useful for counterfactual analysis that will be undertaken in section 5. In the alternative calibration presented in Appendix B.8.1, I suppose that the distribution of φ is stationary across time and is calibrated on the first two years of my sample (1980 and 1981). In such a case, the results are remarkably similar, implying, as we shall see in the result section, that a change in the exogenous capabilities of firms to update their technology isn't the main driver of the fanning out of capital intensity.

4 Aggregation and Quantitative Results

This section displays the results of the model in terms of aggregation and quantification. I first show how firm dynamics materialize in section 4.1. Then, I present results regarding markups in terms of dispersion and aggregate. Third, I display the results for income shares. Fourth, I discuss capital misallocation and the aggregate investment gap. Fifth and lastly, I report results regarding differences in sales and employment concentration dynamics. In Appendix B.7, I detail the algebraic steps to reach the *formulae* displayed in this section.

4.1 Firm Dynamics

There are three drivers of firm dynamics when simulating the model between 1980 and 2019. First, dynamics are introduced in the model by the fall in the RCK driven by a rise in the productivity of the capital good producer. Second, firm heterogeneity originates in their differing cost function to changing their α. Finally, firm heterogeneity also resides in markup endogeneity to market shares which feed-backs to firms' choice of $\Delta\alpha$. In this section, I perform a numerical comparative statics exercise to decompose the relative strength of each of those components. To do so, I simulate my model alternatively shutting down the different channels of the model but keeping the same calibration for each of the variables. Then, I look at the evolution of the market shares associated to each decile of the distribution of firms' capital intensity between 1980 and 2019 for those different simulations. The resulting changes are reported in Figure 4.

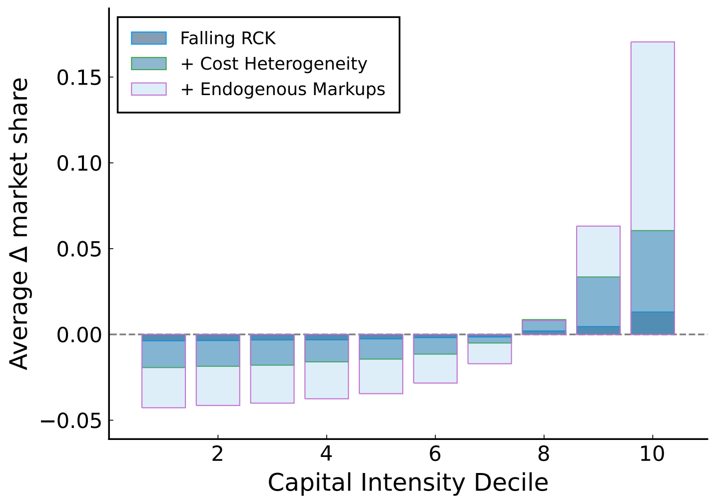


Figure 4: Market share dynamics per capital intensity decile: model comparative statics.
Notes: This graph plots the change in market shares between 1980 and 2019 associated with each decile of the distribution of firms' capital intensity from the simulated model. I include three specifications of the model. (i) "Falling RCK" is the model with a stationary distribution of firms' α calibrated on the distribution of capital intensity in 1980 and the baseline fall in the RCK. The endogenous updating of α channel and the endogenous markup channel are shut down. (ii) "+ Cost Heterogeneity" is the same model with the addition of firms being able to update their α but without the endogenous markup channel. The calibration of distribution of firms' switching cost is the baseline one. (iii) "+ Endogenous Markups" is the fully fledged model: the RCK falls, firms are endowed with different switching costs and markups are endogenous to sales shares.

The first exercise consists in calibrating the model on firms' heterogeneity in capital intensity in 1980 and calibrating the fall in the RCK to the baseline fall between 1980 and 2019. I shut down the other two channels: I suppose that firms cannot update their alpha and that markups are constant. In Figure 4, this corresponds to the "Falling RCK" bars (the dark-most one). In this case, as the RCK falls, all firms increase their capital intensity given their α. Because firms are endowed with different α, their marginal cost is differently elastic to a fall in the RCK (as per equation (12)). Namely, capital-intensive firms (high α) see their marginal cost decrease more than labor-intensive firms (low α) which allows capital-intensive firms to post lower prices and to steal demand from their labor-intensive competition. This translates into reallocation of market shares from labor-intensive towards capital-intensive firms with the latter gaining around 2pp of market shares.

The second exercise consists in allowing firms to update their α and introducing the heterogeneity in firms' switching costs without allowing for markup endogeneity.[27] As the

[27] Recall that this is a comparative statics exercise and not a counterfactual. The calibration of the distribution of the switching costs is the baseline one. A counterfactual exercise without markup endogeneity and where only the distribution of switching costs generates the full dispersion in capital intensity will be

RCK falls, all firms adopt the cheaper capital good, pushing up their capital-to-labor ratio. Now, firms also decide to update their technology (their α) in order to make their output more elastic to the cheaper input. Firms with lower switching costs update their technology more and see their capital-to-labor ratio rise faster than the aggregate. The dispersion of the elasticity of firms' marginal costs to the fall in the RCK is therefore amplified. This results into wider dispersion in firms' marginal costs, prices and hence market shares. This exercise corresponds to the "+ Cost Heterogeneity" in Figure 4. Qualitatively, the pattern is similar to when the distribution of α is stationary but the amplitudes are larger. This time, the most capital-intensive firms see their average market shares increase by 5pp.

The third exercise is the fully-fledged model. The mechanism is similar to the one in the previous exercise up to one extra sub-mechanism: strategic interactions. When firms have dispersed market shares, they stand at different points on their markup curves, which is increasing and convex in market shares. As such, the marginal return to increasing ones' market share by increasing α is higher for firms that are already capital-intensive as they are on a steeper segment of their markup curve. The difference in the intensity of the update of firms' technologies is hence magnified and α, market shares and markups disperse at a faster rate. As the model unveils and the RCK declines secularly, the fanning-out of the distribution of firms' α, market shares and markups accelerates. In such a case, represented by the "+ Endogenous Markups" (light-most bars) in Figure 4, the same qualitative pattern materializes compared to the two other exercises but amplitudes are magnified further. This time, the most capital-intensive firms see their market shares increase by an average of 15pp, in line with the amplitude documented in Figure 3 in the empirics section.

In Appendix B.8, I repeat this exercise with the alternative calibration: supposing that the distribution of switching costs is stationary along time. I find remarkably similar distributional patterns for capital intensity, market shares and markups. This implies that intrinsic differences across firms need not to have widened along this period despite rising concentration of market shares. Rather, the competitive advantage of firms that are able to update their technology more easily is magnified as the RCK falls and by imperfect competition.

4.2 Markups

As firms' market shares disperse in the model, so do markups. Here I show that this translates into a rising aggregate markup by deriving three aggregation identities from the model, one of which differs from the literature. The results are in line with the empirical estimates of Appendix section A.5.2 that build on de Loecker et al. (2020).

explored in the counterfactual section 5.

Firms' markups. Figure 5 plots the correspondence between the model-implied distribution of firms' capital intensity and firms' markups at the beginning of the calibration, in 1980, and the end of the calibration, in 2019. These moments are generated from the fully-fledged model on its baseline calibration. The change in the distribution of markups follows that of market shares seen in the previous section: highly capital-intensive firms witness the highest rise in markups while all firms strictly below the eighth decile see their markup fall. Overall, between 1980 and 2019, the variance of firms' markups increases by a factor of more than 4.

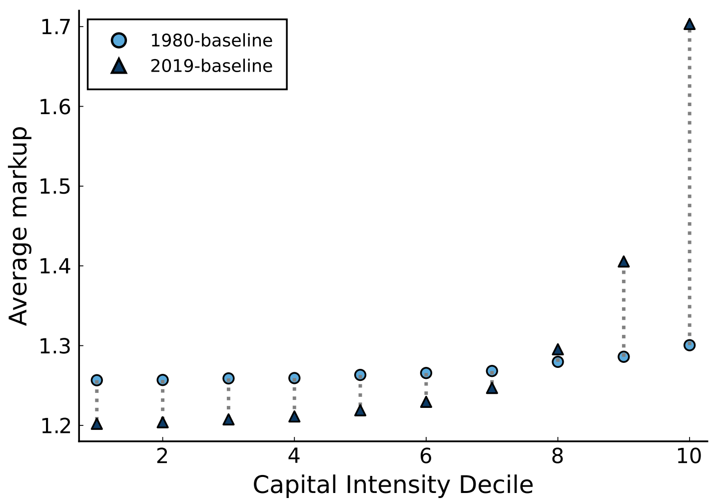


Figure 5: Markups dynamics per capital intensity decile in the model.

Notes: This graph plots the change in markups between 1980 and 2019 associated with each decile of the distribution of firms' capital intensity from the simulated baseline model. The "1980-baseline" distribution of firms' markups corresponds to the beginning of the model where the distribution of firms' α is calibrated on the observed distribution of firms capital intensity in 1980 and the RCK is set to its 1980 level. The "2019-baseline" distribution of firms' markups corresponds to the end of the model where the distribution of firms' α is calibrated on the observed distribution of firms capital intensity in 2019 and the RCK is set to its 2019 level.

Aggregate markup. The model implies three aggregation identities for the markup that I derive in Appendix B.7.2. In this section, I focus on two: one that is the most widely used in the literature and one that is new and corrects one sometimes used in the literature.

If aggregate returns to scale are constant (which I prove in Appendix B.7.1), then one can define an aggregate markup, $\mathcal{M}_t$, over the aggregate marginal cost mc_t, s.t: $p_t = \mathcal{M}_t mc_t$. This implies that aggregate profits can be written: $\pi_t = (1 - \frac{1}{\mathcal{M}_t})p_t Y_t$. From the market-clearing condition on profits: $\pi_t = \int_0^1 \sum_{i=0}^{N} \pi_{i,j,t} dj$, we get an aggregation formula common to monopolistic (see Edmond et al. (2015)) and oligopolistic economies (see Grassi (2017)):

$$\frac{1}{\mathcal{M}_t} = \int_0^1 s_{j,t} \sum_{i=0}^{N} \frac{s_{i,j,t}}{\mu(s_{i,j,t})} dj \tag{24}$$

The aggregate markup is the sales-share-weighted harmonic average of firm-level markups. Notice that concentration of sales depresses the ratio of $s_{i,j}$ to $\mu(s_{i,j})$ due to the convexity of μ in s, translating into a rising aggregate $\mathcal{M}$ (see Grassi (2017) for an analytical formula linking the aggregate markup to the HHI index).

One can also derive an aggregation identity for markups using input weights that spells:

$$\mathcal{M}_t = \int_0^1 \sum_{i=0}^{N} \frac{\mathcal{I}_{i,j,t}}{\mathcal{I}_t} \frac{\zeta_t^{\mathcal{I}}}{\zeta_{i,j,t}^{\mathcal{I}}} \mu_{i,j,t} dj \tag{25}$$

For $\mathcal{I} \in \{K, L\}$ and where $\zeta^{\mathcal{I}}$ denotes the cost share of input $\mathcal{I}$.[28] Here, the aggregate markup is a weighted arithmetic average of firms' markups where the weight is the product of a firm's input share times the inverse of the ratio of the firm's cost share of input $\mathcal{I}$ to the aggregate cost share of the same input. In models where firms' heterogeneity has a factor-neutral source, the aggregate markup can be written: $\mathcal{M}_t = \int_0^1 \sum_{i=0}^{N} \frac{\mathcal{I}_{i,j,t}}{\mathcal{I}_t} \mu_{i,j,t} dj$ where $\mathcal{I}$ is any input. Intuitively, this does not hold in my framework since the choice of input implies different weights, i.e: $\frac{K_{i,j,t}}{K_t} \neq \frac{L_{i,j,t}}{L_t}$ under heterogeneous α.[29]

These aggregation identities deliver an increase in the aggregate markup of 6.71 % between 1980 and 2019; from 1.26 to 1.35 which corresponds to half of the increase in the aggregate markup implied by de Loecker et al. (2020).[30]

4.3 Micro and Aggregate Income Shares

In this section, I present the results of the model in terms of income shares and show that they match the empirical regularities documented by Autor et al. (2020) and Kehrig and Vincent (2021). The fall in the aggregate labor share is due to a reallocation from firms with increasing labor shares towards firms with decreasing labor shares due to heterogeneous updating of alpha. This fall in the labor share is mostly due to a rise in the profit share.

4.3.1 The Labor Share

While aggregate income shares can be directly retrieved from the simulated model, it is useful to derive the analytics of their aggregation to decipher the different mechanisms at

[28] For example, considering L, $\zeta_{i,j,t}^{L} = (1-\alpha_{i,j,t})\left(\frac{Y_{i,j,t}}{L_{i,j,t}}\right)^{\frac{1-\psi}{\psi}} = \frac{w_t L_{i,j,t}}{TC_{i,j,t}}$. At the firm level, the cost share of input $\mathcal{I}$ and the output elasticity to input $\mathcal{I}$ coincide.

[29] In Desazars (2025), I show that supposing $\mathcal{M}_t = \int_0^1 \sum_{i=0}^{N} \frac{\mathcal{I}_{i,j,t}}{\mathcal{I}_t} \mu_{i,j,t} dj$ yields different results empirically depending on the input under consideration.

[30] Note that had capital-intensive firms been smaller in 1980, their rising market share at the expense of labor-intensive (bigger) firms would have depressed the markup. The minimum would have been reached when the distribution of firms' market shares converged to the uniform distribution, before picking up again.

play.

Firms' labor share To do so, consider the spelling of a firm's labor share. For a firm i in sector j, the micro labor share in time t, $\lambda^L_{i,j,t}$, writes:

$$\lambda^L_{i,j,t} = \frac{w_t L_{i,j,t}}{p_{i,j,t} Y_{i,j,t}} = \frac{1-\alpha_{i,j,t}}{\mu(s_{i,j,t})} \Big(\frac{Y_{i,j,t}}{L_{i,j,t}}\Big)^{\frac{1-\psi}{\psi}} \tag{26}$$

The firm-level labor share depends on three components that are all control variables of the firm.[31] Because those three components react differently to a fall in the RCK, the direction of firms' labor share dynamics is ex-ante ambiguous. The different effects are the following:

(i) A complementarity effect encapsulated in the component $(Y_{i,j,t}/L_{i,j,t})^{\frac{1-\psi}{\psi}}$. Because capital and labor are gross complements ($\psi < 1$), this component increases following a fall in the RCK for every firm. I.e, the capital-to-labor ratio does not increase enough to offset the decline in the RCK pushing firms' labor share up.

(ii) A markup effect represented by $\mu(s_{i,j,t})$. Capital-intensive firms see their markup increase while labor-intensive firms see their markup decrease. The increase in the micro labor share inherited from the complementarity effect is hence dampened (or reversed) for capital-intensive firms while it is reinforced for labor-intensive firms.

(iii) A technology updating effect that changes $\alpha_{i,j,t}$. When the RCK declines, firms seek to adjust their capital-biased productivity to be more exposed to the more abundant/cheaper input. Capital-intensive firms increase their α more because the net return for such an investment is higher for them due to both advantageous switching costs and strategic interactions. Because every firm chooses to increase its alpha, this puts downward pressure on the labor share at the benefit of the capital share of every firm.

Overall, consistently with Autor et al. (2020) and Kehrig and Vincent (2021), micro labor share dynamics are heterogeneous across firms. To show this, I stick to the breakdown of my distribution of firms in terms of deciles of capital intensity and plot the average labor share change for all of these deciles in Figure 6.

Top capital-intensive firms see their labor share fall while labor-intensive firms witness a rising labor share. This is because, for capital-intensive firms, the markup and technology updating effects dominate, leading to a rise in profit and capital income shares and a decline in the labor share. For labor-intensive firms, the complementarity and markup-loss effects dominate, resulting in an increase in their labor share. Because labor-intensive firms are more numerous than capital-intensive firms, when looking at the distribution of labor share

[31] In the Cobb-Douglas case, the firm level labor share boils down to: $\lambda^L_{i,j,t} \underset{\psi \to 1}{\longrightarrow} \frac{1-\alpha_{i,j,t}}{\mu(s_{i,j,t})}$. This case can be found in Kehrig and Vincent (2021) for example. In such a case, the level of α is different and coincides with the variable ζ^K defined in empirical section 3, as it captures on its own the elasticity of output to capital, that is, all the ex-RCK determining forces of capital intensity.

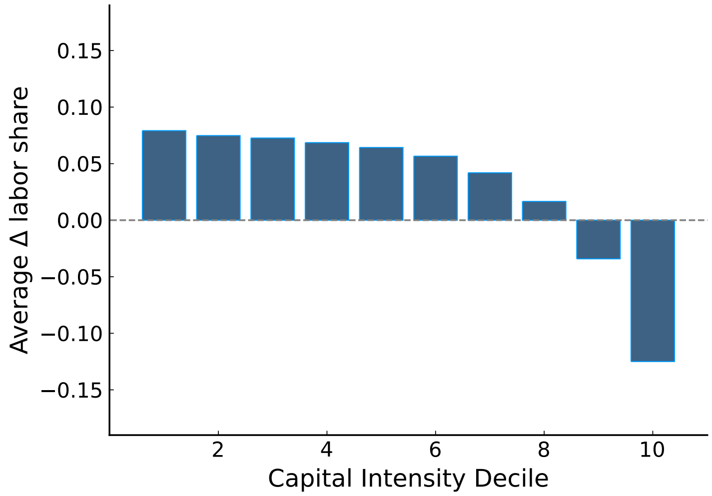


Figure 6: Labor share dynamics per capital intensity decile in the model.

Notes: This graph plots the average change in firms' labor share between 1980 and 2019 given their decile of capital intensity from the simulated baseline model. To obtain this result, firms are clustered into capital intensity deciles within their sectors. The change corresponds to within-sector deciles' labor shares in 2019 minus their labor shares in 1980. These changes are then averaged within deciles.

changes across the whole economy, the median labor share appears to increase by 2.11pp while the unweighted arithmetic average labor share declines by 0.05pp.

Aggregate labor share. Defining λ_t^L as the aggregate labor share, firms' labor shares can be aggregated as such:

$$\lambda_t^L = \frac{w_t L_t}{p_t Y_t} = \int_0^1 s_{j,t} \sum_{i=1}^{N} s_{i,j,t} \lambda_{i,j,t}^L dj \tag{27}$$

The finite sum aggregates the firm-level labor shares into sectoral labor shares and is the sales-weighted arithmetic average of firm-level labor shares. The integral proceeds from the same logic as it computes the sales-weighted arithmetic mean of sectoral labor-shares.[32] Overall, in the baseline model, a change in the aggregate labor share results either from a change in firm-level labor shares, from composition effects within or across sectors or from the interaction between these separate phenomena.[33]

In the model, the aggregate labor share declines by 5.59pp, which corresponds to slightly less than 70% of the actual decline of 8pp reported by the BLS. This aggregate fall is due to the cross-dynamics of firms market shares, plotted in Figure 4, and firms labor shares, plotted in Figure 6. To quantify how these cross-dynamics affect aggregate labor share dynamics, I

[32] Equations (26) and (27) are standard results detailed in Appendix B.7.

[33] Here, I have supposed only one sectoral nest, because, as we shall see, cross-sector reallocation is negligible, but equation (27) can be generalized to as many sectoral layers as wanted.

follow Kehrig and Vincent (2021) and turn to a shift-share decomposition with interaction on equation (27). Denoting ΔX the difference $X_{2019} - X_{1980}$ for any variable X, and $\overline{X}$ the 1980 value of variable X, one can decompose changes in the aggregate labor income share as follows:

$$\begin{aligned}\Delta\lambda = & \underbrace{\int \overline{s}_j \sum \overline{s}_{i,j} \Delta\lambda_{i,j}\, dj}_{\text{Shift}} \\ & + \underbrace{\int \overline{s}_j \sum \Delta s_{i,j} \overline{\lambda}_{i,j}\, dj}_{\text{Share-firms}} + \underbrace{\int \Delta s_j \sum \overline{s}_{i,j} \overline{\lambda}_{i,j}\, dj}_{\text{Share-sectors}} + \underbrace{\int \Delta s_j \sum \Delta s_{i,j} \overline{\lambda}_{i,j}\, dj}_{\text{Share-interaction}} \\ & + \underbrace{\int \overline{s}_j \sum \Delta s_{i,j} \Delta\lambda_{i,j}\, dj}_{\text{Interaction-firms}} + \underbrace{\int \Delta s_j \sum \overline{s}_{i,j} \Delta\lambda_{i,j}\, dj}_{\text{Interaction-sector\&firm}} + \underbrace{\int \Delta s_j \sum \Delta s_{i,j} \Delta\lambda_{i,j}\, dj}_{\text{Triple interaction}}\end{aligned} \tag{28}$$

Typically, the "Share-firms" component designates the effect of the change in firms' market share *ceteris paribus*. Instead, the "Interaction-firms" term represents how firms' simultaneous changes in market share and in labor share affect the aggregate labor share. Finally, the last term captures the composition of every change. Equation (28) can be directly used on the data simulated by my model, which is balanced since there is no entry and exit in the model.[34] The result of the shift-share decomposition is displayed in Table 2. In the first two columns, I report the decomposition of the change in the aggregate labor share implied by my model for the aggregate economy and for manufacturing. I include manufacturing in order to compare the model-implied moments to those of Kehrig and Vincent (2021) who perform the same decomposition on the manufacturing firms of the Census.

Because the fall in the aggregate labor share is, both in the data and in the model, mostly carried by the fall in the manufacturing labor share, the decomposition for manufacturing and the aggregate are qualitatively similar. In the model, the total decline in the aggregate and manufacturing labor shares are driven solely by one component and dampened by two others. The leading component of the decline is the one outlined empirically by Kehrig and Vincent (2021): the joint dynamics of firms' market and labor shares. In other words, the decline of the aggregate labor share is mainly driven by the reallocation of market shares from firms who saw their labor share increase towards firms who saw their labor share decrease. In the model, as shown in Figure 6, falling-labor-share firms are the most capital-intensive firms in their sector for whom the markup and technology updating effects dominate the complementarity effect. Crucially, and as outlined by Kehrig and Vincent (2021), reallocation alone actually drives the labor share up when controlling for interaction (the share component). Similarly, the sole changes in firm-level labor shares (the shift component) also takes the labor share up. Note that in the model, there is negligible cross sectoral reallocation because sectoral

[34] I don't model entry and exit because, as shown by Autor et al. (2020), entry and exit play a negligible role in the decline of the aggregate labor share.

	Model		Kehrig and Vincent (2021)
	Aggregate	Manufacturing	Manufacturing
Shift			
Total Shift	+1.24pp	+2.35pp	+3.5pp
Share			
Share-firms	+0.52pp	+1.67pp	
Share-sectors	−0.05pp	0pp	
Share-interaction	+0.04pp	0pp	
Total Share	+0.51pp	+1.67pp	+2.9pp
Interaction			
Interaction-firms	−7.2pp	−20.62pp	
Interaction-sector&firm	−0.04pp	−0.13pp	
Triple interaction	−0.1pp	−0.14pp	
Total Interaction	−7.34pp	−20.89pp	−26.4pp
Total	−5.59pp	−16.87pp	−20pp

Table 2: Decomposition of the decline of the aggregate labor share: Model and Data.

Note: This table contains the results of the shift-share decomposition of the decline in the aggregate labor income share presented in equation (28). The Model column corresponds to the data simulated from my model for the universe of all firms or only manufacturing firms. The column Kehrig and Vincent (2021) contains the empirical decomposition made by these authors on the manufacturing firms contained in the establishment-level CMF database of the US Census Bureau. Changes are changes between 2019 values and 1980 values for the Model and 1982 and 2012 for Kehrig and Vincent (2021).

composition effects are impeached by a low elasticity of substitution of demand across sectors ($\sigma = 1.01$).

4.3.2 Capital and Profit Shares

Because market shares, markups and technology are drivers of firms' labor shares and are endogenous to the RCK, whether the decline in the labor-share is driven by an increase in the aggregate capital or profit share is ex-ante ambiguous.[35] Overall, in aggregate, a 5.59pp

[35]Note that this is not true if the driver of the decline in the labor share is factor-neutral, i.e if it does not affect capital intensity. This is typically the case in a model that would study the effect of heterogeneous TFP growth or changing competition intensity on the labor share. In such a case, only a rise in the profit share can generate a fall in the labor share. Conversely, a model of automation without endogenous markups would only let changes in the capital share explain changes in the labor share.

decline in the labor share is passed to a 0.39pp increase in the capital share and a 5.20pp increase in the profit share as shown in the first row of Table 3. For manufacturing, a fall of 16.87pp in the sectoral labor share is explained by a rise of 4.32pp in the capital share and of 12.55pp in the profit share, as shown in the second row of Table 3.

Sector	Δ Labor share	Δ Capital share	Δ Profit share
Aggregate	−5.59pp	+0.39pp	+5.20pp
Manufacturing	−16.87pp	+4.32pp	+12.55pp

Table 3: Changes in aggregate income shares in the model: Aggregate and Manufacturing.
Note: These results correspond to the data simulated from my model for the universe of all firms or only manufacturing firms.

The rise in the aggregate capital share is small, because it is affected by two counterfeiting forces. On the downward pressure side; first, capital and labor are complementary, therefore a decline in the RCK takes the firm-level capital share down. Second, capital-intensive firms have structurally the highest capital shares but see their profit shares rise, taking down their capital shares. On the upward pressure side; First, reallocation towards capital-intensive firms implies reallocation towards high capital-share firms. Second, endogenous choice of alpha leads every firm to become more capital intensive yielding a generalized rise in capital shares. The two main drivers for the capital share are the complementarity effect and the endogenous alpha updating effects which results in the relative stability of the capital share.[36] For manufacturing, the updating of α is more aggressive for top-firms, consistently with the stylized fact presented in Figure 2, and reallocation is stronger, consistently with the stylized fact presented in Figure 3. This yields a more pronounced increase in the capital share.

For both the aggregate and manufacturing, the fall in the labor share is therefore mostly due to a rise in the profit share. Firms that update their technology the most gain market shares and build markups, pushing up the aggregate markup and down the aggregate labor share.[37] Heterogeneous ability of firms to update their α therefore allows to account for an important potion of the fall in the aggregate labor share while respecting the empirical

[36]To see this, in a counterfactual exercise in Appendix B.9.3, I simulate a model where the distribution of α is the same for every period (I calibrate it to the one prevailing in 1980). In this case, the domination of the complementarity effect is clear as the labor share increases while the capital share plunges. The increase in the capital share in the baseline model with technological updating hence results from introducing endogenous α adjustments. This is consistent with Oberfield and Raval (2021) who show that aggregate elasticity of substitution between capital and labor is a convex combination of firm-level elasticity of substitution between capital and labor and reallocation. They estimate that firm-level elasticity of substitution between capital and labor is below 1 and confirm the finding of Antras (2004) that reallocation is not big enough to make capital and labor substitutes in the aggregate. In my baseline model, this is overcome by the fact that firms can update their α.

[37]Note that the profit share is the accounting profit share, that gathers both pure economic profits and sunk costs. In Appendix B.8, Figure B.4, I plot the evolution of the share of sunk costs in accounting profits per capital intensity deciles. For top and bottom firms, the share of sunk costs decreases. For top firms this comes from comparative growth between accounting profits that are taken up by convex markups while

regularities detailed by Autor et al. (2020) and Kehrig and Vincent (2021). Similarly to Rognlie (2015) and unlike Barkai (2020), the model predicts a stable capital share. The difference between this calibration and Barkai's is that the model does not include the fall in the interest rate. I discuss this case as an extension in the last section of the paper.

The predictions of the baseline model can be interpreted as a basic synthesis of the debate regarding the labor share. Among the many possible drivers of the decline in the aggregate labor share, two main arguments have been traditionally opposed: capital-deepening, which implies a rising capital share, or rising markups, which posit a rising profit share. My model implies that both channels play together as capital intensity microfounds markups. This joint mechanism matters quantitatively for the aggregate capital-to-labor ratio: a capital-only account requires an unrealistically large surge in the aggregate capital-to-labor ratio, whereas a profits-only account implies a mild rise in the aggregate capital-to-labor ratio. I discuss this comparison in Appendix B.8.2.

4.4 Capital misallocation and the Investment gap

The model implies that firms have dispersed marginal returns to the productivity of capital ($MRPK$), that is, in the words of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) that capital is "misallocated". Given its technology (i.e its α), a firm's marginal return to the productivity of capital can be written:

$$MRPK_{i,j,t} = \mu(s_{i,j,t}) r_t^K$$

Since input prices are supposed homogeneous in the model, the implied distribution of $MRPK_{i,j,t}$ coincides exactly with that of markups reported in Figure 5, as in Baqaee and Farhi (2020) or Edmond et al. (2023). As markups disperse along time, so do MRPKs, implying higher capital misallocation, in line with David and Venkateswaran (2019) and Andrei, Mann, and Moyen (2019).

Rising dispersion in firms' MRPKs in turn affect the aggregate efficiency of the economy. In appendix B.7.4, I show that, even if firms have heterogeneous output elasticity to capital, one can obtain an aggregate $MRPK$ that reads:

$$MRPK_t = \left(\int_0^1 s_{j,t} \sum_{i=1}^{N} \frac{s_{i,j,t}}{MRPK_{i,j,t}} dj \right)^{-1} \tag{29}$$

That is, the aggregate marginal return to the productivity of capital is the inverse of the sales-weighted harmonic average of firms' $MRPK_{i,j,t}$. This result implies that: $MRPK_t / r_t^K = \mathcal{M}_t$. In other words, as the aggregate markup rises, the aggregate econ-

the sunk cost is linear. For the bottom firms, it results from the fact that they increase their alpha less as market shares disperse because the marginal return to updating falls.

omy is pushed away from its efficient frontier defined by: $MRPK_t = r_t^K$.

This aggregation result relates to the empirical finding documented in Abel et al. (1989), Gomme et al. (2011) Gutiérrez and Philippon (2016), and Philippon (2019) who show that the Net Operating Return to Capital ($NORK$), that is defined as the ratio between net operating profits and capital expenditures, has been rising with time. This fact is often referred to as a growing "investment gap" as a rising $NORK$ proxies a rising marginal Tobin's q. In the model, this ratio is simply the ratio between accounting profits and capital expenditures and relates to firm-level $NORK$ in the following manner:

$$NORK_t = \int_0^1 \sum_{i=1}^{N} NORK_{i,j,t} \frac{K_{i,j,t}}{K_t} dj \tag{30}$$

The aggregate Net Operating Return to Capital is the capital-stock-weighted arithmetic mean of firms' NORK. In the model, it rises for two reasons. First, larger firms charge higher markups, implying a rise in large firms' $NORK_{i,j,t}$. Second, larger firms are also more capital-intensive, implying a high capital weight $K_{i,j,t}/K_t$. The covariance between the rise of $NORK_{i,j,t}$ and $K_{i,j,t}/K_t$ for large firms is what drives the aggregate $NORK_t$ in the model to increase by 17.9% between 1980 and 2019, roughly a third of the total rise of 50% documented by Gutiérrez and Philippon (2016).

Importantly, in the absence of markups, the model would imply no capital misallocation and no growing investment gap as firms and the aggregate economy would persistently be on their efficient frontier. Furthermore, in the model, misallocation is carried by large firms (whose $MRPK$ is further away from r_t^K), as suggested empirically by Gutiérrez and Philippon (2016). This is key to discriminate a potential alternative explanation for rising dispersion in capital intensity: borrowing constraints. I show this in the extension and counterfactual section 5.

4.5 Sales and Employment Concentration

In the model, large firms grow through capital, leading their output, capital, and employment to grow at different rates. This connects to the stylized fact documented by Autor et al. (2020), Kwon et al. (2024) and Ma et al. (2024): in the aggregate, the sales share of the largest firms has outgrown their employment share (employment of large firms to total employment). Precisely, Autor et al. (2020) focus on firms within 4D sectors in the US between 1982 and 2012 using Census data. Kwon et al. (2024) document rising assets, receipts, net income and equity concentration for corporations in the US across various granularity levels using SOI data between 1918 and 2018. Ma et al. (2024) expand to scope of their analysis to developed economies and add employment as a measure of concentration.[38] In the model, the relationship between the employment share $\lambda_{i,j,t}^{EMP} = L_{i,j,t}/L_{j,t}$ and the sales share $s_{i,j,t}$ of a firm i in sector j at time t can be written (see Appendix B.7.5 for derivations):

$$s_{i,j,t} = \underbrace{\frac{\mu(s_{i,j,t})}{\mathcal{M}_{j,t}} \frac{\zeta_{j,t}^L}{\zeta_{i,j,t}^L}}_{\varrho_{i,j,t}} \lambda_{i,j,t}^{EMP} \tag{31}$$

Where $\zeta_{i,j,t}^L$ is the output elasticity to labor of firm i equal to its cost share of labor and $\mathcal{M}_{j,t}$ is the aggregate markup of sector j. The sales share of a firm is proportional to its employment share with a proportion coefficient: $\varrho_{i,j,t}$. For the ratio of sales share to employment share to rise, it must be that $\varrho_{i,j,t}$ increases for the largest firms.

In the baseline model an increase in $s_{i,j,t}$ is due to a firm increasing its output elasticity to capital $(1-\zeta_{i,j,t}^L)$ more than its competitors, or than the sectoral aggregate $(1-\zeta_{j,t}^L)$, that is, precisely, to an increase in the ratio: $\zeta_{j,t}^L/\zeta_{i,j,t}^L$. Furthermore, as $s_{i,j,t}$ increases, a firm's markup increases as well compared to the aggregate markup, that is the ratio $\mu(s_{i,j,t})/\mathcal{M}_{j,t}$ increases. The co-variance of the two ratios is what drives the sharp rise in $\varrho_{i,j,t}$ between 1980 and 2019 in the model for the largest firms. Intuitively, as firms increase their output elasticity to capital, their demand for workers per unit of output decreases. In Figure 7, I plot the evolution ϱ_i, the ratio of the sales and employment share, for the 1% largest firms implied by the model and obtained from the Compustat data.

In the aggregate, the model seems to represent well the divergence in sales and employment concentration despite a final value slightly higher for the model (1.16) than for the data (1.13). Note, however, that the model is off in terms of levels. Indeed, for comparison, I have normalized the series to one in 1980. Without this normalization, the initial level in the model is 1.17 while it is 1.09 in the data. In Appendix B.8, I look at different measures of sales and employment concentration in the model and compare them to the empirical findings of Autor et al. (2020), Kwon et al. (2024), and Ma et al. (2024).

[38] Kwon et al. (2024) also includes employment concentration for the US in the Appendix. Ma et al. (2024) also include the US in the employment concentration section of their paper.

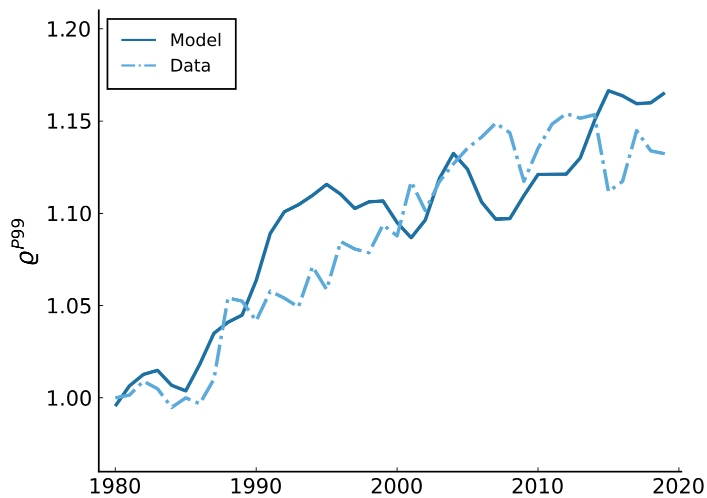


Figure 7: Ratio of sales to employment concentration for top 1% firms: Model vs. Data.

Note: This figure plots the ratio between the sales share and the employment share of the 1% largest firms each year in the model and in Compustat. In both case, the measure of a firm's size is its total sales consistently. In the model, the series is obtained by looking at the 1% firms with largest sales across the whole economy and computing the ratio of their total sales and total employment share. In Compustat, the series is obtained in the same manner: by dividing the total sales share of the 1% biggest sellers with their employment share. Both series are normalized so that their starting value in 1980 is 1.

5 Counterfactual and Extension

In this section, I look at 4 counterfactual exercises and 2 extensions and compare their implied moments to my baseline's. For the counterfactual, I first look at a case where firms are atomistic and hence markups are exogenous. Second, I suppose that firms share the same switching cost of technology. Third, I calibrate the model so as to have a stationary distribution of α along time. Fourth, I compare my baseline model to a model where heterogeneity is factor-neutral and located in TFP. For the extensions, I first build a model where firms are heterogeneous in both capital intensity and TFP and finally augment my model with a fall in the interest rate. Details of those models are located in Appendix B.9 and B.10.

5.1 Counterfactual

Monopolistic competition. In section B.9.1, I compare my baseline model to a version of the model without markup endogeneity. That is, I suppose that a continuum of atomistic firms compete monopolistically within sectors. In such a case, heterogeneity in firms' switching cost is the only driver of the heterogeneity in α. This model results is a milder decline of the labor share: -1.15pp explained fully by a symmetric rise in the aggregate capital share. The small amplitude of this decline is a simple corollary to the result of Appendix section B.8.2 where I showed that the necessary increase in capital intensity to reach a fall of 5.59pp in the labor share in a case where markups are fixed is counterfactually high. Because the profit share is flat while the capital share rises, the investment gap tightens. In this counterfactual economy, the rise in the sales-to-employment share of top firms is qualitatively in line with the data as top firms' output elasticity to labor drops below the economy's aggregate.

Homogeneous switching cost. In this counterfactual exercise, I compare the predicted moments of my baseline model to a version where all firms face the same cost of adapting their production function. To do so, I only change my baseline calibration by supposing that all firms face the same φ; implying that the supply-side of the model collapses to a representative firm. I do two sub-exercises: in the first one, I set φ to the average of the one estimated for the calibration in section 4.2, in the second one, I calibrate φ such that the representative firm's α grows by 99.16%, the growth rate of the average firm's α. Results from these two exercises are reported in Table B.3 and B.4. The key difference with the baseline model is that since firms are symmetric and can be aggregated to a representative firm, there is no dispersion in market shares and markups. As such, the aggregate labor share follows the median labor share and rises as the complementarity effect dominates and the profit share is stable. I also discuss the growth decomposition of the baseline and of this counterfactual in this section.

Stationary distribution of α. In section B.9.3, I simulate my model on a stationary distribution of α to show that its updating is key to match some critical moments. Results are visible in Table B.5. Shutting-down endogenous updating of α prevents the widening of the distribution of α, and hence of market shares and markups. This tames almost all results quantitatively and reverses a few qualitatively. Markups' dispersion still increase but by a third of the baseline amplitude, leading to a 1.6% increase in the aggregate markup. This pushes the profit share up by 1.28pp, leaving the labor share on an upward trajectory and the capital share on a downward one. This increases the NORK and makes the ratio between sales to employment concentration negligibly different from one.

Borrowing constraints. A model with collateral constraint on capital investment *à la* Kiyotaki and Moore (1997) or Cooley and Quadrini (2001) could also replicate dispersion in firms' capital intensity and larger firms being more capital intensive. Intuitively, suppose that firms share the same production technology but differ in TFP, after a decline in the RCK, larger firms face a less tight collateral constraint and can increase their capital stock more than smaller, more constrained firms. However, such a model produce s a counterfactual implication in terms of misallocation. In such a model, small firms increase their capital stock less and hence feature a higher marginal product of capital than large firms. This means that small firms stand further away from their efficient frontier and large firms closer, the opposite of the fact documented by Gutiérrez and Philippon (2016) among others. I derive this in section B.9.5.

Heterogeneous α vs. Heterogeneous TFP. Here, I compare my model to a "classic" Atkeson and Burstein (2008) where firm dynamics are driven by TFP. α is homogeneous and set to 0.12 (the average α in 1980), but firms feature heterogeneous Total Factor Productivity. I calibrate firms' TFP with estimates of TFPR calculated with equation (9) in Appendix 4. This measure is the Solow residual of firms' production function when ψ and α are supposed fixed across firms and time[39]. The rest of the calibration is unchanged. I simulate the model under this calibration and report the resulting changes in the main moments in Table B.6 of section B.9 together with the baseline.

With heterogeneous TFP, the key predictions of the model are a huge surge in the markup and an almost equally huge fall in the capital share. Because α is homogeneous and exogenous, the capital share of every firm drops due to complementarity between capital and labor. The surge in the profit share hence only eats up 1.17pp of the labor share, implying that only an unrealistic rise in markups can deliver the full decline in the labor share. Equally unrealistic is the surge in the Net Operating Return to Capital that is multiplied by almost five. This extreme increase is driven by the amplified rise in profits together with the sharp fall of the capital share (recall that $NORK_{i,j,t} = \lambda^{\pi}_{i,j,t}/\lambda^{K}_{i,j,t}$). Finally the change in the ratio between the sales and the employment share of the biggest firms is compressed because, as discussed in section 4.5, the proportionality coefficient between sales and employment is much less dampened when firms have homogeneous output elasticity to labor.

5.2 Extensions

Heterogeneous α and Heterogeneous TFP. In this extension, I add to the baseline model a factor-neutral layer of heterogeneity at the firm level in the form of total factor productivity (TFP). Intuitively, this matters because some highly labor-intensive firms might display high and growing TFP, countering the reallocation effects resulting from my baseline model. It might also be that capital-intensive firms display high TFP, potentially amplifying the baseline results. To do so, I stick to the calibration of the baseline model with the addition of firm-specific TFP this time estimated using the Solow residual equation (9) in section 4 but under the heterogeneous α case. For each sector, I construct a bivariate kernel density estimate of the joint distribution of firm-level TFP and α per sector from which each 10 firm in each 100 sector draws a pair. The resulting changes in our moments of interest are reported in Table B.7 in Appendix B.10.

Adding the heterogeneous TFP dimension has three principal effects. First it introduces dispersion in sectors that have little relative variance in capital intensity (Trade and Service-related sectors in particular). Second, it amplifies the dominant position of some firms who

[39]This measure isn't Total Factor Productivity for two reasons. First, I do not possess quantity data and hence output is proxied by sales. Second, I have factorized all source of non-homotheticity in the log-linear term of the production function. I abuse language by calling it TFP for the rest of this section for coherence with the related literature.

have both high TFP and high capital intensity. Because TFP is calibrated with TFPR, which captures non-homothetic terms, this could suggest possibly no scale invariance, consistently with a "scaling without mass" phenomenon. Third, it restores the greatness of some firms that are big despite being labor-intensive. The net effect of integrating heterogeneous TFP is hence ex-ante ambiguous for almost all our variables. The aggregate markup happens to grow more than twice as fast as in the baseline. This leads to a decline of the aggregate labor share of 7.22pp, close to its entire actual decline. The capital share also suffers from the jump in the profit share and declines as well. The Net Operating Return to Capital increases by 56.92%, close to the total estimate reported by Gutiérrez and Philippon (2016). Finally the sales to employment share of the biggest firms rises by 4.66% in average (compared to 10% in the baseline) because non-capital-intensive sectors feature concentration as well without large growth through capital. For this exercise, I also look at the evolution of market and employment shares for different degrees of granularity and report my results in Table B.8 of Appendix B.10. Unlike for the baseline model, the model now undershoots empirical estimates but does much better at replicating concentration measures for other sectors than manufacturing.

Falling interest rates. In the baseline model, I have followed a part of the literature and supposed that the interest rate paid by firms is fixed (Fisher, 2015; Campbell, Fisher, Justiniano, and Melosi, 2017; Farhi and Gourio, 2018). Yet, a fall in the required return to capital would further amplify the fall in the RCK and hence matter for my results. In Appendix B.10.2, I extend my model to include a fall in the interest rate calibrated on the estimates of Barkai (2020) by supposing that households become increasingly patient (rising β) following Laubach and Williams (2003), Mian, Straub, and Sufi (2021) or Favero, Krgović, Melone, and Tamoni (2021). Practically, this implies adding a moment to my calibration as well as re-estimating the distribution of firms' α. From the estimation, the average α rises less, because the fall in the RCK is steeper. Firms' estimated switching costs are hence shifted upwards. Yet, all results are similar qualitatively and in magnitude, except for the capital share which, consistently with Barkai (2020), declines.

6 Conclusion

This paper suggests that firm outcomes hinge on how effectively they adopt increasingly productive capital goods. This translates into firms using technologies with different capital intensities and into capital intensity to be a determinant of firms' market shares. I formalize this in a general-equilibrium model with oligopolistic competition in which firms update their output elasticity to capital; heterogeneity arises from differences in adoption capability and is amplified by strategic interactions. Calibrated to U.S. public firms, the framework jointly accounts for four trends: rising dispersion in markups resulting in a rising aggregate markup; a decline in the aggregate labor share driven by reallocation toward capital-intensive firms; greater capital misallocation and a wider aggregate investment gap; and a divergence between large firms' sales and employment shares.

However, the model is mute about the feedback from firms' capacity to adopt innovative capital goods towards capital-embodied technical change. Firms' adoption capacity might incentivize capital-goods producers to invest in innovation and hence explain the direction of technical change towards capital goods (Gordon, 1990; Cummins and Violante, 2002; Jones and Liu, 2024). Such a reality could make policies targeted at encouraging adoption an efficient tool to foster growth.

A necessary condition to design such policies would be to understand what restrains firms from adoption. In this paper, I show that there are two possible causes: imperfect competition and firms' intrinsic capabilities. The role of imperfect competition is ambivalent as markups act as an incentive for pioneer adopters (as in Aghion and Howitt (1992) for pioneer innovators) but also confer market power to the biggest players who might use it to build "defensive positions" (as suggested by Eeckhout (2022)). On the other hand, firms' intrinsic differences in their capacity to adopt efficient capital goods might stem from managerial qualities (Bresnahan et al., 2002; Bloom et al., 2012), difficulties to reorganize production (Chari and Hopenhayn, 1991; Aghion and Tirole, 1997) or simply a lack of information or understanding about the availability of efficient capital goods.

The model could also speak to puzzling international differences such as the diverging growth rate between Europe and the U.S. that coincides with a divergence in the growth rate of their respective capital intensity. Are Americans only better at producing innovation or are they also better at integrating and adopting it? Similarly, the puzzling reallocation of savings from high-MPK to low-MPK countries (Lucas, 1990; Caselli and Feyrer, 2007) might be biased by omitting that an economy's MPK scales with the capital intensity of its firms.

Finally, I also hope that this paper can modestly participate to reflections about the implications of ongoing technological transformations. If the advent of AI and Data magnifies heterogeneous capital intensity, the aggregate labor share ought to keep on declining. In such a case, what will happen to the vast majority of people who rely mostly on labor income to live? Will retail investing expose their income to capital and profit shares or should we encourage, as hypothesized by Karabarbounis (2024), universal dividends?

Appendix

A Empirics

A.1 Data

Compustat Firm-level variables are obtained from Compustat, which includes balance sheet and income statement data for all publicly listed firms in the U.S. The sample period is 1980 to 2019. My empirical variables are constructed as follows:

Sales (pY) measured as *"Sales/Turnover - Total"* in Compustat.

Capital (K) is the total capital stock. It is measured by summing the stocks of tangible capital (*"Property, Plant and Equipment - Total (Gross)"*) and intangible capital (*"Intangible Assets - Total"*).

Labor (L) is measured with *"Employees"* . In robustness tests *"Cost of Goods Sold"* (COGS) and "Selling, General and Administrative expenses" (SGA) are used as alternatives.

A.2 Additional evidence

In this section I provide additional evidence to which I refer to in the main text.

A.2.1 Capital intensity: aggregate and distribution

Aggregate. Figure A.1 shows four measures of aggregate capital intensity: with and without intangibles. With employees or COGS as a denominator.

Dispersion. Figure A.2 displays percentiles of the distribution of firms' capital intensity (capital-to-labor ratio) for 3 2D NAICS sectors.

I build two measures of dispersion of capital intensity; one across sectors and one within sectors. I report these measures in Table A.1. To estimate dispersion of capital intensity across sectors, I calculate the average capital intensity of each sector and take the standard deviation across those sectoral averages. This measure is reported in the third row of Table A.1. In column 1, I report the measure for a 2D NAICS sector granularity for 1980. In column 2, I report results for 2019. Columns 3 and 4 proceed of the same logic at the 4D NAICS sector granularity. In row 4, I reverse my moments, that is I compute the average across sectors of the standard deviations of firms' capital-to-labor ratio within their sectors. This measure captures dispersion in K/L within sectors. Comparing line 3 with line 4

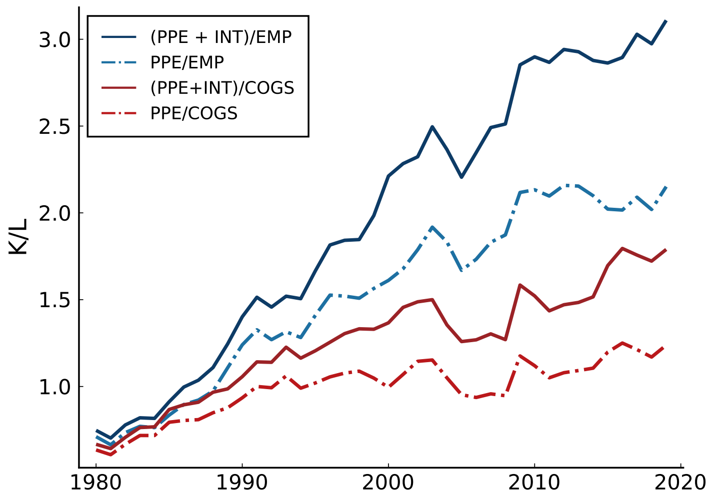
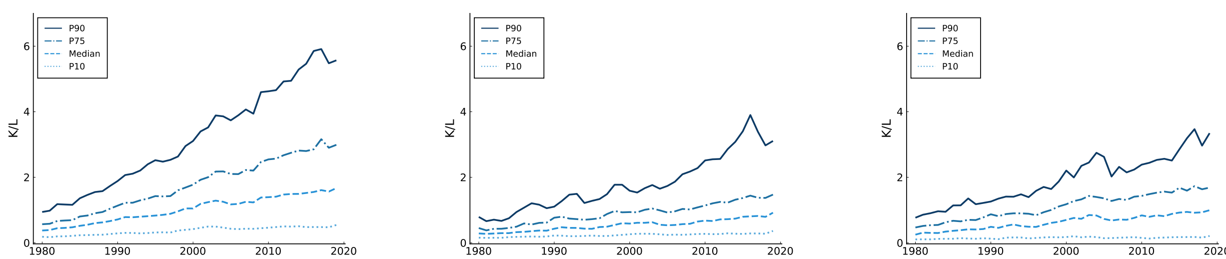


Figure A.1: Measures of Aggregate capital intensity.

Note: Source is Compustat. Sample is 1980-2019. PPE stands for "Property Plant and Equipment", L for "Employees", INT for "Intangible Assets - Total" and COGS for "Cost of Goods Sold". EMP is expressed in number of Employees, PPE, INT and COGS are expressed in 100'000s of 2000 dollars.



(a) Manufacturing (b) Trade (c) Services

Figure A.2: Percentiles of the Distribution of firms' capital intensity for 3 biggest 2D sectors

Note: Source is Compustat. Sample is 1980-2019. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Manufacturing gathers 2-digit NAICS sectors 31, 32 and 33. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).

shows that within sector heterogeneity in capital intensity is big no matter the granularity degree. Mechanically, because 4D sectors are more numerous than 2D sectors, the cross-sector standard deviation is higher for 4D sectors than for 2D while the inverse is true for the within-sector std. In lines 5 and 6 I switch standard deviation for inter-quartile range to show that heterogeneity is not carried solely by a handful of firms but that, as suggested by Figure 1, firms located in the upper tail of the distribution seem to be heavy drivers of the rise in the aggregate capital-to-labor ratio. In the earlier years, only a small fraction of 4D sectors count more than 10 observations. This, together with the fact that my cost data for inputs is observed at the 2D level of granularity, makes me focus on the 2D level for the rest of the main text while measures at thinner levels of granularity are reported in the Appendix.

K/L	2D Sectors		4D Sectors	
Year	1980	2019	1980	2019
Average	1.27	11.05	1.1	11.21
Std across sectors	1.4	17.7	1.8	30.9
Std within sector	1.3	31.5	1.0	18.5
Inter-quartile range across sectors	0.6	2.7	0.3	3.6
Inter-quartile range within sector	1	5.5	0.7	8.3
Number of Sectors	16	16	27	76

Table A.1: Descriptive statistics of the distribution of capital intensity

Note: Source is Compustat. Measure of capital intensity is (PPE + INT)/L. PPE stands for "Property, Plant and Capital (Gross)" and measures physical capital while INT stands for "Intangible Assets" and measures intangible capital. EMP is "Employees" and is the count of employees. Sectors are NAICS. Units are 100'000s of 2000 dollars per employee. Sectors with less than 10 observations are dropped.

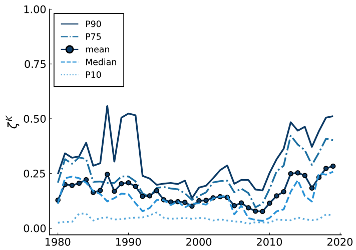
A.2.2 Output elasticity to capital and share of capital in production: ζ and α

Figure A.3 displays percentiles of the distribution of firms' output elasticity to capital of 2D sectors not displayed in the main text.

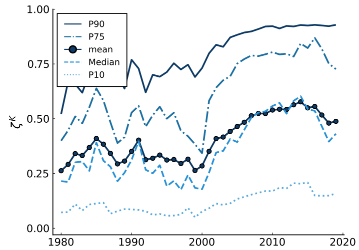
Figure A.4 displays percentiles of the distribution of firms' output elasticity to capital within the 9 biggest 4D sectors within Manufacturing.

Figure A.20 plots some moments of the distribution of firms' α for the aggregate (public) economy. Panel (a) refers to the case where capital and labor are complements while panel (b) is the case where they are substitutes (in the Cobb-Douglas case, α and ζ coincide).

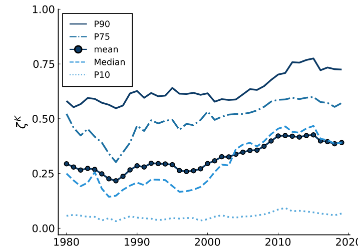
Figure A.21 plots some moments of the distribution of firms' α for manufacturing, trade and services for the baseline case where capital and labor are complements.



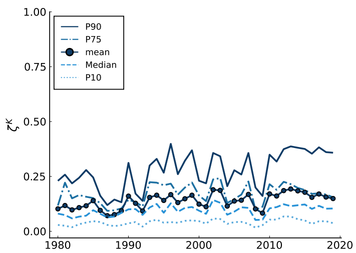
(a) Agriculture



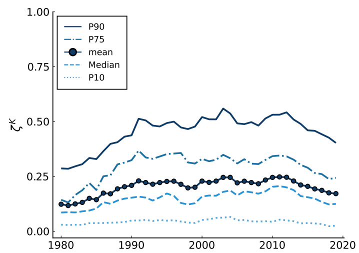
(b) Mining & Extraction



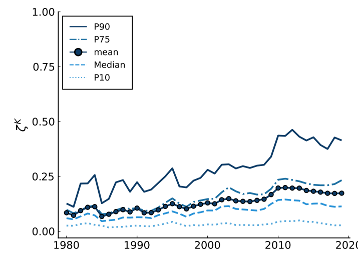
(c) Utilities & Transportation



(d) Construction



(e) Information



(f) Finance

Figure A.3: Percentiles of the Distribution of firms' output elasticity to capital by 2-digit NAICS sector

Note: Source is Compustat. Sample is 1980-2019. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Agriculture is 2-digit NAICS sectors 11. Mining, Quarrying, Oil and Gas Extraction is code 21. Utilities & Transportation is codes 22, 48 and 49. Construction is code 23. Information is code 51. Finance and Insurance is code 52.

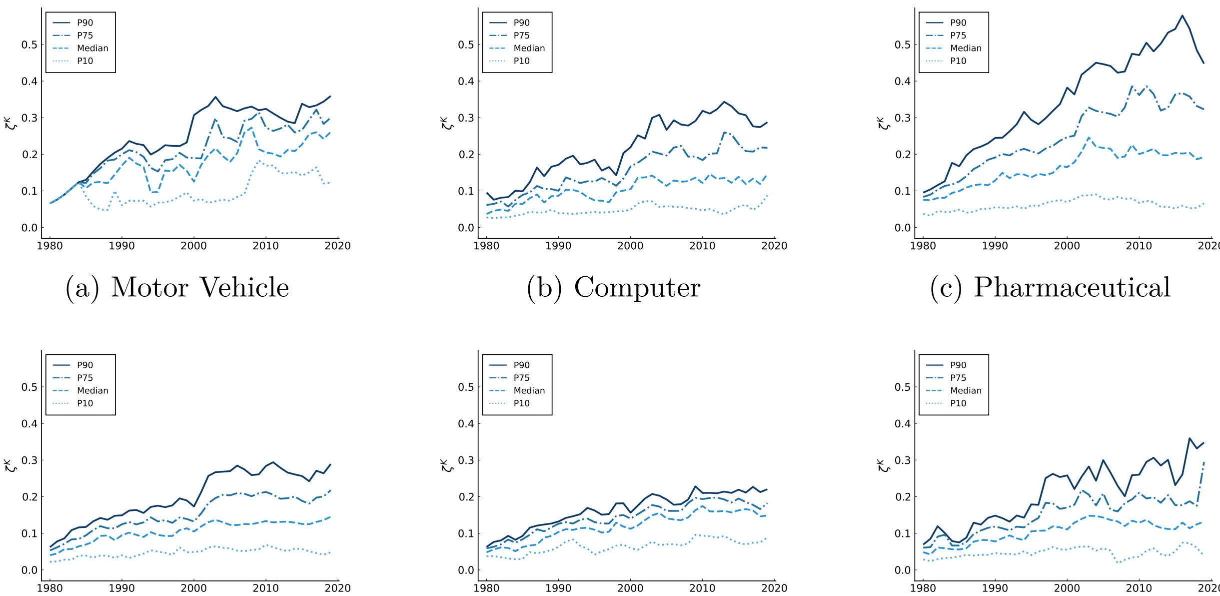
Estimating α. Suppose that firms produce using the CES production function (2), the identity for capital intensity (1) becomes:

$$\frac{K_{i,t}}{L_{i,t}} = \left(\frac{\alpha_{i,t}}{1-\alpha_{i,t}}\frac{w_t}{r_t^K}\right)^{\psi}$$

Isolating for alpha, this gives:

$$\alpha_{i,t} = \frac{r_t^K K_{i,t}^{\frac{1}{\psi}}}{w_t L_{i,t}^{\frac{1}{\psi}} + r_t^K K_{i,t}^{\frac{1}{\psi}}} \tag{32}$$

Which in the Cobb-Douglas limit converges to the cost share of capital $\zeta_{i,t}^K$. In Figure A.20, I display some percentiles of the distribution of firms' α across years obtained using equation (10). In panel (a), I plot my baseline where ψ is set to a usual value in the literature: 0.5. In panel (b) I plot a case where capital and labor are substitutes with $\psi = 1.5$ (for an estimate of ψ for the aggregate economy, see Antras (2004). For an estimation at the firm-level for manufacturing and a proposed aggregation method, see Oberfield and Raval (2021). For a more complete review on studies that estimate ψ, see Rognlie (2015)).



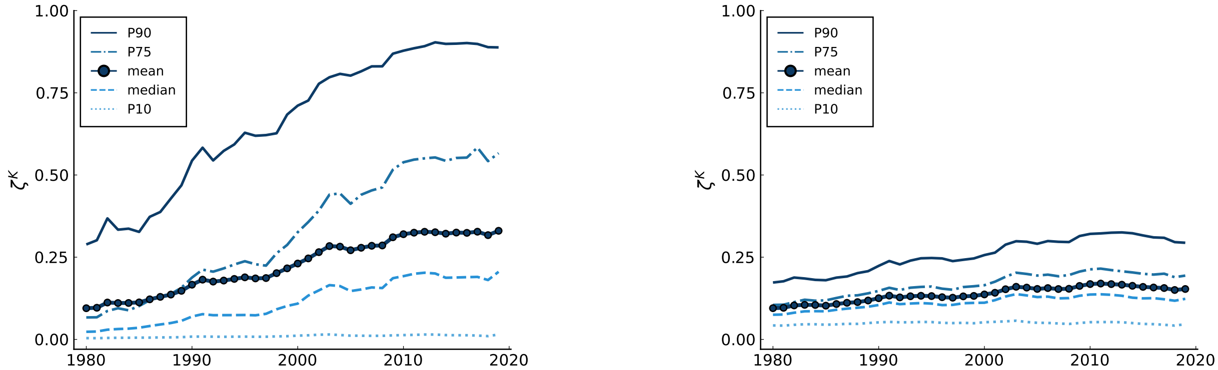
(a) Motor Vehicle (b) Computer (c) Pharmaceutical

(d) Navigation & Control (e) General Machinery (f) Electrical Equipment

Figure A.4: Percentiles of the Distribution of firms' output elasticity to capital by 4-digit NAICS sectors within Manufacturing

Note: Source is Compustat. Sample is 1980-2019. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Motor Vehicle is 4-digit NAICS sectors 3361. Computers is code 3341. Pharmaceutical is codes 3254. Navigation is code 3345. General Purpose Machinery is code 3339. Electrical Equipment is 3359.

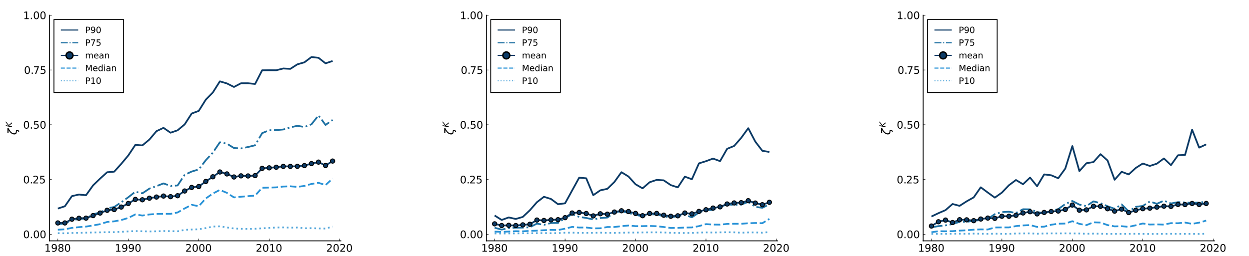
In the baseline case where $\psi = 0.5$, similarly to the distribution of firms' capital-to-labor ratio depicted in Figure 1, the distribution of firms' alpha is non-stationary; both the mean and the variance increase with time. The magnitude of these increase is higher than that of the output elasticity to capital depicted in 2 because heterogeneity in capital intensity is augmented by the geometric weight $\frac{1}{\psi} = 2$. Conversely, in the substitutes case, heterogeneity in capital intensity is dampened and the average and variance of α increase less. In Figure A.21, I plot some percentiles of the distribution of firms' α across years obtained using equation (10) for manufacturing, trade and services when $\psi = 0.5$.



(a) Complements ($\psi = 0.5$) (b) Substitutes ($\psi = 1.5$)

Figure A.5: Moments of the distribution of alpha conditional on ψ

Note: Source is Compustat. Sample is 1980-2019. Moments from the distribution implied by the estimation using equation (10) on Compustat firms. Measure of Capital is (PPE + INT). PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". Measure of Labor is EMP: number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 is the 90th percentile.



(a) Manufacturing (b) Trade (c) Services

Figure A.6: Percentiles of the Distribution of firms' α for 3 biggest 2D sectors

Note: Source is Compustat. Sample is 1980-2019. α is measured using equation (10) for $\psi = 0.5$. Measure of Capital is (PPE + INT). PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". Measure of Labor is EMP: number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Manufacturing gathers 2-digit NAICS sectors 31, 32 and 33. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).

A.2.3 Capital intensity and market shares

Figure A.7 shows the capital intensity of the 5% biggest firms by sales and compares it to the capital intensity of the whole economy and of the median firm.

Figure A.8 shows the market share of each decile of the distribution of capital intensity within the aggregate, trade and services.

Figure A.9 "zooms-in" Figure 3 and plots the evolution of the market share of percentiles: P90, P95 and P99 of the distribution of capital intensity within manufacturing.

Figure A.10 plots the 5-year persistence rate of firms located in the 90th, 95th and 99th percentile of the capital intensity distribution within manufacturing. The persistence rate is conditional on the number of firms present in both year Y and year Y-5.

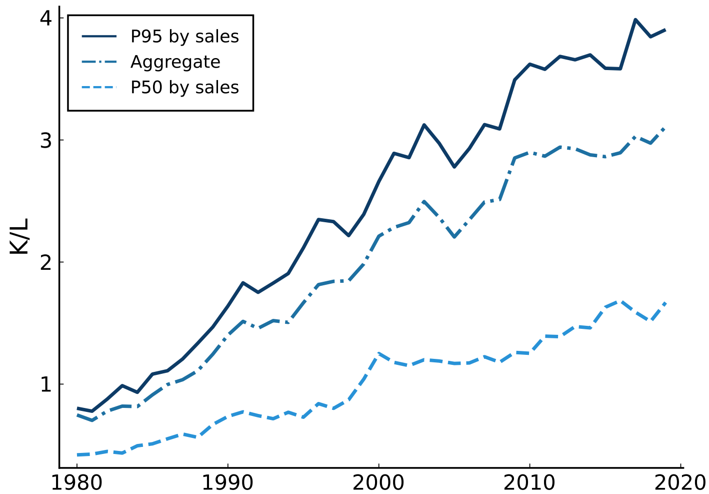
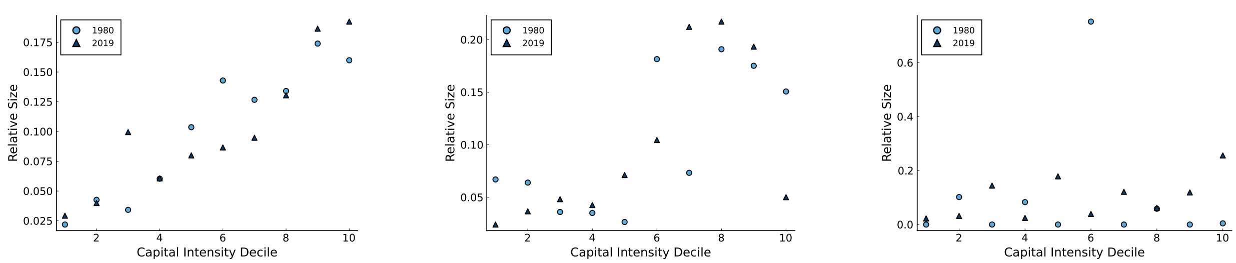


Figure A.7: Capital intensity: biggest firms, aggregate, median firm.

Note: Source is Compustat. Sample is 1980-2019. PPE stands for "Property Plant and Equipment", L for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. Biggest firms are 5% highest sales.



(a) Aggregate (b) Trade (c) Services

Figure A.8: Correspondence between capital intensity decile and market shares.

Note: Source is Compustat. Sample is 1980-2019. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Relative size is sales of the decile nest over total sales of the sectoral aggregate. For example, the denominator of relative size in the aggregate is total sales across all sectors while for Trade it is sales in the Trade sector only. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).

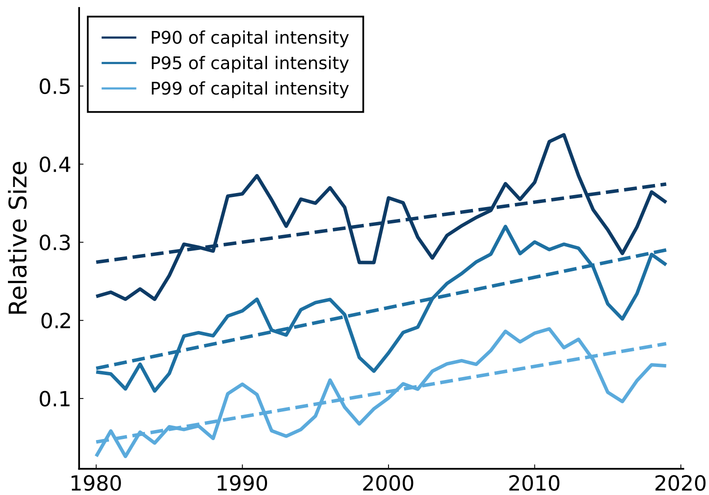


Figure A.9: Evolution of relative size of top percentiles of the capital intensity distribution of manufacturing.

Note: Source is Compustat. Sample is 1980-2019. Manufacturing is 2D NAICS sectors 31, 32 and 33. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Relative size are sales over total sales.

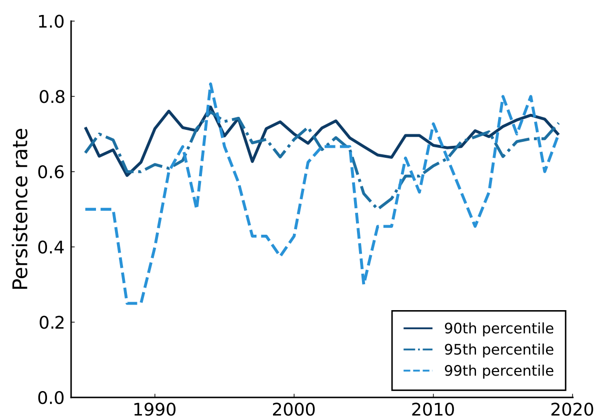


Figure A.10: Persistence rate of top percentiles of the distribution capital intensity in manufacturing.

Note: Source is Compustat. Sample is 1980-2019. Manufacturing is 2D NAICS sectors 31, 32 and 33. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Persistence rate is defined as the number of firms that belong to PX in year Y which also belonged to PX is year Y-5 divided by the total number of firms listed in both year Y and Y-5 that belong in PX is year Y.

A.3 Other stylized facts

In this section, I present additional stylized facts mentioned in the paper.

A.3.1 Capital intensity and firms' labor shares

In this section, I look into the joint distribution of labor shares and capital intensity in the cross section. The measure for a firm's labor share is cost of goods sold divided by sales. Figure A.11 plots the average of the distribution of firms' labor share within each decile of capital intensity for the beginning (1980) and end (2019) of my sample.

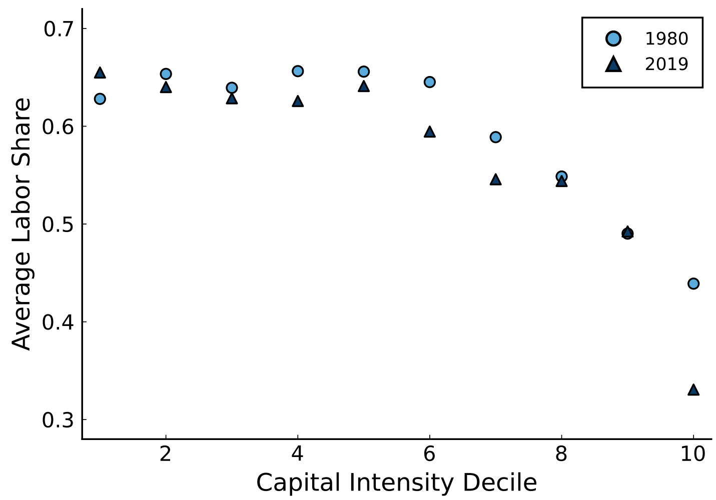


Figure A.11: Average labor share per capital intensity decile: 1980 and 2019

Note: Source is Compustat. Sample is 1980-2019. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. A firm's labor share is its cost of goods sold over its sales. Average is taken within each yearly decile of capital intensity.

A.3.2 The fall in the Relative Cost of Capital

The rise in aggregate capital intensity has been shown to be due to the decline in the relative cost of capital (RCK)[40]. The relative cost of capital is most often defined as the ratio between the cost of capital goods to the cost of consumption goods or the cost of labor. It has declined for two reasons. First, capital good producers have seen their productivity increase more than that of consumption good producers. Second, when adjusting for quality, the decline is even more salient as the quality of capital goods has been increasing faster than that of

[40]See for example: Greenwood et al. (1997), Greenwood et al. (2000), Fisher (2006), Primiceri (2006), Gourio and Kashyap (2007), Gourio and Klier (2015), Karabarbounis and Neiman (2014), Jones and Liu (2024),Dechezleprêtre et al. (2025), Hartley et al. (2025).

consumption goods. The decline in the RCK is hence the result of capital good producers producing more efficiently goods of higher quality. The path of the RCK is plotted in Figure A.12, the construction of the measure is reported in the Data section 2.1. In previous models, where firms are homogeneous, this translates into all firms adopting equally those new vintages of capital goods hence pushing up their capital intensity symmetrically.

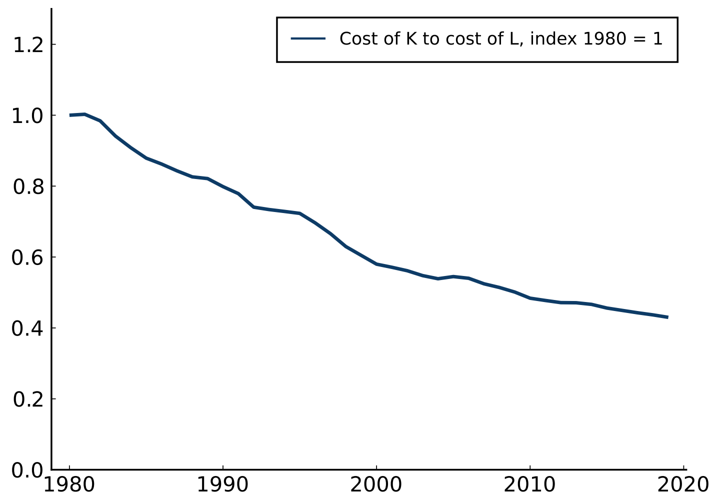


Figure A.12: Declining relative cost of capital

Note: Source is BEA. Sample is 1980-2019. Constant-quality cost of capital is current-cost investment in fixed assets divided by quantity index investment in fixed assets from Fixed Asset tables. Cost of labor is wages and salaries per full-time equivalent employee by industry from NIPA tables.

A.4 Robustness

In this section, I report the robustness exercises related to the empirical section of the main text.

A.4.1 Alternative measure of firms' capital intensity

Figure A.13 plots percentiles of the distribution of firms' capital intensity where the denominator of the measure of capital intensity is cost of goods sold instead of employees in the main text.

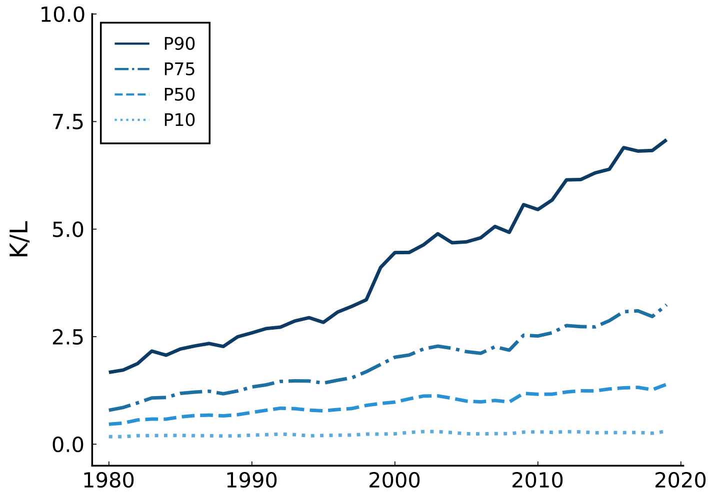


Figure A.13: Percentiles of the distribution of firms' capital intensity: robustness with COGS.

Note: Source is Compustat. Sample is 1980-2019. Measure is (PPE+INT)/COGS. PPE stands for "Property, Plant and Capital (Gross)" and measures physical capital while INT stands for "Intangible Assets" and measures intangible capital. COGS is "Cost of Goods Sold" that is the sum of labor and other intermediate inputs costs. P90 is the 90th percentile, P10, the 10th. Units are 100'000s of 2000 dollars.

A.4.2 Heterogeneous rental rate of capital

Suppose that firms face heterogeneous rental rates of capital due, for example, to differing risk premia, instead of heterogeneous output elasticity to capital. In such as case the capital intensity identity writes:

$$\frac{K_{i,t}}{L_{i,t}} = \frac{\zeta^K}{\zeta^L} \frac{w_t}{(r_{i,t} + \delta) p_t^X} \tag{33}$$

Where p_t^X is the cost of investment goods and $r_{i,t}$ the firm- and time-specific interest rate faced by firms to finance their investment. Inverting equation (33) allows to estimate $r_{i,t}$ for each pair $\{i, t\}$. In Figure A.14, I plot the difference, in percentage point, between the interest rate faced by different firms under such a case. In dark blue, I plot the difference between the interest rate paid by the 10% least capital-intensive firms and the 10% most capital-intensive firms. In light blue, I plot the difference between the interest rate of the 10% least capital-intensive firms and the median firm. If heterogeneous risk premia were to drive the difference in firms' capital intensity, the premium paid by the least capital-intensive firms would be multiplied by 6 over our sample for both measures (from 1.65 to 10 for the least-to-most capital-intensive premium and from 1.06 to 6.35 for least-to-median capital-intensive premium) and by 3 between 1997 and 2019. In comparison, the spread between BBB and AAA corporate Bond Yields from Bank of America (obtained from FRED) was multiplied by 1.1 between 1997 (the earliest date available on FRED) and 2019: from 0.48

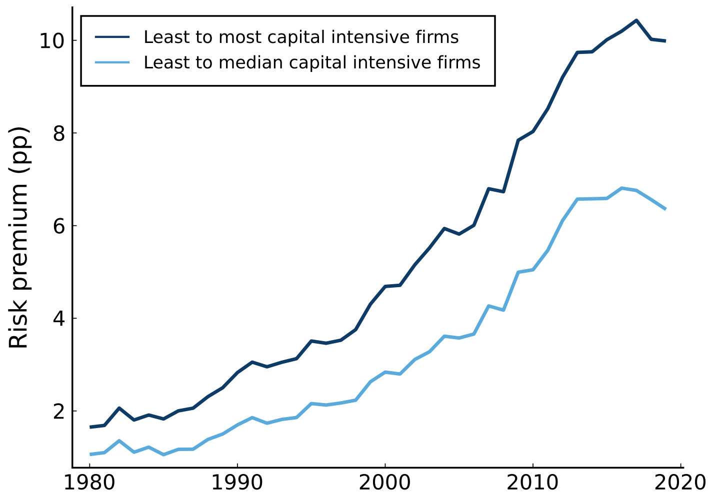


Figure A.14: Implied risk premium across differently capital intensive firms.

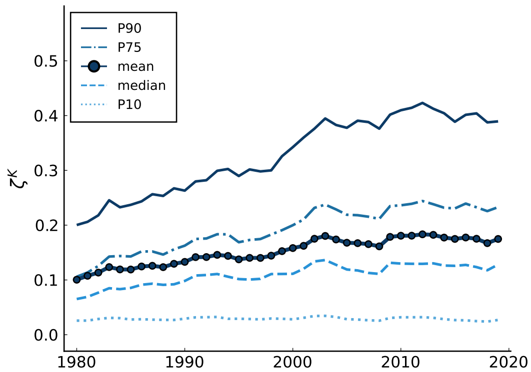
Note: Source is Compustat. Sample is 1980-2019. Measures of risk-premia are obtained from equation (33). It is calculated as the difference between the interest rate paid by two types of firms if interest rates differences were the only explanation for heterogeneous capital intensity, i.e, supposing that firms have homogeneous output elasticity to inputs. In the case "Least to most capital intensive firms", the premium is the difference between the interest rate of the 10% least capital-intensive firms and 10% most capital intensive firms. In the case "Least to median capital intensive firms", the premium is the difference between the interest rate of the 10% least capital-intensive firms and the median firm.

to 0.53. Similarly, the spread between CCC and AAA bonds was multiplied by 1.43: from 6.31 to 7.59. This suggests that riskiness is not enough to explain differing capital intensities across firms.

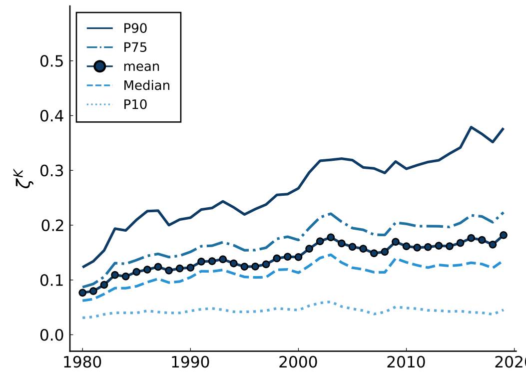
A.4.3 Alternative measures of output elasticity to capital

Figure A.15 replicates Figure 2 but with Cost of Goods Sold as a measure of labor costs.

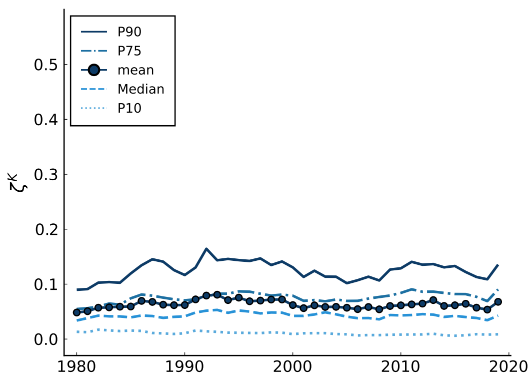
Figure A.16 replicates Figure 2 but includes falling interest rates in the measure of the rental rate of capital following Barkai (2020).



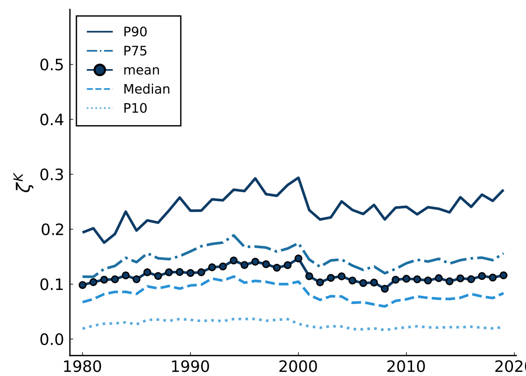
(a) Aggregate



(b) Manufacturing



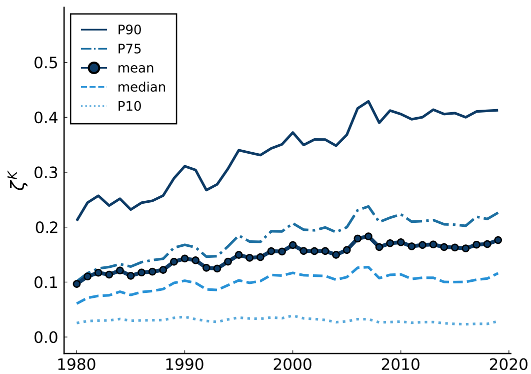
(c) Trade



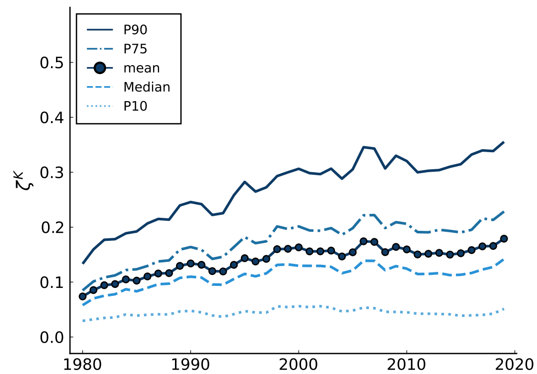
(d) Services

Figure A.15: Percentiles of the distribution of firms' output elasticity to capital per sector: COGS for labor cost.

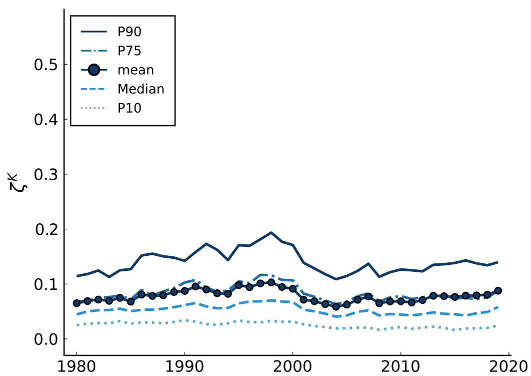
Note: Source is Compustat. Sample is 1980-2019. $\zeta_{it}^K = (r_t^K K_{it})/TC_{it}$ is the output elasticity to capital, or equivalently, the cost share of capital of firm i at time t. K_{it} is the sum of physical capital (PPE in Compustat) and intangible capital (INT in Compustat). TC_{it} is total cost and is the sum of cost of capital and cost of labor defined as Cost of Goods Sold. P90 is the 90th percentile, P10, the 10th. Manufacturing gathers 2-digit NAICS sectors 31, 32 and 33. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).



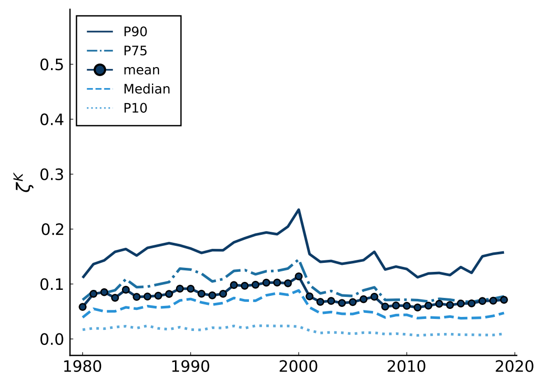
(a) Aggregate



(b) Manufacturing



(c) Trade



(d) Services

Figure A.16: Percentiles of the distribution of firms' output elasticity to capital per sector: including falling interest rates.

Note: Source is Compustat. Sample is 1980-2019. $\zeta_{it}^K = (r_t^K K_{it})/TC_{it}$ is the output elasticity to capital, or equivalently, the cost share of capital of firm i at time t. K_{it} is the sum of physical capital (PPE in Compustat) and intangible capital (INT in Compustat). TC_{it} is total cost and is the sum of cost of capital and cost of labor defined as the wage bill. r_t^K is $r_t + \delta$ times the cost of investment good from the BEA. r_t is the interest rate for BBB companies following Barkai (2020). P90 is the 90th percentile, P10, the 10th. Manufacturing gathers 2-digit NAICS sectors 31, 32 and 33. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).

A.4.4 Aggregate output elasticity to capital from BEA.

Figure A.17 calculates the aggregate output elasticity to capital for manufacturing and the aggregate American economy using BEA data.

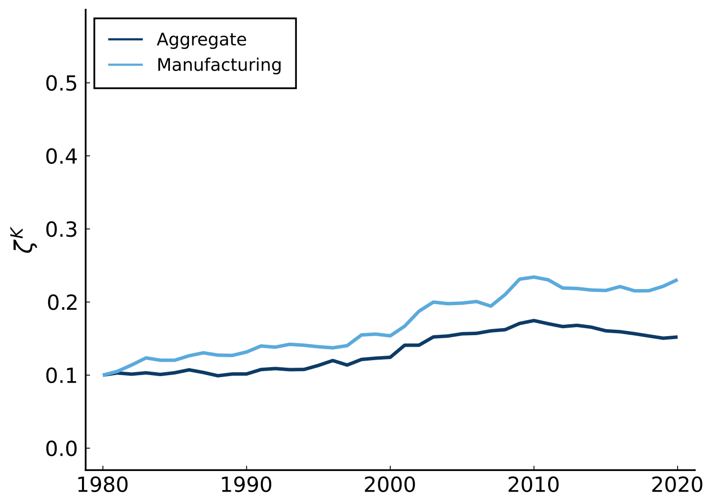


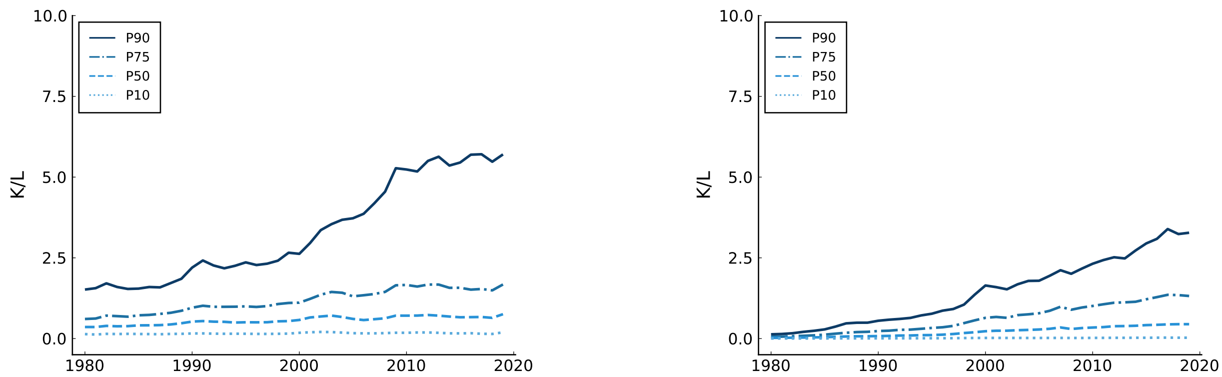
Figure A.17: Output elasticity to capital for Aggregate and Manufacturing using BEA data.
Note: Source is BEA. Sample is 1980-2020. $\zeta^K_{it} = (r^K_t K_{it})/TC_{it}$ is the output elasticity to capital, or equivalently, the cost share of capital of nest i at time t. K_{it} is Private Fixed Asset from the BEA. TC_{it} is total cost and is the sum of cost of capital and cost of labor defined as the wage bill from the BEA. r^K_t is $r_t + \delta$ times the cost of investment good from the BEA. r_t is set to 0.05 and δ to 0.1. Output elasticity in 1980 is normalized to 0.1 for both measures.

A.4.5 Tangible and Intangible

In this section, I repeat the exercises of section **??** but report stylized facts for physical and intangible capital separately. In Figure A.18, I plot the evolution of some percentiles of the distribution of capital intensity where capital is either physical capital only -Panel (a)- or intangible capital only -Panel (b)-. For both measures the patterns are the same as when the two types of capital are aggregated, as in Figure 1 of section **??** and no specific type of capital drives the fanning out. Naturally, Intangibles become more significant post-2000.

I then test weather intangible- and tangible- intensity correlate. To do so, I simply regress physical intensity on intangible intensity across the panel of my firms along the 1980-2019 sample. I find a coefficient of correlation of 6.227. I then include year fixed effects because my original correlation may be biased by common aggregate trends, the coefficient stays the same. Finally I include firm fixed effects to account for their specificity and find a coefficient of 6.105, significant at the 1% confidence threshold. These results are reported in table A.2.

I then move to estimating separately the output elasticity to tangible and intangible capital. To do so, suppose that a firm i in time t produces using the following nested prodution function: $Y_{i,t} = \mathcal{F}_{i,t}(\mathcal{G}_{\rangle,\sqcup}\left(K^P_{i,t}, K^I_{i,t}\right), L_{i,t})$ where $K^P_{i,t}$ is physical capital, $K^I_{i,t}$ is



(a) Physical Capital

(b) Intangible Capital

Figure A.18: Moments of the distribution of capital intensity per type of capital.

Note: Source is Compustat. Sample is 1980-2019. Measure is PPE/L for left figure (a) and INT/L for right figure (b). PPE stands for "Property, Plant and Capital (Gross)" and measures physical capital while INT stands for "Intangible Assets" and measures intangible capital. EMP is "Employees" and is the count of employees. P90 is the 90th percentile, P10, the 10th. Units are 100'000s of 2000 dollars per employee.

	Physical intensity		
	(1)	(2)	(3)
Intercept	-1.270**		
	(0.452)		
Intangible intensity	6.227***	6.228***	6.105***
	(0.025)	(0.025)	(0.027)
Fixed Effects		Year	Year & Date
N	107,369	107,369	105,854
R^2	0.368	0.368	0.609

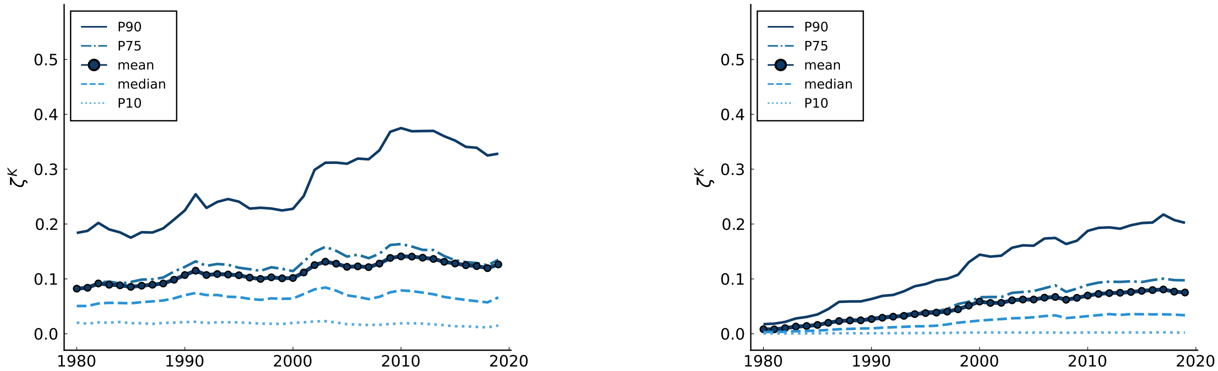
Table A.2: Firm level regressions: Physical on Intangible intensities.

Note: Source is Compustat. Sample is 1980-2019. Table displays estimators of OLS regression of physical intensity on intangible intensity. Physical intensity is PPE/L where PPE stands for "Property, Plant and Equipment" and L is "Employees". Intangible intensity is "INT/L" where INT stands for "Intangible Assets". Note that N is smaller when fixed effects are included because some firms exist only for one year in Compustat. *** indicate the the coefficient is significant at the 1% threshold.

intangible capital and $L_{i,t}$ is labor. Suppose further that there exists a technology $G_{i,t}$ that transforms physical and intangible capital into productive capital. At time t, for a given triplet of costs $\{w_t, r_t^P, r_t^I\}$ of labor physical capital and intangible capital respectively, cost-minimization yields the following identity for a firm's tangible-to-intangible ratio:

$$\frac{K_{i,t}^P}{K_{i,t}^I} = \frac{\zeta_{i,t}^P}{\zeta_{i,t}^I}\frac{r_t^I}{r_t^P} \tag{34}$$

Where $\zeta_{i,t}^P = \frac{\partial \ln(F_{i,t}(.))}{\partial \ln(K_{i,t}^P)}$ is the elasticity of output to physical capital (and conversely for



(a) Physical Capital (b) Intangible Capital

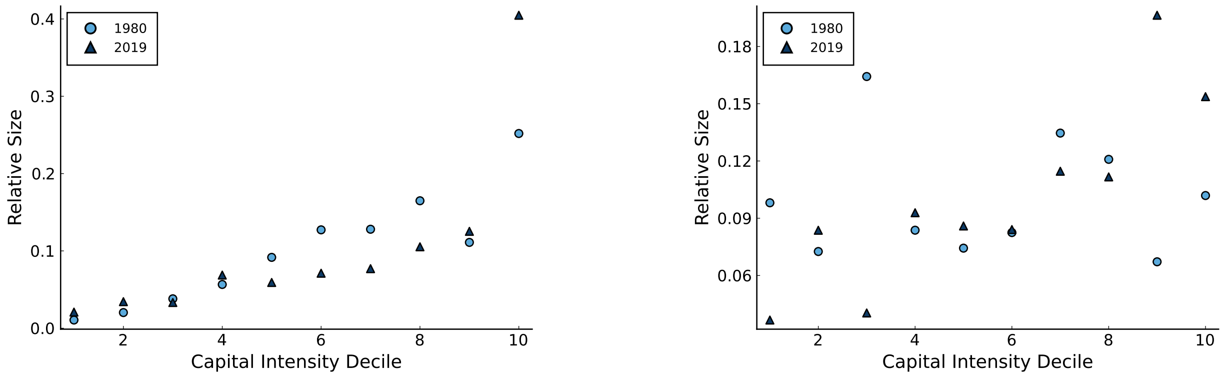
Figure A.19: Moments of the distributions of output elasticity to physical vs. intangible capital.

Note: Source is Compustat. Sample is 1980-2019. Measure is PPE for physical capital and INT for intangible capital. PPE stands for "Property, Plant and Capital (Gross)" and measures physical capital while INT stands for "Intangible Assets" and measures intangible capital. Output elasticities are estimated according to equation (34). EMP is "Employees" and is the count of employees. P90 is the 90th percentile, P10, the 10th. Units are 100'000s of 2000 dollars per employee.

$\zeta^I_{i,t}$ and intangible). Using the separate measure of the cost of physical capital and intangible capital at the 2D sectoral granularity furnished by the BEA (see Data section 2.1) as well as the stock of physical capital and intangible capital from firms' yearly financial statements, I estimate firm- and year-specific $\zeta^P_{i,t}$ as the cost share of physical capital to total costs: $\zeta^P_{i,t} = \frac{r^P_t K^P_{i,t}}{r^I_t K^I_{i,t} + r^P_t K^P_{i,t} + w_t L_{i,t}}$ and similarly for intangible capital. Moments of the distribution of $\zeta^P_{i,t}$ and $\zeta^I_{i,t}$ are plotted in Figure 34.

Finally, I look at the evolution of the market share of the most tangible- and intangible-intensive firms among public manufacturing firms and compare it with the results of Figure 3. I repeat the exact same exercise: the market share of each decile of the distribution of capital intensity is computed for each type of capital. The result is reported in Figure A.20. The picture for Physical Capital is very similar to Figure A.20, in line with the fact, unveiled in Figure A.18, that physical capital is a bigger portion of total capital (a fact also unveiled by Lashkari et al. (2024) for the case of ICT in France). The Figure for intangible reveals an expected change; in 1980, when intangible is of negligible volume, intangible intensity is uncorrelated to capital intensity. In 2019 however a correlation appears, although less strong than for physical capital.

Those differences are however difficult to interpret as the classification between physical and intangible capital is left to the firm (Crouzet and Eberly, 2023). As such, a purchased software can be classified as "Property" if it is considered a long-lived asset and hence enter "Property, Plant and Equipment", our measure of physical capital. In the main body of the paper, I therefore don't focus on differences between tangible and intangible capital, but rather on their sum as total capital.



(a) Physical Capital (b) Intangible Capital

Figure A.20: Market share of capital-intensive firms per type of capital within manufacturing.
Note: Source is Compustat. Sample is 1980-2019. Measure is PPE/L for left figure (a) and INT/L for right figure (b). PPE stands for "Property, Plant and Capital (Gross)" and measures physical capital while INT stands for "Intangible Assets" and measures intangible capital. EMP is "Employees" and is the count of employees. Measure of market share is the ratio of individual sales to total sales of the sector. Manufacturing is 2D NAICS codes 31, 32 and 33.

A.5 Empirical Implications

In this section, I show how accounting for heterogeneity in firms' output elasticity to capital and share of capital in production affects the estimation of firms' Total Factor Productivity of Revenue (TFPR) and markups of prices over marginal costs. Specifically, I show that when one accounts for heterogeneity in structural capital intensity: firms' Solow residuals are much less dispersed and the aggregate and average markup are dampened. Nonetheless, I find that capital-intensive firms charge a significantly higher and faster-growing markup than their labor-intensive competition.

A.5.1 Estimating TFPR with firm-specific α

When firms are supposed to have homogeneous production technology except for their TFP, differences in α are typically ingrained in differences in TFP when the latter is estimated as a Solow residual. A natural question is therefore to ask what proportion of heterogeneous TFP can be attributed to heterogeneous structural capital intensity? Or in other words, in the face of output, how important is capital intensity to explain firm-level differences?

I cannot answer these questions strictly because I do not observe firms' prices (nor quality differences) and the log-linear term captures TFPQ but also potentially non-homotheticity and non-homogeneity of degree one. However, I can estimate the Solow residual of a "revenue" production function and compare the result when α is allowed to vary across firms and time to when it isn't. The Solow residual of the revenue production function takes the form:

$$\ln(p_{i,t}Z_{i,t}) = \ln(p_{i,t}Y_{i,t}) - \frac{\psi}{\psi-1}\ln\left(\alpha_{i,t}K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,t})L_{i,t}^{\frac{\psi-1}{\psi}}\right) \tag{35}$$

In order to assess how much of the Solow residual is "eaten-up" by heterogeneous capital intensity, I perform two estimations and compare the key moments of the resulting distributions. The first estimation procedure is equation (9) with the firm-year alpha estimated following equation (10). The second one is the counterfactual and is hence also equation (9) at the firm-year level but with homogeneous α as an input. In that second case, I set α to the unweighted arithmetic average of the firm-level alphas used in the first estimation: 0.12. For both equations, ψ is set to 0.5. Moments of the distribution of pZ in 1980 and 2019 under homogeneous structural capital intensity (Hom. α) and heterogeneous structural capital intensity (Het. α) are reported in Table 4.12.

Year	Specification	Mean	Median	Variance
1980	Hom. α	2.43	1.93	7.34
	Het. α	1.62 (-33%)	1.33 (-31%)	3.65 (-50%)
2019	Hom. α	3.78	2.01	123.42
	Het. α	2.11 (-44%)	1.43 (-28%)	20.81 (-83%)

Table A.3: Moments of the distribution of the Solow residual per specification.

Note: Source is Compustat. Sample is 1980-2019. Solow residuals are estimated according to equation (9). pY is "Sales/Turnover (Net)". K is the sum of "Property, Plant and Equipment (Gross)" and "Intangible Assets" and L is "Employees". Het. alpha stands for heterogeneous α that is the firm-year estimation obtained in equation (10) while Hom. alpha stands for homogeneous α and is set to 0.12. ψ is set to 0.5. The parenthesis below the value in lines Het. alpha is the percentage difference between the Hom. alpha moment and its Het. alpha corresponding value.

Naturally, since we allow for an extra degree of heterogeneity, all moments appear to be depressed; in 1980, average firm-level Solow residual is 33 % lower with heterogeneous α than without. In 2019, that number climbs to 44 %. The increasing importance of heterogeneous structural capital intensity seems to be concentrated at the tails however, since the percentage difference declines slightly for the median (31 % to 28%). More importantly, the variance of firms' Solow residual, the key moment to generate reallocative effects in heterogeneous firm models, is drastically reduced. It is halved at the beginning of the sample and collapses to 17 % of its homogeneous alpha value when heterogeneous alphas are introduced at the end of the sample. In the baseline quantitative model, firms will have no TFP and rather α will be a control variable of the firm, allowing us to assess how far variation in alpha can go in explaining various trends. In an extension exercise, I'll compare the moments produced by a model with heterogeneity in α to one with heterogeneity in TFP to one with both dimensions of heterogeneity. For completeness, the Kernel densities estimation of the distribution of the Solow residuals under the two specifications are reported in Figure 4.18.

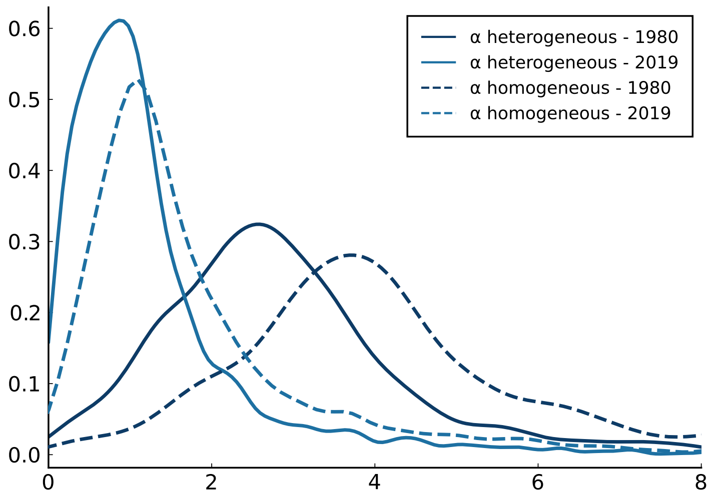


Figure A.21: Solow residual: homogeneous vs. heterogeneous α.

Note: Source is Compustat. Sample is 1980-2019. Solow residual is estimated according to equation (9). Heterogeneous corresponds to firm-year-specific α. Homogeneous corresponds to α set to 0.12 for all firms and all years.

A.5.2 Markup Estimation

Output elasticity to inputs is a critical variable to estimate markups following the production-function approach pioneered by Hall (1988). With this method, if at least one input is set statically and if firms are price takers for the said input, then the first-order condition that determines demand for the input can be inverted to isolate (and estimate) the markup. Considering labor as the variable input yields the expression: $\mu_{i,t} = \zeta^L_{i,t} \frac{p_{i,t} Y_{i,t}}{w_t L_{i,t}}$ where μ is the markup, the first component on the right-hand side, $\zeta^L_{i,t} = \frac{\partial \ln(Y_{i,t})}{\partial \ln(L_{i,t})}$, is the elasticity of output to labor and the second component is the inverse labor share. Under my CES specification detailed in equation (8) this translates into:

$$\mu_{i,t} = \frac{\zeta^L_{i,t}}{\lambda^L_{i,t}} \tag{36}$$

Where I denoted the labor share λ^L. Supposing that λ^L is observable this leaves only ζ to be estimated. Like de Loecker and Warzynski (2012), they propose to rely on the two-stage GMM method of Ackerberg et al. (2015). It consists in estimating the parameters of a production function by first purging it from unobserved productivity shocks and second to retrieve an estimator for ζ using an instrumental-variable general method of moments[41].

[41] This method has encountered a few critiques. First, Klette and Griliches (1996) show that when the two-stage GMM method is applied using revenue data rather than quantity the estimators of output elasticities to inputs are biased downwards. Bond, Hashemi, Kaplan, and Zoch (2021) claim that this bias is large enough to render any estimation of markups uninformative. However, de Ridder et al. (2022) show theoretically and using quantity measures from French administrative data that markups estimated using revenue data correlate positively with true markups. Second, the two-stage GMM retrieves a weighted average of firms'

This method allows to retrieve a sector-specific measure of ζ. Instead, I show that firms' output elasticity to capital differ. I therefore propose to estimate firms' markups using the identity of equation (36) with the firm- and year-specific ζ estimated in section **??** and to compare my resulting measure for $\mu_{i,t}$ with the baseline measure of de Loecker et al. (2020).

In Figure A.22 I report the evolution of the top decile, the top quartile and the median of the distribution of firms' markups under the specification of de Loecker et al. (2020) (homogeneous ζ - red lines) and under mine (heterogeneous ζ - blue lines). The conclusion of rising dispersion in firms' markups resists remarkably. However, the three moments reported are shifted downwards when firm-specific ζ are accounted for and their growth rates are slightly diminished. The sharpest dampening (in level and growth rate) is for the firms located at the upper tail of the distribution. This mechanically translates into a dampening of the sales-weighted average markup as pictured in Figure 3.15.

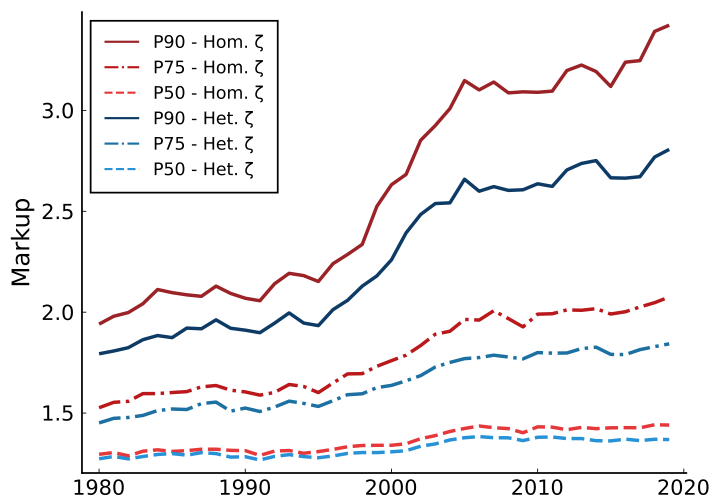


Figure A.22: Distribution of firms' markups: Homogeneous vs. Heterogeneous ζ.

Note: Source is Compustat. Sample is 1980-2019. Markups are estimated following equation (36). Hom. ζ stands for the case where ζ in equation (36) is set to the median ζ. Het. ζ stands for the case where ζ^L in equation (36) is estimated at the firm-level as the cost share of labor. P90 stands for 90th percentile.

Perhaps a more interesting variable to look at is the aggregate markup as it is the variable that impacts misallocation and potential inefficiencies in the aggregate. The aggregation of markups will be studied in detail when we turn to the model (in section 4.2). Note already that under the heterogeneous ζ specification, the aggregation identity derived in Atkeson and Burstein (2008), Grassi (2017) or Edmond et al. (2015) that stipulates that the inverse of the aggregate markup is equal to the domar-weighted harmonic average of firm-level markups holds true. However, the alternative identity sometimes put forward does not. This alternative stipulates that the aggregate markup is equal to the sum of firms'

true ζ as an estimator, mechanically biasing upwards the markups of firms with high true ζ and downwards the markups of firms with low true ζ (Foster, Haltiwanger, and Tuttle, 2022).

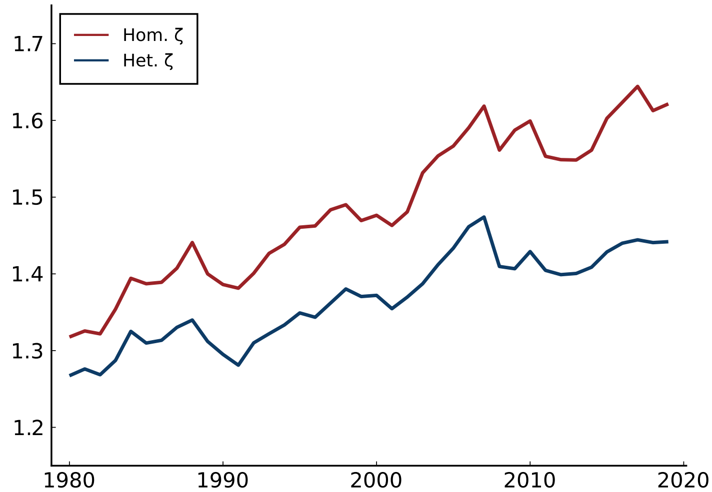


Figure A.23: Sales-weighted average markup: Homogeneous vs. Heterogeneous ζ.

Note: Source is Compustat. Sample is 1980-2019. Markups are estimated following equation (36). Hom. ζ stands for the case where ζ in equation (36) is set to the median ζ. Het. ζ stands for the case where ζ^L in equation (36) is estimated at the firm-level as the cost share of labor.

markups weighted by their input share (capital or labor indifferently). Intuitively, this latter aggregation identity does not hold because when ζ is heterogeneous across firms, input shares aren't the same depending on the input in question.[42] The aggregation identity that I use here is hence the first one. Denoting the aggregate markup as $\mathcal{M}_t$ and supposing a CES demand structure it reads:

$$\frac{1}{\mathcal{M}_t} = \int_0^1 \frac{s_{i,t}}{\mu_{i,t}} di \tag{37}$$

Where $s_{i,t} = \frac{p_{it}Y_{it}}{p_t Y_t}$ is the sales share of firm i to the aggregate. I plot the resulting aggregate markup under the homogeneous ζ specification of de Loecker et al. (2020) (in red) and under my specification (in blue) in Figure 3.16. Again, my measure of the aggregate markup is a downward shifted version of that of de Loecker et al. (2020) but both happen to display similar positive growth rate.

A priori, accounting for firm-level ζ could modify the relationship between market shares and markups. Indeed, allowing for firm-specific ζ mechanically makes my measure of markups for highly capital-intensive firms lower compared to a measure where all firms are supposed to share a common (lower) ζ (see equation (36)). Yet we have seen that capital-intensive firms have seen their market share rise since the 1980s. This matters because market shares are the standard microfoundation for endogenous markups (in models that build on Atkeson and Burstein (2008), Kimball (1995) or Klenow and Willis (2016)).

[42]I discuss further this issue in the results section 4.2 where I provide aggregation identities and prove this claim in Appendix B.7

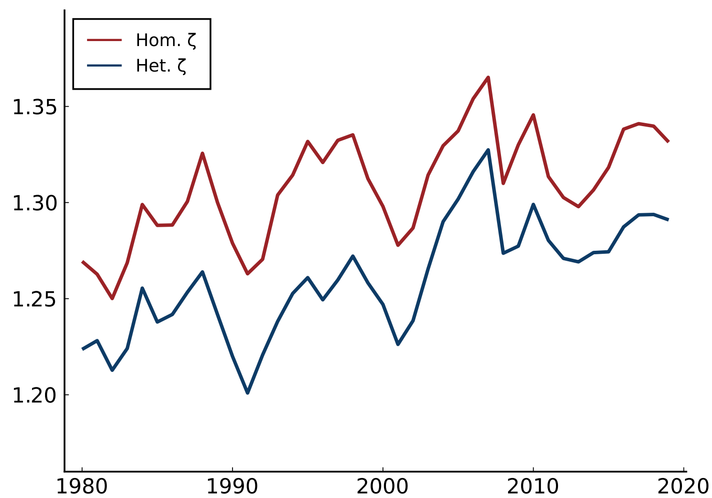


Figure A.24: Aggregate Markup: Homogeneous vs. Heterogeneous ζ.

Note: Source is Compustat. Sample is 1980-2019. Markups are estimated following equation (36). Hom. ζ stands for the case where ζ in equation (36) is set to the median ζ. Het. ζ stands for the case where ζ^L in equation (36) is estimated at the firm-level as the cost share of labor. Aggregate markup is obtained from equation (37).

I test this possibility by regressing, in logarithm, firm-level markups estimated using equation (36) on firm-level capital intensity. The regression equation takes the form:

$$\ln(\mu_{i,t}) = \beta \ln(K_{i,t}/L_{i,t}) + \theta \ln(X_{i,t}) + FE_{i,t} + \nu_{i,t} \tag{38}$$

Where FE denotes fixed effects and ν is the residual. β is the regression coefficient and $X_{i,t}$ are controls. Regressions are conducted over the panel of Compustat firms clustered by years. Results are presented in Table A.4.

In the first three columns, $\mu_{i,t}$ is measured with firm-specific technologies, i.e firm-specific output elasticities to inputs. The last three are the baseline measures of de Loecker et al. (2020) and de Ridder et al. (2022) for which the output elasticity to output is 2D-NAICS sector-specific.

Column (1) of Table A.4 presents the result from a univariate OLS regression admitting a constant term and without fixed effects. It indicates a significantly positive correlation between firm-level log-markups and log-capital intensity. Column (2) adds firm and year fixed-effects, slightly increasing the correlation coefficient but maintaining strong significance. Industry level fixed effects are not displayed because they are subsumed by firm fixed effects. Column (3) includes market share as a control, which as shown in section 2.3, is endogenous to capital intensity. Interestingly, market-share and capital intensity are both significantly and positively related to firm-level markups. In the last 3 columns, the results of the same regressions are reported but with markups estimated with sector-specific output elasticities to inputs. In such a case, the correlation between capital intensity and markups is larger as the bias on firms' output elasticity implies a higher markup for capital-intensive firms and

	Markup (firm-specific technology)			Markup (sector-specific technology)		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.566***			1.005***		
	(0.005)			(0.005)		
K/L	0.108***	0.122***	0.154***	0.271***	0.288***	0.319***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Market share			0.215***			0.210***
			(0.002)			(0.002)
Fixed Effects		Year&Firm	Year&Firm		Year&Firm	Year&Firm
N	105,185	105,185	105,185	105,185	105,185	105,185
R^2	0.027	0.754	0.796	0.146	0.778	0.812

Table A.4: Firm-level regressions: markups and capital intensity.

Note: Source is Compustat. Sample is 1980-2019. Table displays estimators of equation (38). Dependent variable is the firm-level markup measured according to equation (36). K/L is a firms stock of capital per employee measured as the ratio between the sum of "Property, Plant, and Equipment" and "Intangible Assets" per "Employees". Market share is a firms sales to total sales within a year. *** indicate the the coefficient is significant at the 1% threshold.

lower for labor-intensive firms (see Desazars (2025) for proof and discussion).

Importantly, a cross firm regression of markups on capital intensity carried only on the year 1980 also results in a strong positive correlation. This flows from the fact detailed in section 2.3, in 1980 capital-intensive firms are already bigger than labor-intensive firms. Intuitively, this means that the increase in the markup of capital-intensive firms takes the aggregate markup up. If the situation had been reversed, i.e labor-intensive firms were bigger in 1980, the rise in the markup of capital-intensive firms (and concomitant decline in that of the labor-intensive firms) would have resulted into a decrease of the aggregate markup possibly followed by an increase. The inflection point of the aggregate markup dynamic being the moment when capital-intensive and labor-intensive firms have the exact same market shares.

B Theory

B.1 Capital intensity under factor-neutral heterogeneity: TFP, demand and returns to scale

Suppose that a firm produces using the following standard production function: $Y_{i,t} = z_{it}\left(\alpha K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha)L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\gamma_{it}\frac{\psi}{\psi-1}}$ where Y_{it} is the firm's output, K_{it} its capital, L_{it} its labor, α is the share of capital in production, ψ is the elasticity of substitution between capital and labor and γ_{it} are firm and time specific returns to scale. Suppose further that the firm faces the demand curve: $\frac{Y_{it}}{Y_t} = D_{it}(\frac{p_{it}}{p_t})^{-\epsilon}$ where Y_t and p_t are aggregate output and price, p_{it} is a firm's price, D_{it} a demand-shifter, and ϵ the elasticity of substitution of demand across varieties. This firm maximizes its profits by solving the following program:

$$\begin{aligned} \underset{p_{it},Y_{it},K_{it},L_{it}}{max} \quad & p_{it}Y_{it} - r_t^K K_{it} - w_t L_{it} \\ \text{s.t: } & Y_{i,t} = z_{it}\left(\alpha K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha)L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\gamma_{it}\frac{\psi}{\psi-1}} \\ & \frac{Y_{it}}{Y_t} = D_{it}(\frac{p_{it}}{p_t})^{-\epsilon} \end{aligned}$$

The input first-order conditions spell:

$$\begin{cases} r_t^K & = \gamma_{i,t}\alpha mc_{i,t}\left(\frac{Y_{i,t}^{\frac{1}{\gamma_{it}}}}{K_{i,t}}\right)^{\frac{1}{\psi}} Y_{it}^{\frac{\gamma_{it}-1}{\gamma_{it}}} \\ w_t & = \gamma_{i,t}(1-\alpha)mc_{i,t}\left(\frac{Y_{i,t}^{\frac{1}{\gamma_{it}}}}{L_{i,t}}\right)^{\frac{1}{\psi}} Y_{it}^{\frac{\gamma_{it}-1}{\gamma_{it}}} \end{cases}$$

Where mc_{it} is the marginal cost of firm i in time t and is the multiplier of the Lagrangian of the cost-minimization problem. In this case, the output elasticity to capital is: $\zeta_{it}^K = \gamma_{i,t}\alpha\left(\frac{Y_{i,t}^{\frac{1}{\gamma_{it}}}}{K_{i,t}}\right)^{\frac{1-\psi}{\psi}} Y_{it}^{\frac{\gamma_{it}-1}{\gamma_{it}}}$. When taking the ratio of these two conditions, Y_{it} and γ_{it} simplify and the capital intensity identity reads:

$$\frac{K_{i,t}}{L_{i,t}} = \left(\frac{\alpha}{(1-\alpha)}\frac{w_t}{r_t^K}\right)^{\psi}$$

Clearly, capital intensity is orthogonal to demand shifters, TFP and returns to scale as

$Y_{i,t}$ and γ_{it} have simplified.

B.2 Heterogeneous labor-augmenting productivity

In this section I show that if capital and labor are complementary, heterogeneous labor-augmenting productivity can yield heterogeneous capital intensity but cannot satisfy the correlation between capital intensity and size.

Suppose that firms operate the same technology: $Y_{i,t} = \left(\alpha K_{i,t}{}^{\frac{\psi-1}{\psi}} + (1-\alpha) L_{i,t}{}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$, but can satisfy their labor demand using two types of workers. The first type is the numeraire, has labor augmenting productivity of one so that he supplies $L_{i,t} = N_t$ for a wage w_t. The second type is endowed with labor-augmenting productivity B_t so that he supplies: $L_{i,t} = B_t N_t^B$ at wage w_t^B. The choice of worker of firm i affects the firm's capital intensity:

$$\begin{cases} \frac{K_{i,t}}{L_{i,t}} = \frac{K_{i,t}}{N_t} = \left(\frac{\alpha}{(1-\alpha)} \frac{w_t}{r_t^K}\right)^{\psi} \\ \frac{K_{i,t}}{L_{i,t}} = \frac{K_{i,t}}{B_t N_t^B} = \left(\frac{\alpha}{(1-\alpha)} \frac{w_t^B/B_t}{r_t^K}\right)^{\psi} \end{cases}$$

Recall that observed capital intensity is capital per worker, that is $\frac{K_{i,t}}{N_t}$ and $\frac{K_{i,t}}{N_t^B}$. As demonstrated by Acemoglu (2002), under complementarity of inputs ($\psi < 1$), if $B_t > 1$, then $N_t^B < N_t$ and so firms that hire more productive workers feature a higher capital intensity. However, for the two equalities above to hold, the incentive compatibility condition $w_t^B = B_t w_t$ must hold, that is the productive worker is paid a skill premium equal to its productivity relative to the numeraire. And, denoting as $mc_{i,t}^B$ the marginal cost of a firm that hires the worker endowed with productivity B_t we have:

$$\begin{aligned} mc_{i,t}^B &= \left(\alpha^{\psi} r_t^{K(1-\psi)} + (1-\alpha)^{\psi} \left(w_t^B/B_t\right)^{(1-\psi)}\right)^{\frac{1}{1-\psi}} \\ &= \left(\alpha^{\psi} r_t^{K(1-\psi)} + (1-\alpha)^{\psi} w_t{}^{(1-\psi)}\right)^{\frac{1}{1-\psi}} = mc_{i,t} \end{aligned}$$

So firms share the same marginal cost, irrespective of the type of worker that they hire: $mc_{i,t}^B = mc_{i,t}$ and heterogeneity in market shares therefore cannot result from heterogeneous labor-augmenting productivity.

B.3 Monopsony

Suppose that firm i has monopsonistic power so that it pays a wage $\theta w_t < w_t$. Then, its true $\hat{\alpha}_{i,t}$ is given by:

$$\frac{\hat{\alpha}_{i,t}}{1-\hat{\alpha}_{i,t}} = \frac{r_t^K}{\theta w_t}\left(\frac{K_{i,t}}{L_{i,t}}\right)^{\frac{1}{\psi}}$$

And the estimation bias is:

$$\hat{b}_{i,t} = \frac{\frac{\hat{\alpha}_{i,t}}{1-\hat{\alpha}_{i,t}}}{\frac{\alpha_{i,t}}{1-\alpha_{i,t}}} = \frac{1}{\theta} > 1$$

If large firms have monopsonistic power, as per empirical evidence (see for example Benmelech, Bergman, and Kim (2022) or Yeh, Macaluso, and Hershbein (2022)), then my estimate of α is biased downwards and dispersion is even higher. Note though that in Figure A.15 in Appendix A.4, I use a measure of firm-specific wages to compute the output elasticity to capital and find a distribution very close to the one where wages are sector-specific. On the other hand, small and labor-intensive firms having monopsonitic power would contradict empirical evidence.

B.4 Equilibrium equations

1. The Household.
   Supposing a Greenwood et al. (1988) utility function, the household solves.

$$\max_{\{C_t,L_t,X_t,B_{t+1},K_{t+1}\}_{t\in[0:+\infty[}} \sum_{t=0}^{+\infty} \frac{\left(C_t - \frac{L_t^{1+\eta}}{1+\eta}\right)^{1-\gamma}}{1-\gamma}$$

$$\begin{aligned}\text{s.t: } & p_t(C_t + \xi_t X_t) + B_{t+1} = (1+r_t)B_t + w_t L_t + r_t^K K_t + \pi_t \\ & K_{t+1} = (1-\delta)K_t + X_t \\ & K_0, B_0\end{aligned}$$

Solving for FOCs, yields 6 equations for the equilibrium:

$$\text{BC:} \quad p_t(C_t + \xi_t X_t) + B_{t+1} = (1+r_t)B_t + w_t L_t + r_t^K K_t + \pi_t \tag{H.1}$$

$$\text{Accumulation:} \quad K_{t+1} = (1-\delta)K_t + X_t \tag{H.2}$$

$$\text{Euler:} \quad \beta(1+r_{t+1})\frac{p_t}{p_{t+1}^C}\left(C_{t+1} - \frac{L_{t+1}^{1+\eta}}{1+\eta}\right)^{-\gamma} = \left(C_t - \frac{L_t^{1+\eta}}{1+\eta}\right)^{\gamma} \tag{H.3}$$

$$\text{Intra:} \quad L_t^\eta = \frac{w_t}{p_t} \tag{H.4}$$

$$\text{NA:} \quad r_{t+1}^K = (1 + r_{t+1})\xi_t p_t + (1-\delta)\xi_{t+1} p_{t+1}^C \tag{H.5}$$

2. Market structure and market clearing.
The economy aggregates from atomistic sectors:

$$C_t = \left(\int_0^1 c_{j,t}{}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} \tag{MS.1}$$

$$X_t = \frac{1}{\xi_t} \left(\int_0^1 x_{j,t}{}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} \tag{MS.2}$$

$$p_t^X = \xi_t \left(\int_0^1 p_{j,t}{}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} = \xi_t p_t \tag{MS.3}$$

Defining:

$$Y_t = C_t + \xi_t X_t \quad \text{and} \quad Y_{j,t} = c_{j,t} + x_{j,t} \tag{MS.4}$$

Yields:

$$\frac{p_{j,t}}{p_t} = \left(\frac{Y_{j,t}}{Y_t} \right)^{-\frac{1}{\sigma}} \tag{MS.5}$$

Sectors aggregate from non-atomistic intermediate firms:

$$Y_{j,t} = \left(\sum_{i=1}^N Y_{i,j,t}{}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{MS.6}$$

$$p_{j,t} = \left(\sum_{i=1}^N p_{i,j,t}{}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \tag{MS.7}$$

$$\frac{p_{i,j,t}}{p_{j,t}} = \left(\frac{Y_{i,j,t}}{Y_{j,t}} \right)^{-\frac{1}{\epsilon}} \tag{MS.8}$$

Market clears for labor, capital, profits, bonds:

$$\int_0^1 \sum_{i=0}^N K_{i,j,t} dj = K_t \tag{MS.9}$$

$$\int_0^1 \sum_{i=0}^{N} L_{i,j,t} dj = L_t \tag{MS.10}$$

$$\pi_t = \int_0^1 \sum_i \pi_{i,j,t} dj \tag{MS.11}$$

And zero-net supply of bonds:

$$B_t = 0, \forall t \tag{MS.12}$$

3. Firms.

$$\text{Tech}: \quad Y_{i,j,t} = \left(\alpha_{i,j,t} K_{i,j,t}{}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,j,t}) L_{i,j,t}{}^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}} \tag{F.1}$$

$$mc_{i,j,t} = \left(\alpha_{i,j,t}{}^{\psi} r_t^{K(1-\psi)} + (1-\alpha_{i,j,t})^{\psi} w_t{}^{(1-\psi)} \right)^{\frac{1}{1-\psi}} \tag{F.2}$$

$$\text{K-dem}: \quad r_t^K = \alpha_{i,j,t} mc_{i,j,t} \left(\frac{Y_{i,j,t}}{K_{i,j,t}} \right)^{\frac{1}{\psi}} \tag{F.3}$$

$$\text{L-dem}: \quad w_t = (1-\alpha_{i,j,t}) mc_{i,j,t} \left(\frac{Y_{i,j,t}}{L_{i,j,t}} \right)^{\frac{1}{\psi}} \tag{F.4}$$

$$\text{pricing}: \quad p_{i,j,t} = \frac{\varepsilon(s_{i,j,t})}{\varepsilon(s_{i,j,t}) - 1} mc_{i,j,t} = \mu(s_{i,j,t}) mc_{i,j,t} \tag{F.5}$$

$$\text{markup}: \quad \varepsilon(s_{i,j,t}) = \left(\frac{1}{\epsilon} - (\frac{1}{\epsilon} - \frac{1}{\sigma}) s_{i,j,t} \right)^{-1} \tag{F.6}$$

$$\text{profits}: \quad \pi_{i,j,t} = \left(1 - \frac{1}{\mu(s_{i,j,t})} \right) p_{i,j,t} y_{i,j,t} \tag{F.7}$$

And value-added shares are defined as:

$$s_{i,j,t} := \frac{p_{i,j,t} Y_{i,j,t}}{p_{j,t} Y_{j,t}} \tag{F.8}$$

B.5 Solving the model

First, given a normalized value $w_t = 1, \forall t$, calibrate r_t^K for $\frac{r_t^K}{w_t}$ to follow the path of the RCK in the data with one point as the reference point (for example $r_{1980}^K = 0.1$). Then deduce the distribution of alpha from Compustat using the K/L equation of the model.

Second, with r^K, w_t and the exact distribution of α in hand, build Kernel densities of α per sector and build the oligopolistic structure as detailed in the calibration; 100 sectors with 10 firms each drawing from the Kernel density of α attached to its sector. The marginal cost of each firm can be deduced and the intermediate firm block can be solved numerically. Using equations (F.5), (F.6), (F.8) and (MS.8) and The system of equation boils down to:

$$\mu_{i,j,t} = \frac{\varepsilon(s_{i,j,t})}{\varepsilon(s_{i,j,t}) - 1}$$
$$s_{i,j,t} = \Big(\frac{\mu_{i,j,t} mc_{i,j,t}}{p_{j,t}}\Big)^{1-\epsilon}$$

$p_{j,t}$ has to be guessed prior to solving and is then updated using (MS.7) until convergence. This allows to have a solution for: $s_{i,j,t}, p_{i,j,t}, p_{j,t}, \mu(s_{i,j,t}), p_t, p_t^X, s_{j,t}$ (and deduce the implied ξ_t to be compared to the RCK with cost of consumption series).

Third, isolating $L_{i,j,t}$ in (F.4), integrating and summing and merging it with (H.4) yields:

$$\left(\frac{w_t}{p_t}\right)^{\frac{1}{\eta}} = \int_0^1 \sum_i \left(\frac{(1-\alpha_{i,j,t}) mc_{i,j,t}}{w_t}\right)^{-\psi} Y_{i,j,t} dj$$

Replacing $Y_{i,j,t}$ using the demand functions (MS.5) and (MS.8) gives a value for aggregate Y:

$$Y_t = \frac{w_t^{\frac{1}{\eta}+\frac{1}{\psi}}}{p_t^{\frac{1}{\eta}+\sigma} \int_0^1 \sum_i ((1-\alpha_{i,j,t}) mc_{i,j,t})^{-\psi} \, p_{i,j,t}^{-\epsilon} p_{j,t}^{(\sigma-\epsilon)} dj}$$

Fourth, aggregate K can be deduced from the BC of the household. To get there, use (F.7), (F.8), (MS.4), (MS.11), (MS.12), (H.1), (H.4):

$$r_t^K K_t = \frac{p_t}{\mathcal{M}_t} Y_t - w_t L_t$$

From there, X_t is given by (H.2) and every firm- or sector- level quantity can be retrieved using (F.3) or (F.4) coupled with (MS.5), (MS.8) and aggregate Y. Alternatively, we could have passed ξ_t first to the model and then verified r_t^K. The first way is more convenient in that the link to the empirical exercise is more straightforward.

B.6 The choice of $\alpha_{i,j,t}$

The fully-fledged problem of the firms when choosing $\alpha_{i,j,t}$ in time t writes:

$$
\begin{aligned}
&\underset{\alpha_{i,j,t}}{max} \quad \sum_{\tau=0}^{+\infty} \beta^{\tau} \left(\pi_{i,j,t+\tau} - sc_{i,j}(\Delta\alpha_{i,j,t})\right) \\
\text{s.t:}\ &\pi_{i,j,t} = \left(1 - \frac{1}{\mu(s_{i,j,t})}\right) p_{i,j,t} Y_{i,j,t} \\
&\mu(s_{i,j,t}) = \frac{\varepsilon(s_{i,j,t})}{\varepsilon(s_{i,j,t}) - 1} \\
&\varepsilon(s_{i,j,t}) = \begin{cases} \epsilon - (\epsilon - \sigma) s_{i,j,t} & \text{under Bertrand competition} \\ \left(\frac{1}{\epsilon} - (\frac{1}{\epsilon} - \frac{1}{\sigma}) s_{i,j,t}\right)^{-1} & \text{under Cournot competition} \end{cases} \\
&s_{i,j,t} = \left(\frac{p_{i,j,t}}{p_{j,t}}\right)^{1-\epsilon} \\
&p_{i,j,t} = \mu(s_{i,j,t}) mc_{i,j,t} \\
&p_{j,t} = \left(\sum_{i=1}^{N} {p_{i,j,t}}^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} \\
&mc_{i,j,t} = \left({\alpha_{i,j,t}}^{\psi} {r_t^K}^{(1-\psi)} + (1 - \alpha_{i,j,t})^{\psi} {w_t}^{(1-\psi)}\right)^{\frac{1}{1-\psi}} \\
&Y_{i,j,t} = \left(\frac{p_{i,j,t}}{p_{j,t}}\right)^{-\epsilon} Y_{j,t} \\
&Y_{j,t} = \left(\sum_{i=1}^{N} {Y_{i,j,t}}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}} \\
&\alpha_{i,j,0},\ \ \alpha_{i,j,t} \in [\underline{\alpha}_j, \overline{\alpha}_j]
\end{aligned}
$$

Which can be rewritten as:

$$
\underset{\alpha_{i,j,t}}{max} \quad \sum_{\tau=0}^{+\infty} \beta^{\tau} \left(\frac{p_{i,j,t} D(p_{i,j,t})}{\varepsilon(s_{i,j,t})} - \varphi_{i,j} d_t(\Delta\alpha_{i,j,t})\right)
$$

For $D(p_{i,j,t}) = \left(\frac{p_{i,j,t}}{p_{j,t}}\right)^{-\epsilon} Y_{j,t}$ and a pick of $\varepsilon(s_{i,j,t})$. This yields the FOC:

$$-\frac{\partial \varepsilon(s_{i,j,t})}{\partial \alpha_{i,j,t}} \frac{1}{\varepsilon^2_{i,j,t}} p_{i,j,t} D(p_{i,j,t}) + \frac{1}{\varepsilon(s_{i,j,t})} \frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}} D(p_{i,j,t}) + \frac{1}{\varepsilon(s_{i,j,t})} p_{i,j,t} \frac{\partial D(p_{i,j,t})}{\partial \alpha_{i,j,t}} = \varphi_{i,j}(d_t - \beta d_{t+1})$$

Leaving us with three derivatives to determine: $\frac{\partial \varepsilon(s_{i,j,t})}{\partial \alpha_{i,j,t}}$, $\frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}}$, $\frac{\partial D(p_{i,j,t})}{\partial \alpha_{i,j,t}}$. I rewrite everything in terms of definitions and $\frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}}$ which I characterize first:

$$\frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}} = \frac{\partial \mu(s_{i,j,t})}{\partial \alpha_{i,j,t}} mc_{i,j,t} + \mu(s_{i,j,t}) \frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}$$

Where $\frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}$ can be computed. For $\frac{\partial \mu(s_{i,j,t})}{\partial \alpha_{i,j,t}}$:

$$\frac{\partial \mu(s_{i,j,t})}{\partial \alpha_{i,j,t}} = \frac{\partial \mu(s_{i,j,t})}{\partial s_{i,j,t}} \frac{\partial s_{i,j,t}}{\partial p_{i,j,t}} \frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}}$$

So that:

$$\frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}} = \frac{\mu(s_{i,j,t}) \frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}}{1 - \frac{\partial \mu(s_{i,j,t})}{\partial s_{i,j,t}} \frac{\partial s_{i,j,t}}{\partial p_{i,j,t}} mc_{i,j,t}}$$

Where every derivative can be computed from the definitions. Turning to: $\frac{\partial \varepsilon(s_{i,j,t})}{\partial \alpha_{i,j,t}}$:

$$\frac{\partial \varepsilon(s_{i,j,t})}{\partial \alpha_{i,j,t}} = \frac{\partial \varepsilon(s_{i,j,t})}{\partial s_{i,j,t}} \frac{\partial s_{i,j,t}}{\partial p_{i,j,t}} \frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}}$$

Which can all be solved for from the definitions. Finally:

$$\begin{aligned} \frac{\partial D(p_{i,j,t})}{\partial \alpha_{i,j,t}} &= \frac{\partial D(p_{i,j,t})}{\partial p_{i,j,t}} \frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}} \\ &= \frac{\partial D(p_{i,j,t})}{\partial p_{i,j,t}} \frac{\partial \ln(p_{i,j,t})}{\partial \ln(p_{i,j,t})} \frac{\partial \ln(D(p_{i,j,t}))}{\partial \ln(D(p_{i,j,t}))} \frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}} \\ &= \underbrace{\frac{\partial \ln(D(p_{i,j,t}))}{\partial \ln(p_{i,j,t})}}_{-\varepsilon(s_{i,j,t})} \underbrace{\frac{\partial \ln(p_{i,j,t})}{\partial p_{i,j,t}}}_{\frac{1}{p_{i,j,t}}} \underbrace{\frac{\partial D(p_{i,j,t})}{\partial \ln(D(p_{i,j,t}))}}_{D(p_{i,j,t})} \frac{\partial p_{i,j,t}}{\partial \alpha_{i,j,t}} \end{aligned}$$

The marginal return to changing α can hence be rewritten:

$$\underbrace{\frac{\mu(s_{i,j,t})\frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}}{1-\frac{\partial \mu(s_{i,j,t})}{\partial s_{i,j,t}}\frac{\partial s_{i,j,t}}{\partial p_{i,j,t}}mc_{i,j,t}}}_{\frac{\partial \mu(s_{i,j,t})}{\partial \alpha_{i,j,t}}}\left(\frac{\partial \varepsilon(s_{i,j,t})}{\partial s_{i,j,t}}\frac{\partial s_{i,j,t}}{\partial p_{i,j,t}}\frac{p_{i,j,t}D(p_{i,j,t})}{\varepsilon^2_{i,j,t}}+\frac{D(p_{i,j,t})}{\varepsilon(s_{i,j,t})}+\frac{p_{i,j,t}}{\varepsilon(s_{i,j,t})}\frac{\partial D(p_{i,j,t})}{\partial p_{i,j,t}}\right)$$

There are hence 4 derivatives to calculate from the definitions listed in the constraint of the fully-fledged problem: $\frac{\partial \mu(s_{i,j,t})}{\partial \varepsilon(s_{i,j,t})}$, $\frac{\partial \varepsilon(s_{i,j,t})}{\partial s_{i,j,t}}$, $\frac{\partial s_{i,j,t}}{\partial p_{i,j,t}}$, $\frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}$:

$$\frac{\partial \mu(s_{i,j,t})}{\partial \varepsilon(s_{i,j,t})}=-\frac{1}{\left(\varepsilon(s_{i,j,t})-1\right)^2}$$

$$\frac{\partial \varepsilon(s_{i,j,t})}{\partial s_{i,j,t}}=\begin{cases}-(\epsilon-\sigma) & \text{Under Bertrand competition}\\ -(\frac{1}{\sigma}-\frac{1}{\epsilon})\varepsilon(s_{i,j,t})^2 & \text{Under Cournot competition}\end{cases}$$

$$\frac{ds_{i,j,t}}{dp_{i,j,t}}=(1-\epsilon)\frac{s_{i,j,t}}{p_{i,j,t}}\left(1-s_{i,j,t}\right)$$

$$\frac{dmc_{i,j,t}}{d\alpha_{i,j,t}}=\frac{\psi}{1-\psi}mc^{\psi}_{i,j,t}\left(\alpha^{\psi-1}_{i,j,t}r_t^{K(1-\psi)}-(1-\alpha_{i,j,t})^{\psi-1}w_t^{(1-\psi)}\right)$$

The final expression for the marginal return to α can be written, under Cournot competition:

$$\frac{Y_{i,j,t}}{\mu(s_{i,j,t})}\frac{\psi}{1-\psi}mc^{\psi}_{i,j,t}\left(\alpha^{\psi-1}_{i,j,t}r_t^{K(1-\psi)}-(1-\alpha_{i,j,t})^{\psi-1}w_t^{(1-\psi)}\right)\frac{(\frac{1}{\sigma}-\frac{1}{\epsilon})(\epsilon-1)s_{i,j,t}(1-s_{i,j,t})\mu(s_{i,j,t})-1}{(\frac{1}{\sigma}-\frac{1}{\epsilon})(\epsilon-1)s_{i,j,t}(1-s_{i,j,t})\mu(s_{i,j,t})+1}$$

Note that in a case where markups aren't endogenous and fixed to μ, the FOC boils down to:

$$(1-\epsilon)\,Y_{i,j,t}\mu\frac{\partial mc_{i,j,t}}{\partial \alpha_{i,j,t}}=\varphi_{i,j}(d_t-\beta d_{t+1})$$

Which yields:

$$(1-\epsilon)\,Y_{i,j,t}\mu\frac{\psi}{1-\psi}mc^{\psi}_{i,j,t}\left(\alpha^{\psi-1}_{i,j,t}r_t^{K(1-\psi)}-(1-\alpha_{i,j,t})^{\psi-1}w_t^{(1-\psi)}\right)=\varphi_{i,j}(d_t-\beta d_{t+1})$$

B.7 Aggregation

In this section, I derive some aggregation properties of the model. First, I show that while one cannot derive an aggregate production function, it is possible to show that aggregate output features constant returns to scale with respect to aggregate capital and labor. This is a necessary condition to derive the aggregate markup. I then derive the aggregate labor and capital shares and present three different methods to aggregate markups.

B.7.1 Constant aggregate returns to scale

Because firms don't share the same output elasticities to inputs, an aggregate production function cannot be derived analytically. However, some characteristics can. In particular, one can show that the relationship between aggregate output and aggregate capital and labor is homogeneous of degree one. To see this suppose that K_t and L_t are both scaled by a factor of λ. Then, by market clearing, we have:

$$\begin{cases} \lambda K_t & = \int_0^1 \sum_{i=0}^{N} \lambda K_{i,j,t} dj \\ \lambda L_t & = \int_0^1 \sum_{i=0}^{N} \lambda L_{i,j,t} dj \end{cases}$$

Because every firm's production technology is supposed to feature constant returns to scale, this implies that every $Y_{i,j,t}$ is scaled by the scalar λ. Because the CES aggregators are homogeneous of degree 1, we obtain aggregate output as:

$$\left(\int_0^1 \sum_{i=0}^{N} \left(\lambda Y_{i,j,t}\right)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}} = \lambda \left(\int_0^1 \sum_{i=0}^{N} Y_{i,j,t}^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}} = \lambda \left(C_t + \xi_t X_t\right) = \lambda Y_t$$

That is, there exists a function F_t such that: $Y_t = F_t(K_t, L_t)$ where F_t is homogeneous of degree one. This implies that $mc_t = AC_t$ and hence that aggregate profits can be written: $\pi_t = \left(1 - \frac{1}{\mathcal{M}_t}\right) p_t^c Y_t$ where $\mathcal{M}_t$ is the aggregate markup.

B.7.2 Aggregate markup.

There are three ways to express the aggregate markup: sales-weighted, cost-weighted and input-weighted. I show that the first two identities in this model are consistent with the ones already derived in the literature (Grassi, 2017; Edmond et al., 2015; Baqaee and Farhi, 2020; Edmond et al., 2023). However, the third one differs as it is a generalization of the one sometimes used in the literature.

In the preceding section, I have shown that there exists an aggregate markup $\mathcal{M}_t$, defined implicitly as $\pi_t = (1 - \frac{1}{\mathcal{M}_t}) p_t Y_t$. Where π_t are aggregate profits, p_t aggregate price and Y_t aggregate output.

Aggregate markup using sales weights. By market-clearing we have: $\pi_t = \int_0^1 \sum\limits_{i=0}^{N} \pi_{i,j,t} dj$, then:

$$\begin{aligned}
\left(1-\frac{1}{\mathcal{M}_t}\right) p_t Y_t &= \int_0^1 \sum_{i=0}^{N} \left(1-\frac{1}{\mu_{i,j,t}}\right) p_{i,j,t} Y_{i,j,t} \\
\Leftrightarrow 1-\frac{1}{\mathcal{M}_t} &= 1-\int_0^1 s_{j,t} \sum_{i=0}^{N} \frac{s_{i,j,t}}{\mu_{i,j,t}} \\
\Leftrightarrow \frac{1}{\mathcal{M}_t} &= \int_0^1 s_{j,t} \sum_{i=0}^{N} \frac{s_{i,j,t}}{\mu(s_{i,j,t})} dj
\end{aligned}$$

The aggregate markup is the inverse of the sales-weighted harmonic average of individual markups. This aggregation result can already be found in Grassi (2017), Edmond et al. (2015) or Edmond et al. (2023) and holds under oligopolistic competition or monopolistic competition with endogenous markups (with a Kimball (1995) aggregator for example).

Aggregate markup using cost weights. Market clearing on output, inputs and profits implies that:

$$\begin{aligned}
p_t Y_t &= \int_0^1 \sum_{i=0}^{N} p_{i,j,t} Y_{i,j,t} dj = \int_0^1 \sum_{i=0}^{N} r_t^K K_{i,j,t} + \int_0^1 \sum_{i=0}^{N} w_t L_{i,j,t} + \int_0^1 \sum_{i=0}^{N} \pi_{i,j,t} \\
&= r_t^K K_t + w_t L_t + \left(1-\frac{1}{\mathcal{M}_t}\right) p_t Y_t \\
\Leftrightarrow \mathcal{M}_t &= \frac{TC_t}{p_t Y_t}
\end{aligned}$$

But notice that $p_{i,j,t} Y_{i,j,t} = \mu(s_{i,j,t}) mc_{i,j,t} Y_{i,j,t} = \mu(s_{i,j,t}) TC_{i,j,t}$, hence we get:

$$\mathcal{M}_t = \frac{\int_0^1 \sum\limits_{i=0}^{N} p_{i,j,t} Y_{i,j,t} dj}{TC_t} = \int_0^1 \sum_{i=0}^{N} \mu(s_{i,j,t}) \frac{TC_{i,j,t}}{TC_t} dj$$

The aggregate markup is the total-cost-share arithmetic average of individual markups.

Aggregate markup using input weights. A firm's individual labor share is defined as: $\lambda_{i,j,t}^L = \frac{w_t L_{i,j,t}}{p_{i,j,t} Y_{i,j,t}}$. This can be rewritten: $\lambda_{i,j,t}^L = \frac{\zeta_{i,j,t}^L}{\mu(s_{i,j,t})}$, where $\zeta_{i,j,t}^L$ is both the cost share of labor ($\frac{w_t L_{i,j,t}}{TC_{i,j,t}}$) and the output elasticity to labor ($\frac{\partial \ln(Y_{i,j,t})}{\partial \ln(L_{i,j,t})}$). Similarly, the aggregate labor share can be written: $\lambda_t^L = \frac{w_t L_t}{p_t Y_t} = \frac{\zeta_t^L}{\mathcal{M}_t}$ where ζ_t^L is the aggregate cost share of labor: $\frac{w_t L_t}{p_t Y_t}$.

Taking the ratio of the aggregate and firm-level labor share and summing over the firms and sectors yields:

$$\mathcal{M}_t = \int_0^1 \sum_{i=0}^{N} \frac{\zeta_t^L}{\zeta_{i,j,t}^L} \frac{L_{i,j,t}}{L_t} \mu(s_{i,j,t}) dj$$

That is, the aggregate markup is the weighted arithmetic average of firms' markups where the weights are the product of the relative cost share of labor and the employment share. The symmetric holds for capital because $\int_0^1 \sum_{i=0}^{N} \frac{\zeta_t^L}{\zeta_{i,j,t}^L} \frac{L_{i,j,t}}{L_t} \mu(s_{i,j,t}) dj = \int_0^1 \sum_{i=0}^{N} \frac{\zeta_t^K}{\zeta_{i,j,t}^K} \frac{K_{i,j,t}}{K_t} \mu(s_{i,j,t}) dj$. Notice that if firms have homogeneous $\zeta_{i,j,t}^L$ then $\zeta_{i,j,t}^L = \zeta_t^L, \forall i, j, t$ and the aggregation identity collapses to: $\mathcal{M}_t = \int_0^1 \sum_{i=0}^{N} \frac{L_{i,j,t}}{L_t} \mu(s_{i,j,t}) dj$. Here, this critically does not hold because: $\int_0^1 \sum_{i=0}^{N} \frac{L_{i,j,t}}{L_t} \mu(s_{i,j,t}) dj \neq \int_0^1 \sum_{i=0}^{N} \frac{K_{i,j,t}}{K_t} \mu(s_{i,j,t}) dj$. In the data, these two aggregation identities indeed don't coincide (see Desazars (2025)).

B.7.3 The micro and aggregate labor shares.

A firm's labor share is defined by: $\lambda_{i,j,t}^L = \frac{w_t L_{i,j,t}}{p_{i,j,t} Y_{i,j,t}}$. Using (F.4), and plugging the ratio of (F.3) to (F.4) in (F.1) we obtain:

$$\lambda_{i,j,t}^L = \frac{(1 - \alpha_{i,j})}{\mu(s_{i,j,t})} \left(\frac{Y_{i,j,t}}{L_{i,j,t}} \right)^{\frac{1-\psi}{\psi}}$$

Which can be simplified using equation (8) and the capital intensity identity to:

$$\lambda_{i,j,t}^L = \frac{1}{\mu(s_{i,j,t}) \left(1 + \left(\frac{\alpha_{i,j,t}}{1-\alpha_{i,j,t}} \right)^{\psi} RCK^{(1-\psi)} \right)} \tag{F.9}$$

Where, in the baseline $Z_{i,j,t} = 1, \forall \{i, j, t\}$. To find the aggregate labor share, use (MS.10)

$$\begin{aligned} \lambda_t^L = \frac{w_t L_t}{p_t Y_t} &= \frac{\int_0^1 \sum_{i=0}^{N} \frac{w_t L_{i,j_t}}{p_{i,j,t} Y_{i,j,t}} p_{i,j,t} Y_{i,j,t} dj}{p_t Y_t} \\ &= \int_0^1 s_{j,t} \sum_{i=0}^{N} s_{i,j,t} \lambda_{i,j,t}^L dj \end{aligned}$$

As shown in the previous subsection, the aggregate labor share can also be written: $\lambda_t^L = \frac{\zeta_t^L}{\mathcal{M}_t}$ where $\zeta_t^L = \frac{w_t L_t}{TC_t}$ is the aggregate cost share of labor. This allows to aggregate as:

$\lambda_t^L = \int_0^1 \sum_{i=0}^{N} \zeta_{i,j,t}^L \frac{TC_{i,j,t}}{TC_t} dj.$

B.7.4 Aggregate marginal productivities

I have shown in Section B.7.1 that aggregate production has constant returns to scale. This means that the aggregate marginal cost is equal to the aggregate average cost which allows to define the aggregate markup, $\mathcal{M}_t$ as $\pi_t = \left(1 - \frac{1}{\mathcal{M}_t}\right) p_t Y_t$. From market-clearing we have:

$$
\begin{aligned}
p_t Y_t &= \int_0^1 \sum_{i=0}^{N} p_{i,j,t} Y_{i,j,t} dj = \int_0^1 \sum_{i=0}^{N} r_t^K K_{i,j,t} + \int_0^1 \sum_{i=0}^{N} w_t L_{i,j,t} + \int_0^1 \sum_{i=0}^{N} \pi_{i,j,t} \\
&= r_t^K K_t + w_t L_t + \left(1 - \frac{1}{\mathcal{M}_t}\right) p_t Y_t \\
\Leftrightarrow p_t Y_t &= \mathcal{M}_t \left(r_t^K K_t + w_t L_t\right)
\end{aligned}
$$

Defining $MPK_t = \frac{\partial Y_t}{\partial K_t}$ and $MPK_t = \frac{\partial Y_t}{\partial K_t}$, this yields:

$$
\begin{cases}
p_t MPK_t & = \mathcal{M}_t r_t^K \\
p_t MPL_t & = \mathcal{M}_t w_t
\end{cases}
$$

From the aggregation of $\mathcal{M}_t$, we have:

$$
p_t MPK_t = \left(\int_0^1 \sum_{i=0}^{N} \frac{s_{i,j,t}}{\mu_{i,j,t} r_t^K} dj\right)^{-1} = \left(\int_0^1 \sum_{i=0}^{N} \frac{s_{i,j,t}}{p_{i,j,t} MPK_{i,j,t}} dj\right)^{-1}
$$

Defining, similarly to Hsieh and Klenow (2009), the marginal return to the productivity of capital as: $MRPK_t = p_t MPK_t$, this means that the aggregate marginal return to the productivity of capital is the the inverse of the sales-share-weighted harmonic average of firms' MRPK. Alternatively, to get an expression for the marginal productivities purified of prices, one can write:

$$
\begin{cases}
MPK_t & = \frac{1}{mc_t} r_t^K \\
MPL_t & = \frac{1}{mc_t} w_t
\end{cases}
$$

But, since we have seen that aggregate returns to scale are constant, we have: $TC_t = mc_t Y_t = \int_0^1 \sum_{i=0}^{N} TC_{i,j,t} = \int_0^1 \sum_{i=0}^{N} mc_{i,j,t} Y_{i,j,t}$. Therefore:

$$
MPK_t = \frac{r_t^K}{\sum_{i=0}^{N} mc_{i,j,t} \frac{Y_{i,j,t}}{Y_t}} = \left(\sum_{i=0}^{N} \frac{1}{MPK_{i,j,t}} \frac{Y_{i,j,t}}{Y_t}\right)^{-1}
$$

The aggregate marginal productivity of capital is the inverse of the output-share-weighted harmonic average of firms' MPK.

B.7.5 Sales and employment share

From the definition: $\lambda^{EMP}_{i,j,t} = \frac{L_{i,j,t}}{L_{j,t}}$ we have:

$$\begin{aligned}
\lambda^{EMP}_{i,j,t} &= \frac{w_t L_{i,j,t}}{w_t L_{j,t}} \\
&= \frac{\lambda^L_{i,j,t}}{\lambda^L_{j,t}} s_{i,j,t} \\
&= \frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \frac{\zeta^L_{i,j,t}}{\zeta^L_{j,t}} s_{i,j,t}
\end{aligned}$$

Then:

$$\frac{\partial \lambda^{EMP}_{i,j,t}}{\partial s_{i,j,t}} = \varrho_{i,j,t} + \left(\frac{\partial \left(\frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \right)}{\partial s_{i,j,t}} \left(\frac{\zeta^L_{i,j,t}}{\zeta^L_t} \right) + \frac{\partial \left(\frac{\zeta^L_{i,j,t}}{\zeta^L_t} \right)}{\partial s_{i,j,t}} \left(\frac{\mathcal{M}_t}{\mu(s_{i,j,t})} \right) \right) s_{i,j,t}$$

Where:

$$\begin{aligned}
\frac{\partial \left(\frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \right)}{\partial s_{i,j,t}} &= -\left(\frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \right)^2 \left(1 + \frac{\partial \mu(s_{i,j,t})}{\partial s_{i,j,t}} \left(\frac{1}{\mathcal{M}_{j,t}} - \frac{s_{i,j,t}}{\mu(s_{i,j,t})} \right) \right) \\
&= \begin{cases} -\left(\frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \right)^2 \left(1 + (\epsilon - \sigma) \frac{\mu(s_{i,j,t})^2}{\varepsilon(s_{i,j,t})^2} \left(\frac{1}{\mathcal{M}_{j,t}} - \frac{s_{i,j,t}}{\mu(s_{i,j,t})} \right) \right) & \text{under Bertrand.} \\ -\left(\frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \right)^2 \left(1 + \left(\frac{1}{\sigma} - \frac{1}{\epsilon} \right) \mu(s_{i,j,t})^2 \left(\frac{1}{\mathcal{M}_{j,t}} - \frac{s_{i,j,t}}{\mu(s_{i,j,t})} \right) \right) & \text{under Cournot.} \end{cases}
\end{aligned}$$

In both case, $\frac{\partial \left(\frac{\mathcal{M}_{j,t}}{\mu(s_{i,j,t})} \right)}{\partial s_{i,j,t}} < 0$ because $\frac{1}{\mathcal{M}_{j,t}} - \frac{s_{i,j,t}}{\mu(s_{i,j,t})} > 0$. To see this consider:

$$\begin{aligned}
\frac{1}{\mathcal{M}_{j,t}} - \frac{s_{i,j,t}}{\mu(s_{i,j,t})} &= \sum_i \frac{s_{i,j,t}}{\mu(s_{i,j,t})} - \frac{s_{i,j,t}}{\mu(s_{i,j,t})} \\
&= \sum_{k \neq i} \frac{s_{k,j,t}}{\mu(s_{k,j,t})} \geq 0
\end{aligned}$$

Where the last inequality is strict if the distribution of firms is non-degenerate.

B.8 Extra results

B.8.1 Alternative calibration

In this alternative calibration, I modify the second step of the baseline calibration (section 4.2), when the slope of firms' switching cost function is estimated. Instead of supposing that φ is firm-year specific, I now suppose that it is fixed across time, that is, that the distribution of $\varphi_{i,j}$ is stationary. To calibrate this stationary distribution, I focus on the first two years of my sample. Similarly to the baseline calibration, I rely on the optimal alpha identity (equation (19)) for the years 1980 and 1981 to retrieve $\Xi_{i,j,t}$ and calibrate firms' $\varphi_{i,j}$.

With this calibration in hand, the model can be simulated for the subsequent years with the parameters $\varphi_{i,j}$ as ex-ante parameters. Now, $\alpha_{i,j,t}$ is a control of the firm. Supposing Cournot competition, this boils down to solving numerically for $s_{i,j,t}$, $\mu(s_{i,j,t})$ and $\alpha_{i,j,t}$ the system:

$$\begin{cases} \mu_{i,j,t} = \frac{\varepsilon(s_t^{i,j})}{\varepsilon(s_t^{i,j})-1} \\ \varepsilon(s_t^i) = \Big(\frac{1}{\epsilon} - \big(\frac{1}{\epsilon} - \frac{1}{\sigma}\big)\Big(\frac{\mu_{i,j,t}mc_{i,j,t}}{p_t^j}\Big)^{1-\epsilon}\Big)^{-1} \\ \Xi_{i,j,t} = \varphi_{i,j} \end{cases}$$

In Figure B.1, I compare the distribution of α, the primitive distribution of the model, at the end of the sample (in 2019) for the baseline calibration and for this alternative calibration. The key difference being that in the baseline calibration such distribution is targeted while in this alternative calibration, it is an endogenous result. Qualitatively, the Kernel densities are similar. Two key differences stand out: the two modes of the distribution are heavier in the alternative calibration. This is because, if $\varphi_{i,j}$ is stationary, small firms don't update their α at all and stay at the initial α. Instead, more of the large firms converge to the upper bound ($\alpha = 0.9$). These differences matter little for the aggregate results because, all else equal, a firm with an α of 0.1 of 0.2 is small compared to a firm with an α of 0.8 or 0.9.

To see this, in Figure B.2, I plot the market share and markup change of firms ordered in capital-intensity deciles, mirroring Figures 4 and 5 from the main text. The numbers are very close. For example, in both the baseline and the alternative calibration, the 10% most capital-intensive firms see their market share rise by almost 17% while their markup moves from 1.28 to 1.72 for the baseline and from 1.28 to 1.71 in the alternative. Since these three distributions (α, s and μ) are sufficient to deduce all the other results, they are also close to the ones obtained in the baseline calibration.

The fact that the capabilities of firms to update their α is critical initially but needs not to change along time is reminiscent of the fact documented recently by Guntin and Kochen (2025): superstar firms are more capital-intensive at birth compared to their cohort and stay persistently more capital-intensive over their life-cycle. As such, it seems that the key

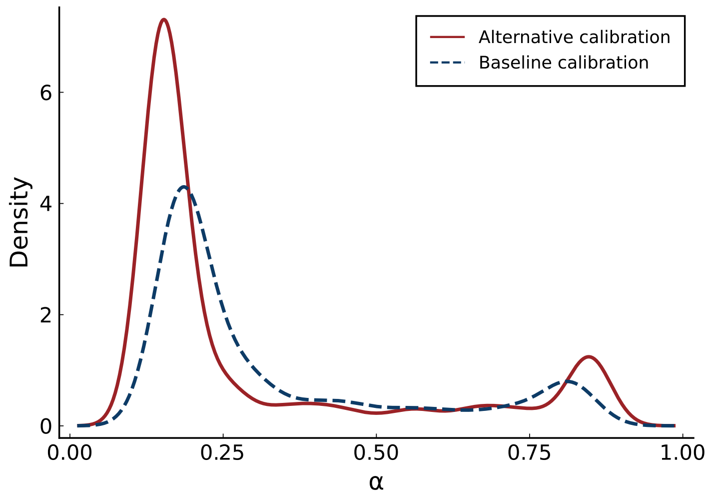


Figure B.1: Distribution of firms' α: baseline vs. alternative calibration.

Note: This graph plots the Kernel density estimation of the distribution of firms' α for the baseline and the alternative calibration. The procedure to calibrate α in the baseline calibration is reported in section 4.2.

difference across firms lies at their birth and their comparative advantage in using capital goods materializes when the RCK falls.

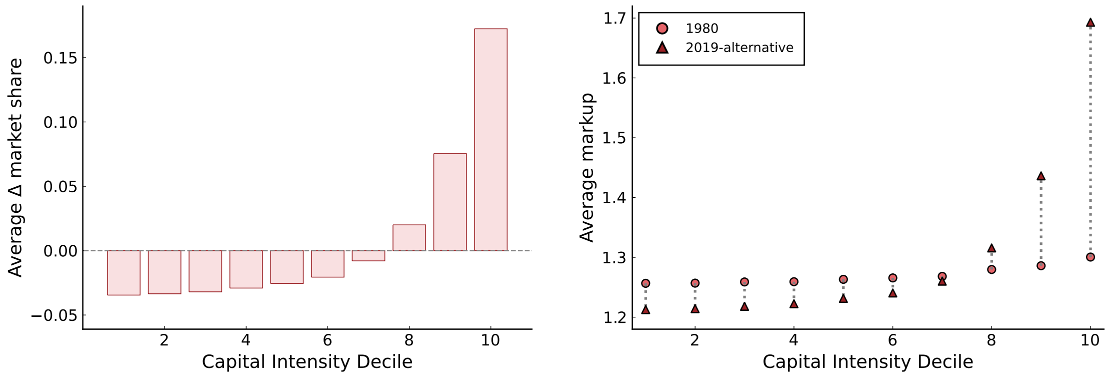


Figure B.2: Market shares' (left) and Markups' (right) dynamics per capital intensity decile with alternative calibration.

Note: The left-hand-side graph plots the change in market shares between 1980 and 2019 associated with each decile of the distribution of firms capital intensity from the simulated model under the alternative calibration. The right-hand-side graph plots the change in markups between 1980 and 2019 associated with each decile of the distribution of firms capital intensity from the simulated baseline model. The "1980" distribution of firms markups corresponds to the beginning of the model where the distribution of firms α is calibrated on the observed distribution of firms capital intensity in 1980 and the RCK is set to its 1980 level. The "2019-alternative" distribution of firms markups corresponds to the end of the model where the distribution of firms is an endogenous result of the model and the RCK is set to its 2019 level.

B.8.2 Taking Stock

In this section, I compare the aggregate capital-to-labor ratio in: (i) the Data, (ii) the baseline model, (iii) a model with changes in capital intensity only, (iv) a model with markups increase only. For (iii) and (iv), I start from the following spelling of the aggregate labor share:

$$\lambda_t^L = \frac{\zeta_t^L}{\mathcal{M}_t} \tag{F.10}$$

Where ζ_t^L is the aggregate cost share of labor. In an environment with no markup change and only increasing capital intensity, a decline of λ_t^L must be explained solely by a decline in ζ_t^L. Given the path of λ_t^L implied by the model (a 5.59pp decline), I retrieve the implied ζ_t^L from equation (F.10) and derive the implied increase in the aggregate capital-to-labor ration using the capital-to-labor ratio identity (equation (1)). On the contrary, supposing no input-biased technological change implies that only a rise in $\mathcal{M}_t$ can explain the fall of λ_t^L in equation (F.10). In such a case, ζ_t^L is fixed in time and only the falling RCK drives up the aggregate capital-to-labor ratio. In Figure B.3, I plot the aggregate capital-to-labor ratio implied by these two scenarii together with my baseline model and the data and clearly, an only-markup and an only-capital account of the decline of the aggregate labor share yield counterfactual paths for the aggregate capital-to-labor ratio.

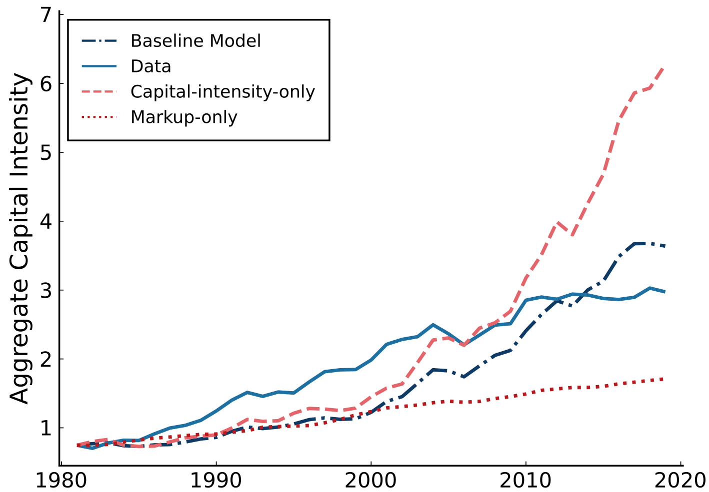


Figure B.3: Aggregate capital-to-labor ratio: data vs. model specifications.

Note: This figure compares four measures of the aggregate capital-to-labor ratio. The "Data" series plots the aggregate capital-to-labor ratio from Compustat data defined as the ratio between the sum of total "Property, Plant and Equipment" and total "Intangible Assets" divided by total "Employees". The "Baseline Model" series plots the evolution of the aggregate capital stock to aggregate labor implied by the model defined in section 3. "capital intensity-only" refers to the aggregate capital-to-labor ratio implied by a scenario where a 5.59pp decline in the aggregate labor share is explained solely by a fall in the cost share of labor in equation (F.10). The implied ζ_t^L is then used in equation (1) together with the baseline decline in the RCK to deduce the implied aggregate capital-to-labor ratio. The series "Markups-only" draws the capital-to-labor ratio for a fixed ζ_t^L (calibrated to the 1980 one) given the baseline decline in the RCK from equation (1).

B.8.3 Sunk cost share of profits

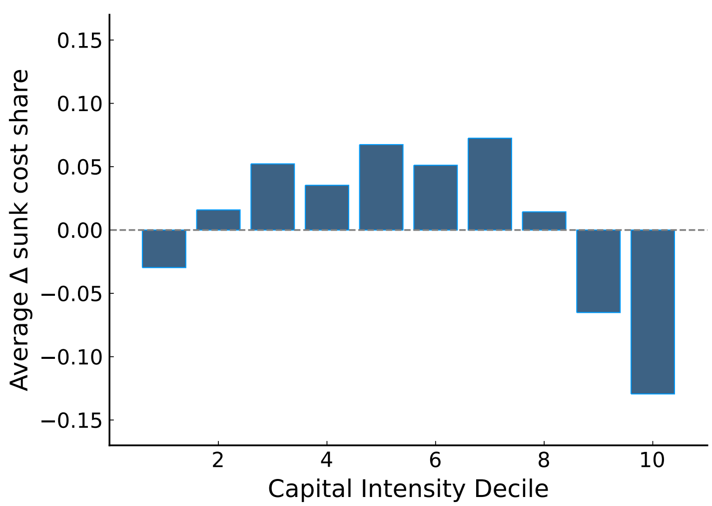


Figure B.4: Change in share of sunk costs in profits.

Note: This figure plots the percentage point change in the share of sunk costs in total accounting profits of firms per capital intensity deciles. One minus that change is the change in the share of pure economic profits out of accounting profits.

B.8.4 Granular results about sales and employment concentration

In Table B.1, I report the percent change in the ratio $\varrho_{i,j,t}$ (or $s_{i,j,t}/\lambda^{EMP}_{i,j,t}$) along my sample for different degrees of granularity in my model and compare these results to the corresponding empirical estimates from Autor et al. (2020), Kwon et al. (2024) and Ma et al. (2024).

	Model	**Data**
Top 1% Aggregate Market Share		
Aggregate	+16.21%	+6.25%[†]
Average of Top Firm within Sector		
Aggregate	+10.66%	
Manufacturing	+18.02%	+10.81%[§]

Table B.1: Change in sales to employment share of biggest firms.

Note: This table contains the evolution of the ratio between sales and employment share $(1/\varrho_{i,j,t})$ of largest firms for different measures. The top panel looks at the Top 1% Aggregate Market Share, that is the 1% firms with the highest aggregate market share: $s_{i,t} = \frac{p_{i,j,t}Y_{i,j,t}}{p_t Y_t}$. External results labeled with † are taken from Kwon et al. (2024) and Ma et al. (2024). Specifically, the number 6.25% is obtained by taking the ratio between the Receipts time series of Figure 1 in Kwon et al. (2024) and the US (BDS) time series from Figure 11 Panel A of Ma et al. (2024). The bottom panel looks at the average sales and employment share of the top firm within each 4D sectors in the model. External results labeled with § are taken from Figure IV in Autor et al. (2020) focusing on CR4.

The top line of Table B.1 contains the evolution of the ratio of the sales to employment share of the 1% biggest firms in the economy. The next two lines look at the change in the ratio of the average market share to average employment share of the top firm within each 4D sector. Overall, the model generates a faster rise in sales share than in employment shares for large firms, consistently with empirical evidence. However, quantitatively, the model systematically overshoots empirical estimates. This can be due to a few factors. First selection effects between Compustat and SOI or Census data. The model is calibrated in Compustat data and matches the dynamics of Compustat firms pretty well as per Figure 7 in the main text. On the other hand the estimates of Autor et al. (2020), Kwon et al. (2024) and Ma et al. (2024) seem to point at differing magnitudes when using Census and SOI data. A second possibility is the fact that, in the baseline calibration, capital intensity is the only dimension of firm heterogeneity while, in reality, factor-neutral dimensions such as TFP or demand shifters also matter. Because factor-neutral drivers of heterogeneity don't affect the ratio $\zeta^L_{i,j,t}/\zeta^L_{j,t}$, they increase much less the ratio between sales share and employment share. This can also explain why the model does poorly to explain concentration of sectors where the distribution of output elasticity is stationary, such as trade (as seen in Figure 2 in the stylized fact section 3). Indeed, as documented by Autor et al. (2020) and Ma et al. (2024), contrarily to the aggregate and to manufacturing, trade has seen the sales share and employment share of its biggest firms rise almost one-to-one, instead the baseline model predicts a 6.62% increase in the ratio of the two. For this reason, in the extension section 5 to come, I calibrate my model on two dimensions of heterogeneity: capital intensity and TFPR. This allows to get a better picture for concentration among all sectors. For trade for example, the increase in the ratio of sales to employment shares of the biggest firms drops to 1.67%.

B.9 Counterfactual

B.9.1 Monopolistic competition

What if markups are homogeneous and fixed? In this section, I answer this question by replacing the oligopolistic industrial organization of the baseline model with monopolistic competition with exogenous markups. That is, there is now continuum of atomistic firms within sectors. The the sectoral-aggregator (equation (19)) now spells:

$$Y_{j,t} = \left(\int_0^1 (Y_{i,j,t})^{\frac{\epsilon-1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon-1}}$$

And a firms price is given by:

$$p_{i,j,t} = \mu mc_{i,j,t}$$

Where $\mu = \epsilon/(\epsilon - 1)$. In such a case, the only source of heterogeneity in firms' choice of α is the slope of their switching cost function. The results for the main moments implied by such a model are reported in Table B.2 together with the baseline results. In such a case, the capital share si flat since markups are fixed. The labor share displays a small decline captured by an increase in the capital share. The NORK declines as the capital share increases while the profit share is fixed and the sales-to-employment share of largest firms increases by roughly half of the baseline.

Symbol	Variable	Baseline	Monopolistic competition
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	99.16%
Y	Growth rate of output (annualized)	1.25%	0.71%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	0
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	0
λ^L	Aggregate labor share (pp change)	-5.59pp	-1.15pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	4.72pp
λ^K	Aggregate capital share (pp change)	0.39pp	1.15pp
λ^π	Aggregate profit share (pp change)	5.20pp	0
$NORK$	Net Operating Return to Capital (% change)	17.9%	-8.96%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	5.61%

Table B.2: Change in main moments for baseline vs. monopolistic competition

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Monopolistic competition corresponds to firms having a fixed and homogeneous markup fixed to 1.26 (the aggregate markup in 1980 in the baseline model). This implies that switching costs alone capture heterogeneity in α. The rest of the calibration is the same as for the baseline. Sales share to employment share is the ratio between the average sales share and the average employment share of the biggest firm in each sector.

B.9.2 Homogeneous switching cost

In Table B.3, I report the results of the baseline model and of the counterfactual model where all firms share the same switching cost calibrated to the average of firms' switching cost in the baseline model. Table B.4 contains the results of the same exercise where the switching cost is calibrated so as to deliver a rise in the representative α equal to the rise of the average α in the baseline model.

The baseline economy grows by a 1.25% *per anum* average which corresponds to slightly less than a fifth of the actual growth of gross output in the US over 1980 to 2019. In the model, growth is the result of the increase in the productivity of capital good producers,

Symbol	Variable	Baseline	Homogeneous switching cost
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	70%
Y	Growth rate of output (annualized)	1.25%	0.71%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	0
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	0
λ^L	Aggregate labor share (pp change)	-5.59pp	3.92pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	3.92pp
λ^K	Aggregate capital share (pp change)	0.39pp	-3.92pp
λ^π	Aggregate profit share (pp change)	5.20pp	0
$NORK$	Net Operating Return to Capital (% change)	17.9%	73.2%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	0

Table B.3: Change in main moments for baseline vs. homogeneous switching cost

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Homogeneous switching cost corresponds to the same model where all firms share the same shape for the graph of their switching cost function. Sales share to employment share is the ratio between the average sales share and the average employment share of the biggest firm in each sector.

which increases at an average of 6.5% per year between 1980 and 2019[43] as well as firms updating their α, almost doubling it in average between 1980 and 2019. For comparison, the Bureau of Labor Statistics (BLS), in their report "Total Factor Productivity - Major Industry Contributions to Output"[44] undertake a decomposition of input contribution to output growth using a classic growth accounting methodology. They find that capital is responsible for an average of 1.17% annual growth in gross output between 1990 and 2019.

In the counterfactual economy, where every firm faces the same updating cost function, the annualized growth rate can theoretically be higher or lower than the baseline one. Higher

[43] TFP growth for capital good producers translates non linearly into aggregate growth due to the concavity of firms' production functions.

[44] https://www.bls.gov/productivity/highlights/contributions-of-total-factor-productivity-major-industry-to-output.htm

because with homogeneous firms, the distribution of market share and markups is uniform and hence the aggregate markup is minimized, involving lower distortion of aggregate output. Lower because firms' α increases less and because the most efficient firms that participate the most to output in the baseline model are absent in the counterfactual. Overall, the second effect dominates. To disentangle the relative weight of the two downward drivers on output growth compared to the baseline model, I repeat this exercise by calibrating the switching cost so that the increase in the average/aggregate α is also 99.16%, purifying out the difference in the growth of α and leaving only the absence of big-efficient firms as the sole downward driver of output growth compared to the baseline. The moments implied by such an economy are reported in Table B.4. I find that still output growth is lower (0.98% per year). As such, in line with Edmond et al. (2015), Edmond et al. (2023) or Aghion et al. (2023) the net effect on output of large firms is positive, even if they charge higher markups, if their size is micro-founded by productivity (in my case, efficiency in adopting external innovation). Because there is no reallocation across firms, the aggregate labor share follows the path of the median labor share and hence rises due to complementarity between labor and capital in production at the micro level.[45] Interestingly, aggregate net operating return to capital surges in the counterfactual economy because, while markups don't change, profits increase 1-to-1 with output (profit share is stable) due to a volume effect while the capital share plunges. Finally, the sales share to employment shares of the biggest firm trivially does not change. Indeed, all firms have the same production technology at every period and hence the same sales and employment share. So any 10% firms will be responsible for 10% of both sales and employment.

[45]Note that since firms' production function are homogeneous, they can be aggregated into an aggregate production function that shares the same parameters. As such, the aggregate elasticity of substitution is the same as the micro one: 0.5.

Symbol	Variable	Baseline	Homogeneous switching cost
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	99.16%
Y	Growth rate of output (annualized)	1.25%	0.98%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	0
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	0
λ^L	Aggregate labor share (pp change)	-5.59pp	2.84pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	2.84pp
λ^K	Aggregate capital share (pp change)	0.39pp	-2.84pp
λ^π	Aggregate profit share (pp change)	5.20pp	0
$NORK$	Net Operating Return to Capital (% change)	17.9%	44.15%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	0

Table B.4: Change in main momentss for baseline vs. low homogeneous switching cost

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Homogeneous switching cost corresponds to the same model where all firms share the same shape for the graph of their switching cost function. Sales share to employment share is the ratio between the average sales share and the average employment share of the biggest firm in each sector.

B.9.3 Stationary distribution of α

In this section, I simulate my model over a stationary distribution of α. That is, I suppose that every year, the distribution of α is the same as the one in 1980. This preserves the heterogeneity in capital intensity but kills the endogenous updating margin. Changes in key moments are reported in Table B.5.

Symbol	Variable	Baseline	Stationary distribution of α
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	0
Y	Growth rate of output (annualized)	1.15%	0.31%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	1.60%
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	+169%
λ^L	Aggregate labor share (pp change)	-5.59pp	+1.89pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	2.21pp
λ^K	Aggregate capital share (pp change)	0.39pp	-3.17pp
λ^π	Aggregate profit share (pp change)	5.20pp	1.28pp
$NORK$	Net Operating Return to Capital (% change)	17.9%	+142.88%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	0.95%

Table B.5: Change of main moments for baseline vs. stationary distribution of α

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Stationary distribution refers to the model simulated on the distribution of α of 1980 for every year, i.e shutting down the endogeneous updating of alpha component.

B.9.4 Heterogeneous α vs. Heterogeneous TFP

Here, I compare my model to a "classic" Atkeson and Burstein (2008) where firm dynamics are driven by TFP. The "classic" Atkeson and Burstein (2008) model is calibrated with homogeneous α, set to 0.12 (the average α over my full sample), but firms feature heterogeneous Total Factor Productivity, that is $Y_{i,j,t}$ becomes:[46]

$$Y_{i,j,t} = Z_{i,j,t}\left(\alpha K_{i,j,t}^{\frac{\psi-1}{\psi}} + (1-\alpha)L_{i,j,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$$

I then calibrate firms' TFP with my estimate of TFPR calculated with equation (9) in Appendix 4. I use the estimation made under the assumption of homogeneous α. This

[46]I do not report all the implied identities as they are standard.

measure is the Solow residual of firms' production function when ψ and α are supposed fixed across firms and time.[47] I keep the same market structure (100 sectors each populated by 10 firms) although the simulated aggregate markup in 1980 is 1.30, higher than that estimated in section A.5.2 under the hypothesis of homogeneous α: 1.27. I do this to make comparison to the baseline more direct and leave out market structure changes. I simulate the model under this calibration and report the resulting changes in the main moments in Table B.6 together with the baseline.

In such a case, the key predictions are a huge surge in the markup and an almost equally huge fall in the capital share. Note that despite a sharper rise in the aggregate markup, the variance of markups increases less under heterogeneous TFPR than under heterogeneous α. This is because the original variance is higher under TFPR micro-foundation of firm dynamics and hence large firms are already on the steeper segment of their markup curves. Any increase in the variance of markups therefore implies a bigger rise in the aggregate markup: all the way to 1.47 compared to the 1.34 estimated in section A.5.2. Because α is homogeneous and exogenous, the capital share of every firm drops due to complementarity between capital and labor. The surge in the profit share hence only eats up 1.17pp of the labor share, implying that only an unrealistic rise in markups can deliver the full decline in the labor share. Equally unrealistic is the surge in the Net Operating Return to Capital that is multiplied by almost five. This extreme increase is driven by the amplified rise in profits together with the sharp fall of the capital share. Finally the change in the ratio between the sales and the employment share of the biggest firms is compressed because, as discussed in section 4.5, the proportionality coefficient between sales and employment is much less dampened when firms have fixed output elasticity to labor.

[47] Remember that this measure isn't Total Factor Productivity for two reasons. First, I do not possess quantity data and hence output is proxied by sales. Second, I have factorized all source of non-homotheticity in the log-linear term of the production function (see section **??**). I abuse language by calling it TFP for the rest of this section for coherence with the related literature.

Symbol	Variable	Baseline	Heterogeneous TFP
$mean(TFP)$	Average firm TFP (% change)	0	55.6%
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	0
Y	Growth rate of output (annualized)	1.25%	0.98%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	12.00%
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	232%
λ^L	Aggregate labor share (pp change)	-5.59 pp	-1.17pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	7.04pp
λ^K	Aggregate capital share (pp change)	0.39pp	-7.48pp
λ^π	Aggregate profit share (pp change)	5.20pp	8.65pp
$NORK$	Net Operating Return to Capital (% change)	17.90%	378.15%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	1.34%

Table B.6: Change in main moments for baseline vs. heterogeneous TFPR

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Heterogeneous TFPR corresponds to the same model where all firms share the same same α but different Solow residuals estimated using equation (9) of section 4. Sales share to employment share is the ratio between the average sales share and the average employment share of the biggest firm in each sector.

B.9.5 Collateral constraints

An alternative possibility for dispersion in capital intensity is collateral constraint. Suppose a Hopenhayn (1992) model augmented with constrained capital investment as in Kiyotaki and Moore (1997). Suppose now that firms differ in TFP $Z_{i,j,t}$ and wield the following production technology:

$$Y_{i,j,t} = Z_{i,j,t}\left(\alpha K_{i,j,t}^{\frac{\psi-1}{\psi}} + (1-\alpha)L_{i,j,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$$

Where α is a homogeneous and exogenous parameter. Suppose that firms own their capital stock and that it evolves according to:

$$K_{i,j,t+1} = (1-\delta)K_{i,j,t} + X_{i,j,t}$$

But that firms are constrained in the amount of investment ($X_{i,j,t}$) they can undertake by their current stock of capital. That is, given a cost of the investment good q_t, a firm faces the following collateral constraint:

$$q_t X_{i,j,t} \leq \theta q_t K_{i,j,t}$$

Suppose, for simplicity, monopolistic competition, such that demand aggregator is defined following Dixit and Stiglitz (1977):[48]

$$Y_t = \left(\int_0^1 Y_{j,t}^{\frac{\sigma-1}{\sigma}} dj\right)^{\frac{\sigma}{\sigma-1}}$$

And:

$$Y_{j,t} = \left(\int_0^1 Y_{i,j,t}^{\frac{\epsilon-1}{\epsilon}} di\right)^{\frac{\epsilon}{\epsilon-1}}$$

Under such assumptions, the problem of the firm can be written as a Bellman operator:

$$V(K_{i,j,t}, Z_{i,j,t}) = \max_{Y_{i,j,t}, X_{i,j,t}, L_{i,j,t}, K_{i,j,t+1}} \quad p_{i,j,t}Y_{i,j,t} - w_t L_{i,j,t} - q_t X_{i,j,t} + \beta E_t(V(K_{i,j,t+1}, Z_{i,j,t+1})$$

subject to the five equations defined in this section (production function, demand functions, law of motion of capital stock and collateral constraint). The first order conditions on

[48]Note that I stick to the nested-CES demand only to maximize comparability with the baseline model.

inputs and the envelope condition give the following identity for a firm's capital intensity:

$$\frac{K_{i,j,t}}{L_{i,j,t}} = \frac{\zeta^K}{\zeta^L}\left(\frac{w_t}{r^K_t + \lambda_{i,j,t}/\beta}\right)$$

Where, like in the baseline model, $r^K_t = q_{t-1}/\beta - (1-\delta)q_t$, $\zeta^K = \frac{\delta ln(Y_{i,j,t})}{\delta ln(K_{i,j,t})}$ and $\lambda_{i,j,t}$ is the Lagrange multiplier of the collateral constraint. If firms are of heterogeneous size (because of initial TFP shocks for example), a smaller firm, with a smaller K_t, will face a tighter collateral constraint and hence a higher multiplier $\lambda_{i,j,t}$ than a large firm. As such, firms will differ in capital intensity due to financial frictions and it will hold that larger firms have a higher capital intensity.

However, such a model predicts a counterfactual implication for misallocation. In such a case, a firm's marginal return to the product of capital writes:

$$MRPK_{i,j,t} = \mu_t(r^K_t + \lambda_{i,j,t}/\beta)$$

On top of homogeneous markups, firms deviate from their efficient frontier because of their collateral constraint. In particular, a firm with a highly binding constraint (high λ) will have a $MRPK$ more distant to r^K_t than a firm with a lowly binding constraint as it lies further below its optimal capital stock. Therefore, in such a model, misallocation of capital is driven by small firms (those with a tight budget constrain), a prediction at odds with the empirical finding of Gutiérrez and Philippon (2016) who demonstrate that large firms lie further away from their efficient frontier.

B.10 Extensions

B.10.1 Heterogeneous α and Heterogeneous TFP

I now propose to turn to an exercise where firms' heterogeneity is bi-dimensional: in TFP and in α. In such a case, the calibration of the model is the one of the baseline (1) with the addition of firm-specific TFP [49] this time estimated using equation (9) in section 4 but under the heterogeneous α case. For each sector, I construct a bivariate kernel density estimate of the joint distribution of firm-level TFP and α per sector from which each 10 firms in the sector draw a pair. The resulting changes in our moments of interest are reported in Table B.7.

Adding the heterogeneous TFP dimension has three principal effects. First it allows to amplify dispersion in sectors that have little relative variance in capital intensity (Service-related sectors in particular). Second, it amplifies the dominant position of some firms who

[49] The production function becomes: $Y_{i,j,t} = Z_{i,j,t}\left(\alpha_{i,j,t}K_{i,j,t}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,j,t})L_{i,j,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}}$

Symbol	Variable	Baseline	Baseline ⊕ Heterogeneous TFP
$mean(TFP)$	Average firm TFP (% change)	0	30.2%
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	99.16%
Y	Growth rate of output (annualized)	1.25%	2.36%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	14.36%
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	403%
λ^L	Aggregate labor share (pp change)	-5.59 pp	-7.22pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	2.06pp
λ^K	Aggregate capital share (pp change)	0.39pp	-1.28pp
λ^π	Aggregate profit share (pp change)	5.20pp	8.50pp
$NORK$	Net Operating Return to Capital (% change)	17.90%	56.92%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	4.66%

Table B.7: Change in main moments for baseline vs. baseline + heterogeneous TFPR

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Baseline + Heterogeneous TFP corresponds to the same model where all firms share the same same α but different Solow residuals estimated using equation (9) of section 4. Sales share to employment share is the ratio between the average sales share and the average employment share of the biggest firm in each sector.

have both high TFP and high capital intensity. Because TFP is calibrated with TFPR, which as shown in section 4 captures non-homotheticity, this could suggest possibly no scale invariance, consistently with a "scaling without mass" phenomenon. Third, it restores the greatness of some firms that are big despite being labor-intensive. The net effect of integrating heterogeneous TFP is hence ex-ante ambiguous for almost all our variables. One variable for which there is no ambiguity is growth. On top of the TFP of the capital good producer and firms' adapting their production function, the model now includes TFP growth for firms. Average yearly gross output growth climbs to 2.36%. For comparison, the Bureau of Labor Statistics (BLS), in their report "Total Factor Productivity - Major Industry Contributions to Output" [50] find that the average output growth due to both TFP and Capital is 2.04% between 1990 and 2019. The aggregate markup happens to grow more than twice as fast as in the baseline while the variance of firms' markups rises only modestly. This is again due to the fact that the variance of markups is higher in 1980 when accounting for heterogeneous TFP. Any percent increase in the variance of markups therefore translates into a higher aggregate markup in the extension economy than in the baseline. This leads to a decline of the aggregate labor share of 7.22pp, close to its entire decline. The capital share also suffers from the jump in the profit share and declines as well. The Net Operating Return to Capital increases by 56.92%, close to the total estimate reported by Gutiérrez and Philippon (2016). Finally the sales to employment share of the biggest firms rises by 4.66%.

	Model	**Data**
Top 1% Aggregate Market Share		
Aggregate	+4.91%	+6.25%†
Average of Top Firm within Sector		
Aggregate	+4.66%	
Manufacturing	+6.26%	+10.81%§

Table B.8: Change in sales to employment share of biggest firms including α and TFP.

Note: This table contains the evolution of the ratio between sales and employment share ($1/\varrho_{i,j,t}$) of largest firms for different measures. The top panel looks at the Top 1% Aggregate Market Share, that is the 1% firms with the highest aggregate market share: $s_{i,t} = \frac{p_{i,j,t} Y_{i,j,t}}{p_t Y_t}$. External results labeled with † are taken from Kwon et al. (2024) and Ma et al. (2024). Specifically, the number 6.25% is obtained by taking the ratio between the Receipts time series of Figure 1 in Kwon et al. (2024) and the US (BDS) time series from Figure 11 Panel A of Ma et al. (2024). The bottom panel looks at the average sales and employment share of the top firm within each 4D sectors in the model. External results are taken from Figure IV in Autor et al. (2020) focusing on CR4.

[50] https://www.bls.gov/productivity/highlights/contributions-of-total-factor-productivity-major-industry-to-output.htm

B.10.2 Falling interest rates

A key determinant of the capital share is the required return to capital (Barkai (2020), **?**, Karabarbounis and Neiman (2019), Karabarbounis (2024)). I have followed the literature in not including the fall in the risk-free interest rate in my measure of the RCK because of differences in risk and liquidity attributes (Fisher (2015), Campbell et al. (2017), Farhi and Gourio (2018)). In this section, I look at the outcome of the model if interest rates decline on top of the relative cost of capital goods, amplifying the fall in the rental rate of capital. I model the decline in the interest rate following Laubach and Williams (2003), Mian et al. (2021), Favero et al. (2021) and many others, supposing that agents have become increasingly patient (rising β). This corresponds to adding the moments reported in Table B.9 to the baseline calibration displayed in Table 1 as well as to re-estimating the distribution of firms' α. I report these in Figures B.5 (complementary capital and labor, $\psi = 0.5$), B.6 (alternatives: Cobb-Douglas $\psi \to 1$ and substitutable capital and labor, $\psi = 1.5$).

Symbol	Parameter	Value	Moment
r	Interest rate	[0.16:0.02]	Federal Feds fund rate
β	Discount factor	$\frac{1}{1+r}$	

Table B.9: Model Calibration with falling interest rates.

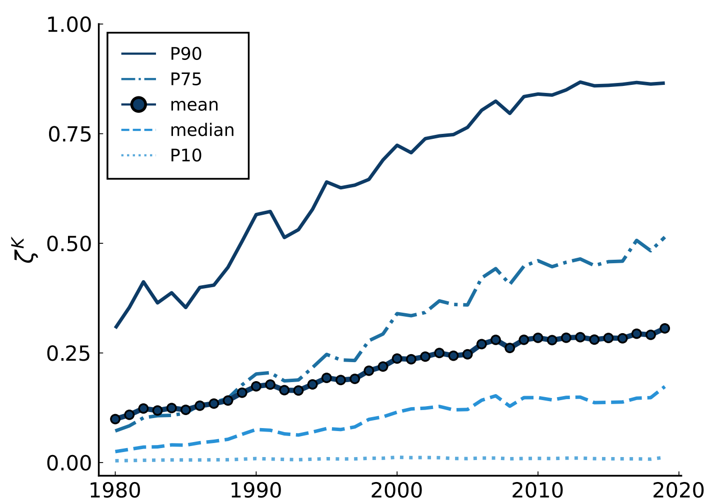
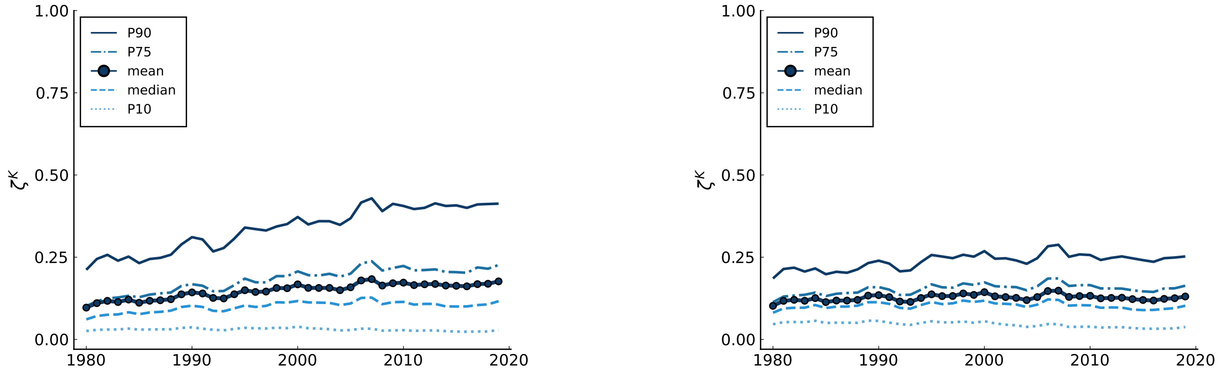


Figure B.5: Moments of the distribution of alpha with falling interest rates

Note: Source is Compustat. Sample is 1980-2019. Moments from the distribution implied by the estimation made using equation (10) on Compustat firms. P90 is the 9th decile.

Results under this new calibration are displayed in Table B.10. The rise in average α is lower because a steeper fall in the RCK (now including the fall in the interest rate) translates into a lower necessary increase in α to match firms' rising capital intensity pattern. This



(a) Cobb-Douglas ($\psi \to 1$) (b) Substitutes ($\psi = 1.5$)

Figure B.6: Moments of the distribution of alpha with falling interest rates for different elasticity of substitution between capital and labor

Note: Source is Compustat. Sample is 1980-2019. Moments from the distribution implied by the estimation made using equation (10) on Compustat firms. P90 is the 9th decile.

dampening effect is stronger for the most capital-intensive firms. This is only partly offset by the steeper fall in the RCK, yielding a compression of the distribution in capital intensity which in turn affects the distribution of market shares and markups. This results in a lower annual growth rate and aggregate markup. While all the other results hold in a same magnitude, the capital share is notably different. The aggregate capital share is now declining, consistently with the differing results of Rognlie (2015) and Barkai (2020). This is because the upward driver of the capital share, rising α, is now lower, while one of its downward driver, falling RCK, is now steeper. Finally, the fall in the Net Operating Return to Capital is amplified by the fall in the capital share.

Symbol	Variable	Baseline	Falling interest rate
$mean(\alpha_{i,j})$	Average firm α (% change)	99.16%	83.01%
Y	Growth rate of output (annualized)	1.25%	1.16%
$\mathcal{M}$	Aggregate markup (% change)	6.71%	5.81%
$VAR(\mu_{i,j})$	Variance of firms' markups (% change)	335%	298%
λ^L	Aggregate labor share (pp change)	-5.59pp	-3.74pp
$median(\lambda^L_{i,j})$	Median firm labor share (pp change)	2.11pp	2.58pp
λ^K	Aggregate capital share (pp change)	0.39pp	-0.99pp
λ^π	Aggregate profit share (pp change)	5.20pp	4.73pp
$NORK$	Net Operating Return to Capital (% change)	17.9%	57.65%
$\frac{s^{P99}}{e^{P99}}$	Sales share to employment share (% change)	10.66%	13.43%

Table B.10: Change of main moments for baseline vs. stationary distribution of α

Note: Change is 2019 compared to 1980. Baseline refers to the baseline model and calibration referred in section 3. Stationary distribution refers to the model simulated on the distribution of α of 1980 for every year and every other parameter is the same as the baseline.

The Bulging Cost Channel and the Flattening Phillips Curve*

Geraud Desazars de Montgailhard

March 13, 2025

Abstract

As an economy becomes more capital intensive, the effect of monetary policy shocks on the rental rate of capital becomes more important to determine the response of the marginal cost of firms. In an environment where the utilization rate of capital is endogenous and where the rental rate of capital is an increasing function of the interest rate, a "cost channel" dampens the traditional "Taylor channel" that implies a negative correlation between interest rates and marginal costs and therefore prices due to demand effects. This happens even if the structural relationship between marginal cost and inflation, *the primitive New Keynesian Phillips curve*, is unchanged. On the contrary, because marginal costs are less elastic to interest rate shocks, the slope of the *conventional New Keynesian Phillips curve*, which relates prices to the output gap, declines as the economy becomes more capital intensive.

*Bocconi University. `geraud.desazars@phd.unibocconi.it`. I am grateful to Basile Grassi, Dmitriy Sergeyev, Nicolas Crouzet and Yueran Ma for their extensive feedback and valuable discussions. I am further grateful for discussions with: Lawrence Christiano, Martin Eichenbaum, Carlo Favero, François Gourio, Greg Kaplan, Chen Lian, Nicola Pavoni, Matthew Rognlie, and Vincent Sterk.

1 Introduction

How does rising capital intensity affect the pass-through of monetary policy shocks to real variables, and, more specifically, the structure of the New Keynesian Phillips Curve (NKPC)? Theoretically, the effect is ambiguous, because capital and labor markets differ in their exposure to interest rate shocks in two manners. First, in terms of cost, the "cost channel" theory recognizes that interest rates determine directly the rental rate of capital whereas it does not affect directly wages. Second, in terms of quantity, capital is often considered stickier in the short run, leaving labor as the adjustment variable for firms to account for demand shocks. In this paper, after deriving theoretically this ambiguity, I show empirically, using local projections on firm-level quarterly data, that the "cost channel" effect seems to dominate qualitatively. In other words, for a given demand shock, a capital-intensive economy sees its marginal cost react less than a labor-intensive economy, resulting in lower inflationary pressure. Using this fact, I build a simple continuous-time New Keynesian model with endogenous utilization rates of capital and a rental rate of capital that depends positively on the real interest rate. This allows me to show that the rise in the capital-intensity of the American economy can explain the fact documented by **?**: while the primitive NKPC (where deviation in the real marginal cost is the antecedent) has remained stable, the conventional NKPC (where deviation in output is the antecedent) has seen its slope flatten. The reason is that, as capital-intensity increases, a demand shock increases the marginal cost less, as the latter depends more heavily on changes in the rental rate of capital than on changes in the wage.

To understand why a rising elasticity of output to capital has ambiguous effects on the structure of the NKPC, consider the example of an expansionary monetary policy shock. Through a demand effect, the cost of both capital and labor rise because such a shock translates into a positive demand shock on the goods market which, in turn, translates into a positive demand shock on the input markets. However, both costs react with different amplitudes due to two key structural differences. First, the rental rate of capital is an increasing function of the interest rate, dampening (or reverting) its rise as a reaction to an interest rate cut (the "cost channel" of monetary policy). If the output elasticity to capital increases, so does the marginal cost elasticity to the rental rate of capital. This implies that marginal costs and hence prices would increase less following an expansionary monetary policy shock. Second, capital is typically supposed to be less adjustable than labor in the short run. This creates short-run diminishing returns to scale of production. If the output elasticity to capital increases, short-run returns to scale -which are equal to the output elasticity to labor- decline, implying that higher labor adjustments are needed to match the new demand. This generates higher wage pressure which in turn creates more inflation. Rising aggregate capital-intensity therefore affects the slope of the conventional NKPC, but the presence of a cost channel, that is the degree to which capital can be adjusted at the monetary policy frequency, determines the direction of the change.

After showing this ambiguity theoretically in section 2, I perform an empirical exercise to determine which of the two cases dominates in section 3. I rely on the fact that firms feature heterogeneous capital intensity, even within sector. As such, if a cost channel exists, capital-intensive firms should have a marginal cost that is more elastic to a rental rate of capital cut implied by an interest rate cut and hence be able to produce more efficiently, implying a rise in their market share. On the contrary, in the absence of a cost channel, capital-intensive firms display lower short-run returns to labor and hence produce less efficiently and should lose market shares at the expense of their labor-intensive competition. I find that the first effect dominates using local projections on quarterly firm-level data from Compustat and quarterly estimates of interest rates shocks from the work of Romer and Romer (2004) extended to 2018.

In the last section, I ask what is the quantitative decline in the slope of the conventional NKPC given the observed rise in the American Economy's capital intensity. To do so, I build a continuous-time New Keynesian model inspired by Christiano, Eichenbaum, and Evans (2005) and **?**. The model includes an endogenous utilization rate of the capital stock to calibrate the relative stickiness in the adjustment of capital to that of labor. The model also includes a rental rate of capital to allow for a cost channel of monetary policy. I simulate my model on calibrations of aggregate capital-intensity for the four decades between the 1980s and the 2010s and find that, while the slope of the primitive New Keynesian Phillips curve stays constant, as unveiled in the theoretical exercise, the slope of the conventional New Keynesian Phillips curve flattens by a factor of four.

Literature. This paper contributes to two strands of the literature: the one that proposes explanations for the flattening of the Phillips curve and the one that studies the effects of the cost channel for the pass-through of monetary policy to real variables.

The empirical disconnect between different measures of economic slack and inflation has been interpreted as a flattening of the Phillips curve by **?**, Blanchard, Cerutti, and Summers (2015), **?** among others. These estimation have been updated to account for changes in inflation expectations (**?**Hazell, Herreno, Nakamura, and Steinsson, 2020) and simultaneity between the output gap and inflation when the central bank seeks to stabilize economic slack (**?**). These papers typically find remarkably small estimates of the slope of the conventional Phillips curve. **?** show that those small estimates are due to a weak pass-through from marginal costs to the output gap rather than from the output gap to inflation. This paper proposes an explanation for this phenomenon. This explanation relies on the growing strength of the "cost channel" of monetary policy and the key role played by reallocation across heterogeneous firms to deliver these real effects (which have already been established by Meier and Reinelt (2020), Baqaee, Farhi, and Sangani (2021), Rubbo (2020) or Hartley et al. (2025)). The cost channel has been extensively studied in the literature since, at least, Barth and Ramey (2001), Ravenna and Walsh (2006) or Christiano et al. (2005). Recently, Beaudry, Hou, and Portier (2022) consider its incidence when the Phillips curve is "quite flat", **?** look at its implications when firms are organized in a production network and Ottonello and Winberry (2020) consider how it affects firms heterogeneously when firms differ in their "distance to default".

2 Intuition

In this section, I show, using canonical identities, how rising capital intensity alters the pass-through of nominal shocks to real variables. I do so through the example of the New Keynesian Phillips curve (NKPC), both in its primitive form (where marginal cost is the antecedent) and in its conventional form (where the output gap is the antecedent) because the impact on the two is crucially different. Consider a non log-linearized primitive NKPC expressed in continuous time (all equations in this section will be derived when the full model is exposed in section 4 as well as in Appendix B):

$$\rho\pi_t = \dot{\pi}_t + \lambda\left(\mu mc_t - 1\right) \tag{1}$$

Where π_t is the inflation rate at time t, $\dot{\pi}_t$ is inflation change, ρ is the time discount rate, μ is an exogenous markup, mc_t is the real marginal cost and λ is the slope of the primitive NKPC which depends on nominal frictions and the elasticity of demand to prices. Suppose that the production technology is Cobb-Douglas with constant returns to scale:

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha} \tag{2}$$

Where Y_t is output, K_t is capital and L_t is labor and α is the output elasticity of capital that I use to measure an economy's capital intensity. Because λ is orthogonal to firms' production technology, the primitive NKPC isn't affected by a change in α. This is however untrue for the conventional NKPC that links inflation to the output gap. To see this notice that, if capital can be adjusted at the monetary policy frequency, as in the the medium-scale New Keynesian model of Christiano et al. (2005) or **?**, then the primitive NKPC can be written:[51]

$$\rho\pi_t = \dot{\pi}_t + \lambda\left(\frac{\mu}{Z_t}\left(\frac{r_t^K}{\alpha}\right)^{\alpha}\left(\frac{w_t}{1-\alpha}\right)^{1-\alpha} - 1\right) \tag{3}$$

Where r_t^K is the real rental rate of capital and w_t is the real wage. As before, inflation is a function of the distance between the flexible price marginal cost and the one implied by nominal frictions, or equivalently, the distance between implied and desired markup (μ). In the biggest bracket, the first element, with geometric weight α refers to the unit cost of

[51] Derivations are reported when the baseline model is presented in section 4 and in Appendix B.3

capital. The second element, with weight $1-\alpha$, refers to the unit cost of labor. If labor isn't affected by wealth effects, and the rental rate of capital is pinned down by a Hall and Jorgenson (1967) no-arbitrage condition between bonds and capital ($r_t^K = i_t - \pi_t + \delta$), a log-linearized version of the conventional NKPC can be derived:[52]

$$\rho\widehat{\pi}_t = \dot{\pi}_t + \lambda\left(\underbrace{\frac{\phi(1-\alpha)}{1+\alpha\phi}\widehat{y}_t}_{\text{Taylor channel}} + \underbrace{\frac{\alpha(1+\phi)}{1+\alpha\phi}\widehat{r}_t^K}_{\text{Cost channel}}\right) \tag{4}$$

Where the first component is a usual Taylor channel of monetary policy that links the output gap to inflation while the second component is the cost channel of monetary policy that isolates the particular effect of an interest rate shock on the rental rate of capital. Importantly, the two gaps react to an interest rate shock in opposite directions. An expansionary monetary policy shock translates into a positive demand shock, i.e in an increase in the output gap which results in inflationary pressure. However, the reaction of the cost of capital is the opposite because $r_t^K = r_t + \delta$. A decline in r_t thus generate a negative cost of capital gap as long as the real interest rate gap is negative, yielding deflationary pressure.

The secular rise in the capital intensity of developed economies might then matter because it results in a composition effect between those two channels. To see this, consider that following a secular change in α (that affects steady-state values and hence leaves gaps unchanged), we get the following change in inflation pass-throughs:

$$\frac{\partial\widehat{\pi}_t}{\partial\alpha} = \lambda\left(-\frac{\phi(1+\phi)}{(1+\alpha\phi)^2}\widehat{y}_t + \frac{1+\phi}{(1+\alpha\phi)^2}\widehat{r}_t^K\right) \tag{5}$$

A change in α reallocates weight from the Taylor channel to the cost channel of monetary policy. If the two channels react differently to an interest rate shock, this implies that the reaction of inflation to a monetary policy shock changes as an economy becomes more capital intensive. Note that this is true even if the slope of the primitive NKPC, λ, that depends only on nominal rigidities and the elasticity of demand to prices, is unchanged.

In Appendix B.1, I investigate the case where capital is fixed, which corresponds to the so-called "small scale" New Keynesian model of Woodford (2003) or Gali (2015). In such a case, production features short-run decreasing returns to scale (homogeneity of degree $1-\alpha$). If α increases, returns to scale decrease and firms must hire more labor to respond to a certain positive demand shock. This results into higher wage pressure and hence higher marginal costs which yields higher inflation. In the "small scale" New Keynesian model, rising capital intensity therefore implies a steepening of the slope of the conventional NKPC.

Unlike changes in nominal or real frictions, a change in the capital intensity of an economy therefore does not affect the slope of the primitive Philipps curve but affects the slope of

[52] Algebra without log-linearization as well as steps to reach equation (4) are provided in Appendix B.3.

the conventional Philipps curve, in line with the empirical result of **?**. The degree to which capital can be adjusted in turn determines the direction of the change in the slope of the conventional Philipps curve. If capital can be adjusted, a cost channel appears that flattens the Phillips curve. If not, returns to increasing labor diminish, resulting in higher wage and inflationary pressure. I now move to an empirical exercise in order to assess which one of these two scenario seems to be prevalent.

3 Empirics

In this section, I propose to look at the respective reactions of differently capital intensive firms to a homogeneous and exogenous monetary policy shock. By showing that capital-intensive firms seem to be gaining market shares at the expense of labor-intensive firms, I show that there exists a cost channel.

3.1 Testable implication of the Cost Channel

To assess whether monetary policy shocks pass to the real economy through a cost channel, I study how market shares reallocate across firms with different capital intensities after a common monetary policy shock. To see the relevance of this measure, suppose that firms feature differently capital-intensive technologies, as documented in **?**. That is, a firm's i production technology at time t is given by:

$$Y_{i,t} = Z_t K_{it}^{\alpha_i} L_{it}^{1-\alpha_i} \tag{6}$$

which implies real marginal costs:

$$mc_{it} = \frac{1}{Z_t} \left(\frac{r_t^K}{\alpha_i} \right)^{\alpha_i} \left(\frac{w_t}{1-\alpha_i} \right)^{1-\alpha_i} \tag{7}$$

A capital-intensive firm (with high α_i) has a marginal cost that is more elastic to a shock to r^K than w compared to a labor-intensive firm (low α_i), that is, to the cost channel of monetary policy. Then, supposing monopolistic competition and a Dixit and Stiglitz (1977) aggregator for demand:

$$Y_t = \left(\int_0^1 Y_{i,t}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}}$$

Yields firms' demand curve which allow to pin down their relative size. A firm's sales share spells: $s_{i,t} = p_{i,t} Y_{i,t} / p_t Y_t$ and so the relative size of a firm i to a firm ι can be written:

$$\frac{s_{i,t}}{s_{\iota,t}} = \left(\frac{mc_{i,t}}{mc_{\iota,t}}\right)^{1-\epsilon}$$

If $\alpha_i > \alpha_\iota$ and a cost channel exists on r^K, then the marginal cost of firm i should increase less that that of firm ι following an expansionary interest rate shock and the capital-intensive firm i should gain market shares. I use this prediction as a test for the presence of a cost channel in the data.[53]

3.2 Data and Variable construction

For firm-level data, I rely on the publicly available Quarterly North American Fundamentals cut of the Compustat database. It provides the financial statements (balance sheets, income statements and cash flow statements) of publicly traded firms in the United States. The dataset is retrieved from the Wharton Research Data Services. The sample starts in 1980, because data prior to that date is sparse and because this paper focuses on trends that start manifesting in the 1980s. The sample ends in 2019 to ignore the COVID episode specificity. I exclude observations with missing sector codes and aggregate the three manufacturing codes (3133), retail (4445), and transport/warehousing (4849), and exclude public administration. The sample counts 162,282 unique firm-year observations, covers 17,708 firms across 19 2-digit NAICS industries. For a complete description of the data, see Appendix A.1.

To build a measure of firms' capital intensity, I compute the capital to labor ratio where capital is the sum of physical capital ("Property, Plant and Equipment (Gross)") and intangible capital ("Intangible Assets") and labor is the count of employees ("Employees"). That is, the capital-intensity of a firm i in sector j at time t is:

$$\text{Kint}_{ijt} \equiv \frac{\text{PPE}_{ijt} + \text{Intangibles}_{ijt}}{\text{Employees}_{ijt}},$$

Because of listing and de-listing which limits the horizons of firms, I build clusters of firms per capital intensity. Within each 2-digit NAICS sector j and calendar year t, I form deciles $d$of the cross-sectional distribution of Kint_{ijt}.[54] For each (j, d, t) cell I aggregate firm sales and construct the deciles within-sector market share:

$$s_{jdt} \equiv \frac{\sum_{i\in(j,d,t)} \text{Sales}_{iht}}{\sum_{i\in(h,t)} \text{Sales}_{ijt}}.$$

For interest rate shocks data, I use the measure of Romer and Romer (2004) updated until

[53] For the documentation of heterogeneity in capital intensity at the firm level, see **?**.

[54] I cluster firms within 2D NAICS sector to account for firm-specific capital intensity and avoid having utility and manufacturing firms as most capital-intensive while service firms are labor intensive.

December 2018.[55] Because my firm-level data is at the quarterly frequency while the shocks are reported monthly, I follow Ottonello and Winberry (2020) and **?** and sum monthly shocks per quarters to obtain a measure for quarterly monetary policy shocks. The resulting series are plotted in Figure A.11 in Appendix A.1.

3.3 Local projections: market-share response by capital intensity

I estimate medium-run impulse responses of market shares to monetary shocks using local projections (**?**). To parsimoniously capture heterogeneity across the capital-intensity distribution, each decile $d \in \{1, \ldots, 10\}$ is mapped to its midpoint on $[0, 1]$, $\text{decile\_mid}_{jd} \in \{0.05, 0.15, \ldots, 0.95\}$, and I define the treatment intensity:

$$x_{jdt} \;\equiv\; \text{shock}_t \times \text{decile\_mid}_{jd}.$$

For horizons $h = 0, 1, \ldots, H$ (baseline $H = 5$ quarters), I estimate

$$s_{jd,\,t+h} \;=\; \theta_{jd} \;+\; \gamma_t \;+\; \beta_h \, x_{jdt} \;+\; \rho \, s_{jd,\,t-1} \;+\; \varepsilon_{jd,\,t+h}, \tag{8}$$

on the panel of sectordecile cells. The specification includes sector×decile fixed effects θ_{jd} (absorbing time-invariant differences in average market shares within sector) and quarter fixed effects γ_t (absorbing aggregate conditions common to all cells). Standard errors are clustered at the sector level. In Figure 3.7, I report $\hat{\beta}_h$ and its 95% confidence bands. Because x_{jdt} is scaled by decile_mid_{jd}, $\hat{\beta}_h$ measures the effect of a one-unit quarterly policy shock on market share *per unit of capital-intensity rank*; the implied topminusbottom decile effect is hence $0.90\,\hat{\beta}_h$ (since $0.95 - 0.05 = 0.90$).

The coefficients are negative at all horizons and most pronounced around three quarters, indicating that when $\text{shock}_t < 0$ (monetary easing), more capital-intensive deciles gain market share relative to less capital-intensive ones. The estimate at $h = 0$ is already negative, implying an *immediate* within-quarter reallocation of market shares following policy easingconsistent with the fact that the quarterly market-share observation can embed shocks occurring within the same quarter and that firms adjust sales rapidly. The magnitude is economically meaningful. Scaling by 0.90 gives the implied D10–D1 response; the IRF peaks around three quarters and then partially reverses. Quantitatively, using the peak estimate $\hat{\beta}_3 \approx -2.2$ (in percentage points per unit shock × rank) and our convention that shock < 0 denotes easing, a typical -25 bp surprise (shock $= -0.25$) implies a D10–D1 increase of $0.90 \times (-2.2) \times (-0.25) \approx +0.50$ percentage points. Interpreting the heterogeneity around the within-quarter mean, the top decile moves by $(0.95 - 0.50) \times (-2.2) \times (-0.25) \approx +0.25$ pp, while the bottom decile moves by $(0.05 - 0.50) \times (-2.2) \times (-0.25) \approx -0.25$ pp, so gains

[55] I use the dataset of Miguel Acosta that can be found here: https://github.com/miguel-acosta/RomerRomer2004/tree/master.

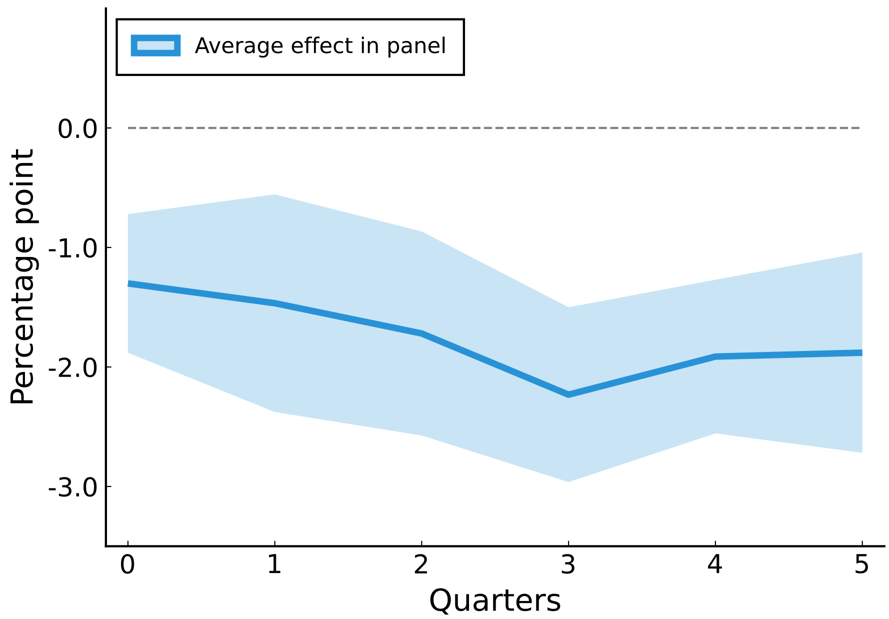


Figure 3.7: Market-Share Reallocation after Monetary Policy Shocks.

Notes: The figure plots impulse responses from the local-projection specification in equation (8), estimated on sectordecile panels at quarterly frequency with horizons $h = 0, \dots, 5$. The solid line shows the estimator $\hat{\beta}_h$; the shaded band is the 95% confidence interval. Quarter fixed effects γ_t and sector×decile fixed effects θ_{jd} are included; standard errors are clustered at the sector level. The treatment variable is the product of the interest rate shock and a firm's capital-intensity decile within sector-quarter. Outcomes are *percentage points* of sector sales. Because quarter fixed effects absorb common movements, the coefficients are identified from *cross-decile heterogeneity* and can be interpreted as deviations around the within-quarter mean; high-K deciles move in the opposite direction of low-K deciles. The implied topminusbottom decile response is $0.90\,\hat{\beta}_h$. Capital intensity is the sum of physical and intangible capital per employee. Deciles are formed within sectorquarter; market share s_{jdt} is decile sales over sectorquarter total sales. Sample is 1980Q12019Q4. Sources are Compustat North American Fundamentals (sales, capital, employees) and Romer and Romer (2004) shocks aggregated to quarters.

at the top are mirrored by losses at the bottom.[56]

Overall, an expansionary interest rate shocks therefore seems to reallocate market shares from labor-intensive to capital-intensive firms, in line with the prediction implied by the presence of a cost channel of monetary policy. To understand the quantitative effect of rising capital intensity on the slope of the conventional NKPC, I now turn to a full-fledged New Keynesian model that I will calibrate on different states of capital intensity and deduce the slope of the conventional NKPC for these different calibrations.

4 Baseline Model

In this section, I build a standard non-linearized New Keynesian model in continuous time. I use this model to simulate the reaction of an economy to a monetary policy shock for different levels of capital-intensity. Namely, I compare the reaction of an economy with the capital-intensity of the American economy in the 1980s, 1990s, 2000s and 2010s. This allows me to retrieve the reaction of inflation and output, among others, and hence to deduce the slope of the conventional NKPC, given a certain level of capital-intensity.

4.1 Environment

Households. The economy is populated by a continuum of measure 1 of symmetrical and infinitely-lived households. Time is continuous. The representative households enjoys consumption flows ($C_t > 0$) and dislikes labor flows ($L_t > 0$) where $L_t \in [0, 1]$ represents the fraction of time endowed to the household that he spends working. Time endowment is normalized to 1. The utility function is strictly increasing and strictly concave in consumption, and strictly decreasing and strictly convex in hours worked. Preferences are time-separable and the future is discounted at rate $\rho \geq 0$. The representative household seeks to maximize the objective[57]:

$$\int_0^{+\infty} e^{-\rho t} \frac{\left(C_t - \frac{L_t^{1+\phi}}{1+\phi}\right)^{1-\gamma}}{1-\gamma} dt \tag{9}$$

Where ϕ is the inverse Frisch elasticity and γ is the constant relative risk aversion coefficient. The objective is maximized subject to a budget constraint that spells:

[56] Equivalently, one can define $x_{jdt} = \text{shock}_t \times (\text{decile\_mid}_{jd} - 0.5)$; with quarter fixed effects this is numerically identical to our baseline but makes the zero-sum property explicit.

[57] The Greenwood et al. (1988) utility function is essential to derive a closed form solution conventional NKPC when capital is endogenous because, since $Y_t = C_t + I_t$ having wealth effects in the intra-temporal condition for the household of the form $C_t L_t^{\phi} = \frac{W_t}{P_t}$ allows only for partial identification of the link between real marginal cost an output.

$$P_tC_t + P_tI_t + \dot{B}_t = i_tB_t + W_tL_t + R_t^K u_t K_t + R_t + T_t \quad (10)$$

The household purchases the final good at price P_t. The household can use his revenue to consume or to save using a bond B_t that returns the nominal interest rate i or investing (I_t) in the capital good. $\dot{B}_t$ denotes adjustment in bonds holding at time t. The investment good and the consumption come at the same price. The household's revenue is also affected by his labor revenue W_tL_t, his claim on firms' profits: R_t, and transfers T_t. The household accumulates capital according to:

$$\dot{K}_t = I_t - \delta K_t \quad (11)$$

And faces an increasing and convex cost function $a(u_t)K_t$ to transform a proportion u_t of its capital stock into rentable capital services, similarly to Christiano et al. (2005). The first order condition of the household yields an Euler equation that relates consumption and saving:

$$\dot{C}_t \; - \; L_t^{\phi}\,\dot{L}_t \; = \; \frac{i_t - \rho - \pi_t}{\gamma}\left(C_t \; - \; \frac{L_t^{1+\phi}}{1+\phi}\right), \quad (12)$$

Where $\pi_t = \frac{\dot{P}_t}{P_t}$ is the inflation rate. It also yields an optimality condition for labor supply:

$$L_t^{\phi} = \frac{W_t}{P_t} \quad (13)$$

A no arbitrage condition *à la* Hall and Jorgenson (1967) relates the allocation of savings between the bond and the capital stock and gives the identity for the real rental rate of capital:

$$r_t^K = (i_t + \delta - \pi_t) \quad (14)$$

This equation is a no arbitrage type; firms invest in the capital good until its return adjusted for depreciation is equal to that of the bond. Finally, I suppose that $a(u_t) = (\sigma_u/2)u_t^2$ so that the optimality condition for the utilization rate is simply:

$$u_t = \frac{r_t^K}{\sigma_u} \quad (15)$$

Firms One competitive representative *final-good producer* aggregates a continuum of measure 1 of intermediate inputs indexed by i *à la* Dixit and Stiglitz (1977):

$$Y_t = \left(\int_0^1 Y_{it}^{\frac{\epsilon-1}{\epsilon}}\, di\right)^{\frac{\epsilon}{\epsilon-1}} \quad (16)$$

Where Y_i is the output of intermediate firms, used as inputs by the final-good producers. Market clearing for the final good implies:

$$Y_t = I_t + C_t \tag{17}$$

That is that the household can chose how to allocate final output into consumption and investment in the capital stock. The final good producers takes the prices of their inputs as given and minimizes their costs, yielding, respectively, the price aggregators and the inverse demand function per variety i:

$$P_t = \left(\int_0^1 P_{it}^{1-\epsilon} di\right)^{\frac{1}{1-\epsilon}} \tag{18}$$

$$\frac{Y_{it}}{Y_t} = \left(\frac{P_{it}}{P_t}\right)^{-\epsilon} \tag{19}$$

The *intermediate-goods producers* are symmetric and share the same constant-returns-to-scale production technology[58]:

$$Y_{it} = Z_t K_{it}^{\alpha} L_{it}^{1-\alpha} \tag{20}$$

Note that I don't index α with a time subscript because I suppose that α isn't a control variable of the firm within the frequency of a shock. This yields the individual nominal marginal cost:

$$MC_t = \frac{1}{Z_{it}} \left(\frac{R_t^k}{\alpha}\right)^{\alpha} \left(\frac{W_t}{1-\alpha}\right)^{1-\alpha} \tag{21}$$

Capital and labor markets clear so that:

$$u_t K_t = \int_0^1 K_{it} di \tag{22}$$

$$L_t = \int_0^1 L_{it} di \tag{23}$$

Intermediate firms compete in a monopolistic industrial organization, allowing them to charge a markup $\mu = \frac{\epsilon}{\epsilon - 1}$ over their marginal costs. Were prices flexible, a firms price would be: $P_{it}^* = \mu MC_{it}$. However, prices are sticky and their adjustment is subject to a **?** adjustment cost. Specifically, when a firm desires to change its price by $\dot{p}_{it}$, it pays a quadratic cost Θ defined by:

[58] In Appendix B.5 I derive the model in the case where firms are heterogeneous in α. Intuitively, this adds a reallocation effect to real variables. I leave this to further work.

$$\Theta(\frac{\dot{P}_{it}}{P_{it}}) = \frac{\theta}{2}(\frac{\dot{P}_{it}}{P_{it}})^2 P_t Y_t \tag{24}$$

Where $\theta > 0$, ensuring that the cost function is increasing in the distance between the targeted price and its current level. To eliminate real resource costs of inflation, I suppose that $\int \Theta(\frac{\dot{P}_{it}}{P_{it}})d_i = T_t$, that is adjustment costs are rebated to households as transfers [59]. A firm maximizes its profit by solving the optimal control problem:

$$V_0(P_{j0}) = \underset{\{P_{jt}\}_{t\in[0,+\infty[}}{max} \int_0^{+\infty} e^{-\int_0^t \rho(s-t)ds} \left(R_{it}(P_{it}) - \Theta(\frac{\dot{P}_{it}}{P_{it}}) \right) dt \tag{25}$$

Where:

$$R_{it} = (P_{it} - MC_{it})\, Y_{it} \tag{26}$$

Is the nominal flow revenue of firm i. Assuming that firms are symmetric, this yields an inflation identity for the firm similar to that retrieved by **?**:

$$(r_t - \frac{\dot{Y}_t}{Y_t})\pi_t = \dot{\pi}_t + \frac{\epsilon - 1}{\theta}(\mu mc_t - 1) \tag{27}$$

Where $mc_t := \frac{MC_t}{P_t}$ is the real marginal cost or the inverse of the implicit markup.

Monetary Policy Finally, the model is closed by supposing that a monetary authority decides on the nominal interest rate following a Taylor rule that takes as antecedent the natural rate of interest i_t^* (the one that prevails under flexible prices), current inflation and the output gap and that converges progressively towards its target (at rate ρ):

$$\frac{i_t}{i_t^*} = \left(\frac{i_{t-1}}{i_t^*}\right)^{\rho} \left(\left(\frac{\pi_t}{\pi_t^*}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_t^*}\right)^{\phi_y} \right)^{1-\rho} e^{\nu_t} \tag{28}$$

And ν_t is a shock.

4.2 Calibration

The solution concept is very simple. I compute one steady-state by decade. These steady-states differ along one exogenous dimension: the output elasticity to capital α. All other parameters remain fix for better isolation of my mechanism. I then shock these decade-specific steady-states with the exact same interest rate shock and compare the reactions of the economy. There are hence two steps to the calibration exercise. First, some usual parameters

[59] Note that $\Theta(\frac{\dot{P}_{it}}{P_{it}}) \sim 0$ anyways for small π_{it}

are simply set to their standard value in the literature in order to make comparison of results direct. I will track variables quarterly. The parameters are summarized in Table 4.11. I suppose that firms share the same TFP, $Z_{it} = 1, \forall\{i, t\}$ and markups are fixed such that $\mu = \frac{\epsilon}{\epsilon-1}$.

Variable	Name	Value	Source
$1/\phi$	Frisch Elas.	0.5	
$1/\sigma_u$	Elas. of K utilization to rental rate of capital	100	Christiano et al. (2005)
ρ	Time discount factor	0.01	Standard quarterly
δ	Depreciation rate	0.025	Gutiérrez and Philippon (2016)
ϵ	Elas. of substitution Between int. goods	6	
φ	Taylor rule: Sensitivity to inflation	1.1	
θ	Price adjustment cost 60	60	Slope of PC = 0.1

Table 4.11: External parameters for the model

Second, I calibrate the aggregate output elasticity to capital and labor per decade. I first compute the yearly aggregate α and then take the average of α per decade. In order to estimate yearly output elasticities to capital, I follow a standard IO way (see for example Syverson (2004) or Foster et al. (2022)) and invert the marginal rate of transformation to obtain:

$$\alpha_t = \frac{R_t^k K_t}{R_t^k K_t + W_t L_t} \tag{29}$$

That is, under a Cobb-Douglas specification with constant returns to scale, the output elasticity to capital is the cost share of capital to total cost. K_t, L_t, R_t^k and W_t are aggregate and year-specific and obtained from the Bureau of Economic Analysis (BEA) Fixed Asset and NIPA tables. K_t is the sum of physical (Structures and Equipment) and intangible (Intellectual Property Products) capital. L_t is a count of employees. To build the cost of constant-quality capital, I follow Gourio and Rognlie (2020) and take the ratio between current-cost investment in fixed assets per sector (Table 2.7) and the quantity index for investment in fixed assets(Table 2.8). The quantity index is built in two steps. First, the historical-cost investment in fixed asset measure is deflated. Second, using hedonic methods pioneered by Gordon (1990) and Cummins and Violante (2002) the measure is adjusted for changes in the quality of capital goods.[60]. Similarly, the cost of labor is "Wages and Salaries

[60]Importantly, as noted by Greenwood et al. (1997), the cost of capital goods and the deflator for capital

Per Full-Time Equivalent Employee by Industry" (Tables 6.6 B, C and D). For a complete description of the data, see Appendix A.1. In Figure 4.8, I plot the evolution of the aggregate output elasticity to capital. Clearly, it displays a secular increase from 0.1 to 0.3 between 1980 and 2019[61].

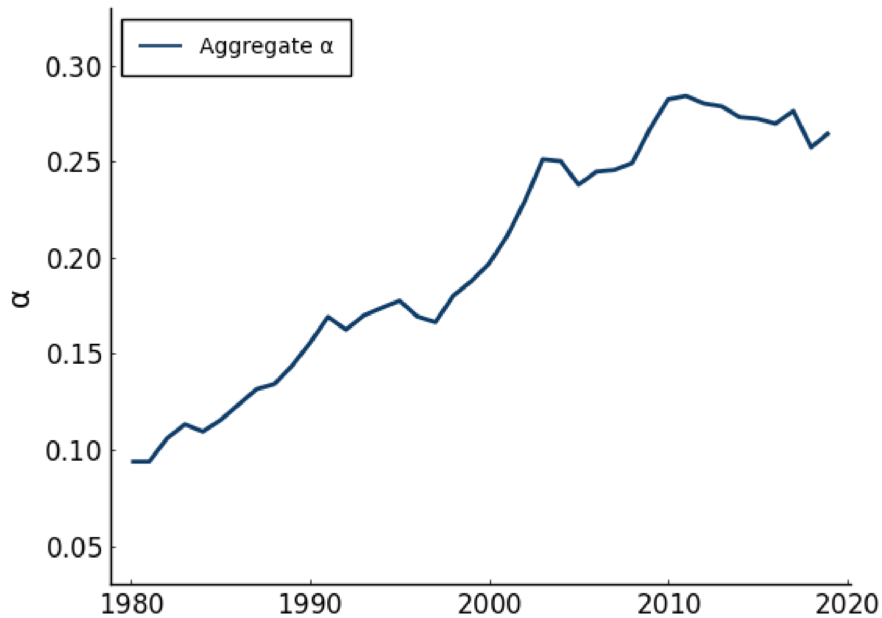


Figure 4.8: Evolution of aggregate output elasticity to capital

Note: Source is BEA. Sample is 1980-2019. α is calculated according to equation (10).

5 Results

In this section, I show that the rising aggregate capital intensity of the American economy implies a flattening of the conventional NKPC in line with the numbers estimated by the empirical literature. I also report impulse response functions for different levels of capital intensity and comment on their difference.

5.1 The Phillips curve across time

In this section I propose to estimate, from the simulations of my model, the evolution of the slope of the Phillips curve in the model and to compare it with empirical findings in the

goods don't coincide because the capital stock and the investment flow don't have the same weights in terms of capital goods types and vintages.

[61]See Appendix B.5 for the Kernel density of the distribution of firms' α for the 1980s and the 2010s.

literature. To do so, consider the *conventional* New Keynesian Phillips curve where current inflation is a function of expected inflation and deviation of output from its steady-state level: $\tilde{Y}_t$ (I write its discrete formulation to match the empirical exercise to be performed on discrete data):

$$\pi_t = \beta E_t(\pi_t) + \kappa_t \tilde{Y}_t$$

Y_t and Y_t^* (the steady-state value of aggregate output prior to a shock[62]) can be retrieved from the model simulation. Indeed, estimating the model-implied slope of the conventional New Keynesian Phillips curve boils down to estimating:

$$\kappa_t = \frac{(\pi_t - \beta E_t(\pi_t))}{\tilde{Y}_t} \tag{30}$$

That is, the deviation of inflation from its anchor to the deviation of output from the steady-state. In order to get how κ_t changes across decades in my model, I shock four different steady-state calibrations of my model with a 100bp decline in the nominal interest rate. Each calibration differs from the other under one single dimension; the underlying distribution of firms' output elasticity to capital (depicted in Figure 4.8). Calibrations are decade-specific as exposed in section 4.2. As a result, I obtain four sets of impulse response functions that contain, in particular $(\pi_t - \beta E_t(\pi_t))$ and $\tilde{Y}_t$, allowing me to retrieve κ_t. The decade specific κ_t are reported in Figure 5.9. For completeness, I also plot $\lambda = \frac{\epsilon-1}{\theta}$, the slope of the primitive NKPC.

Overall, the slope coefficient of the conventional New Keynesian Phillips declines by a factor of almost 4 between the 1980s and the 2010s, within the range of empirical estimates. This corresponds for example to half of the decline estimated by **?** who compare components of PCE inflation to year-on-year changes in unemployment, and to twice the decline estimated by Hazell et al. (2020) who provide estimates of the Phillips curve slope from the cross-section of US states[63]. Interestingly, in line with Blanchard et al. (2015) and **?**, the sharpest decline happens between the 1980s and the 1990s.

[62] Notice that I index steady-state values with subscript t because they depend on the distribution of firms' α if r^k and w differ in the steady-state.

[63] They construct state-level price indices for non-tradeable goods back to 1978 in order to overcome identification problems related to shifting long-run inflation expectations and mitigating supply/demand shocks.

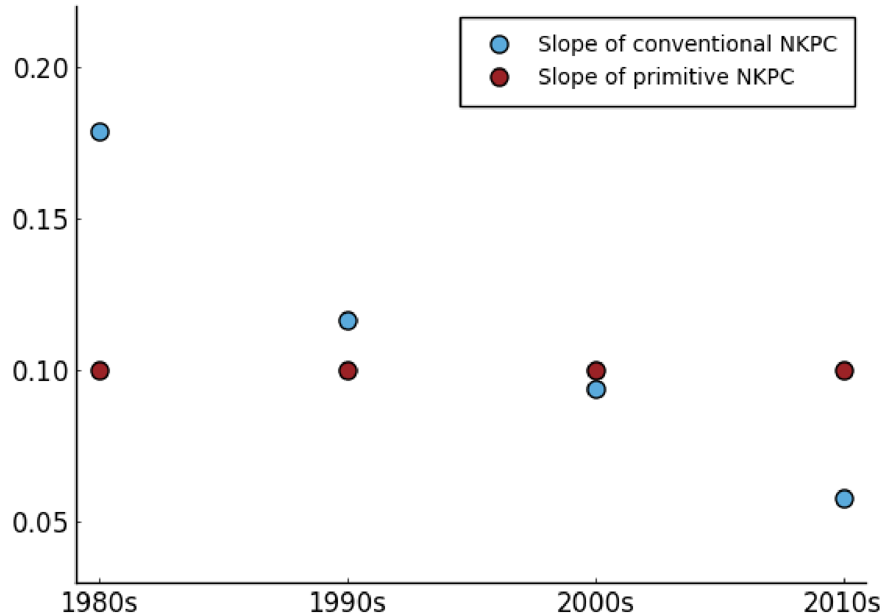


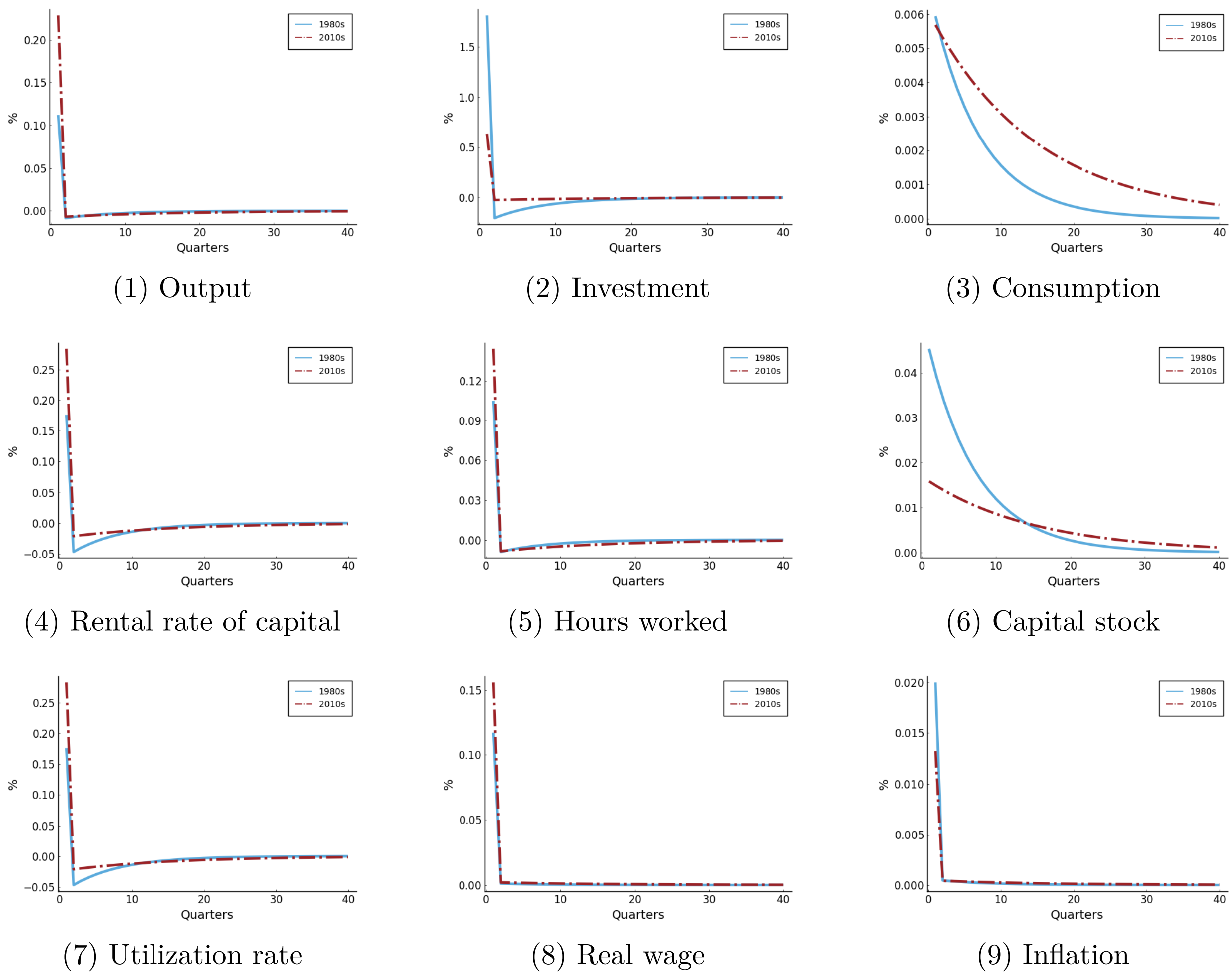
Figure 5.9: Model-implied slopes of Phillips curves per decade

Note: The slope of the Phillips curve is calculated according to equation (30). $(\pi_t - \beta E_t(\pi_t))$ and $\tilde{Y}_t$ are log deviations from steady state. The shock that generated a deviation is a 100 basis point monetary policy shocks. The 1980s calibration differs from the 2010s calibration only in the distribution of firms output elasticity to capital displayed in Figure B.13.

5.2 Impulse responses

In this section, I report the impulse response functions (IRFs) of endogenous variables approximated at the first order to a 100bp shock to the nominal interest rate for the 1980s calibration and the 2018s calibration. Because those two calibrations differ only in the underlying distribution of capital-intensity; this allows to isolate the effects of increasing average and dispersion of capital intensity. The IRFs are displayed in Figure 5.10.

Panel (1) reports the respective reaction of output in percent deviation from the steady-state. To facilitate comparisons, both steady-state output have been calibrated to 1. Clearly, in the 2010s, output reacts more to an expansionary monetary policy shock. To understand why, consider the sub-components of output. Panel (3) shows that consumption reaction is roughly similar in level on impact and rather small. Panel (2) shows that the more labor-intensive economy increases investment more (in % deviation) which results in a higher rise in the capital stock (panel (6)) while panel (7) shows that the utilization rate hikes more for the capital-intensive economy. Panel (2) and (6) seem to be the more puzzling facts here, but recall that variables are reported as deviations from steady-state. In the labor-intensive economy (1980s), steady-state investment is lower than in the capital-intensive economy (2010s) making a similar increase in level of investment bigger in elasticity terms for the labor-intensive economy.



(1) Output (2) Investment (3) Consumption

(4) Rental rate of capital (5) Hours worked (6) Capital stock

(7) Utilization rate (8) Real wage (9) Inflation

Figure 5.10: Comparison of impulse responses: 1980s vs. 2010s calibrations

Note: IRFs are in percentage deviations from steady state. Shocks are 100 basis point monetary policy shocks. The 1980s calibration differs from the 2010s calibration only in the distribution of firms output elasticity to capital displayed in Figure B.13.

6 Conclusion

This paper investigates how the secular rise in the capital intensity of U.S. production structurally altered the transmission of nominal shocks to real variables. In particular, I show that such a transformation is orthogonal to the slope of the primitive NKPC but has ambiguous effects on the slope of the conventional NKPC, depending on modeling hypothesis. In the data, the effect is unambiguous: the conventional NKPC flattens as an economy becomes more capital intensive. I then build a simple New Keynesian model with variable utilization rates of capital and a rental rate of capital that depends positively on the real interest rate and show that, as the output elasticity to capital rises, the slope of the conventional NKPC flattens.

This is because, when production is labourintensive, an interest rate shock influences firms marginal costs through its effect on real wages. As the economy becomes more capitalintensive, however, the rental rate of capital-itself an increasing function of the interest rate-emerges as the dominant component of marginal cost, thereby strengthening the "cost channel" that attenuates the conventional negative comovement between the policy rate and inflation. This shift occurs without altering the primitive NKPC which links marginal cost and inflation; instead, it diminishes the response of marginal costs to a demand shock and therefore the slope of the conventional NKPC. This is, to the best of my knowledge, the only mechanism so far that can account for a stable primitive NKPC slope and a declining conventional NKPC slope over the last decades documented by **?**.

This paper needs several adjustments. First, for the empirical exercise, data on market shares at a higher frequency would help ensure other confounding affects are neutralized. Second, for the theoretical exercise, because these findings are derived from a firstorder, loglinear solution, they capture only local dynamics. A fuller assessment requires higherorder approximations capable of accommodating nonlinearities, level effects, and potential asymmetries between rate hikes and cuts. Future work should also integrate reallocation stemming from heterogeneity in capital-intensity across firms to gauge how heterogeneity in firms capital intensity redistributes the gains and losses of monetary interventions, in particular through the lens of stock prices. Finally, exploring alternative model ingredients -capital adjustment costs, nominal wage rigidities, and habit formation in consumption- will help isolate which assumptions are essential.

Appendix

A Empirics

A.1 Data

Compustat North American Fundamentals (Annual). To build firm- and year-specific measures I use the publicly available Annual North American Fundamentals cut of the Compustat database. It provides the financial statements (balance sheets, income statements and cash flow statements) of publicly traded firms in the United States. The dataset is retrieved from the Wharton Research Data Services. The sample starts in 1980, because data prior to that date is sparse and because this paper focuses on trends that start manifesting in the 1980s. The sample ends in 2019 to ignore the COVID episode specificity. I focus mainly on firms' sales, physical and intangible capital and employment. I exclude observations with missing or negative sales, capital and employees. I exclude observations not nested within a sector (no information or "Other"). The sample counts 162,282 unique firm-year observations, covers 17,708 firms across 19 2-digit NAICS industries.[64] I build the following three variables:

Sales (pY) measured as *"Sales/Turnover - Total"* in Compustat.

Capital (K) is the total capital stock. It is measured by summing the stocks of tangible capital (*"Property, Plant and Equipment - Total (Gross)"*) and intangible capital (*"Intangible Assets - Total"*).

Labor (L) is measured with *"Employees"* . In robustness tests *"Cost of Goods Sold"* (COGS) and "Selling, General and Administrative expenses" (SGA) are used as alternatives.

Bureau of Economic Analysis (BEA). I build three sets of sector-specific measures using data from the BEA: real gross output, deflators for output and capital and the constant-quality relative cost of capital.

Real gross output by sector is retrieved from the "GDP-by-industry" tables of the BEA which provides a series for real gross output (in 2017 billions of dollars) per 3-digit NAICS sector. Because some 3D sectors don't have data prior to 1997 (every 3D sectors in the 2D sector "Retail trade" for example), I focus on 2D sectors.

To deflate firms' output, I compute a deflator for each 2D sector by dividing real gross output by nominal gross output from the "GDP-by-industry" tables of the BEA. Similarly, for capital goods I use data from the "Fixed Asset Tables" of the BEA. I divide the Current-Cost Stock of Capital (Table 2.1) with the Historical-Cost Stock of Capital (Table 2.3) which gives me 3 separate deflators for "Structures", "Equipment" and "Intellectual Property Products"

[64] Public administration is not considered. Manufacturing codes 31, 32 and 33 are consolidated. Same for Retail Trade (44 and 45) and Transportation and Warehousing (48, 49)

for each 2D sector. I then use a weighted sum (by relative stock) of the deflator for Structures and Equipment as the deflator for physical capital and the deflator for Intellectual Property Products as the deflator for intangible capital.

Finally, I use data from the BEA to construct the relative cost of *constant-quality* capital (RCK) following Gourio and Rognlie (2020). To build sector-specific cost of constant-quality capital, I take the ratio between current-cost investment in fixed assets (Table 2.7) and the quantity index for investment in fixed assets (Table 2.8). The quantity index is built in two steps. First, the historical-cost investment in fixed asset measure is deflated for each sector. Second, using hedonic methods pioneered by Gordon (1990) and Cummins and Violante (2002) the measure is adjusted for changes in the quality of capital goods.[65]. The cost of consumption is the price index for non-durable goods obtained from Table 1.2.4 from National Income and Product Accounts (NIPA) tables. The cost of labor per industry is "Wages and Salaries Per Full-Time Equivalent Employee by Industry" (Tables 6.6 B, C and D).

In the baseline, I do not consider the fall in the interest rate as an additional driver for the decline in the relative cost of capital (in the form of a rental rate) because, as pointed by Farhi and Gourio (2018) for example, risk and liquidity differences between the required return to capital and the risk-free interest rate makes their correlation uncertain. For completeness, in the Appendix, I repeat my estimations and results including the fall of the interest rate.

Monetary policy shocks. For interest rate shocks data, I use the measure of Romer and Romer (2004) updated until December 2018[66]. Because my firm-level data is at the quarterly frequency while the shocks are reported monthly, I follow Ottonello and Winberry (2020) and **?** and sum monthly shocks per quarters to obtain a measure for yearly monetary policy shocks. The resulting series are plotted in Figure A.11.

[65]Importantly, as noted by Greenwood et al. (1997), the cost of capital goods and the deflator for capital goods don't coincide because the capital stock and the investment flow don't have the same weights in terms of capital goods types and vintages.

[66]I use the dataset of Miguel Acosta that can be found here: https://github.com/miguel-acosta/RomerRomer2004/tree/master.

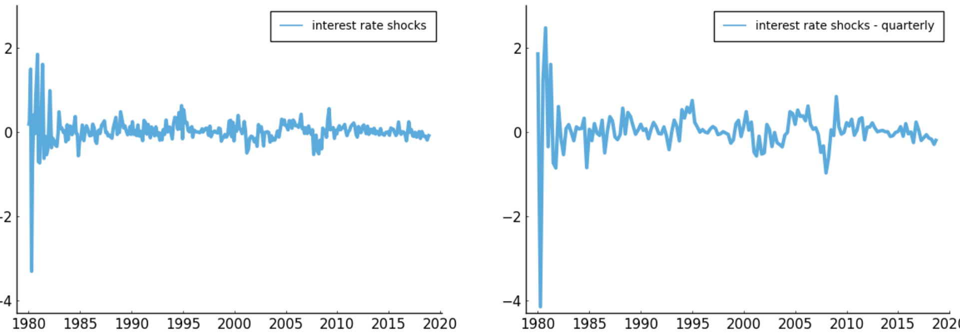


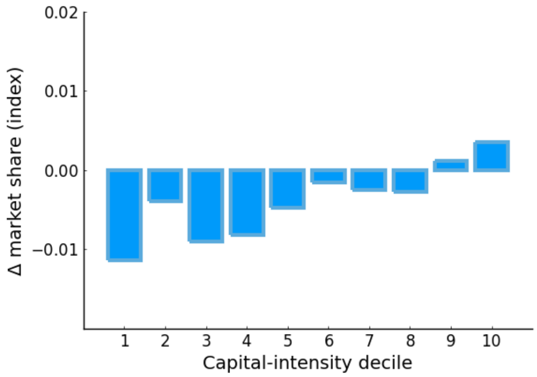
Figure A.11: Monetary Policy shocks à la Romer and Romer (2004) between 1980 and 2018: monthly (left) vs. quarterly (right) aggregation.

Note: Source is https://github.com/miguel-acosta/RomerRomer2004/tree/master.

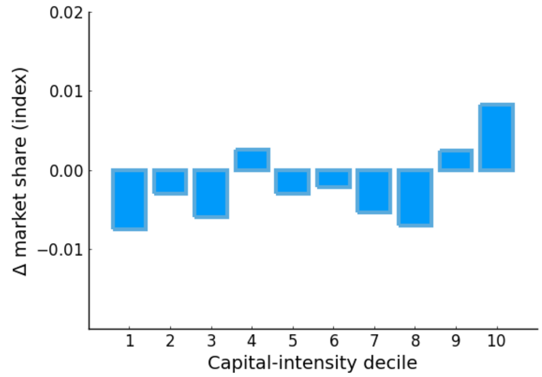
A.2 Market share response per capital-intensity

In this exercise, I compare the quarter-over-quarter change in within-sectors market shares of firms following a negative interest rate shock. Within sectors, I nest my firms in deciles of the capital-intensity distribution. For each sectors, this gives me a vector of 10 market share changes following negative shock. I then average those deciles-specific changes across sectors. That is, I take the mean of the change in market shares of the 10% most capital-intensive firms within Manufacturing, Information, Retail Trade ...etc (and similarly for each decile). This leaves me with a vector of 10 changes per quarters. I then normalize these changes by the size of the shock by dividing the change by the absolute value of the shock. Finally, I average out those changes over all the quarters within a decade and report the results for the 1980S, 1990s, 2000s and 2010s. The results are reported in Figure A.12.

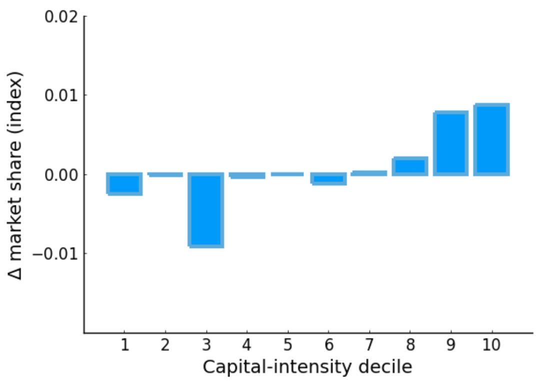
For every decade, most capital-intensive firms (right-most deciles) seem to gain market shares, suggesting that their marginal cost increases less than that of their labor intensive competition. Since capital-intensive firms have a marginal cost that is more elastic to the rental rate of capital than their labor-intensive competition, this suggests that the rental rate of capital is less elastic to a negative interest rate shock that the wage. With time, as aggregate capital-intensity rises and the distribution of firms' capital-intensity within sectors fans out, effects appear to become stronger. Note that the market share changes don't sum up to one for two reasons. First, because they are averaged without weighting across sectors and across years. Second, because the market share changes are normalized by the shock sizes. This is essential to make the 1980s, where shocks were on average larger, to the subsequent decades.



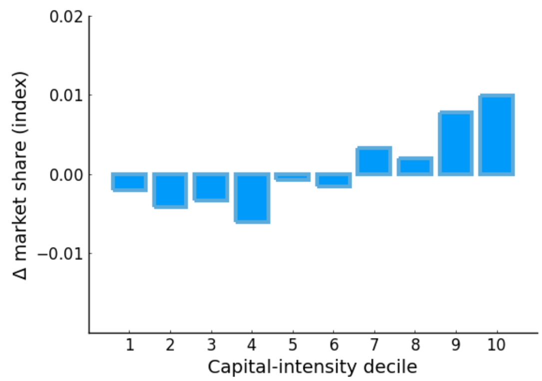
(1) 1980s



(2) 1990s



(3) 2000s



(4) 2010s

Figure A.12: Reallocation of market shares across firms following a negative shock to the interest rate.

Note: Source is Compustat and shocks from Romer and Romer (2004) aggregated quarterly. Sample is 1980-2019. Measure of Capital intensity is (PPE + INT)/EMP. PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". EMP is expressed in number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. Market share is a firm's sales ("Sales/Turnover (Net)") to the total sales of its 2D-NAICS sector.

B Theory

B.1 Fixed capital stock

If capital is "clay" within the frequency of a monetary policy shock, as discussed in Chapter 5 of Woodford (2003), production features short-run decreasing returns to scale and equation (2) can be rewritten (algebra related to this section is presented in Appendix B.2):

$$Y_t = A_t L_t^{1-\alpha} \tag{31}$$

Where $A_t = Z_t \overline{K}_t^{\alpha}$ where $\overline{K}_t$ is the nonadjustable capital stock. In such a case, the marginal cost simply spells: $mc_t = \frac{w_t}{1-\alpha}\left(\frac{Y_t^{\alpha}}{A_t}\right)^{\frac{1}{1-\alpha}}$ and accounting profits can be written: $R_t = p_t Y_t - TC_t = (p_t - (1-\alpha)mc_t)Y_t$. In the frictionless equilibrium, profit maximization yields: $p_t = \frac{\epsilon}{\epsilon-1}(1-\alpha)mc_t$[67]. Then, the primitive NKPC is:

$$\rho\pi_t = \dot{\pi}_t + \lambda(\mu mc_t - 1)$$

Where λ, the slope of the primitive NKPC is orthogonal to α and depends on nominal frictions and the elasticity of demand to prices. To get to the conventional NKPC, one can consider a usual intra-temporal condition from the household's problem to replace w to get the conventional NKPC:

$$\rho\pi_t = \dot{\pi}_t + \lambda\left(\left(\frac{Y_t}{Y_t^*}\right)^{\frac{\phi-1}{1-\alpha}} - 1\right) \tag{32}$$

Where ϕ is the inverse Frisch elasticity[68]. If $\phi \neq 1$, the relationship between the output gap and inflation depends on α. Namely, for the standard case where $\phi > 1$, as an economy becomes more capital-intensive (α rises), returns in terms of Y from increasing L decrease, and since capital is forced to be fixed, a demand shock that results in an increase in the output gap needs a higher adjustment in terms of L to be fulfilled. This higher adjustment in L implies a higher rise in the wage which passes to a higher increase in the marginal cost and finally inflation. Even if the slope of the primitive NKPC is unchanged, the slope of the conventional NKPC is affected by a change in α. In a case were capital is fixed, the slope increases. This difference in slopes' dynamics can be expressed by the changing relationship

[67]Note that $(1-\alpha)mc_t$ is often rewritten as ψ_t where ψ is called the "cost function" (this is the case in Gali (2015) for example) and the "markup" is defined as $\mu = \frac{\epsilon}{\epsilon-1}$. There, ψ is the average cost and μ is hence a markup over the average cost rather than the marginal cost. Here, since α and the marginal cost NKPC are our object of interest, I don't use this short cut. Then, $\mu = \frac{\epsilon}{\epsilon-1}(1-\alpha)$ is the desired markup over the nominal marginal cost under flexible-price equilibrium, or equivalently, the inverse of the flexible-price real marginal cost.

[68]Here the intra-temporal condition of the household is one without wealth effects. I use such a specification as it will help derive the conventional NKPC in the case where K is endogenous.

between marginal cost and output gap. Here, in log-deviations from steady-states (variables with hats):

$$\frac{\partial \widehat{mc}_t}{\partial \widehat{Y}_t} = \frac{\phi - 1}{1 - \alpha}$$

And hence:

$$\frac{\partial \left(\frac{\partial \widehat{mc}_t}{\partial \widehat{Y}_t} \right)}{\partial \alpha} = \frac{\phi - 1}{(1 - \alpha)^2} \tag{33}$$

If capital is fixed in the short-run, a more capital-intensive economy, with higher α, sees its marginal cost increase more following a demand shock and this translates, by the primitive NKPC into higher inflation.

I now move to a case where capital can be adjusted, and show that the relationship can be the opposite: higher alpha flattens the conventional NKPC.

B.2 Derivations for fixed capital

In this case, the production technology features short-run decreasing returns to scale and can be written: $Y_{it} = A_{it} L_{it}^{1-\alpha_i}$. The marginal cost is: $mc_{it} = \frac{w_t}{1-\alpha_i} \left(\frac{Y_{it}^{\alpha_i}}{A_{it}} \right)^{\frac{1}{1-\alpha_i}} = \frac{TC_{it}}{(1-\alpha_i)} Y_{it}$ for TC the total cost. So, accounting profits can be spelled: $R_{it} = p_{it} Y_{it} - TC_{it} = (p_{it} - (1 - \alpha_i) mc_{it}) Y_{it}$. In the frictionless equilibrium, profit maximization yields: $p_{it} = \frac{\epsilon}{\epsilon - 1}(1 - \alpha_i) mc_{it}$[69] so $\mu = \frac{\epsilon}{\epsilon-1}(1 - \alpha_i)$ is the desired markup under flexible-price equilibrium, or equivalently, the inverse of the flexible-price real marginal cost.

If the Rotemberg adjustment cost is defined by: $\Theta(\frac{\dot{p}_{it}}{p_{it}}) = \frac{\theta}{2}(\frac{\dot{p}_{it}}{p_{it}})^2 P_t Y_t$, then a firm solves the optimal control problem:

$$V_0(p_{j0}) = \underset{\{p_{jt}\}_{t \in [0,+\infty[}}{max} \int_0^{+\infty} e^{-\int_0^t \rho(s-t) ds} \left(R_{it}(p_{it}) - \Theta(\frac{\dot{p}_{it}}{p_{it}}) \right) di$$

subject to its demand curve (equation (19)), the adjustment cost function (equation (24)) and the revenue identity (equation (26)). This translates into the Hamiltonian:

$$\mathcal{H}(p_{it}, \dot{p}_{it}, \eta) = (p_{it} - (1 - \alpha_i) mc_{it}) \left(\frac{p_{it}}{P_t} \right)^{-\epsilon} Y_t - \frac{\theta}{2} \left(\frac{\dot{p}_{it}}{p_{it}} \right)^2 P_t Y_t + \eta \dot{p}_{it}$$

Where $\dot{p}_{it}$ is the control variable, p_{it} the state and η the co-state. This yields the following

[69] Note that $(1 - \alpha_i) mc_{it}$ is often rewritten as ψ_{it} where ψ is called the "cost function" (this is the case in Gali (2015) for example) and the "markup" is defined as $\mu = \frac{\epsilon}{\epsilon-1}$. There ψ is the average cost and μ is hence a markup over the average cost rather than the marginal cost. Here, since α and the marginal cost Phillips curve are our object of interest, I don't use this short cut.

Phillips curve if firms are symmetric:

$$\rho\pi_t = \dot{\pi}_t + \frac{\epsilon - 1}{\theta}(\mu mc_t - 1)$$

Where $mc_t := \frac{mc_t}{p_t}$ is the inverse induced markup and $\rho = (r_t - \frac{\dot{C}_t}{C_t}) = (r_t - \frac{\dot{Y}_t}{Y_t})$ because we ignore investment.

To get to the conventional New Keynesian Phillips, consider a usual intra-temporal condition with GHH preferences (equation (13) in the baseline model): $L_t^\phi = \frac{W_t}{P_t}$. One can hence rewrite the real marginal cost:

$$mc_t = \frac{1}{1-\alpha}\left(\frac{Y_t}{A_t}\right)^{\frac{\phi-1}{1-\alpha}} \tag{34}$$

Recall that in the flexible price equilibrium, $mc_t = \frac{1}{\mu}$. Denoting Y_t^* the natural output (the one that prevails in the flexible-price equilibrium), we can hence rewrite:

$$\mu mc_t = \left(\frac{Y_t}{Y_t^*}\right)^{\frac{\phi-1}{1-\alpha}} \tag{35}$$

And the conventional New Keynesian Phillips curve can be written:

$$\rho\pi_t = \dot{\pi}_t + \frac{\epsilon-1}{\theta}\left(\left(\frac{Y_t}{Y_t^*}\right)^{\frac{\phi-1}{1-\alpha}} - 1\right) \tag{36}$$

B.3 Derivations for endogenous capital

In this case, the production technology features constant returns to scale and can be written: $Y_{it} = Z_{it}K_{it}^{\alpha_i}L_{it}^{1-\alpha_i}$. The marginal cost is: $mc_{it} = \frac{1}{Z_{it}}\left(\frac{r_t^K}{\alpha_i}\right)^{\alpha_i}\left(\frac{w_t}{1-\alpha_i}\right)^{1-\alpha_i} = TC_{it}Y_{it}$ for TC the total cost. So, accounting profits can be spelled: $R_{it} = p_{it}Y_{it} - TC_{it} = (p_{it} - mc_{it})Y_{it}$. In the frictionless equilibrium, profit maximization yields: $p_{it} = \frac{\epsilon}{\epsilon-1}mc_{it}$ so $\mu = \frac{\epsilon}{\epsilon-1}$ is the desired markup of price over the marginal cost under flexible-price equilibrium, or equivalently, the inverse of the flexible-price real marginal cost.

If the Rotemberg adjustment cost is defined by: $\Theta(\frac{\dot{p}_{it}}{p_{it}}) = \frac{\theta}{2}(\frac{\dot{p}_{it}}{p_{it}})^2 P_t Y_t$, then a firm solves the optimal control problem:

$$V_0(p_{j0}) = \underset{\{p_{jt}\}_{t\in[0,+\infty[}}{max} \int_0^{+\infty} e^{-\int_0^t \rho(s-t)ds}\left(R_{it}(p_{it}) - \Theta(\frac{\dot{p}_{it}}{p_{it}})\right) di$$

subject to its demand curve (equation (19)), the adjustment cost function (equation (24)) and the revenue identity (equation (26)). This translates into the Hamiltonian:

$$\mathcal{H}(p_{it},\dot{p}_{it},\eta)=(p_{it}-mc_{it})(\frac{p_{it}}{P_t})^{-\epsilon}Y_t-\frac{\theta}{2}(\frac{\dot{p}_{it}}{p_{it}})^2P_tY_t+\eta\dot{p}_{it}$$

Where $\dot{p}_{it}$ is the control variable, p_{it} the state and η the co-state. This yields the following aggregate primitive NKPC if firms are symmetric:

$$(r_t-\frac{\dot{Y}_t}{Y_t})\pi_t=\dot{\pi}_t+\frac{\epsilon-1}{\theta}(\mu mc_t-1)$$

Where $mc_t := \frac{mc_t}{p_t}$ is firm's i inverse induced markup and $r_t = i_t - \pi_t$ is the real interest rate. To get to the conventional NKPC, let us replace w_t using the intra-temporal condition for the household (equation (13) in the baseline model) and r^K using the no-arbitrage between bonds and capital condition of Hall and Jorgenson (1967) (equation (14) in the baseline model), the marginal cost becomes:

$$\begin{aligned} mc_t &= \frac{1}{Z_t}\left(\frac{r_t+\delta}{\alpha}\right)^{\alpha}\left(\frac{L_t^{\phi}}{1-\alpha}\right)^{1-\alpha} \\ &= \frac{1}{Z_t}\left(\frac{r_t+\delta}{\alpha}\right)^{\alpha}\left(\frac{\left(\left(\frac{1-\alpha}{\alpha}\frac{r_t^K}{w_t}\right)^{\alpha}Y_t\right)^{\phi}}{1-\alpha}\right)^{1-\alpha} \end{aligned} \tag{37}$$

Where I have used the marginal rate of transformation: $\frac{K_t}{L_t}=\frac{\alpha}{1-\alpha}\frac{w_t}{r_t^K}$ which yields $Y_t = \left(\frac{\alpha}{1-\alpha}\frac{w_t}{r_t^K}\right)^{\alpha}L_t$. Recall that the desired markup is the inverse of the real natural marginal cost, defined as: $mc_t^* = \frac{1}{Z_t}\left(\frac{r_t^*+\delta}{\alpha_i}\right)^{\alpha_i}\left(\frac{w_t^*}{1-\alpha_i}\right)^{1-\alpha_i}$ and so μmc_t can be written:

$$\mu mc_t = \left(\frac{r_t+\delta}{r_t^*+\delta}\right)^{1+\phi\alpha}\left(\frac{w_t^*}{w_t}\right)^{\alpha\phi(1-\alpha)}\left(\frac{Y_t}{Y_t^*}\right)^{\phi(1-\alpha)} \tag{38}$$

And so the conventional NKPC becomes:

$$\rho\pi_t=\dot{\pi}_t+\frac{\epsilon-1}{\theta}\left(\left(\frac{r_t+\delta}{r_t^*+\delta}\right)^{1+\phi\alpha}\left(\frac{w_t^*}{w_t}\right)^{\alpha\phi(1-\alpha)}\left(\frac{Y_t}{Y_t^*}\right)^{\phi(1-\alpha)}-1\right)$$

As in the fixed capital case, the slope of the conventional NKPC depends on α. Note that here, $\frac{r_t+\delta}{r_t^*+\delta}$ and $\frac{w_t}{w_t^*}$ which correspond to the deviations of the unitary costs of the inputs, are functions of $\frac{Y_t}{Y_t^*}$, defining $f:\frac{Y_t}{Y_t^*}\to\frac{r_t+\delta}{r_t^*+\delta}$ and $g:\frac{Y_t}{Y_t^*}\to\frac{w_t}{w_t^*}$, with $f'>0$ and $g'>0$, we get:

$$\rho\pi_t=\dot{\pi}_t+\lambda\left(f\left(\frac{Y_t}{Y_t^*}\right)^{1+\phi\alpha}g\left(\frac{Y_t}{Y_t^*}\right)^{\alpha\phi(\alpha-1)}\left(\frac{Y_t}{Y_t^*}\right)^{\phi(1-\alpha)}-1\right) \tag{39}$$

Since λ is orthogonal to α, to understand if an increasing α flattens or enhances the slope of the conventional NKPC, one can, again, consider how a change in α alters the pass-through of the output gap to the marginal cost (in logs, where hats denote deviations from steady-state):

$$\frac{\partial \widehat{mc}_t}{\partial \widehat{Y}_t} = \left((1+\phi\alpha) f'\left(\widehat{Y}_t\right) + \alpha\phi(\alpha-1) g'\left(\widehat{Y}_t\right) + \phi(1-\alpha)\right) \tag{40}$$

And now the derivative with respect to α:

$$\frac{\frac{\partial \widehat{Y}_t}{\partial \widehat{mc}_t}}{\partial \alpha} = \left(\alpha f'\left(\widehat{Y}_t\right) + (1+\phi\alpha)\frac{\partial f'\left(\widehat{Y}_t\right)}{\partial \alpha} + \alpha\phi(\alpha-1)\frac{\partial g'\left(\widehat{Y}_t\right)}{\partial \alpha} + \phi(2\alpha-1) g'\left(\widehat{Y}_t\right) - \phi\right) \tag{41}$$

The sign of this derivative is ambiguous. Under standard calibrations, the first three components of the parenthesis are positive while the last two are negative. The second component is positive because, as the economy becomes more capital intensive, a demand shock on the goods market translates into a stronger demand shock on the capital goods market from firms. By symmetry, this means that the shock on the labor market is softer and hence that $\partial g'\left(\widehat{Y}_t\right)/\partial\alpha < 0$ and so that the third term is positive. The fourth term is negative because $\alpha < 0.5$ typically holds. When an economy becomes more capital-intensive, the change in the relationship between the marginal cost and the output gap is hence ambiguous when capital can adjust at the shock frequency. If $(\frac{\partial \widehat{mc}_t}{\partial \widehat{Y}_t})/\partial\alpha < 0$, then the conventional NKPC flattens as an economy becomes more capital-intensive because the pass-through of fluctuations in the output gap to the marginal cost weakens, in line with the evidence unveiled by **?**. This can also be understood through the relative reaction of the marginal cost gap and the output gap to a nominal interest rate shock (see Appendix B.3). The remainder of the paper is dedicated to understanding which forces dominate and to retrieving the quantitative change in the slope of the conventional NKPC given a changing α. In the next section, I test empirically which force dominates using firm-level data. In the model section, I build a New Keynesian model that I simulate on different calibrations of α and show how the slope of the conventional NKPC changes.

The link between marginal cost and interest rates. One can also look at how a shock to the nominal interest rate affects the marginal cost:

$$\frac{\partial mc_t}{\partial i_t} = \left(\frac{\alpha}{r_t^K}\frac{\partial r_t^K}{\partial i_t} + \frac{1-\alpha}{w_t}\frac{\partial w_t}{\partial i_t}\right) mc_t. \tag{42}$$

That is, the total change in the real marginal cost following an interest rate shock is a weighted sum of how the real rental rate of capital and the real wage react to the shock. A change in the capital-intensity of an economy (or a firm) alters all together those pass-through, their respective weights and the marginal cost initial level. To see this, consider the plain derivative of equation (42) with respect to α:

$$\begin{aligned}\frac{\partial\left(\frac{\partial mc_t}{\partial i_t}\right)}{\partial\alpha} \quad = \quad & \overbrace{\left(\frac{\alpha}{r_t^K}\frac{\partial r_t^K}{\partial i_t}+\frac{1-\alpha}{w_t}\frac{\partial w_t}{\partial i_t}\right)\frac{\partial mc_t}{\partial\alpha}}^{\text{Level}} \quad + \quad \overbrace{\left(\frac{1}{r_t^K}\frac{\partial r_t^K}{\partial i_t}-\frac{1}{w_t}\frac{\partial w_t}{\partial i_t}\right)mc_t}^{\text{Weights}} \\ & + \quad \underbrace{\left(\frac{\alpha}{r_t^K}\frac{\partial\left(\frac{\partial r_t^K}{\partial i_t}\right)}{\partial\alpha}+\frac{1-\alpha}{w_t}\frac{\partial\left(\frac{\partial w_t}{\partial i_t}\right)}{\partial\alpha}\right)mc_t}_{\text{Costs}}\end{aligned} \tag{43}$$

The first component, "Level", refers to how the real marginal cost differs if α changes (due to differences in equilibrium r^K and w or due to concavity effects). The second component, "Weights", refers to the reallocation of weights from the real wage channel to the real rental rate channel (if the two differ). The third component, "Costs", refers to demand effects; if the economy is more capital-intensive, demand effects for the capital good will be heavier than for labor. Intuitively, those three components ought to matter. The level effect because $r^k < w$ typically holds in standard calibrations. The weights and costs effect matter *iff* $\partial r_t^K/\partial i_t \neq \partial w_t/\partial i_t$ which can hold true for two reasons. First, in equilibrium, r^k is an increasing function of the real interest rate; in a classic Hall and Jorgenson (1967) setting where the price of the final and capital goods are the same, it spells: $r_t^K = (i_t + \delta - \pi_t)$ with δ the depreciation rate of capital. In such a case, if labor is not rented, the negative relationship between r^K and i_t is dampened if not reversed compared to that between w_t and i_t. This would imply: $|\partial r_t^K/\partial i_t| < |\partial w_t/\partial i_t|$. However, and second, capital is typically thought to not be set statically. If a firm chooses K_t, its choice of K_t is bounded by the available capital stock in t which is typically given by: $K_t = (1-\delta)K_{t-1} + I_t$. As a shock happens, firms must wait for I_t to adjust in order to adjust their use of capital while adjustment for labor is immediate. In such a case, r_t^K captures, in the short-run, all the demand effect for K_t. This would instead imply: $|\partial r_t^K/\partial i_t| > |\partial w_t/\partial i_t|$.

As such, a change in the capital intensity of an economy impacts the pass-through of an interest rate shock to the marginal cost and hence inflation, even if the structural relationship between marginal costs and inflation is unchanged. However, the direction of this impact is theoretically uncertain. I now propose an empirical exercise to test which of the two forces dominates.

A log-linear approximation. Let hats denote log-deviations from steady state. The capital-to-labor ratio identity coupled with GHH preferences and the marginal cost identity can be rewritten:

$$\begin{aligned}(1+\alpha\phi)\,\widehat{l}_t &= \widehat{y}_t + \alpha\widehat{r}_t^K \\ \widehat{mc}_t &= \alpha\widehat{r}_t^K + (1-\alpha)w_t\end{aligned}$$

Therefore:

$$\widehat{mc}_t = \frac{\phi(1-\alpha)}{1+\alpha\phi}\widehat{y}_t + \frac{\alpha(1+\phi)}{1+\alpha\phi}\widehat{r}_t^K$$

And so the log-linearized version of the NKPC spells:

$$\rho\widehat{\pi}_t = \dot{\pi}_t + \lambda\left(\frac{\phi(1-\alpha)}{1+\alpha\phi}\widehat{y}_t + \frac{\alpha(1+\phi)}{1+\alpha\phi}\widehat{r}_t^K\right)$$

This yields the following co-movement in marginal cost and output gaps:

$$\frac{\partial\,\widehat{mc}_t}{\partial\,\widehat{y}_t} = \frac{\phi(1-\alpha)+\alpha(1+\phi)\psi_t}{1+\alpha\phi}$$

Where:

$$\psi_t \equiv \frac{\partial\,\widehat{r}_t^K}{\partial\,\widehat{y}_t}.$$

Is the reaction of the rental rate of capital to a demand shock. This relationship is positive iff:

$$\psi_t > -\frac{\phi(1-\alpha)}{\alpha(1+\phi)}.$$

The relationship between the output and marginal cost gap evolves with α as follows:

$$\frac{\partial}{\partial\alpha}\left(\frac{\partial\,\widehat{mc}_t}{\partial\,\widehat{y}_t}\right) = \frac{1+\phi}{(1+\alpha\phi)^2}\Big(\psi_t - \phi\Big) + \frac{\alpha(1+\phi)}{1+\alpha\phi}\psi_\alpha$$

Where $\psi_\alpha \equiv \partial\psi_t/\partial\alpha$.

B.4 Solving the model

B.4.1 The equilibrium equations

The equilibrium of the model is characterized by 23 equations. The household block contains 5 equations. The firm block contains 8 equations per firms. The market structure 5 equations. The market-clearing 5 equations.

Household The Hamiltonian spells:

$$\mathcal{H} = ln(C_t) + \frac{L_t^{1+\phi}}{1+\phi} + \lambda_t(i_t B_t + W_t L_t + R_t^K K_t + R_t + T_t - P_t C_t - Q_t I_t) + \eta_t(I_t - \delta K_t)$$

Where B_t and K_t are the state variables while C_t, L_t, I_t and B_t are the controls. λ_t, η_t are the shadow value of an extra unit of B_t and K_t respectively and the co-state of the problem. This yields the optimality equations **??** to 14. Together with the budget constraint and the law of motion for capital, these constitute all the optimality equations of the household block:

$$P_t C_t + Q_t I_t + \dot{B}_t = i_t B_t + W_t L_t + R_t^K K_t + R_t + T_t \tag{H.1}$$

$$\dot{K}_t = I_t - \delta K_t \tag{H.2}$$

$$\frac{\dot{C}_t}{C_t} = (i_t - \rho - \pi_t) \tag{H.3}$$

$$C_t L_t^{\phi} = \frac{W_t}{P_t} \tag{H.4}$$

$$R_t^K = (i_t + \delta) Q_t - \dot{Q}_t \tag{H.5}$$

Market structure and demand From the CES aggregator we obtain:

$$C_t = \left(\int_0^1 C_{it}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{MS.1}$$

$$I_t = \frac{1}{\xi_t} \left(\int_0^1 I_{it}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{MS.2}$$

$$P_t = \left(\int_0^1 P_{it}^{1-\epsilon} di \right)^{\frac{1}{1-\epsilon}} \tag{MS.3}$$

$$Q_t = \xi_t P_t \tag{MS.4}$$

$$\frac{Y_{it}}{Y_t} = \left(\frac{P_{it}}{P_t} \right)^{-\epsilon} \tag{MS.5}$$

Firms Using the demand and production function gives us the marginal cost and inverse demand for inputs. With price adjustment function we get the optimal price from the profit maximisation.

$$Y_{it} = Z_{it} K_{it}^{\alpha_i} L_{it}^{1-\alpha_i} \tag{F.1}$$

$$MC_{it} = \frac{1}{z_{it}} \left(\frac{R_t^k}{\alpha_i} \right)^{\alpha_i} \left(\frac{W_t}{1-\alpha_i} \right)^{1-\alpha_i} \tag{F.2}$$

$$\mu_t W_t = P_{it}(1-\alpha_i)\frac{Y_{it}}{L_{it}} \tag{F.3}$$

$$\mu_t R_t^K = P_{it}\alpha_i \frac{Y_{it}}{K_{it}} \tag{F.4}$$

$$P_{it}^* = \mu MC_{it} \tag{F.5}$$

$$\Theta(\frac{\dot{P}_{it}}{P_{it}}) = \frac{\theta}{2}(\frac{\dot{P}_{it}}{P_{it}})^2 P_t Y_t \tag{F.6}$$

The Hamiltonian of the firm spells:

$$\mathcal{H} = (P_{it} - MC_{it})(\frac{P_{it}}{P_t})^{-\epsilon} Y_{it} - \frac{\theta}{2}(\frac{\dot{P}_{it}}{P_{it}})^2 P_t Y_t + \eta_t \dot{P}_{it}$$

Where $\dot{P}_{it}$ is the control variable, P_{it} is the state and η_t is the co-state. The optimal price change of a firm is given by:

$$\rho_{it}\frac{\dot{P}_{it}}{P_{it}} = \dot{\pi}_{it} + \frac{\epsilon - 1}{\theta}(\mu mc_{it} - 1) \tag{F.7}$$

$$R_{it} = (P_{it} - MC_{it})\, Y_{it} \tag{F.8}$$

Market-clearing Defining:

$$Y_{it} = C_{it} + I_{it} \tag{MC.1}$$

We get:

$$Y_t = C_t + \xi_t I_t \tag{MC.2}$$

$$L_t = \int_0^1 L_{it} di \tag{MC.3}$$

$$K_t = \int_0^1 K_{it} di \tag{MC.4}$$

$$B_t = 0 \tag{MC.5}$$

Policy

$$i_t = i_t^* + \varphi \pi_t \tag{P.1}$$

Where:

$$i_t^* = \rho \tag{P.2}$$

B.4.2 The steady state

The steady-state is defined by zero-inflation: $\pi_t = 0$ and stationary variables $\dot{X}_t = X, \forall X$

$$PC + QI + \dot{B} = iB + WL + R^K K + R + T \tag{SS.1}$$

$$\dot{K} = I - \delta K \tag{SS.2}$$

$$i = \rho \tag{SS.3}$$

$$CL^\phi = \frac{W}{P} \tag{SS.4}$$

$$R^K = (i + \delta)Q \tag{SS.5}$$

$$C = \left(\int_0^1 C_i^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{SS.6}$$

$$I = \frac{1}{\xi} \left(\int_0^1 I_i^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{SS.7}$$

$$P = \left(\int_0^1 P_i^{1-\epsilon} di \right)^{\frac{1}{1-\epsilon}} \tag{SS.8}$$

$$Q = \xi P \tag{SS.9}$$

$$\frac{Y_i}{Y} = \left(\frac{P_i}{P}\right)^{-\epsilon} \tag{SS.10}$$

$$Y_i = Z_i K_i^{\alpha_i} L_i^{1-\alpha_i} \tag{SS.11}$$

$$MC_i = \frac{1}{z_{it}} \left(\frac{R^k}{\alpha_i}\right)^{\alpha_i} \left(\frac{W}{1-\alpha_i}\right)^{1-\alpha_i} \tag{SS.12}$$

$$\mu W = P_i(1-\alpha_i)\frac{Y_i}{L_i} \tag{SS.13}$$

$$\mu R^K = P_i \alpha_i \frac{Y_i}{K_i} \tag{SS.14}$$

$$P_i = \mu MC_i \tag{SS.15}$$

$$R_i = (P_i - MC_i)\, Y_i \tag{SS.16}$$

$$Y_i = C_i + I_i \tag{SS.17}$$

$$Y = C + \xi I \tag{SS.18}$$

$$L = \int_0^1 L_i di \tag{SS.19}$$

$$K = \int_0^1 K_i di \tag{SS.20}$$

$$B = 0 \tag{SS.21}$$

B.5 The case of heterogeneous firms

Suppose the same household and aggregator as in the baseline model but consider now the case where firms are heterogeneous in their output elasticity to input.

The *intermediate-goods producers* share the same constant-return-to-scale production technology but differ along two dimensions: TFP, z_{it}, and output elasticity to capital, α_{it} (and, by symmetry, to labor):

$$Y_{it} = Z_{it} K_{it}^{\alpha_i} L_{it}^{1-\alpha_i} \tag{HF.1}$$

These differences carry to the individual nominal marginal cost:

$$MC_{it} = \frac{1}{z_{it}} \left(\frac{R_t^k}{\alpha_i} \right)^{\alpha_i} \left(\frac{W_t}{1-\alpha_i} \right)^{1-\alpha_i} \tag{HF.2}$$

This implies that firms' marginal costs are heterogeneously elastic to changes in input costs. A capital-intensive firm (with high α_i) will see its marginal cost react less to changes in W_t than a labor-intensive firm and vice-versa to changes to R_t^K. Capital and labor markets clear so that:

$$u_t K_t = \int_0^1 K_{it} di \tag{HF.3}$$

$$L_t = \int_0^1 L_{it} di \tag{HF.4}$$

Intermediate firms compete in a monopolistic industrial organization, allowing them to charge a markup $\mu = \frac{\epsilon}{\epsilon - 1}$ over their marginal costs. Were prices flexible, a firms price would be: $p_{it}^* = \mu mc_{it}$. However, prices are sticky and their adjustment is subject to a **?** adjustment cost. Specifically, when a firm desires to change its price by $\dot{p}_{it}$, it pays a quadratic cost Θ defined by:

$$\Theta(\frac{\dot{P}_{it}}{P_{it}}) = \frac{\theta}{2} (\frac{\dot{P}_{it}}{P_{it}})^2 P_t Y_t \tag{HF.5}$$

Where $\theta > 0$, ensuring that the cost function is increasing in the distance between the targeted price and its current level. To eliminate real resource costs of inflation, I suppose that $\int \Theta(\frac{\dot{P}_{it}}{P_{it}}) d_i = T_t$, that is adjustment costs are rebated to households as transfers [70]. A firm maximizes its profit by solving the optimal control problem:

$$V_0(P_{j0}) = \max_{\{P_{jt}\}_{t \in [0,+\infty[}} \int_0^{+\infty} e^{-\int_0^t \rho(s-t) ds} \left(R_{it}(P_{it}) - \Theta(\frac{\dot{P}_{it}}{P_{it}}) \right) dt \tag{HF.6}$$

Where:

$$R_{it} = (P_{it} - MC_{it}) Y_{it} \tag{HF.7}$$

is the flow revenue of firm i. This yields an inflation identity for the firm similar to that

[70] Note that $\Theta(\frac{\dot{P}_{it}}{P_{it}}) \sim 0$ anyways for small π_{it}

retrieved by **?** but where the marginal cost is firm-specific:

$$(r_t - \frac{\dot{Y}_t}{Y_t})\pi_{it} = \dot{\pi}_{it} + \frac{\epsilon - 1}{\theta}(\mu mc_{it} - 1)s_{it} \tag{HF.8}$$

Where $mc_{it} := \frac{mc_{it}}{p_{it}}$ is the inverse of firm's i implicit markup[71] and $s_{it} = \frac{P_{it}Y_{it}}{P_t Y_t}$. A convenient property of the continuous time modeling is that firm-specific inflation identities can be aggregated into (proof in appendix **??**):

$$\pi_t = \int_0^{+\infty} \pi_{it} s_{it} di$$

which allows to decompose aggregate inflation into three components:

$$\Delta\pi_t = \int_0^{+\infty} \Delta\pi_{it}\overline{s_i}di + \int_0^{+\infty} \overline{\pi}_i \Delta s_{i,t} + Cov(\pi_{it}, s_{it})$$

Where a $\overline{x}$ indicates an unchanged variable x (it's pre-shock steady state value for example).

B.5.1 Firms' Phillips curve

If the Rottemberg adjustment cost is defined by: $\Theta(\frac{\dot{p}_{it}}{p_{it}}) = \frac{\theta}{2}(\frac{\dot{p}_{it}}{p_{it}})^2 P_t Y_t$, then a firm solves the optimal control problem:

$$V_0(p_{j0}) = \max_{\{p_{jt}\}_{t\in[0,+\infty[}} \int_0^{+\infty} e^{-\int_0^t \rho(s-t)ds} \left(R_{it}(p_{it}) - \Theta(\frac{\dot{p}_{it}}{p_{it}}) \right) di$$

subject to its demand curve (equation (19)), the adjustment cost function (equation (24)) and the revenue identity (equation (26)). This translates into the Hamiltonian:

$$\mathcal{H}(p_{it}, \dot{p}_{it}, \eta) = (p_{it} - mc_{it})(\frac{p_{it}}{P_t})^{-\epsilon} Y_t - \frac{\theta}{2}(\frac{\dot{p}_{it}}{p_{it}})^2 P_t Y_t + \eta \dot{p}_{it}$$

Where $\dot{p}_{it}$ is the control variable, p_{it} the state and η the co-state. This yields the following firm-level inflation identity:

$$(r_t - \frac{\dot{Y}_t}{Y_t})\pi_{it} = \dot{\pi}_{it} + \frac{\epsilon - 1}{\theta}(\mu mc_{it} - 1)s_{it}$$

Where $mc_{it} := \frac{mc_{it}}{p_{it}}$ is firm's i inverse induced markup and $s_{it} = \frac{p_{it}Y_{it}}{P_t Y_t}$ is the revenue share of firm i.

[71] Note that this is not the real marginal cost any more in that the denominator is firm specific. As is usual, a firm increases its price when its implicit markup falls too far below its desired markup.

B.5.2 Characterization of the Aggregate Phillips curve

Recall the price aggregator: $P_t^{1-\epsilon} = \int_0^1 p_{it}^{1-\epsilon} di$ which can be rewritten:

$$e^{(1-\epsilon)log(P_t)} = \int_0^1 e^{(1-\epsilon)log(p_{it})} di$$

Notice than $\pi_t = \frac{\dot{P}_t}{P_t} = \frac{\partial log(P_t)}{\partial t}$, hence taking the derivative of the above rewriting of the price aggregator with respect to time, we obtain:

$$\begin{aligned}
&(1-\epsilon)e^{(1-\epsilon)log(P_t)} \frac{\partial log(P_t)}{\partial t} = (1-\epsilon) \int_0^1 e^{(1-\epsilon)log(p_{it})} \frac{\partial log(p_{it})}{\partial t} di \\
\Leftrightarrow\ & P_t^{1-\epsilon} \pi_t = \int_0^1 p_{it}^{1-\epsilon} \pi_{it} di \\
\Leftrightarrow\ & \pi_t = \int_0^1 (\frac{p_{it}}{P_t})^{1-\epsilon} \pi_{it} di
\end{aligned}$$

But $(\frac{p_{it}}{P_t})^{1-\epsilon} = \frac{p_{it}y_{it}}{P_t Y_t} = s_{it}$ is the sales share of firm i. As such, aggregate inflation can be written:

$$\pi_t = \int_0^1 s_{it} \pi_{it} di$$

B.5.3 Calibration for heterogeneous firms

Instead of calibrating the economy with a single α, I suppose that the economy counts 1000 firms at every decade. Each firm draws its α from the Kernel density of α that prevails at that time. A decade's Kernel density is the average of the yearly kernel densities within that decade. In order to estimate firm-specific output elasticities to capital, I follow a standard IO way (see for example Syverson (2004) or Foster et al. (2022)) and invert the marginal rate of transformation to obtain:

$$\alpha_i = \frac{r_t^k K_{it}}{r_t^k K_{it} + w_t L_{it}} \tag{HF.9}$$

That is, under a Cobb-Douglas specification with constant returns to scale, the output elasticity to capital is the cost share of capital to total cost. K_{it} and L_{it} are firm and time specific and obtained from Compustat while r_t^k and w_t are 2-digit NAICS sector-specific and obtained from the Bureau of Economic Analysis (BEA) Fixed Asset and NIPA tables. K_{it} is the sum of physical ("Property, Plant and Equipment (Gross)" in Compustat) and intangible ("Intangible Assets") capital. L_{it} is a count of employees. To build the cost of constant-quality capital per 2D NAICS sector, I follow Gourio and Rognlie (2020) and take the ratio between current-cost investment in fixed assets per sector (Table 2.7) and the quantity index for investment in fixed assets per sector (Table 2.8). The quantity index is built in two steps. First, the historical-cost investment in fixed asset measure is deflated. Second, using hedonic methods pioneered by Gordon (1990) and Cummins and Violante (2002) the measure is adjusted for changes in the quality of capital goods.[72]. Similarly, the cost of labor per industry is "Wages and Salaries Per Full-Time Equivalent Employee by Industry" (Tables 6.6 B, C and D). For a complete description of the data, see appendix A.1. In Figure B.13, I plot the Kernel density of the distribution of firms' α for the 1980s, 1990s, 2000 and 2010s. Clearly, the distributions are unimodal and the mode is consistently between 0 and 0.2. However, the density of that node declines secularly from a decade to another. In the 2010s, a firm is approximately three times less likely to be located at the mode than in the 1980s. This flattening of the distribution of alpha is mostly at the benefit of firms with high alpha as the relative mass of firms with an alpha lower than the mode also shrinks with time.

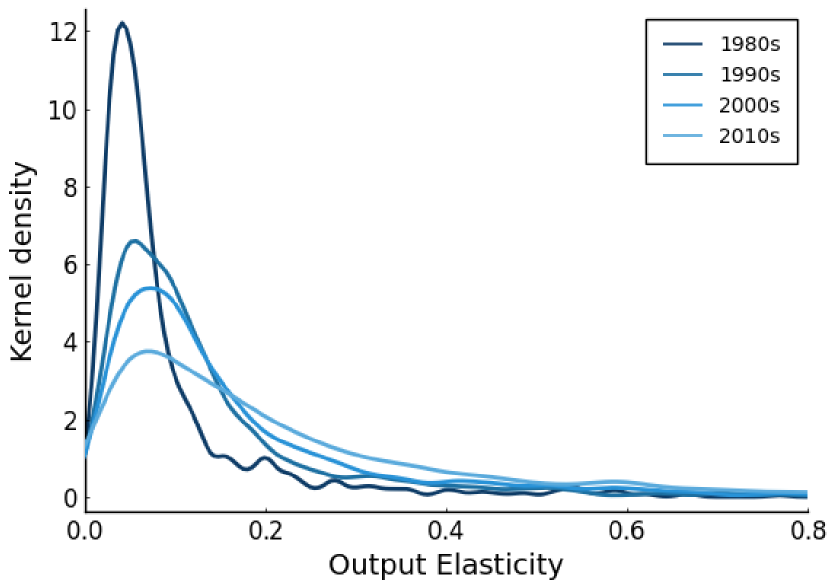


Figure B.13: Kernel density of distribution of firms' output elasticity to capital per decade.

Note: Source is Compustat. Sample is 1980-2019. Firms' α is calculated according to equation (10).

[72]Importantly, as noted by Greenwood et al. (1997), the cost of capital goods and the deflator for capital goods don't coincide because the capital stock and the investment flow don't have the same weights in terms of capital goods types and vintages.

Markups vs. Machines*

Geraud Desazars de Montgailhard

March 13, 2025

Abstract

A common practice when estimating markups using the "production function approach" is to assume homogeneous output elasticity to inputs across firms. In this paper, I first show that this assumption biases upward both the dispersion of firms markups and the estimated increase in average and aggregate markups. I then show that this assumption doesn't hold and propose a method to co-estimate firms' markups and output elasticity to inputs by complementing the production function approach with firms' cost shares. Even after accounting for dispersion in output elasticity to capital, the rise in markups dispersion and in the aggregate markup resist remarkably well. However, I find that levels are significantly lower than previously estimated by the literature.

*Bocconi University. `geraud.desazars@phd.unibocconi.it`. I am grateful to Basile Grassi, Dmitriy Sergeyev, Nicolas Crouzet and Yueran Ma for their continuous advice and support.

1 Introduction

There is increasing evidence of a rise in the dispersion of firms' markups of prices over marginal costs. The implications are many and of first-order importance for academics and policy makers: rising profits at the expense of workers (de Loecker et al., 2020) and capitalists (Gutiérrez and Philippon, 2016; Barkai, 2020), declining incentives to innovate as firms escape competition (**?**Aghion et al., 2023), rising real rigidities in the pass-through of nominal shocks to real variables (Meier and Reinelt, 2020; Baqaee et al., 2021).

Most of this evidence relies on the "production function approach" developed in Hall (1988). With this method, if at least one input is set statically and firms are price takers for this input, then the first-order condition that determines demand for this input can be inverted to isolate -and estimate- the markup. An unobserved variable is necessary to conduct this estimation: the elasticity of output to the input in consideration. Recently, two papers, de Loecker et al. (2020) and de Ridder et al. (2022) have proposed to rely on the two-stage GMM method of Ackerberg et al. (2015) to estimate those production function parameters and have reached the conclusion that firms' markups have become more dispersed in the past decades thus raising both average and aggregate markups.

However, those papers rely on production function estimates that are aggregated, at least at the sectoral level, implying that firms share the same elasticities of output to inputs. Yet, in **?**, I show that firms differ in capital-intensity and show that those differences can be traced to firm-level heterogeneity in the output elasticity to inputs. In this paper, I show that omitting firm heterogeneity in output elasticity to inputs biases upward the estimates of markup dispersion and the average and aggregate markups.

Theoretically, I show that such a bias could cancel completely the conclusions of de Loecker et al. (2020) and de Ridder et al. (2022). I therefore re-estimate firm-level markups under the same set of hypotheses as de Loecker et al. (2020) but with firm-specific output elasticity to inputs estimated using cost-shares. I find that the conclusion of rising markups dispersion and rising aggregate markups resist remarkably well qualitatively but that markup levels are dampened.

Finally, firms' output elasticity to inputs also affect the estimator of firms' Total Factor Productivity of Revenue (TFPR) when the latter is estimated as a Solow residual on revenue data. I show that when allowing for firm-specific input-biased technological differences, key moments of the distribution of TFPR change. Average TFPR is 33% lower in 1980 and 44% lower in 2019 when output elasticity to inputs are firm-specific than when they are homogeneous. Median TFPR is consistently 30% lower over the sample. The dispersion of firms' TFPR is the most affected moment. In 1980, the variance of firms' TFPR is 50% lower while in 2019, it is 83% lower, in line with rising dispersion in firms' capital intensity.

The rest of the paper is organized as follows: Section 2 shows how supposing homogeneity in firms' output elasticity to inputs biases markups estimates and I propose a method to co-estimate markups and firm-specific technologies. Section 3 presents the results of the markups estimation and compares them to the existing literature. In section 4 I document the impact of firm-specific output elasticities to inputs on the estimation of firms' Solow resiudal. Section 5 concludes and outlines the next steps of this project.

2 Co-estimating firms' markups and production functions

This section is divided in two parts. First, I show that supposing that firms share the same output elasticity to input biases the estimate of their markup. Second, I propose a method to estimate, together, firm-specific output elasticity to inputs and firm-specific markups.

2.1 The estimation bias

Suppose that a firm i is endowed with productivity Z_i and chooses a combination of two inputs L_i and K_i in order to produce output Y_i with technology F, such that $F : \{Z_i, K_i, L_i\} \to Y_i$ where F is increasing and concave in K_i and L_i and increasing and linear in Z_i. Its cost minimization problem yields the first order condition for L_i:

$$w = mc_i \frac{\partial F}{\partial L_i}$$

Where w is the cost of input L and mc_i is the marginal cost of firm i and the multiplier of the Lagrangien of its cost minimization problem. This can be rewritten in terms of output elasticity to L_i and markups μ_i. Denote the output elasticity to L_i as $\zeta_i = \frac{\partial F}{\partial L_i} \frac{L_i}{Y_i}$ and the markup of price over marginal cost as: $\mu_i = \frac{p_i}{mc_i}$ to get to the Hall (1988) formula:

$$\mu_i = \frac{\zeta_i}{\lambda_i^L} \tag{1}$$

Where $\lambda_i^L = \frac{wL_i}{p_i Y_i}$ is the income share of input i. Since input shares are typically observable in publicly available datasets with firm-level data (such as financial statements), an econometrician only needs to estimate ζ_i to retrieve μ_i. A common hypothesis in macroeconomic models is that firms share the same production technologies up to a constant, the TFP component (here denoted Z_i). That is, that firms share the same $\zeta_i = \overline{\zeta}$. If this isn't true, the induced bias for estimates of true μ_i, $\hat{\mu}_i$ can be retrieved easily:

$$b_{\mu_i} := \frac{\hat{\mu}_i}{\mu_i} = \frac{\overline{\zeta}}{\zeta_i} \tag{2}$$

That is, for L representing labor, a firm with small labor-intensity and hence small ζ_i will be mistaken for a firm with high markup. This assumption seems particularly critical in light of the fact that manufacturing seems to be the sector that contains the firms whose markups rose the most. Indeed, as shown in **?**, manufacturing is also the sector that has seen its capital intensity and output elasticity to capital rise the most. These firm-level biases can be amplified when computing the aggregate markup. Suppose that the true aggregate markup, $\mathcal{M}$, can be written[73]:

[73] This aggregation identity holds in most cases. See, for example, Grassi (2017), Edmond et al. (2015), Edmond et al. (2023) or **?**.

$$\mathcal{M} = \left(\int_{\mathcal{I}} \frac{s_i}{\mu_i} d_i \right)^{-1} \tag{3}$$

Where $\mathcal{I}$ is the universe of all firms and s_i is the sales share of total sales of firm i: $\frac{p_i Y_i}{pY}$ (with p and Y aggregate price and output). If one considers instead the biased measure: $\hat{\mu}_i$, the resulting biased aggregate markup spells:

$$\hat{\mathcal{M}} = \left(\int_{\mathcal{I}} \frac{\zeta_i}{\bar{\zeta}} \frac{s_i}{\mu_i} d_i \right)^{-1}$$

The bias for the aggregate markup therefore depends on $Cov(\frac{\zeta_i}{\bar{\zeta}}, s_i)$. If bigger firms have smaller ζ_i, as is the case (see **?**), the bias for the aggregate markup is amplified.

2.2 Estimating firm-specific output elasticities

In order to solve this issue, I propose to estimate firm-specific output elasticity to inputs and use this estimate for the estimation of markups. We can rewrite equation (1) as: $\zeta_i = \frac{wL_i}{mc_i Y_i}$. If F is homogeneous of degree 1, $mc_i Y_i = TC_i$ where TC_i is the total production cost of firm i. That is:

$$\zeta_i = \frac{wL_i}{TC_i} \tag{4}$$

The output elasticity to input L of firm i is the cost share of input L for firm i. If costs are observed to the econometrician, one can insert equation (4) into equation (1) to retrieve a firm's true markup:

$$\mu_i = \frac{\frac{wL_i}{TC_i}}{\lambda_i^L} \tag{5}$$

I now turn to estimating firms' markups using this identity. Moments of the distribution of firms' output elasticity to capital are plotted in Figure A.19 in Appendix A.2 and discussed extensively in **?**.

3 Results

3.1 Data

This section presents data sources and measure constructions. Further information is detailed in Empirical Appendix A.1.

Compustat North American Fundamentals (Annual). To build firm- and year-specific measures I use the publicly available Annual North American Fundamentals cut of the Compustat database. It provides the financial statements (balance sheets, income statements and cash flow statements) of publicly traded firms in the United States. The dataset is retrieved from the Wharton Research Data Services. Similarly to de Loecker et al. (2020), the sample starts in 1950 and ends in 2016. I focus mainly on firms' sales, physical and intangible capital and employment. I exclude observations with missing or negative sales, capital and employees. I exclude observations not nested within a sector (no information or "Other"). The sample counts 162,282 unique firm-year observations, covers 17,708 firms across 19 2-digit NAICS industries[74].

Bureau of Economic Analysis (BEA). I build three sets of sector-specific measures using data from the BEA: real gross output, deflators for output and capital and the constant-quality relative cost of capital.

Real gross output by sector is retrieved from the "GDP-by-industry" tables of the BEA which provides a series for real gross output (in 2017 billions of dollars) per 3-digit NAICS sector. Because some 3D sectors don't have data prior to 1997 (every 3D sectors in the 2D sector "Retail trade" for example), I focus on 2D sectors.

To deflate firms' output, I compute a deflator for each 2D sector by dividing real gross output by nominal gross output from the "GDP-by-industry" tables of the BEA. Similarly, for capital goods I use data from the "Fixed Asset Tables" of the BEA. I divide the Current-Cost Stock of Capital (Table 2.1) by the Historical-Cost Stock of Capital (Table 2.3) which gives me 3 separate deflators for "Structures", "Equipment" and "Intellectual Property Products" for each 2D sector. I then use a weighted sum (by relative stock) of the deflator for Structures and Equipment as the deflator for physical capital and the deflator for Intellectual Property Products as the deflator for intangible capital.

Finally, I use data from the BEA to construct the relative cost of *constant-quality* capital (RCK) following Gourio and Rognlie (2020). To build sector-specific cost of constant-quality capital, I take the ratio between current-cost investment in fixed assets (Table 2.7) and the quantity index for investment in fixed assets (Table 2.8). The quantity index is built in two steps. First, the historical-cost investment in fixed asset measure is deflated for each sector. Second, using hedonic methods pioneered by Gordon (1990) and Cummins and Violante (2002) the measure is adjusted for changes in the quality of capital goods.[75]. The cost of consumption is the price index for non-durable goods obtained from Table 1.2.4 from National Income and Product Accounts (NIPA) tables. The cost of labor per industry is "Wages and Salaries Per Full-Time Equivalent Employee by Industry" (Tables 6.6 B, C and D).

I do not consider the fall in the interest rate as an additional driver for the decline in the relative cost of capital (in the form of a rental rate) because, as pointed out by Farhi and Gourio (2018) for example, risk and liquidity differences between the required return to capital and the risk-free interest rate make their correlation uncertain.

[74]Public administration is not considered. Manufacturing codes 31, 32 and 33 are consolidated. Same for Retail Trade (44 and 45) and Transportation and Warehousing (48, 49)

[75]Importantly, as noted by Greenwood et al. (1997), the cost of capital goods and the deflator for capital goods don't coincide because the capital stock and the investment flow don't have the same weights in terms of capital goods types and vintages.

3.2 Firms' markups

In Figure 3.14, I plot moments of the distribution of markups estimated according to equation (5). In red, I replicate the exercise of de Loecker et al. (2020) by setting the output elasticity to labor of every firm to the aggregate cost share of labor. In blue, I compute firms' markups using their own cost share of labor for every year.

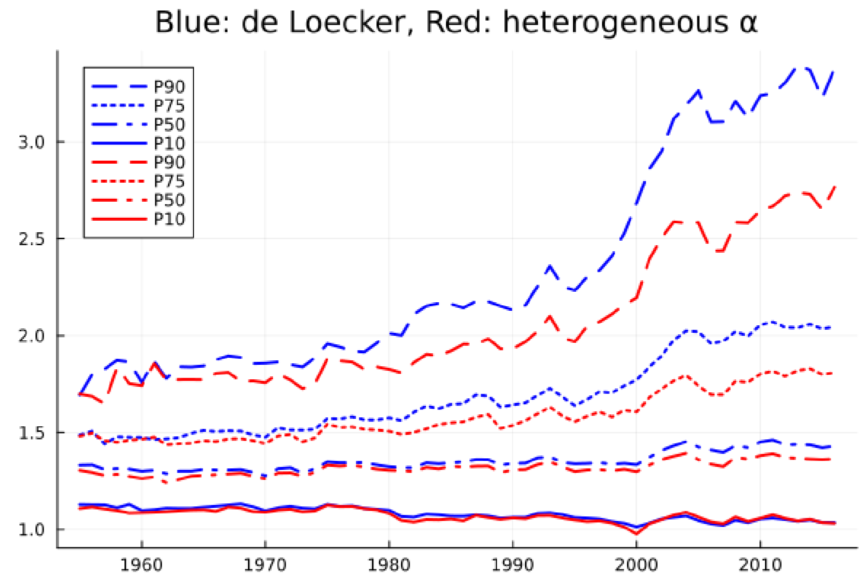


Figure 3.14: Moments of the distribution of firm-level markups: De Loecker, Eeckhout, Unger (2020) vs. heterogeneous α

Similarly to de Loecker et al. (2020), I find that the distribution of firms' markups fans out considerably with a net acceleration around 2000. However, levels are significantly lower and the spread between blue and red measures widens with time. Furthermore, as expected from section 2, firms with high markups are firms with low (and decreasing) elasticity of output to labor. As such, the difference between the red and blue series is the highest for top deciles and inexistent for low deciles. This has a large impacts on the sales-weighted average markups series depicted in Figure 3.15. Because firms with highest markups are those with highest market shares, the bias correction impacts proportionately the sales-weighted average markup. While such a measure increased by 23% in de Loecker et al. (2020), it increases by only 9% with my specification.

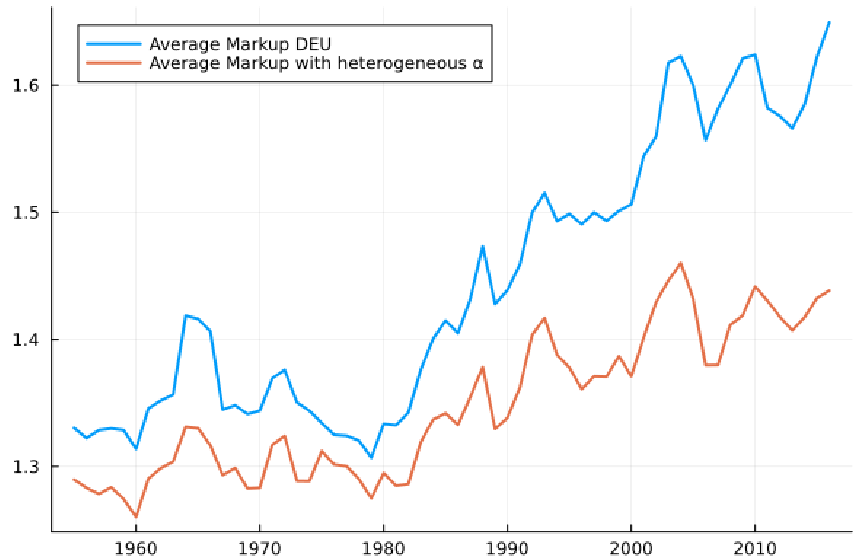


Figure 3.15: Sales-weighted average markup

3.3 Aggregate markup

In Figure 3.16, I plot the aggregate markup computed according to equation (3) under the specification of de Loecker et al. (2020) (in blue) and under mine (in red). The difference in level appears neatly as the red series is consistently below the blue one. Trends are also less pronounced. Between the 1960s and the 1980s, the aggregate markup estimated with homogeneous output elasticity declines while it stays stable when ζ is firm-specific. Similarly, the rise in the 1980s is much less pronounced in the red series than in the blue series. Overall, the aggregate markup under the specification of de Loecker et al. (2020) rises by 3.85% between 1950 and 2016 and by 5.47% between 1980 and 2016. Instead, when output elasticity are firm-specific, the aggregate markup rises by 4% both between 1950 and 2016.

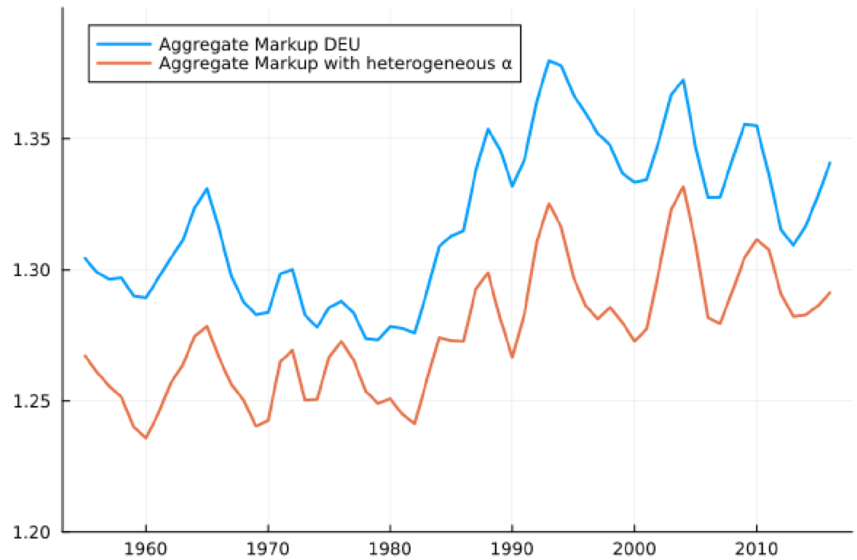


Figure 3.16: Aggregate markup

3.4 Implications for theoretical aggregation

The literature has derived several identities to aggregate firms' markups, on top of the one in equation (3). One of these alternatives stipulates that the aggregate markup is equal to the sum of firms' markups weighted by their input share (capital or labor indifferently). Intuitively, this latter aggregation identity doesn't hold because when ζ is heterogeneous across firms, input shares aren't the same depending on the input in question. In the data, this is crucially the case; aggregation through (3) or using input weights gives very different series as shown in Figure 3.17.

Instead, in **?**, I derive two aggregation formulas besides equation (3). First, one can use:

$$\mathcal{M} = \int_{\mathcal{I}} \frac{\mathcal{L}_i}{\mathcal{L}} \frac{\zeta^{\mathcal{L}}}{\zeta_i^{\mathcal{L}}} \mu_i di \tag{6}$$

For $\mathcal{L} \in \{K, L\}$ or any other input and where $\zeta^{\mathcal{L}}$ denotes the output elasticity to input $\mathcal{L}$. Therefore, the aggregate markup is the weighted arithmetic average of firms' markups where weights are the product of a firm's input share to total input and the inverse of the ratio of the aggregate output elasticity to that input and the one of the firm. Note that $\zeta_i^{\mathcal{L}} = \frac{w^{\mathcal{L}} \mathcal{L}_i}{TC_i}$, that is, as we saw, if the production technology displays constant returns to scale, the output elasticity to an input is equal to the cost share of this input to total costs. This allows to rewrite the markup aggregation identity:

$$\mathcal{M} = \int_{\mathcal{I}} \sum_{i=0}^{N} \frac{TC_i}{TC} \mu_i di \tag{7}$$

The aggregate markup can also be interpreted as the total-cost-weighted arithmetic average of firms' markups. Note that these *formulae* are a generalization of the previous one put forward by the literature.

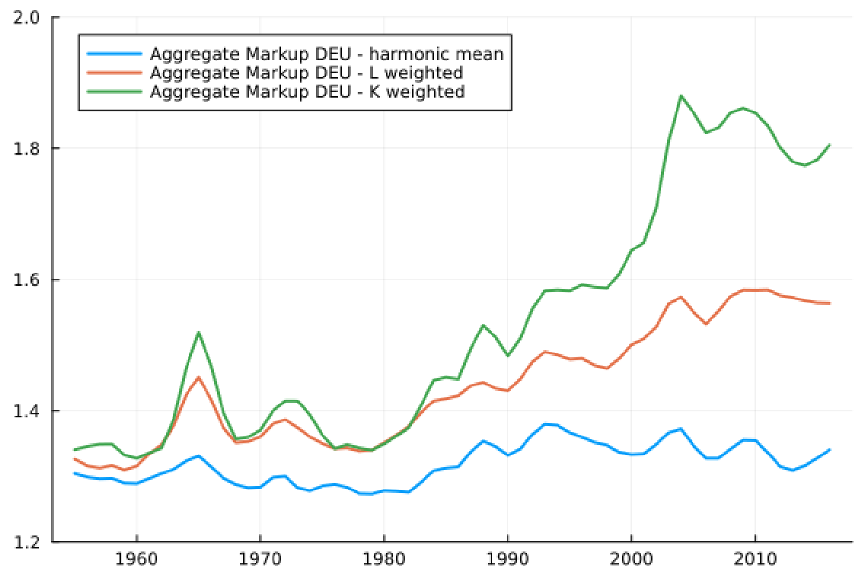


Figure 3.17: The three markup aggregation identity if α are homogeneous

Indeed, supposing that firms share the same ζ implies that equation (6) and (7) collapse to $\mathcal{M} = \int_{\mathcal{I}} \frac{\mathcal{L}_i}{\mathcal{L}} \mu_i di$.

4 Firms' TFPR

The estimation of firms' total factor productivity of revenue (TFPR) also relies on firms' output elasticities to inputs when TFPR is estimated as a Solow residual. To see this, suppose that a firm wields a Constant Elasticity of Substitution (CES) production technology that is homogeneous of degree one an homothetic:

$$Y_{i,t} = \left(\alpha_{i,t} K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,t}) L_{i,t}^{\frac{\psi-1}{\psi}}\right)^{\frac{\psi}{\psi-1}} \tag{8}$$

Where α_i is the share of capital in production, $1-\alpha_i$ that of labor and ψ is the elasticity of substitution between labor and capital. Under such a specification, the output elasticity to labor writes: $\zeta_i = (1-\alpha_i)\left(Y_{i,t}/L_{i,t}\right)^{1/\psi}$. Here, I am choosing to microfound the heterogeneity in ζ_i with α_i rather than ψ but the two imply the same impact on TFPR (see Appendix A.1). The Solow residual of the revenue production function takes the form:

$$\ln(p_{i,t} Z_{i,t}) = \ln(p_{i,t} Y_{i,t}) - \frac{\psi}{\psi-1} \ln\left(\alpha_{i,t} K_{i,t}^{\frac{\psi-1}{\psi}} + (1-\alpha_{i,t}) L_{i,t}^{\frac{\psi-1}{\psi}}\right) \tag{9}$$

Note that here, I focus on TFPR rather than TFPQ (Z_i) because (i) I do not observe firms' prices (nor quality differences), and (ii) I am supposing that the production technology is homothetic and homogeneous of degree one. If this isn't the case, non-homotheticity and returns to scale can be captured in the term Z_i. Nevertheless, this specification allows to compare how heterogenneous output elasticities to inputs modify the estimation of TFPR compares to a case where it is supposed homogeneous.

To see this, I perform two estimations and compare the key moments of the resulting distributions. The first estimation procedure is equation (9) with the firm-year alpha estimated in section A.3 following equation (10). The second one is the counterfactual and is hence also equation (9) at the firm-year level but with homogeneous α as an input. In that second case, I set α to the unweighted arithmetic average of the firm-level alphas used in the first estimation: 0.12. For both equations, ψ is set to 0.5. Moments of the distribution of pZ in 1980 and 2019 under homogeneous structural capital intensity (Hom. α) and heterogeneous structural capital intensity (Het. α) are reported in Table 4.12.

Naturally, since we allow for an extra degree of heterogeneity, all moments appear to be depressed; in 1980, average firm-level Solow residual is 33 % lower with heterogeneous α than without. In 2019, that number climbs to 44 %. The increasing importance of heterogeneous structural capital intensity seems to be concentrated at the tails however, since the percentage difference declines slightly for the median (31 % to 28%). More importantly, the variance of firms' Solow residual, the key moment to generate reallocative effects in heterogeneous firm models, is drastically reduced. It is halved at the beginning of the sample and collapses to 17 % of its homogeneous alpha value when heterogeneous alphas are introduced at the end of the sample. In the baseline quantitative model, firms will have no TFP and rather α will be a control variable of the firm, allowing us to assess how far variation in alpha can go in explaining various trends. In an extension exercise, I'll compare the moments produced by a model with heterogeneity in α to one with heterogeneity in TFP to one with both dimensions of heterogeneity. For completeness, the Kernel densities estimation of the distribution of the Solow residuals under the two specifications are reported in Figure 4.18.

Year	Specification	Mean	Median	Variance
1980	Hom. α	2.43	1.93	7.34
	Het. α	1.62 (-33%)	1.33 (-31%)	3.65 (-50%)
2019	Hom. α	3.78	2.01	123.42
	Het. α	2.11 (-44%)	1.43 (-28%)	20.81 (-83%)

Table 4.12: Moments of the distribution of the Solow residual per specification.

Note: Source is Compustat. Sample is 1980-2019. Solow residuals are estimated according to equation (9). pY is "Sales/Turnover (Net)". K is the sum of "Property, Plant and Equipment (Gross)" and "Intangible Assets" and L is "Employees". Het. alpha stands for heterogeneous α that is the firm-year estimation obtained in equation (10) while Hom. alpha stands for homogeneous α and is set to 0.12. ψ is set to 0.5. The parenthesis below the value in lines Het. alpha is the percentage difference between the Hom. alpha moment and its Het. alpha corresponding value.

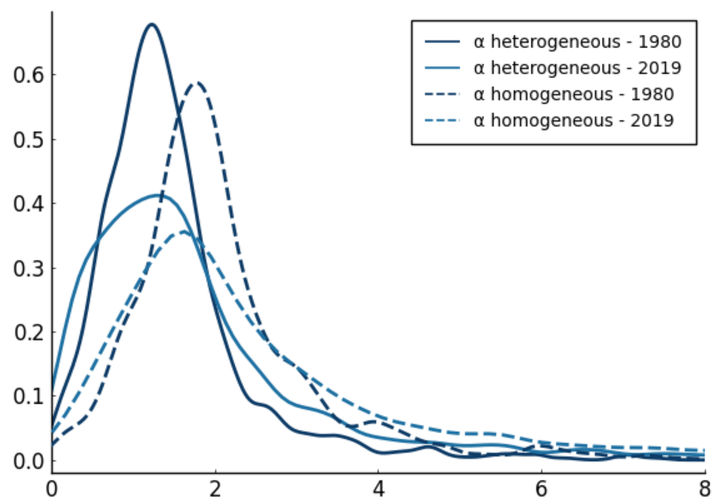


Figure 4.18: Solow residual: homogeneous vs. heterogeneous α.

Note: Source is Compustat. Sample is 1980-2019. Solow residual is estimated according to equation (9). Heterogeneous corresponds to firm-year-specific α. Homogeneous corresponds to α set to 0.12 for all firms and all years.

5 Conclusion

In this paper, I show that omitting the firm specificity of output elasticity to inputs results in biased estimates for markups when using the production function approach. In particular, because bigger firms, who charge higher markups, have lower output elasticity to labor, this biases upward both markup dispersion and the aggregate markup. Still, the bias is not big enough to overcome the stylized fact of de Loecker et al. (2020) and de Ridder et al. (2022) that firms' markups have grown more dispersed and that the aggregate markup has increased over the past decades in the US.

A key hypothesis in this paper is that firms' technologies are homogeneous of degree 1. If that were not the case, a bias in firms' markups would still remain. Estimating firm-specific returns to scale is hence the next step that will enrich this project.

Appendix

A Empirics

A.1 Data

Compustat Firm-level variables are obtained from Compustat, which includes balance sheet and income statement data for all publicly listed firms in the U.S. The sample period is 1980 to 2019. My empirical variables are constructed as follows:

Sales (pY) measured as *"Sales/Turnover - Total"* in Compustat.

Capital (K) is the total capital stock. It is measured by summing the stocks of tangible capital (*"Property, Plant and Equipment - Total (Gross)"*) and intangible capital (*"Intangible Assets - Total"*).

Labor (L) is measured with *"Employees"* . In robustness tests *"Cost of Goods Sold"* (COGS) and "Selling, General and Administrative expenses" (SGA) are used as alternatives.

A.2 Distribution of output elasticity to capital

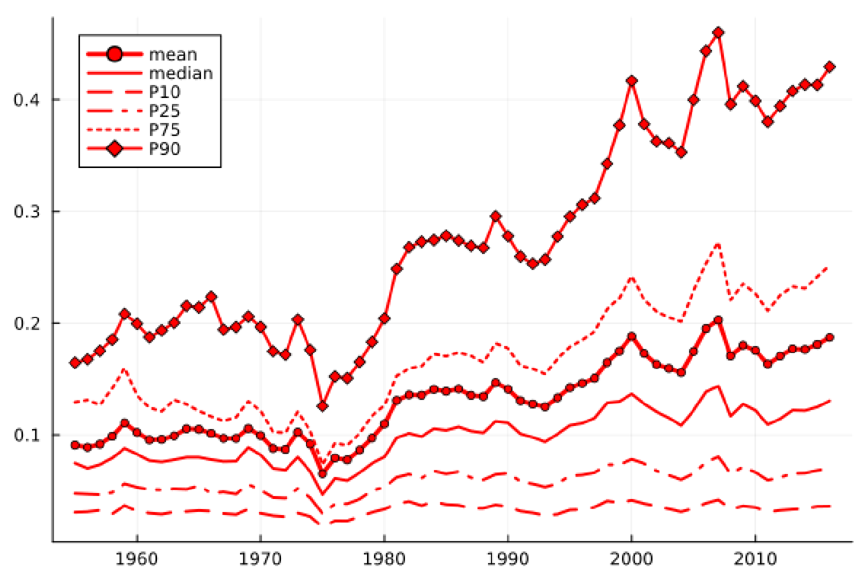


Figure A.19: Moments of the distribution of firm-level output elasticity to capital

A.3 Distribution of α

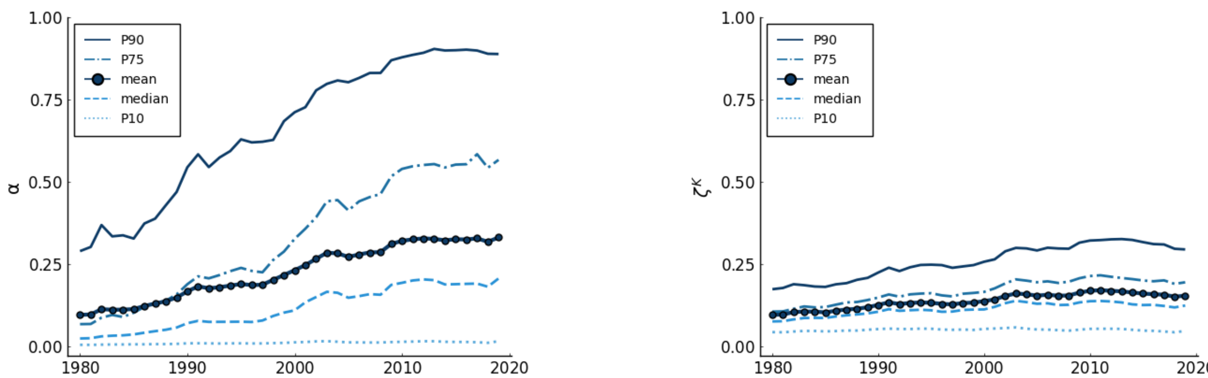
Suppose that firms produce using the CES production function (8), the identity for capital intensity implied by cost minimization writes:

$$\frac{K_{i,t}}{L_{i,t}} = \left(\frac{\alpha_{i,t}}{1-\alpha_{i,t}}\frac{w_t}{r_t^K}\right)^{\psi}$$

Isolating for alpha, this gives:

$$\alpha_{i,t} = \frac{r_t^K K_{i,t}^{\frac{1}{\psi}}}{w_t L_{i,t}^{\frac{1}{\psi}} + r_t^K K_{i,t}^{\frac{1}{\psi}}} \tag{10}$$

Which in the Cobb-Douglas limit converges to the cost share of capital $\zeta_{i,t}^K$. In Figure A.20, I display some percentiles of the distribution of firms' α across years obtained using equation (10). In panel (a), I plot my baseline where ψ is set to a usual value in the literature: 0.5. In panel (b) I plot a case where capital and labor are substitutes with $\psi = 1.5$ (for an estimate of ψ for the aggregate economy, see Antras (2004). For an estimation at the firm-level for manufacturing and a proposed aggregation method, see Oberfield and Raval (2021). For a more complete review on studies that estimate ψ, see Rognlie (2015)).

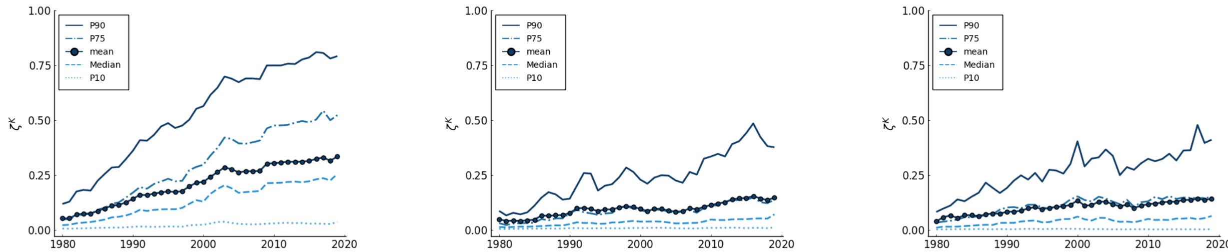


(1) Complements ($\psi = 0.5$) (2) Substitutes ($\psi = 1.5$)

Figure A.20: Moments of the distribution of alpha conditional on ψ

Note: Source is Compustat. Sample is 1980-2019. Moments from the distribution implied by the estimation using equation (10) on Compustat firms. Measure of Capital is (PPE + INT). PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". Measure of Labor is EMP: number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 is the 90th percentile.

In the baseline case where $\psi = 0.5$, the distribution of firms' alpha is non-stationary; both the mean and the variance increase with time. The magnitude of these increase is higher than that of the output elasticity to capital depicted in A.19 because heterogeneity in capital intensity is augmented by the geometric weight $\frac{1}{\psi} = 2$. Conversely, in the substitutes case, heterogeneity in capital-intensity is dampened and the average and variance of α increase less. In Figure A.21, I plot some percentiles of the distribution of firms' α across years obtained using equation (10) for manufacturing, trade and services when $\psi = 0.5$.



(1) Manufacturing (2) Trade (3) Services

Figure A.21: Percentiles of the Distribution of firms' α for 3 biggest 2D sectors

Note: Source is Compustat. Sample is 1980-2019. α is measured using equation (10) for $\psi = 0.5$. Measure of Capital is (PPE + INT). PPE stands for "Property Plant and Equipment", EMP for "Employees" and INT for "Intangible Assets - Total". Measure of Labor is EMP: number of Employees, PPE and INT are expressed in 100'000s of 2000 dollars. P90 relates to the top decile of the distribution of capital intensity. Manufacturing gathers 2-digit NAICS sectors 31, 32 and 33. Trade is Wholesale (code 42) and Retail (code 44-45). Services gathers Professional, Scientific and Technical services (code 54), Management of Companies and Enterprises (code 55), Educational Services (code 61), Health Care and Social Assistance (code 62) and Arts, Entertainment and Recreation (code 71).

References

Abel, A. B., N. G. Mankiw, L. H. Summers, and R. J. Zeckhauser (1989). Assessing dynamic efficiency: Theory and evidence. *The Review of Economic Studies*.

Acemoglu, D. (2002). Directed technical change. *The review of economic studies*.

Acemoglu, D., G. W. Anderson, D. N. Beede, C. Buffington, E. E. Childress, E. Dinlersoz, L. S. Foster, N. Goldschlag, J. C. Haltiwanger, Z. Kroff, et al. (2022). Automation and the workforce: A firm-level view from the 2019 annual business survey.

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics 40*(S1), S293–S340.

Acemoglu, D. and V. Guerrieri (2008). Capital deepening and nonbalanced economic growth. *Journal of political Economy*.

Acemoglu, D. and P. Restrepo (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review*.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from us labor markets. *Journal of political economy*.

Acemoglu, D. and P. Restrepo (2022). Tasks, automation, and the rise in us wage inequality. *Econometrica*.

Acemoglu, D. and P. Restrepo (2023). Automation and rent dissipation: Implications for inequality, productivity, and welfare.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica*.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2023). A theory of falling growth and rising rents. *Review of Economic Studies*.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica*.

Aghion, P. and J. Tirole (1997). Formal and real authority in organizations. *Journal of political economy*.

Akcigit, U. and S. T. Ates (2023). What happened to us business dynamism? *Journal of Political Economy*.

Ali, A., S. Klasa, and E. Yeung (2008). The limitations of industry concentration measures constructed with compustat data: Implications for finance research. *The Review of Financial Studies*.

Andrei, D., W. Mann, and N. Moyen (2019). Why did the q theory of investment start working? *Journal of Financial Economics*.

Antras, P. (2004). Is the us aggregate production function cobb-douglas? new estimates of the elasticity of substitution. *The BE Journal of Macroeconomics*.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*.

Atkeson, A., J. Heathcote, and F. Perri (2022). The end of privilege: A reexamination of the net foreign asset position of the united states.

Atkin, D., A. Chaudhry, S. Chaudry, A. K. Khandelwal, and E. Verhoogen (2017). Organizational barriers to technology adoption: Evidence from soccer-ball producers in pakistan. *The Quarterly Journal of Economics*.

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

Azar, J., M. C. Schmalz, and I. Tecu (2018). Anticompetitive effects of common ownership. *The Journal of Finance*.

Baley, I. and L. L. Veldkamp (2025). *The Data Economy: Tools and Applications*. Princeton University Press.

Baqaee, D., E. Farhi, and K. Sangani (2021). The supply-side effect of monetary policy. *NBER Working Paper*.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*.

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance*.

Barkai, S. and S. G. Benzell (2024). 70 years of us corporate profits. *Journal of Corporate Finance*.

Barth, M. J. and V. A. Ramey (2001). The cost channel of monetary transmission. *NBER macroeconomics annual*.

Beaudry, P., C. Hou, and F. Portier (2022). Monetary policy when the phillips curve is quite flat. *American Economic Journal: Macroeconomics*.

Benmelech, E., N. K. Bergman, and H. Kim (2022). Strong employers and weak employees: How does employer concentration affect wages? *Journal of Human Resources 57*(S), S200–S250.

Bergeaud, A., C. Malgouyres, C. Mazet-Sonilhac, and S. Signorelli (2025). Technological change and domestic outsourcing. *Journal of Labor Economics*.

Berlingieri, G., F. Boeri, D. Lashkari, and J. Vogel (2024). Capital-skill complementarity in firms and in the aggregate economy. *NBER Working Paper*.

Blanchard, O., E. Cerutti, and L. Summers (2015). Inflation and activity–two explorations and their monetary policy implications.

Bloom, N., R. Sadun, and J. Van Reenen (2012). The organization of firms across countries. *The quarterly journal of economics*.

Boehm, C. E., A. Flaaen, and N. Pandalai-Nayar (2020). Multinationals, offshoring, and the decline of us manufacturing. *Journal of International Economics*.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics*.

Burstein, A., V. M. Carvalho, and B. Grassi (2020). Bottom-up markup fluctuations.

Campbell, J. R., J. D. Fisher, A. Justiniano, and L. Melosi (2017). Forward guidance and macroeconomic outcomes since the financial crisis. *NBER macroeconomics annual*.

Caselli, F. and J. Feyrer (2007). The marginal product of capital. *The quarterly journal of economics*.

Chari, V. V. and H. Hopenhayn (1991). Vintage human capital, growth, and the diffusion of new technology. *Journal of political Economy*.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*.

Cho, W. H., F. Manaresi, and R. Martin (2025). Superstars, shooting stars, and the labour share. *Documents de travail de l'OCDE sur la science, la technologie et l'industrie*.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*.

Cohen, W. M. and D. A. Levinthal (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*.

Cooley, T. F. and V. Quadrini (2001). Financial markets and firm dynamics. *American economic review*.

Corrado, C., C. Hulten, and D. Sichel (2009). Intangible capital and us economic growth. *Review of income and wealth*.

Crouzet, N. and J. Eberly (2021). Intangibles, markups, and the measurement of productivity growth. *Journal of Monetary Economics*.

Crouzet, N. and J. Eberly (2023). Rents and intangible capital: A q+ framework. *The Journal of Finance*.

Crouzet, N., J. Eberly, A. Eisfeldt, and D. Papanikolaou (2024). Intangible capital, firm scope, and growth.

Crouzet, N. and J. C. Eberly (2019). Understanding weak capital investment: The role of market concentration and intangibles.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022). A model of intangible capital.

Cummins, J. G. and G. L. Violante (2002). Investment-specific technical change in the united states (1947–2000): Measurement and macroeconomic consequences. *Review of Economic dynamics 5*(2), 243–284.

David, J. M., R. Ranciere, and D. Zeke (2024). International diversification, reallocation, and the labor share.

David, J. M., L. Schmid, and D. Zeke (2022). Risk-adjusted capital allocation and misallocation. *Journal of Financial Economics*.

David, J. M. and V. Venkateswaran (2019). The sources of capital misallocation. *American Economic Review*.

Davis, S. J., J. Haltiwanger, R. Jarmin, J. Miranda, C. Foote, and E. Nagypal (2006). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms [with comments and discussion]. *NBER macroeconomics annual*.

de Loecker, J., J. Eeckhout, and S. Mongey (2021). Quantifying market power and business dynamism in the macroeconomy. *NBER Working Paper*.

de Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*.

de Loecker, J. D. and F. Warzynski (2012). Markups and firm-level export status. *American economic review*.

de Ridder, M. (2024). Market power and innovation in the intangible economy. *American Economic Review*.

de Ridder, M., B. Grassi, and G. Morzenti (2022). The hitchhiker's guide to markup estimation.

Dechezleprêtre, A., D. Hémous, M. Olsen, and C. Zanella (2025). Induced automation innovation: evidence from firm-level patent data. *Journal of Political Economy*.

Decker, R. and J. Williams (2023). A note on industry concentration measurement.

Demirer, M. (2020). Production function estimation with factor-augmenting technology: An application to markups.

Desazars, G. (2025). Machines vs. markups.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American economic review*.

Edmond, C., V. Midrigan, and D. Y. Xu (2015). Competition, markups, and the gains from international trade. *American Economic Review*.

Edmond, C., V. Midrigan, and D. Y. Xu (2023). How costly are markups? *Journal of Political Economy*.

Eeckhout, J. (2022). *The profit paradox: How thriving firms threaten the future of work*. Princeton University Press.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024). Gpts are gpts: Labor market impact potential of llms. *Science*.

Elsby, M. W., B. Hobijn, and A. Şahin (2013). The decline of the us labor share. *Brookings Papers on Economic Activity*.

Fama, E. F. and K. R. French (2001). Disappearing dividends: changing firm characteristics or lower propensity to pay? *Journal of Financial economics*.

Fama, E. F. and K. R. French (2004). New lists: Fundamentals and survival rates. *Journal of financial Economics*.

Farhi, E. and F. Gourio (2018). Accounting for macro-finance trends: Market power, intangibles, and risk premia.

Favero, C. A., S. Krgović, A. Melone, and A. Tamoni (2021). Long-run trends in demographics, income inequality, and the natural rate of interest: Further evidence.

Fisher, J. D. (2006). The dynamic effects of neutral and investment-specific technology shocks. *Journal of political Economy*.

Fisher, J. D. (2015). On the structural interpretation of the smets–wouters risk premium shock. *Journal of Money, Credit and Banking*.

Foster, L. S., J. C. Haltiwanger, and C. Tuttle (2022). Rising markups or changing technology?

Gali, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

Garibaldi, P. and E. D. Turri (2024). Monopsony in growth theory.

Gomme, P., B. Ravikumar, and P. Rupert (2011). The return to capital and the business cycle. *Review of Economic Dynamics*.

Gordon, R. J. (1990). *The measurement of durable goods prices*. University of Chicago press.

Gourio, F. and A. K. Kashyap (2007). Investment spikes: New facts and a general equilibrium exploration. *Journal of Monetary Economics*.

Gourio, F. and T. H. Klier (2015). Recent trends in capital accumulation and implications for investment. *Chicago Fed Letter*.

Gourio, F. and M. Rognlie (2020). Capital heterogeneity and investment prices: How much are investment prices declining.

Gourio, F. and L. Rudanko (2014). Customer capital. *Review of Economic Studies*.

Grassi, B. (2017). Io in io: Size, industrial organization, and the input-output network make a firm structurally important. *Work. Pap., Bocconi Univ., Milan, Italy*.

Grazzini, J. and L. Rossi (2023). Firm ex ante heterogeneity, entry and the labour share. *The Economic Journal*.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review.*

Greenwood, J., Z. Hercowitz, and P. Krusell (1997). Long-run implications of investment-specific technological change. *The American economic review*, 342–362.

Greenwood, J., Z. Hercowitz, and P. Krusell (2000). The role of investment-specific technological change in the business cycle. *European Economic Review.*

Grullon, G. and R. Michaely (2002). Dividends, share repurchases, and the substitution hypothesis. *the Journal of Finance.*

Guntin, R. and F. Kochen (2025). The origins of top firms.

Gutiérrez, G. and T. Philippon (2016). Investment-less growth: An empirical investigation. *NBER technical report..*

Hall, R. E. (1988). The relation between price and marginal cost in us industry. *Journal of political Economy.*

Hall, R. E. and D. W. Jorgenson (1967). Tax policy and investment behavior. *The American economic review.*

Hartley, J., K. Hassett, and J. D. Rauh (2025). Firm investment and the user cost of capital: New us corporate tax reform evidence. *NBER Working Paper*.

Hartman-Glaser, B., H. Lustig, and M. Z. Xiaolan (2019). Capital share dynamics when firms insure workers. *The Journal of Finance.*

Hazell, J., J. Herreno, E. Nakamura, and J. Steinsson (2020). The slope of the phillips curve: evidence from us states.

Hémous, D. and M. Olsen (2022). The rise of the machines: Automation, horizontal innovation, and income inequality. *American Economic Journal: Macroeconomics.*

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica.*

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics.*

Hubmer, J. (2023). The race between preferences and technology. *Econometrica.*

Hubmer, J., M. Chan, S. Ozkan, S. Salgado, and G. Hong (2024). Scalable versus productive technologies.

Hubmer, J. and P. Restrepo (2025). Not a typical firm: The joint dynamics of firms, labor shares, and capital–labor substitution.

Jones, B. F. and X. Liu (2024). A framework for economic growth with capital-embodied technical change. *American Economic Review*.

Jones, C. I. (2005). The shape of production functions and the direction of technical change. *The Quarterly Journal of Economics*.

Kahle, K. and R. M. Stulz (2020). Why are corporate payouts so high in the 2000s? *National Bureau of Economic Research*.

Karabarbounis, L. (2024). Perspectives on the labor share. *Journal of Economic Perspectives*.

Karabarbounis, L. and B. Neiman (2014). The global decline of the labor share. *The Quarterly journal of economics*.

Karabarbounis, L. and B. Neiman (2019). Accounting for factorless income. *NBER Macroeconomics Annual*.

Kaymak, B. and I. Schott (2023). Corporate tax cuts and the decline in the manufacturing labor share. *Econometrica*.

Kehrig, M. and N. Vincent (2021). The micro-level anatomy of the labor share decline. *The Quarterly Journal of Economics*.

Keil, J. (2017). The trouble with approximating industry concentration from compustat. *Journal of Corporate Finance*.

Keynes, J. M. (1930). Economic possibilities for our grandchildren. In *Essays in persuasion*, pp. 321–332. Springer.

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking*.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of political economy*.

Klenow, P. J. and J. L. Willis (2016). Real rigidities and nominal price changes. *Economica*.

Klette, T. J. and Z. Griliches (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of applied econometrics*.

Koh, D., R. Santaeulàlia-Llopis, and Y. Zheng (2020). Labor share decline and intellectual property products capital. *Econometrica*.

Kwon, S. Y., Y. Ma, and K. Zimmermann (2024). 100 years of rising corporate concentration. *American Economic Review*.

Lashkari, D., A. Bauer, and J. Boussard (2024). Information technology and returns to scale. *American Economic Review*.

Laubach, T. and J. C. Williams (2003). Measuring the natural rate of interest. *Review of Economics and Statistics*.

Liu, E., A. Mian, and A. Sufi (2022). Low interest rates, market power, and productivity growth. *Econometrica*.

Lucas, R. E. (1990). Why doesn't capital flow from rich to poor countries? *The American Economic Review*.

Lucas, R. E. and B. Moll (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*.

Ma, Y., M. Zhang, and K. Zimmermann (2024). Business concentration around the world: 19002020. *Working Paper*.

Martinez, J. (2021). *Putty-clay automation.* Centre for Economic Policy Research.

Marx, K. (1867). *Das kapital.* Verlag von Otto Meisner.

McGrattan, E. R. and E. C. Prescott (2010). Technology capital and the us current account. *American Economic Review*.

Meier, M. and T. Reinelt (2020). Monetary policy, markup dispersion, and aggregate tfp. *ECB Working Paper*.

Mian, A. R., L. Straub, and A. Sufi (2021). What explains the decline in r*? rising income inequality versus demographic shifts. *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

Mongey, S. (2022). Market structure and monetary non-neutrality. *Econometrica forthcoming*.

Oberfield, E. and D. Raval (2021). Micro data and macro technology. *Econometrica*.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*.

Ottonello, P. and T. Winberry (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*.

Peretto, P. F. and J. J. Seater (2013). Factor-eliminating technical change. *Journal of Monetary Economics*.

Peters, M. (2020). Heterogeneous markups, growth, and endogenous misallocation. *Econometrica*.

Peters, M. and C. Walsh (2022). Population growth and firm-product dynamics. *NBER Working Paper 29424*.

Philippon, T. (2019). *The great reversal: How America gave up on free markets*. Harvard University Press.

Primiceri, G. E. (2006). Why inflation rose and fell: policy-makers' beliefs and us postwar stabilization policy. *The Quarterly Journal of Economics*.

Ravenna, F. and C. E. Walsh (2006). Optimal monetary policy with the cost channel. *Journal of Monetary Economics*.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*.

Rognlie, M. (2015). Deciphering the fall and rise in the net capital share: accumulation or scarcity? *Brookings papers on economic activity*.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*.

Rubbo, E. (2020). Networks, phillips curves and monetary policy. *Econometrica forthcoming*.

Schumpeter, J. A. (1911). *The theory of economic development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*.

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of political Economy*.

Wang, O. and I. Werning (2022). Dynamic oligopoly and price stickiness. *American Economic Review*.

Woodford, M. (2003). *Interest and prices*. Princeton University Press.

Yeh, C., C. Macaluso, and B. Hershbein (2022). Monopsony in the us labor market. *American Economic Review*.

Zeira, J. (1998). Workers, machines, and economic growth. *The Quarterly Journal of Economics*.