Université de Montréal

Essay on firm dynamics, competition and International trade

Par
SIWE Guy Leonel

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Juin, 2021

Université de Montréal
Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée :
Essay on firm dynamics, Competition and International trade

présentée par :
SIWE Guy Leonel

Devant le jury composé des personnes suivantes :

President du jury:	Guillaume Sublet,
Directeur de recherche:	Immo Schott,
Co-Directeur de recherche:	Baris Kaymak,
Membre du jury:	Mathieu Marcoux,
Examinateur externe:	Matthias Meier, Univ. de Mannheim

A mes parents, monsieur WETCHEBEWE Andre et madame DJAMBOU Elise

Remerciements

J'ai une profonde gratitude envers mon directeur de recherche Immo Schott et mon co-directeur Baris Kaymak, pour sa présence continue, sa patience, ses multiples conseils en tout genre, son implication et sa contribution, pour m'avoir encouragé et d'avoir cru en moi tout au long de cette thèse. Merci infiniment pour tout.

Je remercie infiniment Mathilde Lebrand, pour son soutien, son aide et pour sa disponibilité surtout durant les deux dernières années de ma thèse. Je remercie également Mathieu Marcroux et Guillaume Sublet pour leur disponibilité, leur aide et leur conseil tout au long de cette thèse sans oublier l'ensemble des professeurs, chercheurs ainsi que le personnel administratif du département de sciences économiques de l'Université de Montréal et du CIREQ.

J'exprime également ma reconnaissance à mes collègues de promotion qui ont été d'une aide précieuse tout au long de cette thèse: Marlène Koffi, Fatim Diabagaté, Magnim Farouh, Ngolo Koné, Lucienne Talba, Idriss Tsafack et Samuel Gingras. Je remercie particulièrement mes amis Ghislain Afavi et Felicien Goudou pour leur soutien et leur disponibilité continue. Je ne saurai oublier tous les étudiants de doctorat du département de sciences économiques de l'Université de Montréal dont les discussions enrichissantes m'ont permis d'approfondir mes connaissances et d'améliorer le contenu de cette thèse.

Mes pensées et ma reconnaissance vont à toute ma famille et particulierement à mes parents Wetchebewe Andre et ma mère Djambou Elise pour le soutien moral et encouragements. C'est grace à vous que j'ai parcouru tout ce chemin. À tout mes amis et toutes personnes non-cités recevez mes sincères remerciements pour votre soutien.

Contents

List of Tables

List of Figures

Introduction

Le pouvoir de marché des entreprises a toujours été un sujet d'intérêt tant du point de vue du bien-être économique que de l'allocation des ressources. D'une part, le pouvoir de marché incite les entreprises à moins produire - réduisant le bien-être des consommateurs et d'autre part, la variation de la marge bénéficiaire entre les entreprises induit une friction sur l'allocation des intrants et diminue la productivité agrégée. Suite aux récents débats sur la concentration des parts de marché dans l'économie américaine par une minorité d'entreprises, la question du pouvoir de marché et ses implications ont été largement débattues. Cette thèse, divisée en trois chapitres, contribue à cette vaste littérature d'étude du pouvoir de marché de l'entreprise.

Le premier chapitre examine les causes de la hausse des marges bénéficiaires aux États-Unis dans un environnement où les entreprises accumulent des actifs intangibles. En effet, au cours des quatre dernières décennies, l'économie américaine a connu une augmentation en moyenne des marges bénéficiaires par les entreprises et une augmentation de la concentration des parts de marché. Ces faits ont suscité des inquiétudes quant au pouvoir de marché détenu par certaines entreprises et à l'étendue de la concurrence. Dans ce chapitre, nous proposons une nouvelle méthode d'agrégation des marges bénéficiaires des entreprises utilisant la distribution des parts de marché. Ensuite, nous documentons trois faits empiriques basés sur la distribution jointe des marges et des parts de marché des entreprises cotées en bourse. Premièrement, la hausse de la marge agrégée a été bien inferieure à celle suggerée par la moyenne. Deuxièmement, les plus grandes entreprises en termes de part de marché, appelées "superstars", ne possèdent pas les marges bénéficiaires les plus élevées et troisièmement, la hausse de la marge agrégée résulte principalement d'une réallocation des parts de marché, associée conjointement à une variation des marges des entreprises. Pour expliquer ces faits, nous proposons un modèle où les entreprises accumulent des actifs intangibles sous forme de base de clients le long de leur cycle de vie, en investissant une valeur fixe de leurs ventes. Une baisse dans la rentabilité des entreprises entrainent une réduction temporaire des marges, afin que ces dernières fixent des prix bas et attirent de nouveaux clients. Une hausse des investissements en base de clientèle profite aux entreprises ayant de fortes opportunités de croissance, leur permettant d'accroitre les parts de vente et d'augmenter davantage leur marge bénéficiaire. Ainsi, l'augmentation des investissements en base de clientèle au cours des quatre dernières décennies explique de moitié l'augmentation globale des marges bénéficiaires aux États-Unis.

Le deuxième chapitre examine les implications d'une variation endogène des marges bénéficiaires sur la productivité des usines le long de leur cycle de vie. En effet, la variation des marges bénéficiaires est une potentiel source d'inefficience dans l'allocation des ressources, réduisant la productivité globale. En utilisant l'enquête annuelle sur le secteur manufacturé en Colombie, nous montrons une diminution de la dispersion des marges bénéficiaires avec l'âge, suggérant une convergence des marges bénéficiaires des usines avec l'âge. Nous expliquons et testons cette dynamique à travers d'une théorie d'apprentissage de la demande le long du cycle de vie (à l'exemple de la formation d'une base de la clientèle). Les nouvelles usines obtiennent une productivité et une de demande dès leur entrée et apprenent de leur demande en ajustant leur marge bénéficiaire. Au cours de leur vie, elles accumulent leur demande, ce qui entraine leur croissance et une hausse de leur marge bénéficiaire. La convergence résulte d'une inadéquation entre l'échelle optimale qu'une usine souhaite opérer compte tenu de sa productivité et son échelle de production actuel. Le modèle prédit que la grande dispersion des marges bénéficiaires à l'entrée du marché, entraine une inefficience dans l'allocation des ressources productives et reduit la productivité moyenne à des nouvelles usines. Au fur et à mesure que l'usine prend de l'age, cette friction à l'allocation diminue et contribue à augmenté la productivité moyenne des usines agées.

Dans le troisième chapitre, j'étudie les gains de bien-être associés à des nouveaux investissements en infrastructures de transport. Il s'agit d'un travail de recherche conjoint avec Mathilde Lebrand de la Banque mondiale. En effet, de nombreux travaux ne prennent pas en compte la demande de services de transport lors de l'étude des gains associés à de nouveaux projets d'investissements en infrastructures. Nous développons ainsi un modèle Ricardien multi-pays, multisectoriel avec un secteur des transports dont l'équilibre est déterminer de manière endogène et qui présente deux caractéristiques principales : une diversité de modes de transport et l'absence de concurrence dans le secteur de transport. Les predictions nous montre que une hausse des investissements en infrastructures réduisent les coûts d'expédition selon trois marges : une réduction du coût marginal des transporteurs, une diminution de la marge de profit des transporteurs et une reallocation des parts de marché vers les modes de transport les plus bénefiques. Ces deux dernières marges sont particulièrement affectées par les préférences des importateurs en matière de choix de mode de transport.

Mots clés: Marges bénéficiaires, actifs intangibles, base de clients, agrégation, inefficience allocative, croissance, productivité, coûts de transport, échanges internationaux, investissements en infrastructures.

Introduction

Firms' market power has always been a topic of interest in terms of welfare and resource allocations analysis. On the one hand, market power reduces the firms' incentive to produce, reducing consumer welfare. On the other hand, the variation in firms' markup induces friction to input allocation and lowers aggregate productivity. With the ongoing debate on sales concentration in the US economy by a few firms, there is a lot of question concerning the causes and implications of market power. This thesis, divided into three chapters, contributes to this vast literature of studying the firm's market power.

The first chapter discusses the causes of US markup rise in an environment where firms are building intangibles assets. Indeed, over the last four decades, both the average price markups charged by firms and market shares concentration have increased. This raised concerns about market power and the extent of competition. I develop a new method to aggregate firm-level markups using the sales shares distribution. I then document three facts based on the joint distribution of markups and sales shares among publicly listed firms. First, the rise in the aggregate markup has been much less than that suggested by average markups. Second, the largest sales firms-called "superstar firms" are *not* at the top of the markups distribution and third, the rise in US markup has been driven by a reallocation of sales shares, jointly associated with a change in firm-level markup. To explain these trends, I develop a model where firms accumulate customers base over their life cycle by investing a fixed share of their sales. Changes in firm profitability lead to a temporary reduction in markups as firms attract new customers. An exogenous increase in the share of investment into the customer base benefits firms with high growth opportunities, allowing them to gain more sales share and increase their markup. Therefore, I find that the rise in the share of investment into customer base over the last four decades explains half of the overall increase in US markup.

The second chapter discusses the implications of an endogenous variation in markups to plant productivity over the life cycle. Indeed, markup varies across plants, inducing an inefficiency in allocating resources that lower the aggregate productivity. Using the Annual Columbia Manufacturing survey, I find a decrease in markup dispersion with age, suggesting a convergence of all plant-level markup with age in cross-sectional. I explain and test this dynamic through active learning on demand by plants over the life cycle (e.g. customer base formation). New plants draw a productivity shock and demand shock at entry and actively learn about their demand by adjusting their markup. Over their life, they build their demand and growth in both size and markup. This convergence results from a mismatch between the optimal scale a plant desire to operate and its current scale for a given level of productivity. The model predicts that the high variation in markups at entry induces inefficient input allocation, lowering the average productivity. Along with age, the declines in markups dispersion reduce the allocative inefficiency and foster productivity growth over the life cycle.

In the third chapter, I study the welfare gain from new infrastructures investments. It is joint work with Mathilde Lebrand from the World Bank. Indeed, many frameworks don't take into account the demand for transport services when studying trade policies. We develop a multi-countries and multi-sectoral Ricardian model with an endogenous transport sector. We introduce two main features for endogeneity: the multi-modality and a lack of competition in the transport sector. New infrastructure investments reduce the shipping cost along three margins: a transporters marginal cost reduction, a decline in the transporters profit margins and a reallocation of market shares across shipping modes. The two last margins are driven by the shipping mode preference for a good.

Keywords: Markup, intangible assets, aggregation, allocative inefficiency, growth, productivity, life cycle, transport cost, international trade, infrastructure investments.

Chapter 1

Increasing Markups and the Concentration of Market Shares

1.1 Introduction

Recent macroeconomic trends exhibit an increase in market shares concentration, an increase in the average markup, and an increase in the US economy's profit share. Some authors have interpreted these trends to stem from a decline in competition, driven by high-markup firms (De Loecker et al., 2020). Others argue that given new information technologies, market shares have been reallocated towards highly efficient firms (Autor et al., 2017), which may also charge high-markup (Baqaee and Farhi, 2018).[0] In addition, recent evidence documents an increase in intangible assets accumulation by firms (Crouzet and Eberly, 2018). In this paper, I analyze the causes that drive the rise of markup in an environment where firms are building intangible assets.

The key departure in my approach is to start with the joint distribution of sales shares and markups and then compute a measure of markup for the economy. By considering an aggregation of firm's outputs within an industry, I extend the firm's markup estimation framework proposed by De Loecker et al. (2020) to the industry level. Aggregate markup is defined as the degree of monopoly power held by a representative firm within an industry. It is measured as the sales-weighted *harmonic* mean of firm-level markups - equivalent to the weighted cost average. This approach is based on the fact that firms optimally choose input such that their market shares are proportional to the markups weighted by the input cost. The representative firm will make a similar decision but holds all market shares. This aggregation is not unique to specific industries because it is not based on demand structure and price strategy. Using firm-level data from Compustat, I found an increase in aggregate markup of 13% points above the marginal cost over the last four decades, consistent with the estimates based on macroeconomic variables.

[0] Other references include Gutierrez et al. (2019), Gutierrez and Philippon (2019), De Loecker et al. (2020), Syverson (2019), Basu (2019), Eggertsson et al. (2018).

Because markup at the aggregate level is fully characterized by the joint distribution of markups and sales shares, I documented cross-sectionally and temporally features of that joint distribution. *(i)* Both the rise in the markup and sales concentration happened in two different periods, suggesting they are driving by two distinct shocks. *(ii)* The rise in the markup at the aggregate level is driven at 70% by a reallocation of sales shares from low to high markup firms. However, this sales shares reallocation has been jointly associated with a change in firm-level markup.[1] *(iii)* At the cross-sectional level, the largest sales firms so-called "superstar firms" are *not* the highest markup firms, justifying the *low* and positive correlation between markups and sales shares found in the literature. In addition, there is no significant difference in markup growth between superstar firms and non-superstar firms.

Motivated by these observations, I built a model to study the causes of the rise of markups in an environment where firms built intangible assets. An environment of intangible assets accumulation by firms is a feature observed in the US economy over the last four decades (Crouzet and Eberly, 2018). The model is in the spirit of Hopenhayn (1992) where firms build intangibles assets in the form of a customer base as in Ravn et al. (2004). In the model, firms are both heterogeneous in productivity and the number of customers they serve. They spend a fixed share of their sales to invest in customer base accumulation (for example, marketing spending to acquire a customer base). Invest in customer base scale the future demand and output, increasing the future sale and profit. However, it requires lower prices today to attract new customers. This lowers the current markup as well as profits.

Both markups and sales increase as long as firms build and lock-in their customer base. The most productive firms take advantage of their low marginal cost to produce more and lower their markup to invest in a customer base. In this regard, they charge a low markup relative to less efficient firms. Therefore, the largest firms that are the most productive with a large customer base are not the highest markup firms. Thus, sales shares increase with both customer base and productivity but the negative correlation between markup and firm productivity generates a non-perfect correlation between markups and sales shares.

Therefore, I calibrated the model to the US economy in 1980 using Compustat data and performed some counterfactual experiments. Those experiments analyze the potential causes of the rise of US markup in an economy of intangibles assets accumulation by firms. I focus on three main causes discussed in the literature: a change in the market structure (through a fall in the exit rate), an increase in productivity gap between leader and followers firms, and an increase in the share of customer base investment.

I begin by simulating an increase in the share of investment in the customer base. Such an increase raises both the cost and the benefit of investing in the customer base.

[1] In others words, the shift in sales shares results from a change in firm-level markup. De Loecker et al. (2020) shows that the sales shares reallocation is independent with the change in firm-level markup. The difference between those results is aggregation.

However, that change benefits more firms with more growth opportunities (low customer base). They increase their investment in customer base by reducing their markup which allows them to grow faster in size and markup. The rapid growth in markup and size increases the aggregate markup at the stationary equilibrium without driving the sales share concentration. In addition, the rise in the share of investment in intangible assets contributes to the rise of the share of intangible assets in the economy and lowers consumer welfare as the firms have increased their markup.

Second, I simulate a change in a market structure through a fall in the exit rate. A decline in the exit rate increases the survival probability and thus growth opportunities for all firms. Therefore, all firms lower their markup to benefit from that change, especially firms with more growth opportunities (i.e. a low customer base). All firms grow faster in customer base as a result of the cumulated growth opportunities over their life. Therefore, aggregate markup declines as all firms lower their markup but the decrease in sales concentration represents the low growth opportunities of the largest firms. Such a decline in the entry rate contributes to lower consumer welfare and the job reallocation rate, although the share of intangible assets in the economy increases.

Finally, I simulate an increase in the productivity gap between leader and followers firms by increasing productivity dispersion. The increased productivity dispersion augments the productivity gap between the largest firms and other firms and raises the incentive of the most efficient firms to build a customer base. Therefore sales share concentration increases as the largest firms become more productive and gain more customer base. However, aggregate markup slowly decreases as the largest firms have reduced their markup and increased sales share. Furthermore, that increase in the productivity gap between leader and follower lowers the consumer welfare with an increase in the job reallocation rate.

To quantify the contribution from the rise in the share of investment in intangible assets to the rise of aggregate markup, I exploit the exogenous variation observed in the Selling and General Administrative Spending to sales ratio (SG&A to sales ratio) to match the variation in the share of investment in intangible assets during the transition. Changes in SG&A to sales have been highlighted as evidence of customer base investments by firms (Gourio and Rudanko, 2014). The SG&A to sales ratio has increased from 0.125 to 0.16 over the four last decades and industries that have experienced an increase in the SG&A to sales ratio have grown in markup.

The model predicts that the exogenous rise in SG&A to sales ratio over the four last decades explains half of the observed increase in aggregate markup with no change in the sales shares concentration. This increase has increased the share of intangible assets by 29 % and reduced consumer welfare by 98 %. Solving for the transition dynamic to quantify the contribution from both the within-increase in markup and the reallocation of sales shares to the rise of markup, I found that the reallocation of sales shares from low to high markup associated with a change in firm-level markup is the main driven source in the rise of aggregate markup. The joint change in sales share and markup drive the reallocation in sales shares from low to high markup firms.

This paper contributes to the growing literature on the increase in aggregate markups (Barkai, 2016; De Loecker et al., 2020; Hall, 2018; De Loecker et al., 2020). De Loecker et al. (2020) develop a micro-level estimation of firms' markup that is independent of the price strategy set by firms. They find that the average markup rose from 21% to 61% above the marginal cost between 1980 and 2014. However, this sharp increase in markup is not a consensus. Karabarbounis and Neiman (2018) and Traina (2018) apply a correction to the De Loecker et al. (2020) by introducing overhead labor cost measured by the Selling and General Administrative (SG&A) spending as a variable cost. They found a stable average markup over the same period. My paper differs from the previous papers by starting with the notion of aggregation. I propose a useful metric to aggregate firm-level markup based on any firms' price strategy. My result is consistent with the markup estimation based on macroeconomics variables (Nekarda and Ramey, 2013; Eggertsson et al., 2018) and the rise in profit share. Furthermore, the aggregation offers an answer to the puzzling difference discussed by Basu (2019) between the change in firm-level markup and some observed macro observations.[2]

The paper is also related to the causes of the increase in aggregate markup and market share concentration (De Loecker et al., 2020; Autor et al., 2017; Gutierrez and Philippon, 2017; Baqaee and Farhi, 2018; Reenen, 2018; Aghion et al., 2019). De Loecker et al. (2020) suggest that the increase in markup is consistent with the decline in firm competition. That decline may result from a rise in entry cost (Gutierrez and Philippon, 2017), a decline in antitrust law (Peltzman, 2014; Edmond et al., 2018), increase in merger and acquisition (Grullon et al., 2018), increase in the scalability through intangible asset (Edmond et al., 2018; De Ridder, 2019), the irreversibility of intangible investments (Weiss, 2020). On the other hand, the reallocation of market shares across firms has been suggested as the primary driver for the market share concentration. Baqaee and Farhi (2018) argues that the increase in markup and market share concentration is followed by an increase in firms' efficiency, leading to a greater market share reallocation. Autor et al. (2017) focus their analysis on the market share concentration and argues that new information technology leads to a reallocation of market share from low to high efficient firms. Recently Aghion et al. (2019) argue that the increase in sales shares concentration results from an expansion of large firms to the new market. This paper is linked to this literature on two points. First, I show that recent sales shares concentration is not associated with the rise of markup and is mainly driven by an increase in the productivity gap between leader and follower. Second, the rise of intangible investments has increased the scale of future demand and contribute to the rise of aggregate markup. This framework exploits the scalability effect of intangible assets as in De Ridder (2019) but through a dynamic setting.

[2] Eggertsson et al. (2018); Syverson (2019) develop a model to explain how the increase in markup and the decline in the interest rate could explain recent macroeconomic facts. One key ingredient is the magnitude of the markup growth that I reconcile with firm-level estimation.

The next sections are organized as follows. Section 2 discusses the aggregation of firm-level markup and presents the data on the joint distribution of markups and sales. The third section proposes a model and explains the main causes of the rise in markup and sales shares concentration. Section 4 concludes the paper.

1.2 Empirical

This section presents a method to estimate industry markup based on the aggregation of firm-level markups. I estimate firm-level markup using the production approach (De Loecker et al., 2020). This approach fits with the structure of my dataset structure and does not require any assumptions about the market structure and the firm price strategy. Based on the aggregation and the joint distribution between sales and markups, I provide some new evidence about the rise of markup and the role of superstar firms.

1.2.1 Data

I use the Compustat dataset which reports information about the publicly listed US firms' balance sheets. This dataset has been recently used to estimate firm-level markup (De Loecker et al., 2020; Traina, 2018; Karabarbounis and Neiman, 2018) and to study market share concentration (Gutierrez and Philippon, 2019). The main advantage of the dataset is that it covers all sectors and it is representative in terms of sales. Publicly traded firms account for 41% of the US sales and 1/3 of the US labor market (Asker et al., 2015; Davis et al., 2006), even though they represent a small fraction of US firms.[3]

Price and quantity are not reported separately in Compustat. To compute the firm-level markup, I use the De Loecker and Warzynski (2012) production framework based on Hall (1988). The period for my sample is from 1980 to 2014. The main variables used to estimate firm-level markups are sales, Cost of Goods Sold (COGS) and Selling and General Administrative Spending (SG&A). COGS represents all the direct costs used to produce a good that I use to compute firm-level markup. Traina (2018) argues that the Selling and General Administrative Spending (SG&A) is a variable input and uses to estimate firm-level markup. SG&A usually reports the entire cost of selling and delivering goods and services but also all the cost to manage the company.[4] This paper considers SG&A as a fixed cost of production used to force sales.

Since Compustat tracks publicly traded firms, I define entry as the first time a firm is listed in my dataset. Exit arises when a firm is no longer listed. Because new publicly-listed traded firms can be new firms, private firms or the result of mergers and acquisitions, these definitions of entry and exit should not be confused with the creative-

[3] An important issue from using Compustat to estimate US markup is the selection. Publicly listed firms are the largest firms inducing a bias of representativeness in terms of markup. See Basu (2019) and Syverson (2019) for more comments.

[4] It can also include rent, utilities and supplies that are not part of the manufacturing process.

destructive process of firm dynamics. Other variables used to estimate the production function include capital expenditure and capital stock. I construct capital stock using the perpetual inventory method. I use the Gross Property Plant and Equipment (PPEGT) to measure the book value of capital at the entry in the sample and compute the net investment with the change in the Net Property Plant and Equipment (PPENT). I consolidate the capital expenditure with the change in PPENT to compute the firm investment in capital stock. I deflate all the variables using the Consumer Price Index (CPI).

I measure the sales concentration by the share of total sales held by superstars' firms. Following Autor et al. (2017), I define superstar firms as the four largest firms in terms of sales for each industry (SIC 3 digits) and year. Because firm-sales measures in Compustat contain many outliers, I use alternative definitions procedures for robustness. First, I define "superstar firms" based on the market value instead of sales. I measure the sales shares concentration for the 20 largest firms by sector (2 digits) for a given year or the 100 largest firms by year. Our main findings are robust using alternative definitions of superstars.[5]

1.2.2 Markups estimation and aggregation

Consider a firm i in industry j that use an aggregate input index c_i (reflecting a bundle of labor, materials, energy and other inputs) and capital stock k_i to produce an output y_i. Both inputs are respectively priced at p^c and r in a competitive input market. The firm produces output using a technology defined by $y_i = y_i(a_i, k_i, c_i)$ and sells at the price p_i. a_i refers to an efficiency parameter. I assume that there are no distortions such as financial constraints or overhead labor costs.[6] The firm's objective is to minimize its total spending to produce a level y_i of output.

$$\begin{aligned} &\min_{c_i,k_i} rk_i + p^c c_i \\ s/c \quad &y_i(a_i, k_i, c_i) \leq y_i \end{aligned} \tag{1.1}$$

Let define λ_i the Lagrange multiplier which represents the shadow cost to produce one additional unit of output. The multiplier measures the marginal cost because it represents the variation in the total cost following a marginal increase in output. Equation (1.2) states that the firm would increase its input until the price of the input c_i equals the marginal gain from an additional unit of input. Therefore, I recover the firm's marginal cost from equation (1.2) and estimate the firm's markup (equation 1.3) as the wedge between the input share (α_i^c) and the input elasticity (θ_i^c).[7]

[5]The subset of superstars firms change every year. The probability of being a superstar firm conditional on being a superstar firm last year is about 85%.

[6] Rotemberg and Woodford (1999) and Hall (1988) have shown that a such distortions may affect the estimation of markup and suggest that our markup estimation is biased.

[7]A similar result is obtained under Bertrand competition.

$$p^c = \lambda_i \frac{\partial y_i}{\partial c_i}(a_i, k_i, c_i) \tag{1.2}$$

$$\mu_i = \frac{p_i}{\lambda_i} = (\frac{\partial y_i}{\partial x_i}\frac{x_i}{y_i})(\frac{p_i y_i}{p^x x_i}) = \theta_i^x \frac{p_i y_i}{p^x x_i} \qquad x = k, c \tag{1.3}$$

Assuming a Cobb-Douglas production function ($y_i = a_i k_i^{\alpha_j} c_i^{\beta_j}$) where α_j and β_j are respectively the capital and aggregate input elasticities in the industry j. I estimate those elasticities using the Olley and Pakes (1996) control function approach. The strategy consists of estimating the production function using investment as a control variable. I select this approach due to the data restrictions.[8]

Following Olley and Pakes (1996), I assume that the efficiency parameter (a_{it}) follows an AR(1) process and use deflated sales to proxy the firm's value-added.[9] Investment is a choice variable and depend on efficiency and the level of capital stock ($i_i = f(k_i, a_i)$). f is assumed to be invertible in each component. However, selection distorts the investment and output decision by introducing an exit rule. Firms decide to stay and produce if their productivity is higher than a threshold ($\bar{a}$). I estimate the production function by sectors (SIC 2 digits) using the following specification:

$$\ln(\tilde{s}_{it}) = \alpha_j \ln(k_{it}) + \beta_j \ln(c_{it}) + \gamma_j pr_{it} + \ln(a_{it}) + \varepsilon_{it} \tag{1.4}$$

where $\tilde{s}_{it}$, pr_{it} are respectively the deflated sales and the survival probability. The estimation strategy is as follows. In the first stage, I used a non-parametric approach to identify the aggregate input index elasticity β_j. To do so, I exploited the invertibility property of the control function to recover the capital and substitute it in the production function. Furthermore, I used the polynomial approximation to the fourth degree to proxy the inverse of the control function. In the second stage, I used the $AR(1)$ process of productivity where I controlled for selection to identify α_j. I computed the predicted probability to stay and introduced it as a covariate of the AR(1) process. The table (1.5) in the appendix reports the estimated elasticity of COGS by sectors. There is substantial heterogeneity between elasticity across firms. Using that elasticity, I estimated firm-level markup using equation (1.3) and applied the De Loecker and Warzynski (2012) correction to the estimation to remove noise.

---

[8](Levinsohn and Petrin, 2003) used materials as a control variable. Other approach includes (Wooldridge, 2009; Ackerberg et al., 2015) which required materials as inputs for the estimation. The lumpiness of investment can be a limitation of this approach. However, I found that only 1% of firms have an investment equal to 0. I ran a regression with and without them and there was no significant difference.

[9]This have been recently an issue because the estimated elasticities are biased by the average markup. See Bond et al. (2020). Recent papers propose various ways to estimate the production function by adding some restrictions on the return to scale (Flynn et al., 2019) or using non parametric approach (Gandhi et al., 2020).

Aggregation of firm-level markups

I define industry markup as the wedge between price and marginal cost set by a representative firm in an industry. In other words, it is the degree of monopoly power that a representative firm would have in a given industry.

To extend the De Loecker et al. (2020) framework and compute industry markup (μ_j), I remain in the same environment and compute the industry aggregate variables. The aggregate output (y_j) (resp. the price index p_j) is assumed to be a CES aggregation of the firm's outputs (firm's price) within an industry where η is the elasticity of substitution across the firm's output. I also define the aggregate capital (k_j) and aggregate input index (c_j), respectively, as the sum of capital and input index within each industry.

$$k_j = \int_{i \in j} k_i di \qquad c_j = \int_{i \in j} c_i di$$

Equations (1.5) and (1.6) respectively represent the aggregate output and aggregate productivity (TFP) within the industry j when the firm's objective is to minimize the total cost.[10] Equation (1.6) shows that in the absence of markup heterogeneity, the firm aggregate productivity is the geometric mean of firm productivity. Markup variation across firms distorts the aggregate productivity as described in the misallocation literature.

$$y_j = \left(\int_{i \in j} y_i^{(1-\frac{1}{\eta})} di \right)^{\frac{1}{(1-\frac{1}{\eta})}} = A_j k_j^{\alpha_j} c_j^{\beta_j} \tag{1.5}$$

$$A_j = \left(\int_{i \in j} \tilde{a}_i \left(\frac{(\tilde{a}_i \mu_i^{-1})^{\frac{1}{(1-\tilde{\alpha}_j - \tilde{\beta}_j)}}}{\int_{i \in j} (\tilde{a}_i \mu_i^{-1})^{\frac{1}{(1-\tilde{\alpha}_j - \tilde{\beta}_j)}} di} \right)^{(\tilde{\alpha}_j + \tilde{\beta}_j)} di \right)^{\frac{1}{(1-\frac{1}{\eta})}} \tag{1.6}$$

where $\tilde{a}_i = a_i^{(1-\frac{1}{\eta})}$; $\tilde{\alpha}_j = (1-\frac{1}{\eta})\alpha_j$ and $\tilde{\beta}_j = (1-\frac{1}{\eta})\beta_j$.

The main feature of De Loecker and Warzynski (2012) framework is that it derives the marginal cost through the Lagrange multiplier. Under a competitive input market and aggregate production function, the industry marginal cost can be recovered using the cost minimization setting. Therefore, I derive the marginal cost of our representative firm through the cost minimization setting and estimate the industry markup with a given industry price index. The representative firm problem is defined by:

$$\begin{aligned} &\min_{c_j, k_j} p^c c_j + r k_j \\ s/c \quad & A_j k_j^{\alpha_j} c_j^{\beta_j} \leq y_j \end{aligned} \tag{1.7}$$

[10] See the proof in the appendix.

Solving for the representative firm's problem, I recover the industry marginal cost. Given industry price index p_j, the aggregate industry markup is the wedge between the input elasticity and the industry cost-share.

$$\mu_j = \beta_j \frac{p_j y_j}{p^c c_j}$$

Let's define by s_i, ω_i and s_j as respectively, the firm's sales, firm's sales share, and industry sales. The following proposition describes the main result.

Proposition 1: *Under CES aggregation of outputs, the industry markup based on aggregation is the sale-weighted harmonic mean of firm-level markups. This is equivalent to the cost-weighted arithmetic mean:*

$$\mu_j^{-1} = \int_{i \in j} \frac{s_i}{s_j} \mu_i^{-1} di = \int_{i \in j} \omega_i \mu_i^{-1} di \tag{1.8}$$

The thinking behind the result is the following. Sales shares are optimally allocated across firms based on the markups variation weighted by the input cost. A similar allocation is made by the representative firm that holds all the sales shares. Therefore, the aggregate wedge is a cost-weighted average of firm-level markups which is equivalent to the sales-weighted harmonic means of firm-level markups. A similar result has been suggested by Edmond et al. (2018) under monopolistic and oligopolistic competition with specific demand functions.

Proposition (1) extends the Edmond et al. (2018) result to all industries independently of the industry market structure or the pricing strategy within the industry. The De Loecker et al. (2020) estimation of marginal cost doesn't require any assumptions about the market structure or the demand function. However, the result is sensitive to heterogeneity in the input elasticity across firms and aggregate function.

The harmonic mean gives greater weight to small markup values and is less sensitive to the right tail of the markups distribution. A sharp increase in the right tail of the markups distribution within an industry will not necessarily induce a sharp rise in industry markup unless the highest markup firms increase their sales shares. Therefore, aggregate markup will be less sensitive to the upper tail of the markups distribution than the average markup. Considering this aggregation, the estimated markup based on microdata can reconcile the macroeconomics facts, suggesting a low markup increase in the US economy.[11]

1.2.3 Stylized facts

This section presents some stylized facts about the trends in aggregate markup and the shares of sales held by the largest firms (superstars) in the US economy. Even though

[11] See Basu (2019) and Syverson (2019) puzzle

the aggregation is at the industry level, I measure the US markup as the sales-weighted harmonic mean of industry markup.[12] I measure the superstar's sales shares as the shares of sales held by the four largest firms by industry (to 3 digits) and weight by the industry sales share.

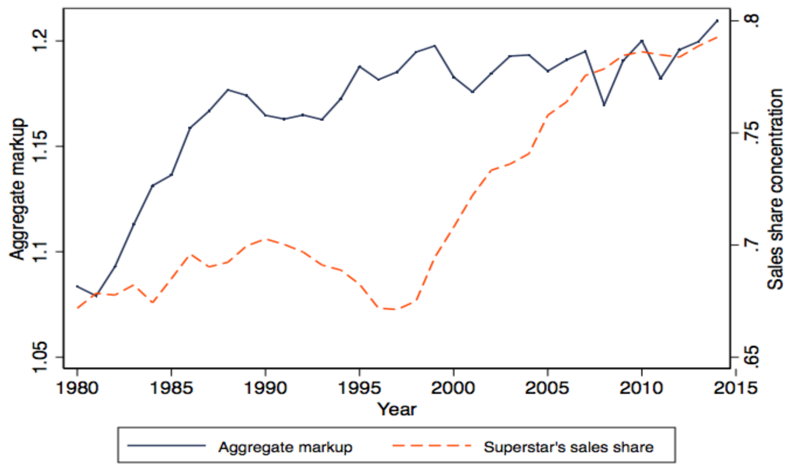

Figure 1.1: Aggregate markup and sales share concentration in US between 1981-2014

Figure (1.1) shows an increase in markup among publicly listed traded firms from 7.12% to 20.45% (a 13% points increase) above the marginal cost between 1980 and 2014. Most of the growth in markup occurred between 1980-1990, where markup increased from 7.12% to 17% above the marginal cost. After 1990, aggregate markup slowly increases and is driven by some business cycle variations (especially in 2001 and 2007).[13] Although the pattern is different, my estimated growth in markup has the same magnitude than Nekarda and Ramey (2013) and Eggertsson et al. (2018) estimations based on macro data during the same time period (see figure (1.6) in the appendix). The observed difference should result from the sample restriction on publicly traded firms and the estimated heterogeneity in input elasticity across industries.

Figure (1.1) also represents the total share of sales held by superstar firms in the US. Superstar firms hold more than 60% of the overall sales shares among publicly-listed traded firms and their sales shares have risen since 1980, reflecting the rise of the sales

[12] Basically, I have applied a similar aggregation at the industry level to estimate the US markup. The main concern is the heterogeneity in input elasticity across industries. Using a sales-weighted harmonic mean instead of a cost-weighted average allows me to analyze implications from the shift in sales shares.

[13] I replicated the De Loecker et al. (2020) estimation using a sales-weighted mean to compute aggregate markup and found a similar pattern. De Loecker et al. (2020) report a moment from the joint distribution between sales and markups which is the average markup. However, this moment is not a measure of market power in the overall economy. In appendix 5.3, I test the consistency in my estimation with the rise in profit share as suggested by Syverson (2019). A rise in markup growth by 13% is consistent with the rise of the profit share from 8% to 13% as in (Barkai, 2016).

concentration. Despite the data selection, the pattern is similar to the Autor et al. (2017) estimation of average sales shares concentration based on the US economic census. There are two periods of growth in sales concentration. The first period is between 1980-1990, followed by a sharp increase in markup and a second period in mid 1990 where markup slowly increases.

I observe a similar relation between markup and sales share concentration across industries. Figure (1.7) in the appendix represents the scatter plot between the average growth in markup and the average growth in superstar's sales share across industries (SIC 3 digits) over the periods 1980-2000 and 2000-2014. Despite the variation across industries, the rise of markup in most industries has been followed by an increase in superstars' sales share during the two first decades (see the left-hand panel in figure 1.7). However, during the last two decades, the rise in superstars' sales share has been followed by a decline in markup in most of the industries (see the right-hand panel in figure 1.7). Those trends suggest that the recent sharp increase in sales concentration after 2000 may not be related to the rise in markup.

This observation contributes to the recent debate on how the rise in sales concentration reflects a lack of competition. Rossi-Hansberg et al. (2018) discuss this question by showing a divergence between national and local concentration in sales, concluding that the sales concentration doesn't result from a decline in competition. Aghion et al. (2019) shows that firms' expansion into new markets has driven the sales concentration but not the change in markup. Although there is an absence of a positive correlation between markup growth and sales concentration after 2000, it doesn't suggest that superstar firms didn't adjust their markup to gain more sales shares.

Most recent studies suggest that superstar firms have the highest markup (Edmond et al., 2018; Baqaee and Farhi, 2018) and may have driven the rise of markup through an increase in sales concentration. However, the sharp increase in sales concentration after 2000 is not followed by an increase in markup during the same period. It suggests that superstar firms are not at the top of markups distribution or that the change in correlation between markup and sales share for both superstars and non-superstars matters for those two aggregate paths.

Change in the joint distribution of sales and markups

This section analyses the joint distribution of markups and sales shares both cross-sectionally and temporally. That joint distribution fully characterizes the aggregate markup. It is computed using sales shares and markups distribution within the industry and is weighted over the industry using the industry sales shares. I begin by documenting the cross-sectional joint distribution of markups and sales shares and therefore use the accounting growth decomposition to analyze the dynamic in that joint distribution.

Figure (1.8) in the appendix represents the joint distribution of markups and sales shares in 1980 and 2014. Notice, first, that sales shares and markups became more disperse over time without any significant shift in both distributions. Second, given that superstars' firms held more than half of the sales shares, one might expect the sales shares distribution to be left-skewed if superstar firms are the highest markup firms. However, the plot shows a high sales concentration at the average markup, suggesting that large sales firms are distributed around the average markup. A similar result is observed in figure (1.9) in the appendix, which represents the joint distribution of markups and sales shares for both superstar and non-superstar firms. Superstar firms are not distributed at the top of the markups distribution but with the non-superstars firms at the average markup. Such features mitigate the implications from the rise in superstars' sales share to the aggregate markup.

Most of the theory of endogenous markup predicts a perfect correlation (both positive or negative) between markups and sales shares, but empirical findings suggest a low correlation between markups and sales shares (Hastings et al., 2017; Burstein et al., 2019). I estimate a low and positive correlation between markups and sale shares within the industry (≈ 0.2). That low correlation results from the fact that the largest sales firms are not the highest markup firms. Because most of the models with endogenous markup fail to generate a low correlation between markups and sales shares, they may not truly identify the role of large firms in the overall increase in markups.

To understand the contribution from those changes in the joint distribution of markups and sales shares to the US markup's growth, I use the Foster et al. (2001) accounting growth decomposition given by equation (1.9). It allows me to decompose the markup's growth into a change in the within firm-level markup (*within*) and a reallocation of sales share across firms. The reallocation of sales share is divided into a shift in sales shares that is independent of the change in firm-level markup (*between*), or correlated with the change in firm-level markup (*cross)* or finally driven by the extensive margin (*net-entry*). I apply this decomposition to the harmonic aggregation of firm-level markups and use a scale factor to recover the markup's growth.

$$\Delta\mu_t^{-1} = \underbrace{\sum_{i\in S}\omega_{it-1}\Delta\mu_{it}^{-1}}_{Within} + \underbrace{\sum_{i\in S}\mu_{it-1}^{-1}\Delta\omega_{it}}_{Between} + \underbrace{\sum_{i\in S}\Delta\omega_{it}\Delta\mu_{it}^{-1}}_{Cross} + \underbrace{\sum_{i\in E}\omega_{it}\mu_{it}^{-1} - \sum_{i\in X}\omega_{it-1}\mu_{it-1}^{-1}}_{Netentry} \quad (1.9)$$

First, I apply the decomposition at the industry level at which point the net-entry component disappears. Table (1.6) in the appendix decomposes the aggregate markup growth into a within-industry increase in markup and an industry reallocation of sales share. The result suggests that the rise in markup has been driven at 79% by an increase in industry markup rather than a reallocation of sales shares from low markup industries to high markup industries (24.9%). However, on average, that increase in industry markup has not been followed by an increase in industry size. Most of that increase in industry markup and reallocation in sales shares across industries arose between 1980-1990.

Second, I apply the decomposition at the firm level within an industry. Table (1.7) in the appendix reports the weighted average of the accounting decomposition across the industry. I run the decomposition over four periods to distinguish between periods of a sharp increase in markups and periods of sales concentration. On average, markup increases by 0.73 pp per year over the last fourth decades driving by a within the increase in the firm-level markup (30%) and a reallocation in sales shares from the low to high markup firms.

The reallocation in sales shares from low to high markup firms is essentially driven by the *cross term* suggesting that on average, the firms' sales share growth has been associated with an increase in firm-level markup. Therefore the correlation between markup and sales share matters. De Loecker et al. (2020) shows that the rise in markup is driven by a reallocation in sales shares which is independent of the change in firm-level markup. The aggregation drives this difference in whether the sales shares reallocation is independent or not to the change in firm-level markup. The use of the sales-weighted harmonic mean of firm-level markup introduces an additional weight in the accounting decomposition that shifts the change in sales shares by the change in the cost.[14]

In addition, both the extensive margin (*net-entry term*) and the *between term* have a negative effect on the sales shares reallocation. The *between term* suggests that without considering a change in firm-level markup, low markup firms are gaining more sales share than high markup which reduces the markup growth. Between 2000-2010 where the sales concentration sharply increases, the *between term* has been more important while the *cross term* didn't change. This suggests that the rise in sales concentration has likely been independent of the change in firm-level markup and has slowed the markup growth. The *net-entry term* plays a marginal effect on the reallocation effect but has contributed to slow the rise in aggregate markup.[15]

Thus, the largest firms are not the highest markup firms and the rise in markup has been driven by a reallocation of sales shares from low to high markup. However, this reallocation in sales shares has been associated with a change in firm-level markup. Furthermore, the rise in sales concentration is likely not associated with a change in firm-level markup and has slowed the US markup growth. Therefore, the role of superstar firms in the rise in US markup is mitigated.

Role of largest firms

Baqaee and Farhi (2018) shows an improvement in the allocative efficiency in the US driven by the reallocation of market shares to high markup firms. They suggest that efficient firms have a high markup and the increase in aggregate markup might reflect

[14]The additional weight induce by the aggregation is such as $(\Delta\mu_{it}^{-1} = \mu_{it}^{-1}\mu_{it-1}^{-1}\Delta\mu_{it})$. That weight changes the contribution of each component of the sales shares reallocation.

[15]The negative sign from the net-entry traduces that exiting firms have a high markup and sales than new firms.

a greater allocative efficiency. Autor et al. (2017) argues that new technology improves automation and raises overhead labor costs. The more efficient firms enjoy the technology advent that reduces their labor demand and expands their production. Therefore, the labor share falls and market share concentration rises. If markups are the inverse of labor share (as suggested by Rotemberg and Woodford (1999)), both aggregate markup and market share concentration will increase at the same time.

Table 1.1: Difference in sales and markup between superstars and non-superstars

	(1) ln(markup)	(2) ln(sale)	(3) ln(markup)	(4) ln(sale)	(5) ln(markup)	(6) ln(markup)
superstar	0.0154***	3.084***			-0.0309***	
	(0.00402)	(0.0160)			(0.00124)	
(superstar=0) × time			0.00173***	0.0277***		0.00130***
			(0.000144)	(0.000568)		(0.0000528)
(superstar=1) × time			0.00174***	0.0293***		0.00128***
			(0.000143)	(0.000567)		(0.0000528)
Constant	0.210***	2.336***	-3.243***	-53.03***	0.144***	-2.452***
	(0.00149)	(0.00587)	(0.286)	(1.133)	(0.00105)	(0.106)
Observations	178443	182083	178443	182083	178443	178443
R^2	0.0681	0.3064	0.0667	0.3048	0.3122	0.3107
Industry FE	√	√	√	√	√	√
Time FE	√	√			√	
Cost weight					√	√

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table (1.1) reports the difference in the average sales and markup between superstars and non-superstars on the cross-sectional level and within the industry. Columns (1) and (2) report the cross-sectional difference in average sales and markups between superstars and non-superstars. Even though both superstars and non-superstars markups distribution are similar, superstar firms have on average, a high markup and large sales share than non-superstar firms. Furthermore, columns (3) and (4) report the average growth for each type of firm. On average, both superstar and non-superstar markups have increased at the same rate (almost 0.17% per year). In contrast, superstars' sales share has grown more than non-superstar firms - explaining the rise in sales concentration driven by those largest firms. This concludes that the rise in markup is driven by an increase in the within firm-level markup by both superstar and non-superstar firms and by a reallocation of sales shares from non-superstar to superstar firms.

However, this approach doesn't account for the aggregation function to measure markups for both types of firms. Proposition (1) shows that aggregate markup can be decomposed into groups where the average markup for a group is the cost-weighted average of firm-level within the group.[16] In others words, I cost-weighted the regression to account for the aggregation. Column (5) in the Table (1.1) reports the cross-sectional difference in cost-weighted average markup between superstar and non-superstar firms. At the aggregate level, superstar firms have a lower markup than non-superstar firms suggesting a difference in correlation between markups and sales shares for both types of firms.

Furthermore, columns (3) and (4) report the average weighted growth for each type of firm and shows that the superstar firm's markup has grown less than the non-superstar firm's markup. Thus, the rise in markup is driven by an increase in the markup by both superstar and non-superstar firms. If sales share concentration is independent of the rise of markup as suggested by Rossi-Hansberg et al. (2018) and Aghion et al. (2019)), it would have contributed to slow the rise in markup with no special role play by superstar firms to the rise of markup.

1.3 Model

In this section, I build a model of endogenous markup, where the joint distribution of sales shares and markups is characterized by a low correlation between markups and sales shares and use the model to quantify the contribution from exogenous shocks to the rise in markup. The environment is characterized by firms' accumulation of intangibles assets (R&D, brands, customer base, etc). Recent evidence documents an increase in intangible assets accumulation by firms (Crouzet and Eberly, 2018). Furthermore, investment in some intangibles assets is a source of a firm's market power (Aghion et al., 2019; Peters, 2020). The model is built on the spirit on Hopenhayn (1992) framework where heterogeneous firms in productivity accumulate intangible assets in the form of customer base. Such accumulation required a costly investment and generated an endogenous dynamic in markup.

1.3.1 Set-up

Preferences

There is a representative consumer with a CES utility function who owns all firms. Each firm produces a differentiated good i and consumers have a habit preference (b_{it}) for each good that may externally change their behaviour (external habit formation). Firms valued habit as an asset and constantly invest to build it over its life. Habit represents the customer base asset holds by firms. There is no competition in customer base accumulation in the sense that an increase in habit preference for a given good

[16] Or sales-weighted harmonic means of firm-level markups within those groups.

may alter the habit preference for other goods.[17] The consumer supplies inelastic labor at wage ω_t, gets dividends D_t from firms but doesn't save income. The representative consumer allocates his income to buy differentiated goods and solve :

$$\begin{aligned} \max_{(c_{it})_i} C_t &= \left[\int_i (y_{it} b_{it}^{-\theta})^{1-1/\eta} di\right]^{1/(1-1/\eta)} \\ \int_i p_{it} y_{it} di &= \omega_t + D_t = R_t \end{aligned} \tag{1.10}$$

where y_{it} is the consumption of goods i purchase at price p_{it} at the time t. η ($\eta > 1$) is the elasticity of substitution between differentiated goods and θ ($\theta < 0$ and $\theta(1-\eta) < 1$) is the deep habit parameters as defined by Ravn et al. (2004). By solving the consumer problem, the demand for each differentiated good and the price index is given by equation (1.16) where p_t the price index (I normalize $p_t = 1$).

$$y_{it} = p_{it}^{-\eta} b_{it}^{\theta(1-\eta)} R_t \qquad\qquad p_t = \left(\int (p_{it} b_{it}^{\theta})^{1-\eta} di\right)^{1/(1-\eta)} \tag{1.11}$$

Production

At the beginning of each period, each firm i holds a level of customer base b_{it} and draws an idiosyncratic productivity shock a_{it} from an $AR(1)$ process. Given the demand, the unit cost of labor ω_t and its characteristics (a_{it}, b_{it}), the firm chooses the level of labor n_{it}, output y_{it}, price p_{it} and the future stock of customer base b_{it+1}. Investing in customer base is costly and requires a fixed cost (φ_t) per unit of sales. The fixed cost is used as investment to build customer base and then φ_t represents the share of investment in intangible assets (customer base) made by firms to increase future sales.[18]

$$\begin{aligned} \pi_{it} &= p_{it} y_{it} - \omega_t n_{it} - \varphi_t p_{it} y_{it} \\ &= (1-\varphi_t) p_{it} y_{it} - \omega_t n_{it} \end{aligned} \tag{1.12}$$

Equation (1.16) represents the law of customer base accumulation, where δ is the habit stock depreciation and $(\varphi_t p_{it} y_{it})$ is the amount of investment. Each period, a firm loses a fraction $(1-\delta)$ of its customer base but pays the fixed cost φ_t as an investment to increase its asset. Accumulating the customer base increases the future demand and profit through a demand shift. However, at the margin, the return from each additional unit declines with the size of the customer base ($\theta(1-\eta) < 1$) inducing a limit to the accumulation.

$$b_{it+1} = (1-\delta) b_{it} + \delta(\varphi_t p_{it} y_{it}) \tag{1.13}$$

[17]This assumption suggests an absence of competition for customer base. It represents the perfect loyalty of the consumer to each good that is produced.

[18]φ_t captures marketing spending made by the firm to increase its future demand.

The firm uses a Cobb-Douglas technology ($y_{it} = a_{it}n_{it}^{\alpha}$) to produce goods. At the end of each period, the firm exogenously exits the market with an exogenous probability $\bar{s}$. I define by $\tilde{\beta} = (1-\bar{s})\beta$ the discount rate adjusted to the survival rate. By assuming a symmetric equilibrium, the firm problem can be described over the state variables (a, b). The recursive formulation of the firm problem is the following:

$$
\begin{aligned}
&v(a,b) = \max_{p,y,n,s'} \left\{(1-\varphi)py - \omega n + (1-\bar{s})\beta \int v(a',b')dG(a'|a)\right\}\\
&y = an^{\alpha}\\
&p = y^{-1/\eta} b^{-\theta(1-1/\eta)} R^{1/\eta}\\
&b' = (1-\delta)b + \delta(\varphi p y)\\
&\ln(a') = (1-\rho)\ln(a) + \sigma\varepsilon'
\end{aligned}
\tag{1.14}
$$

where $v(a, b)$ is the firm's value function at a given state's variables. The optimal price set by a firm is such that it charges a markup on its marginal cost. A firm's markup is the wedge between the labor input elasticity and the labor share. Firms charge a markup for two reasons. First, the differentiation of goods across producers provides a monopoly power and second the customer base accumulation, which allows firms to lock in customers and charge a markup.

$$
\begin{aligned}
(\mu^{-1} - \hat{\mu}^{-1}) &= \tilde{\beta}E_a\left((1-\delta)(\mu'^{-1} - \hat{\mu}^{-1}) - (\theta\delta\varphi\bar{\mu}^{-1})\frac{p'y'}{b'}\mu'^{-1}\right)\\
&= -(\theta\varphi\delta\bar{\mu}^{-1})\tilde{\beta}E_a\left(\sum_{j=0}^{\infty}(\tilde{\beta}(1-\delta))^j\frac{p_{j+1}y_{j+1}}{b_{j+1}}\mu_{j+1}^{-1}\right)
\end{aligned}
\tag{1.15}
$$

The incentive to build a customer base generates a trade-off between setting a high price to increase the short-run profit or setting a low price to attract more customers. That trade-off is represented by equation (1.18), where μ is the firm markup, $\hat{\mu} = \bar{\mu}(1-\varphi)^{-1}$ is the upper bound of the markups distribution and $\bar{\mu}$ the monopolistic markup. The latter represents the upper bound in the markups distribution. The left-hand side is the marginal cost of an additional unit of customer base represented by the short-run profit, and the right-hand side is the expected marginal gain from an additional unit of customer base represented by the expected future marginal profit plus the marginal gain of sales share from the demand shift, conditional on survival.

To build customer base, firms reduce their markup and make more sales at the cost of profit loss. The rise in sales increases the cost of building customer base and the size of the customer base next period. The rise in customer base increases the firm demand, sales shares, and profit the following periods, making firms more profitable as they build customer base. Therefore, firms' markup increase as they are growing. The decrease in the return of an additional unit of customer base induces a decline in the firm's markup growth over the firm's life. A firm builds customer base until it reaches the upper bound markup ($\hat{\mu}$) where it fully exploits its monopoly power.

Entrant's problem

Because entry in Compustat means launching an IPO, a new publicly-listed traded firm can be a private or newly-created company. A new firm will launch an IPO if the expected value after the entry is higher than an outside option. This outside option is the expected value of being either a private firm or the cost paid at the entry for a newly-created firm. I summarize this cost by (k_e).

Prior to entry, a new IPO firm pays the fixed cost (k_e) per unit of wage and draws a new productivity shock (a_0) and customer base (b_0) from two independent distributions $\tau(a_0)$ and $\chi(b_0)$. $\tau(a_0)$ is assumed to be a log normal distribution derived from the stationary distribution of productivity growth motion and $\chi(b_0)$ the stationary distribution of customer base derived from the steady-state distribution.[19] After drawing a new productivity shock and customer base, a new firm starts to produce in the next period as an incumbent.[20] The recursive formulation of the new firm problem is as follows:

$$\begin{aligned}
&v_{new} = -\omega k_e + \beta \int_{b_0,a_0} \left(\max_{p,y,b'} \left\{ (1-\varphi)py - \omega n + \tilde{\beta} E_{a_0}(v(a',b') \right\} \right) \chi(b_0)\tau(a_0) \\
&y = a_0 n^{\alpha} \\
&p = y^{-1/\eta} b_0^{-\theta(1-1/\eta)} R^{1/\eta} \\
&b' = (1-\delta)b_0 + \delta(\varphi p y)
\end{aligned} \tag{1.16}$$

The free entry condition implies that firms will enter until there is no value at the entry ($v_{new} = 0$). That free entry condition allows us to define the mass of new firms at entry and the law of motion of firm's distribution. Given the productive efficiency transition matrix and the policy function, I define the transition probability across the current and future state. With a mass of new plants m_t, the law of motion of firm distribution is defined by:

$$\Gamma_{t+1}(a',b') = \sum_{a,b} \Psi(a',b'|a,b)\Gamma_t(a,b) + m_{t+1}\chi(b_0)\tau(a_0) \tag{1.17}$$

1.3.2 Competitive equilibrium

A stationary equilibrium with entry which consists of a policy function $b'(a,b)$, $y(a,b)$, $n(a,b)$, $p(a,b)$, $d(a,b)$ and positive number R, ω, m such that : *i)* $b'(a,b)$, $y(a,b)$, $n(a,b)$, $p(a,b)$, $d(a,b)$ solves the firm problem given the R and ω. *ii)* $c(a,b)$ solves the consumer problem given R and ω. *iii)* free entry condition: $v_{new} = 0$. *iv)* labor market clear:

[19] Although the two distributions are assumed to be independent, they have common parameters. A high dispersion from productivity at the entry also result in a high dispersion in the customer base at entry. I use this approach to avoid imposing a distribution of customer base where parameters may affect the result.

[20] This time laps is introduced in order to easily compute the transition dynamic.

$\int n(a,b)d\Gamma(a,b) + mk_e = 1$. *v)* output market clear: $y(a,b) = c(a,b) \quad \forall a, b$ and *vi)* The stationary distribution of firms $\Gamma(a,b)$ solves the following equation :

$$\Gamma(a',b') = \textstyle\sum_{(a,b)} \Psi(a',b'|a,b)\Gamma(a,b) + m\chi(b_0)\tau(a_0)$$

1.3.3 Joint distribution and aggregate markup

This section describes the key features of the stationary joint distribution of markups and sales shares. The main feature is the low correlation between markups and sales shares that have been missing in many models of endogenous markup. Lets us define by $z(a,b)$ the sales share at a given state (a,b).

Proposition 2 : *The correlation between markups and sales shares $\rho(\ln z, \ln \mu)$ is such that :*

$$-1 < \rho(\ln z, \ln \mu) = \underset{(>0)}{\gamma_1}\, \rho(\ln a, \ln \mu) + \underset{(>0)}{\gamma_2}\, \rho(\ln b, \ln \mu) - \underset{(>0)}{\gamma_3} < 1 \tag{1.18}$$

where $\gamma_1 = \frac{\gamma_3 - 1}{\alpha}$; $\gamma_2 = \gamma_3\theta(1 - \frac{1}{\eta})$ and $\gamma_3 = \frac{1}{\frac{1}{\alpha} - (1 - \frac{1}{\eta})}$

Proposition (2) characterizes the correlation between markups and sales shares. That correlation depends on how markups varies with the states variables. Figure (1.10) in the appendix represents the firm's policy function for both firm's markup and sales share. Both markup and sales share increase with the customer base. By building their customer base, firms scale their future demand, which reduces their future marginal cost, increasing their sales and markups.

However, sales share (resp. markup) increases (resp. decrease) with the firm's productivity. This is simply because the most productive firms exploit their efficiency to invest more in customer base and then lower their markup. The largest firms are the most productive firms with the highest stock of customer base. Because markup decreases with a firm's productivity, those largest firms are not the highest markup firms. This negative relation between markup and a firm's productivity lowers the correlation between markups and sales shares.

By solving for the aggregate markup defined as the market power charged by a representative firm at the industry level, equation (1.19) shows that aggregate markup is fully characterized by the general equilibrium parameters (ω, R, m). This characterization is useful to follow the trend aggregate in markup although it doesn't show which firms are driving the markup growth.

$$\mu_{agg}^{-1} = \int_{a,b} z\mu^{-1}\Gamma(a,b) = \frac{\omega(1 - mk_e)}{\alpha R} \tag{1.19}$$

1.4 Quantitative analysis

1.4.1 Calibration

The model period is a year. I pre-set some parameters related to preferences and calibrate other parameters to match the key features from the joint distribution of markups and sales shares. The discount factor is set to $\beta = 0.96$, corresponding to a 4% yearly interest rate. Following Foster et al. (2016), both labor elasticity and elasticity of substitution of across differentiated goods are respectively set to $\alpha = 0.67$ and $\eta = 2.1$. Both the fixed cost share φ and the exit rate are chosen to match the share of investment in customer base and the exit rate in 1980. I measure the share of investment in customer base by the share of Selling and General and Administrative (SG&A) spending to the sales following (Gourio and Rudanko, 2014) which is equal to 12.5 % and the exit rate ($\bar{s}$) represents 10%.

Table 1.2: Calibration summary : joint identify parameters

Parameters	Values	Moments	Target	Model
k_e	1.40 x 10^{-6}	Average log markup	0.0740	0.0740
ρ	0.7655	AR(1) log markup	0.7350	0.7500
σ	0.2742	IQR of log markup	0.3340	0.3260
δ	0.3999	AR(1) log sales	0.9640	0.8810
θ	-1.0690	superstars (18.15%) sales shares	0.6720	0.6850

The remaining parameters are jointly calibrated to match the main moments. The estimated parameters are shown in the table (1.3), and they are chosen to minimize the average distance between data and the model.[21] The productivity parameters (ρ, σ) and δ are chosen to match both autocorrelations in sales and dispersion in markup. Due to the high value of some firms' markups, I use the Inter-Quantile Range (IQR) to measure the markup dispersion. Dispersion in markup results from the heterogeneity that we have in both productivity and customer base. This heterogeneity is driven by σ as the customer base distribution is derived at the steady-state dynamic in the customer base.

Both ρ and δ drive the autocorrelation in markup and sales. Those parameters are respectively the source of persistence in productivity and customer base, inducing a dependence between the growth and the level of both markup and sales. The autocorrelation in sales increases with both ρ and δ, as sales increase in productivity and customer base. However, those parameters would have an opposite effect on the autocorrelation of markup.

The deep habit parameter θ is chosen to match the sales share concentration. The sales share concentration is measured by the share of sales held by superstar firms. Superstar firms represent on average 18.15% of our sample of firms by industry and they

[21]The average percentage difference found is 8%.

account for 67% of the share of the overall sales.[22] Because $\theta(1-\eta)$ measures the elasticity of customer base in the demand and superstar firms are the most productive with a high level of customer base, θ likely matches the sales share concentration.

Finally, to pin down the entry cost, I target the average log markup. The average log markup is the sales-weighted harmonic mean of a firm's markup in log. The value is about 0.0776, which represents an average markup of 1.06 in 1980. From the entry condition, a fall in the wages reduces the marginal cost which raises markup and the firm's expected profit at entry. Therefore, I use the change in the wage to target the log average markup and then identify the entry cost.

1.4.2 Causes in the rise of markup

In this section, I use the model to analyze how shock on some parameters changes both the aggregate markup and the sales concentration. I then study the welfare effect and job reallocation rate change.[23] I focus on three shocks: the rise in the share of investment in customer base (φ), the change in the market structure through a decline in the exit rate s and an increase in the productivity gap between leader and follower firms through an increase in productivity dispersion σ.

Those exogenous shocks have been recently discussed as potential causes of the US economy's rise of markup. For example, Foster et al. (2018) and Decker et al. (2018) found an increase in productivity dispersion before 2000, especially in the high-tech sector which drives the productivity slowdown and may explain an increase in the gap between leader and followers firms within the industry. De Loecker et al. (2020) shows a low decline in the output elasticity and the return to scale since 1980. Furthermore, recent evidence suggests an accumulation of intangibles assets such as brands, customer base and R&D by firms (Crouzet and Eberly, 2018). Such an increase may result from a rise in the share of spending allocated to build a customer base or increase the sensitivity of the demand to the customer base.

(a) Increasing in the share of investment in customer base (φ): An increase in (φ) raises both the cost and the benefit of building customer base which unequally impacts all firms. The change benefits firms with high growth opportunities (low customer base firms), increasing their incentive to build more customer base and to grow faster in size and markup (figure 1.2). Along the life cycle, firms with a high customer base increase their markup and gain sales share. Therefore, aggregate markup increases.

[22] The share of superstar firms is high because of the definition of the industry and also the restriction from the dataset to publicly-traded firms.

[23] The job reallocation rate is measured as $RR = (JC + JD)/2$ where JC and JD refers to the jobs creation and jobs destruction rate. In this setting, the welfare change also characterizes the change in the aggregate productivity growth.

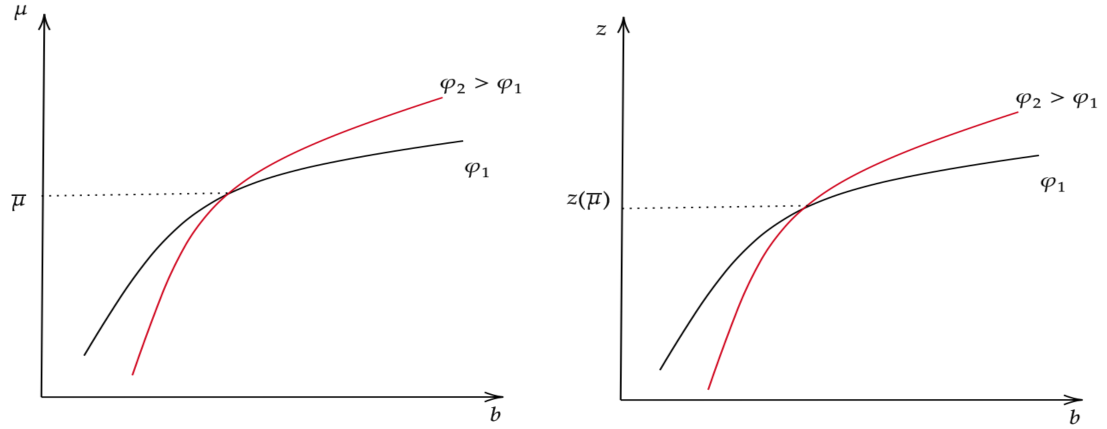

Figure 1.2: Change in φ on markup and market shares for a given productivity (a)

As the average firm at entry is less productive with a low level of customer base, it lowers his markup to grow faster. The expected profit at entry falls, inducing a marginal decline in the mass of firms at entry.[24] As few firms enter, the number of operating firms falls and contribute to lower the labor demand and then the wage.[25] In addition, the decline in the number of operating firms lowers the aggregate revenue R. However, the aggregate revenue fall is less than the wage due to the markup growth along the life cycle. Although high customer base firms have gained individually more sales share, the decline in the number of operating firms has reduced the number of superstar firms, leading the share of the total sales holds by superstar firms constant.

The rise in the share of investment in the customer base has additional implications. Figure (1.3) shows the relative change in the share of intangible assets, welfare and job reallocation. First, the rise in the share of investment in intangible assets contributes to lower consumer welfare as the number of operating firms declines. In addition, it increases the share of intangible assets (customer base) in the economy and makes unchanged the job reallocation rate. Two mechanisms drive the job reallocation rate result. At the intensive margin, the fall in the number of operating firms lowers the job reallocation rate while at the extensive margin, the fast growth in size contributes to increase the job creation and destruction and raise the job reallocation rate. Those two opposite mechanisms have a similar magnitude, making unchanged the job reallocation rate.

[24]The fast growth opportunities contribute to reduce the fall in the expected profit at entry and induce a small decline in the mass of firms at entry.

[25]Although the fast growth in customer base increase the labor demand over the life cycle, the fast growth in markup over the life cycle reduces the labor demand and then mitigate the overall effect on labor demand.

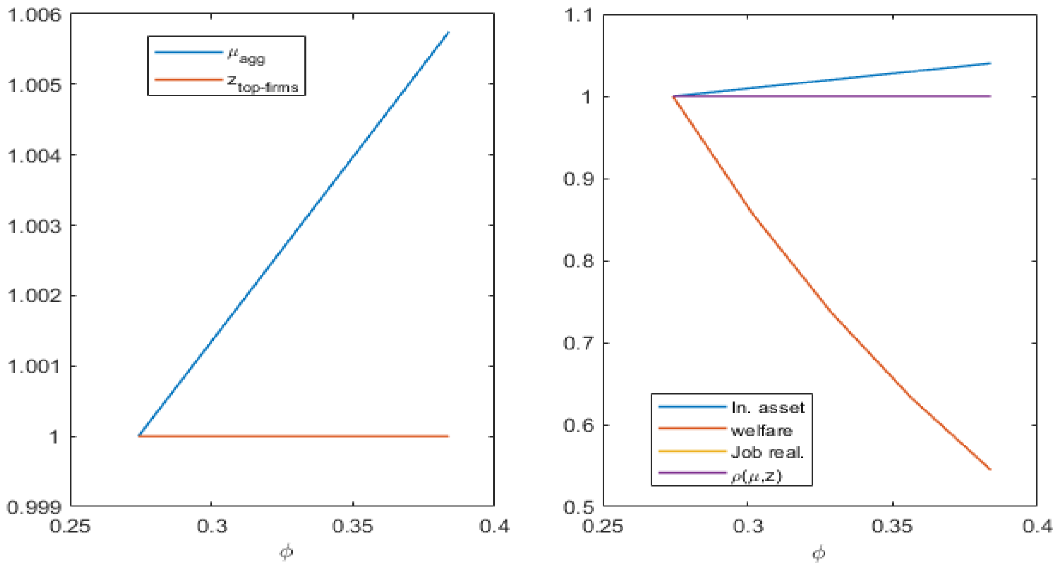

Figure 1.3: Aggregate implication from an increase in φ

(b) Change in the market structure: I simulate the change in market structure through a fall in exit rate (s). Because the equilibrium is stationary, a fall in the exit rate would decline the entry rate. The decline in s raises the survival probability and the growth opportunities of all firms, leading them to lower their markup unequally. The decline benefits firms with high growth opportunities (low customer base) who lower their markup to build more customer base and grow faster in size and markup (figure 1.11 in appendix). The decline in all firms' markup reduces the aggregate markup.

Because all firm-level markup declined, the expected profit at entry falls, reducing the mass of firms at the entry m. Therefore, only a few firms operate which reduces the labor demand and lowers the wage. Because few firms operate, their total sales decline and even decline more as all firms lower their markup. Aggregate markup declines because all firms lower their markup, reducing revenue R more than the wage ω. With the decline in the number of operating firms, the number of superstar firms falls. In addition, firms with a high customer base have fewer growth opportunities and don't gain sales share. Those two mechanisms reduce the sales share concentration.

Figure (1.4) represents other aggregate implications from the fall in the exit rate. The fall in the exit rate increases firms' incentive to build a customer base as they are more likely to survive. Therefore, the share of intangible assets (customer base) increases with the fall in the exit rate. However, the fall in the entry rate combined with increased markup variation lowered the aggregate revenue and reduced consumer welfare. The job reallocation rate is driven by both the intensive and extensive margin. As fewer firms operate and enter, the job reallocation rate declines. However, at the intensive margin, the increased accumulation in customer base drives the size growth-inducing more jobs reallocation. The extensive margin effect dominates as there is less job reallocation following a fall in the exit rate.

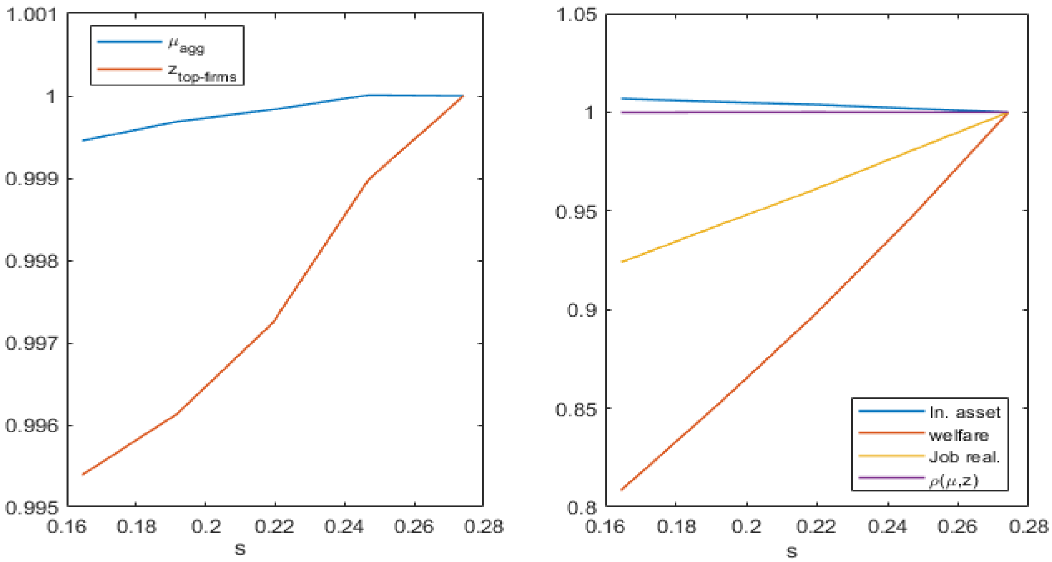

Figure 1.4: Aggregate implication from a fall in s

(c) Increasing in the productivity gap between leader and followers: Foster et al. (2018) and Decker et al. (2018) point to an increase in the productivity gaps between the leader and follower firms as a potential source of US markup's growth. I model that shock simply through an increase in productivity dispersion (σ). An increase in σ raises the productivity gap between the largest firms (superstar firms) and the rest (non-superstar firms). The largest firms become more productive and have more incentive to invest in customer base. They gain more sales share as they become more efficient, lower their markup and increase their customer base. By lowering their markup and increasing their sales shares, superstar firms contribute to lower the average markup and raise the sales share concentration (figure 1.12 in appendix).

As the productivity dispersion increases, the average firm at entry becomes less productive and charges a low markup. The expected profit at entry falls, reducing the mass of new firms. Thus, few firms operate which reduces the labor demand and lowers the wage ω. The decline in labor demand is more important as the average firm becomes less productive. In addition, the decline in the number of operating firms lowers the aggregate revenue R.

The increasing gap in productivity between leader and follower has additional aggregate implications (figure 1.5). First, the share of intangible assets in the economy remains unchanged because the rise in productivity dispersion has increased the dispersion in the customer base, especially at entry without affecting the mean. Despite the unchanged level in customer base share, consumer welfare falls as the total sales decline. Because more productive firms increase their size as they become more efficient, they hire more labor, explaining the job reallocation rate.

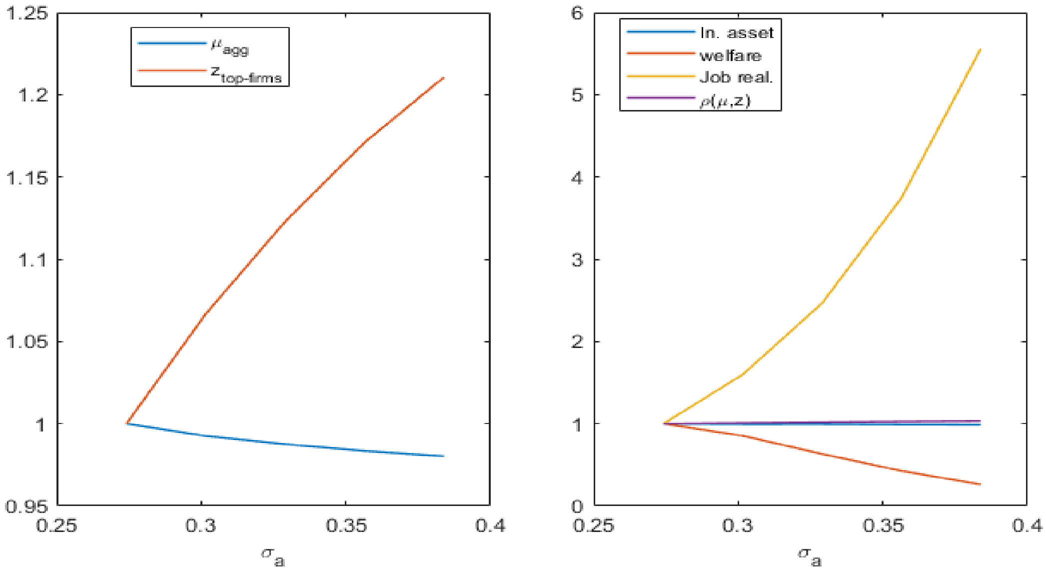

Figure 1.5: Aggregate implications from an increase in σ

In an environment where firms build intangibles assets, especially customer base, the rise in the share of investment in customer base may raise the aggregate markup in the economy. Other shocks such as the change in the market structure through the fall of entry rate (similar to a fall in exit rate) and the increasing productivity gap between leader and follower firms cannot drive the rise in markup. However, their implications are useful to understanding the rise in sales concentration and the decline in the jobs reallocation rate. For example, the model shows that the increasing gap in productivity between leader and follower may drive the sales concentration.

1.4.3 Rise in the share of SG&A spending

This section quantifies the contribution from an observed rise in the share of investment in the customer base to the rise of markup. I measure the share of investment in customer base by the SGA to sales ratio. This variable has been used to assess the firm incentive to build customer base (Gourio and Rudanko, 2014). I compute the transition dynamic to quantify the contribution from the within growth in firm-level markup and the reallocation of sales shares.

The SG&A represents all additional costs paid by firms to force future sales such as marketing spendings, customer data analytics, services.[26] It is usually measured as a

[26]SG&A spending represents the indirect costs to the firms that are not related to his production

fixed or overhead cost. A high SG&A to sales ratio indicates the firm's spending to build a customer base or goodwill and raise the future sales (Gourio and Rudanko, 2014). I am focussing my analysis on the aggregate trend of SG&A to sales. SG&A to sales ratio at the industry level is the sales-weighted average of a firm's SG&A to sales ratio.

Table 1.3: Change in the share of investment in customer base

		Relative change between 1981-2014	
	Benchmark	Change from data	Counterfactual
Aggregate markup	1.0765	1.1144	1.0441
Sales share Concentration	0.6854	1.1683	1.0000
S/NS markup ratio.	1.0155	1.0159	1.0000
Correlation (μ, z)	0.8568	4.8613	1.0000
Share of Intangible assets	0.1057	0.0000	1.2961
Welfare	1.7724	0.0000	0.0180
Job Reallocation Rate	0.1496	0.0000	1.0000
Average Size	0.0000	0.0000	1.0000

The left-hand panel of the figure (1.13) shows a sharp rise in aggregate SG&A to sales ratio between 1980-1990, following by slower growth between 1990-2014. SG&A to sales ratio rose by 4 pp between 1980-1990 (from 0.12 to 0.16 over the four last decades). However, the aggregate trend in SG&A to sales ratio is similar to the aggregate trend in markup, especially the sharp increase between 1980-1990. The right-hand panel of the figure (1.13) represents the scatter plot between markup's growth and the SG&A to sale ratio growth across industries. Markup has grown in industries where the SG&A to sales ratio increased faster. Those two correlations suggest that the SG&A to sales ratio variation has been positively correlated with the rise of US markup.

Table (1.3) presents the results from the counterfactual experiments where I quantify the change in markup and sales concentration followed by an increase in φ. Going from the benchmark economy, I simulate the counterfactual where the share of investment in customer base rise by 4 pp. That increase corresponds to the rise in the SG&A to sales ratio between 1980 and 2014.[27]. At the benchmark economy, the main features from the joint distribution of markups and sales shares at fitted. Intangible assets represent 10% of the total sales and 24% of the labor are reallocated.

The counterfactual experiment shows that an increase in the share of investment in intangible assets by 4 % explains half of the overall increase in aggregate markup (by 4 % relative to the 11 % in the data). As previously presented, there is no effect on the

process (SG&A). That includes advertising costs, selling costs, indirect labor costs, etc. Traina (2018) shows that the rise SG&A shares are correlated with the rise of estimate markup and recently De Loecker et al. (2020) suggest to include SG&A as part of the variable cost to estimate the marginal cost.

[27]Some moments from the benchmark economy are consistent with data, except for the correlation between markup and the sales share (the correlation in the sample is 0.02)

sales shares concentration and the difference in markup growth between superstar and non-superstar firms. Despite the rise in the share of intangibles assets by 29 % driven customer base accumulation, the consumer welfare has fallen by 98 %.

To quantify the contribution from a within-firm growth in markup and the sales reallocation following the rise in the share of investment in the customer base, I compute the transition dynamic between the two steady states taking the path in the share of investment in intangible assets exogenously.[28] Both within and sales reallocation occurs when firms build their customer base. As previously mentioned, I apply the accounting growth decomposition to the sales-weighted harmonic mean of a firm's markup.[29]

Table 1.4: Model simulation: Within vs Reallocation

				Reallocation		
	Growth	within	Reallocation	between	cross	netentry
Data	0.73	0.21	0.52	-3.13	4.53	-0.87
Model	0.35	0.15	0.20	-2.45	2.45	0.2
Contribution	100	42.86	57.14	-	-	-

Table (1.4) shows the contribution from the within-increase in markups and the sales share reallocation following the rise in the share of investment in customer base. For firms with a low customer base, the rise in the share of investment into customer base increases their incentive to build customer base and lowers their markup. As incumbent firms build customer base, they raise their markup and gain more sales share. This mechanism drives the joint effect from the markup and sales share change, which drive the sales share reallocation (positive sign from the cross-component). The within growth in firm-level markup is driven by the firm's productivity and customer base. Because firms with a low customer base are small and charge a low markup, they grow faster in size and markup (negative sign from the between component). Quantitatively, the rise in the share of investment in customer base has induced both a within-increase in markup and a reallocation in sales share from low to high markup firms. The reallocation is mainly driven by a joint change in firm-level markup and firm sales share.

1.5 Conclusion

Recent evidence on the rise in markup, increased profit share, and sales concentration across all industries have raised concerns about market power and a potential lack of

[28] The transition dynamic allows me to track both the sales and markup dynamic of each type of firm between the two steady states.

[29] I solve the transition dynamic by backward induction. I guess the revenue path and iterate the value function backward. Equilibrium conditions determine both the dynamic in the wage and the mass of entrants. Finally, I iterate the firms' distribution forwards to match the revenue equilibrium.

competition in the US economy. Changes in markups and sales by the largest sales firms - so-called "superstars" -have been singled out as a potential source of the increase of the aggregate markup. The present paper contributes methodological, empirical, and theoretical elements to understanding the rise in US markup. The paper analyzes the potential causes of the rise in an environment where firms are building intangible assets. I emphasize the measure of aggregate markup and its implications on the role played by superstar firms, the correlation between markup and sales shares, and the rise in the share of investments in intangibles assets, especially customer base.

Furthermore, the paper models the dynamics of firms in an environment where they have an incentive to build intangible assets in the form of a customer base. Such an environment captures and contextualizes recent evidence on the rise of intangible assets in the US economy and allows me to generate a non-perfect correlation between sales and markup. This last feature has been missing in many models of endogenous markups. While intangible assets accumulation gives more power to firms in terms of price setting and distorts the market competition, it doesn't generate sales concentration as observed in data between 1980-1990. However, the rise in the productivity gap between the leader and follower firms may have raised the sales concentration after 2000.

With the ongoing recession, some superstar firms have exploited their technological advantage and benefited from the recession. The regulation of superstar firms has received more attention. Despite the data limitation resulting from selection, a greater emphasis on the growing role of intangible assets in determining aggregate markup should be added to those competition concerns.

1.6 Appendix

1.6.1 Aggregation

Let us consider an industry where all firms produce differentiated goods and η is the elasticity of substitution across these goods. Industry output is assumed to be a Dixit-aggregation of firm output. I derive a firm demand function from solving for the industry price index based on firm price. The firm's demand and industry price index are derived from the minimization of the total spending.

$$y_j = (\int_{i\in j} y_i^{(1-1/\eta)} di)^{1/(1-1/\eta)} \tag{1.20}$$

$$y_i = \left(\frac{p_i}{p_j}\right)^{-\eta} y_j \qquad p_j = \left(\int_{i\in j} p_i^{(1-\eta)} di\right)^{1/(1-\eta)} \qquad p_j y_j = \int_{i\in j} p_i y_i di$$

Although we have the demand function, the firm's goal is to minimize its total cost. Its suggest that I don't make any assumptions about the price strategy or the type of competition within the industry. The demand is defined only for the purpose of aggregation. Therefore equation (1.3) represents the firm's optimal decision. The firm's sales deflated by the industry price index is defined as:

$$\frac{s_i}{p_j} = y_i^{(1-1/\eta)} y_j^{1/\eta} = \tilde{a}_i k_i^{\tilde{\alpha}} c_i^{\tilde{\beta}} y_j^{1/\eta} \tag{1.21}$$

where $\tilde{a}_i = a_i^{(1-\frac{1}{\eta})}$; $\tilde{\alpha} = (1-\frac{1}{\eta})\alpha$ and $\tilde{\beta} = (1-\frac{1}{\eta})\beta$. From the firm's optimal decision, the level of capital and aggregate input index are given by:

$$k_i = \frac{\alpha}{r}\frac{s_i}{\mu_i} \qquad c_i = \frac{\beta}{p^c}\frac{s_i}{\mu_i} \tag{1.22}$$

By substituting equation (1.22) to (1.21), I can recover the firm's sales deflated by the industry price index:

$$s_i = \left(\left(\frac{\alpha}{r}\right)^{\tilde{\alpha}} \left(\frac{\beta}{p^c}\right)^{\tilde{\beta}} (\tilde{a}_i)(\mu_i^{-1})^{(\alpha+\tilde{\beta})} p_j y_j^{1/\eta}\right)^{1/(1-\tilde{\alpha}-\tilde{\beta})} \tag{1.23}$$

Therefore the aggregate capital and aggregate input index are:

$$k_j = \int_{i\in j} k_i di = \frac{\alpha}{r}\int_{i\in j} s_i \mu_i^{-1} di \qquad c_j = \int_{i\in j} c_i di = \frac{\beta}{p^c}\int_{i\in j} s_i \mu_i^{-1} di \tag{1.24}$$

By substituting (1.23) in (1.24), and (1.22). I can recover that :

$$\frac{k_i}{k_j} = \frac{c_i}{c_j} = \frac{(\tilde{a}_i \mu_i^{-1})^{1/(1-\tilde{\alpha}-\tilde{\beta})}}{\int_{i\in j} (\tilde{a}_i \mu_i^{-1})^{1/(1-\tilde{\alpha}-\tilde{\beta})} di} = \theta_i \tag{1.25}$$

Therefore the aggregate demand is :

$$\begin{aligned}
y_j &= (\int_{i\in j} (a_i k_i^{\alpha} c_i^{\beta})^{(1-1/\eta)} di)^{1/(1-1/\eta)} \\
&= (\int_{i\in j} (\theta_i^{(\alpha+\beta)} a_i)^{(1-1/\eta)} di)^{1/(1-1/\eta)} k_j^{\alpha} c_j^{\beta} \\
&= A_j k_j^{\alpha} c_j^{\beta}
\end{aligned}$$

Following the definition of industry markup as the degree of monopoly power held by a representative firm at the industry level, the markup set by the representative firm using the cost minimisation setting is:

$$\mu_j = \beta_j \frac{p_j y_j}{p^c c_j} = \int_{i\in j} \frac{p^c c_i}{p^c c_j} \beta_j \frac{p_i y_i}{P^c c_i} = \int_{i\in j} \frac{p^c c_i}{p^c c_j} \mu_i$$

That is equivalent to :

$$\mu_j^{-1} = \int_{i\in j} \frac{p_i y_i}{p_j y_j} \mu_i^{-1} = \int_{i\in j} \omega_i \mu_i^{-1}$$

This result holds due to the constant input elasticity within the industry and the separability allowed by the CES aggregation. That is the main reason why this industry aggregation cannot hold at the aggregate level (heterogeneity in input elasticity across the industry).

1.6.2 Test of consistency for aggregation

To test the consistency of our aggregate markup estimation, I follow the Syverson (2019) and Basu (2019) decomposition of aggregate markup.

$$\mu = \frac{P}{MC} = \frac{P}{AC} \frac{AC}{MC} \tag{1.26}$$

The ratio $\frac{AC}{MC}$ is equal to the return to scale κ following our estimation of a firm's markup above. The firm-level markup estimation above is decribed using only one input. If we use all the inputs in order to recover the marginal cost, the elasticity will refer to the return to scale. The ratio $\frac{P}{AC} = \frac{1}{1-s_\pi}$, where s_π is the profit share in revenue. Therefore :

$$\mu = \frac{\kappa}{1 - s_\pi} \tag{1.27}$$

$$\frac{\mu_{2014}}{\mu_{1980}} = \frac{1 - s_{\pi,1980}}{1 - s_{\pi,2014}} \frac{\kappa_{2014}}{\kappa_{1980}} \tag{1.28}$$

I estimate an increase in the aggregate markup from 1.08 in 1980 to 1.20 to 2014.

De Loecker et al. (2020) estimate a return to scale from 1.03 to 1.08 and Barkai (2016) estimate a profit share from 8% to 13%.

$$1.1214 = \frac{1.20}{1.07} \simeq \frac{1 - 0.013}{1 - 0.08} \frac{1.08}{1.03} = 1.1249 \tag{1.29}$$

Our aggregate markup estimation seems to be consistent with the trend in the profit share.

1.7 Tables

Table 1.5: Cost of Goods Sold elasticity by sectors

	COGS elasticity
Agriculture/forestry/fishing	0.92
Mining	0.68
Construction	0.95
Manufacturing	0.79
Transport/communication/gas	0.85
wholesale_trade	0.89
retail_trade	0.91
services	0.74
non_classified	0.84
Total	0.80

Note.- The table reports estimation of Cost of Goods Solds elasticity at the sectoral level, using Olley and Pakes (1996) control function. Sectors reflect a group of industry (sic 2 digits) following the SIC classification.

Table 1.6: Accounting growth decomposition of markup across industry (sic 2 digits)

	Growth	Within	Between	Cross
1980-1989	10.07	6.64	3.93	-0.49
1990-1999	2.38	1.58	0.87	-0.07
2000-2010	0.17	1.35	-1.35	0.17
2010-2014	2.37	2.51	0.51	-0.65
Total	3.71	2.96	0.92	-0.17
Contrib	100.00	79.74	24.90	-4.64

Note.- The table represents the HFK accounting decomposition across industries. The change in markup within the industry is the main source (79.74%) of the overall rise in markup. The shift of sales shares across the industry only explains 24.90% of that overall increase in markup. Most of those effects arise between 1980-1990. The result is similar to De Loecker et al. (2020).

Table 1.7: Accounting growth decomposition at the firm-level

	Growth	Within	Reallocation	Reallocation: Between	Reallocation: Cross	Reallocation: Net-entry
1980-1989	1.37	0.87	0.48	-2.93	4.73	-1.32
1990-1999	0.41	-0.03	0.48	-2.17	4.77	-2.12
2000-2010	0.59	0.03	0.58	-4.70	4.89	0.38
2010-2014	0.43	-0.17	0.61	-1.66	2.44	-0.17
Total	0.73	0.21	0.52	-3.13	4.53	-0.87
Contrib	100.00	29.25	72.00	.	.	.

Note.- The table displays the growth decomposition of real marginal cost. Given the relation between the growth in real marginal cost and growth in markup express by $\Delta\mu_t = -\mu_t^{-1}\mu_{t-1}^{-1}\Delta\mu_t^{-1}$, I scale the decomposition by $-\mu_t^{-1}\mu_{t-1}^{-1}$ to recover the growth in markup. However, the result can be interpret in terms of real marginal costs and the conclusion would be similar.

Table 1.8: Difference in markup's growth between superstars and non-superstars between 1980-1990

	1980-2000			2000-2014		
	(1) ln(markup)	(2) ln(markup)	(3) ln(sale)	(4) ln(markup)	(5) ln(markup)	(6) ln(sale)
(superstar=0) × time	-0.000470	0.00233***	0.0204***	0.00518***	0.00242***	0.0386***
	(0.000252)	(0.0000910)	(0.00102)	(0.000644)	(0.000255)	(0.00244)
(superstar=1) × time	-0.000462	0.00231***	0.0219***	0.00520***	0.00240***	0.0402***
	(0.000252)	(0.0000910)	(0.00102)	(0.000644)	(0.000255)	(0.00244)
Constant	1.131*	-4.495***	-38.36***	-10.17***	-4.703***	-74.82***
	(0.502)	(0.181)	(2.035)	(1.292)	(0.513)	(4.887)
Observations	117385	117385	119068	61058	61058	63015
R^2	0.0578	0.3381	0.3189	0.0880	0.3092	0.2935
Industry FE		✓	✓		✓	✓

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.9: Correlation between selling cost and markup

	(1) ln($markup$)	(2) ln($markup$)	(3) ln($markup$)	(4) ln($markup$)	(5) ln($markup$)
ln($XSGA/Sale$)	0.0931***		0.0168***		0.0493***
	(0.00101)		(0.00111)		(0.00204)
ln($Advert/Sale$)		0.0162***		0.00998***	0.00835***
		(0.00129)		(0.00138)	(0.00120)
Constant	0.414***	0.368***	0.314***	0.352***	0.447***
	(0.00172)	(0.00388)	(0.00176)	(0.00399)	(0.00381)
Observations	161385	64914	161385	64912	60813
R^2	0.0502	0.0024	0.2119	0.1747	0.2710
Industry FE			√	√	√
Advert FE		√		√	√

Standard errors in parentheses

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

1.8 Figures

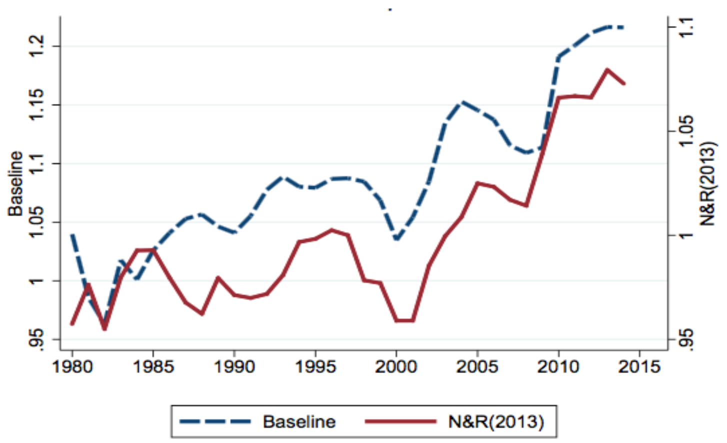

Figure 1.6: Markup estimation using macro data

Source : Eggertsson et al. (2018). The plot represents an estimation of aggregate markup based on macro data. The baseline estimation from Eggertsson et al. (2018), used labor and capital to compute marginal cost. The red plot from Nekarda and Ramey (2013) only used labor to estimate the marginal cost.

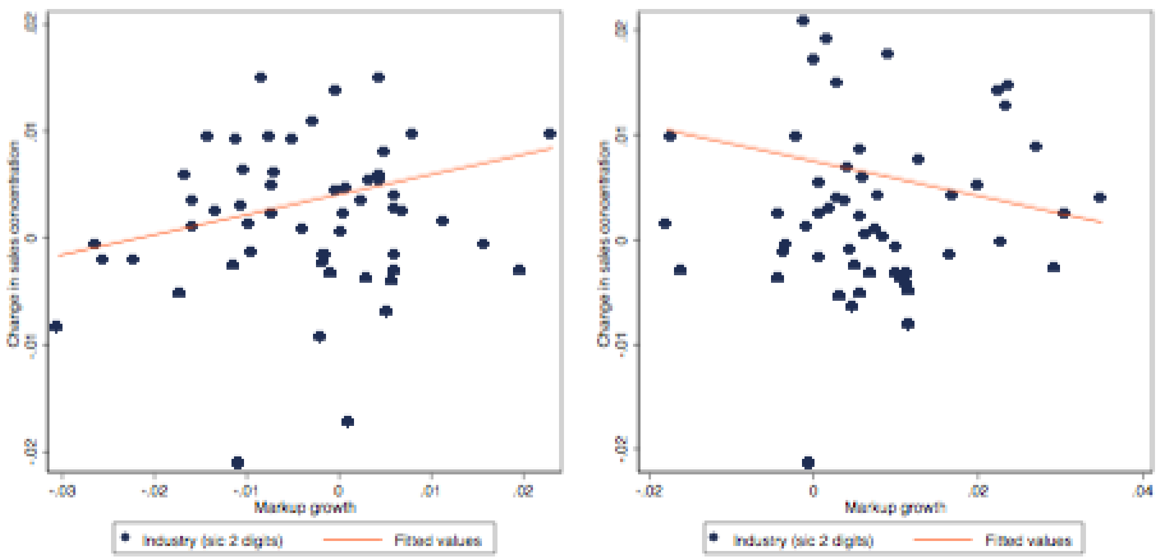

Figure 1.7: Scatter plots between average markup's growth and superstar's sales share growth across industries

Note: - The left panel represents the scatter plot between the average growth in markup and the average growth in sales share across industries between 1980-2000. The right panel represents the scatter between the average growth in markup and the average growth in sales share across industries between 2000-2014.

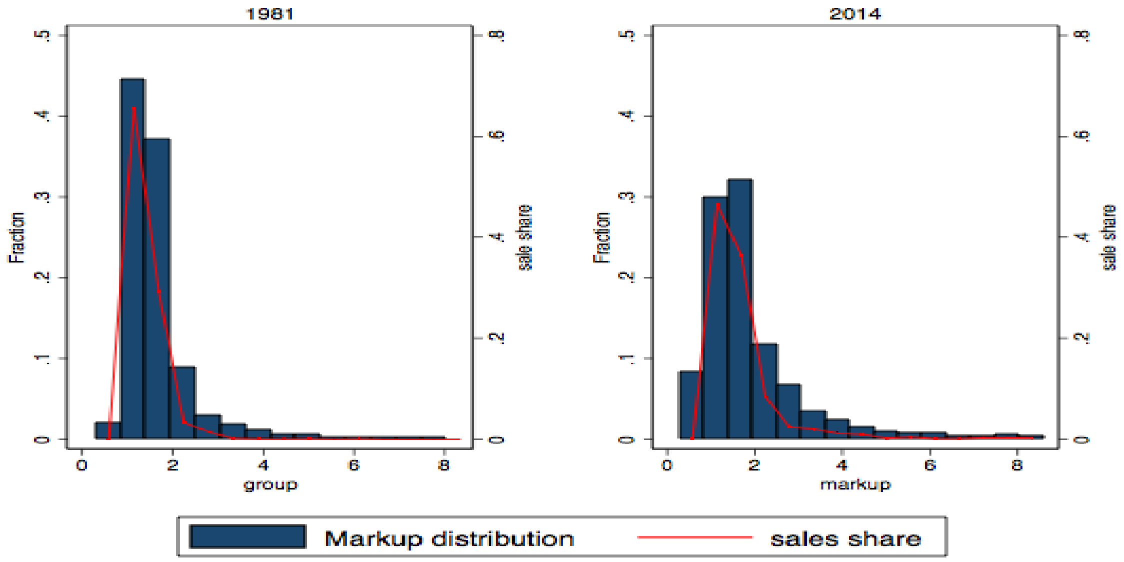

Figure 1.8: Joint distributions of markups and sales shares in 1981 and 2014

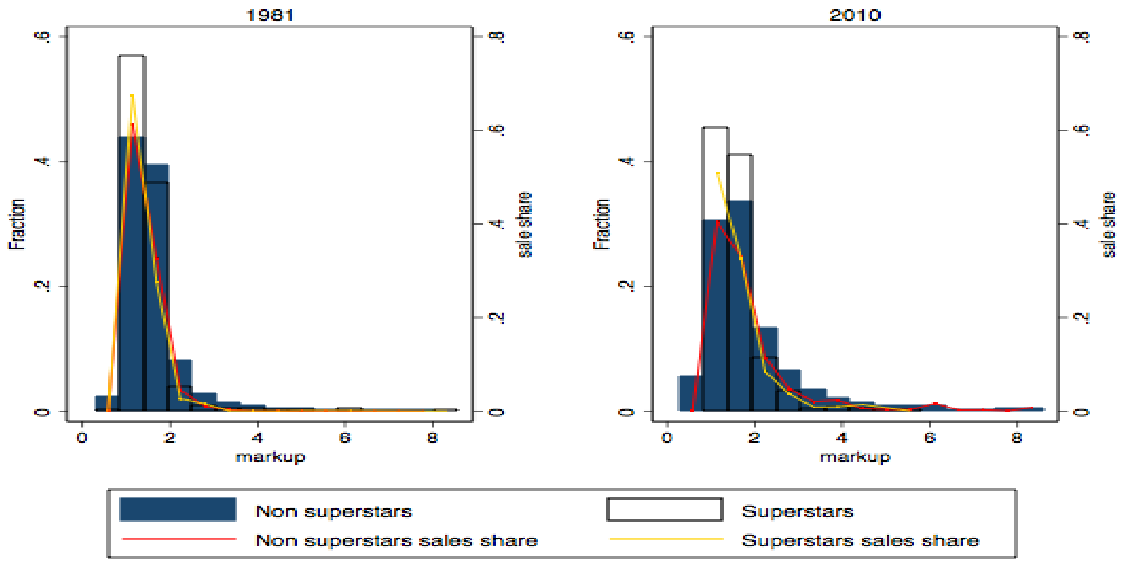

Figure 1.9: Joint distributions of markups and sales shares between superstars and non-superstars

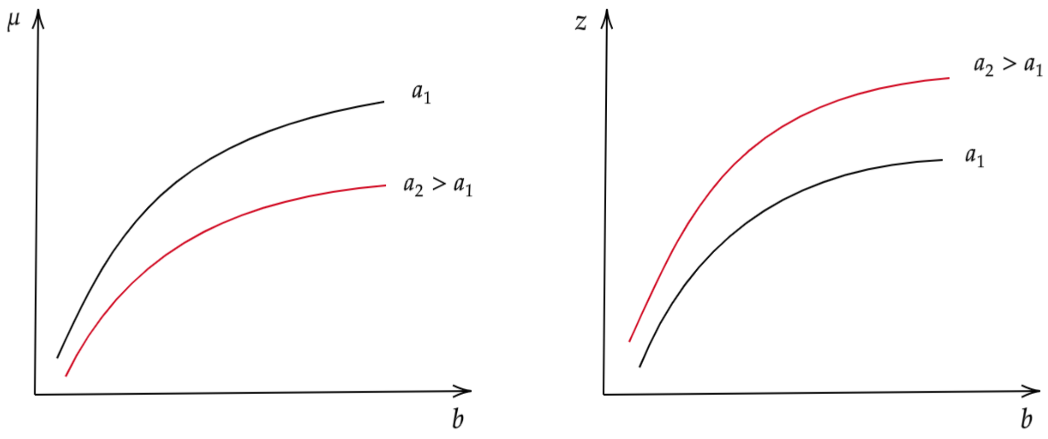

Figure 1.10: Markups and sales share policies function

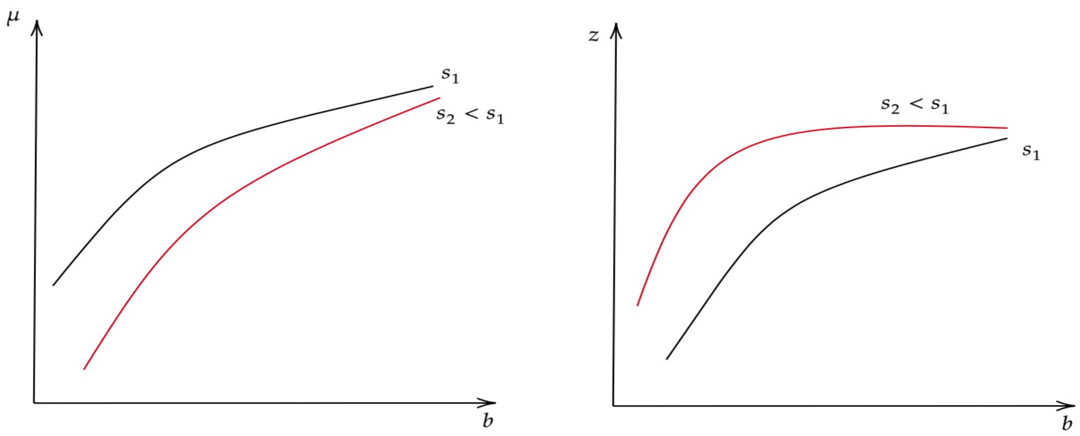

Figure 1.11: Change in s on markup and sales share for a given productivity (a)

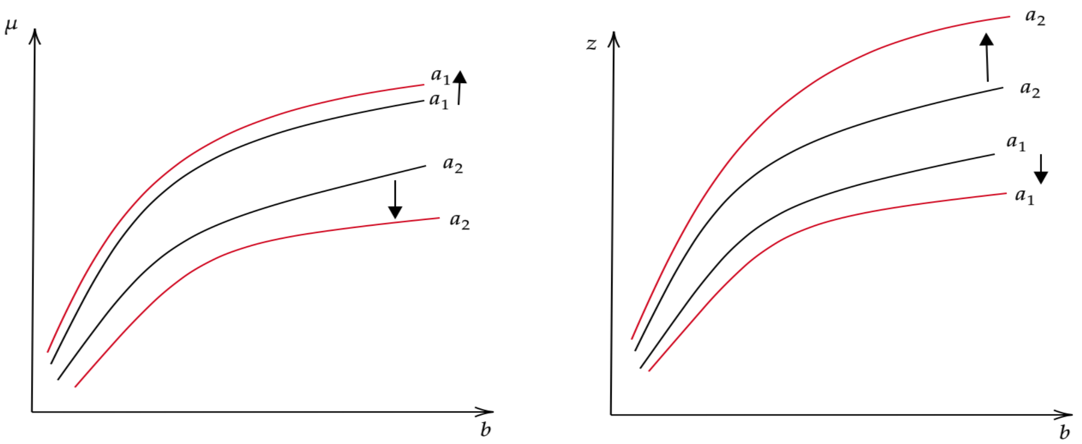

Figure 1.12: Change in σ on markup and sales share for a given productivity (a)

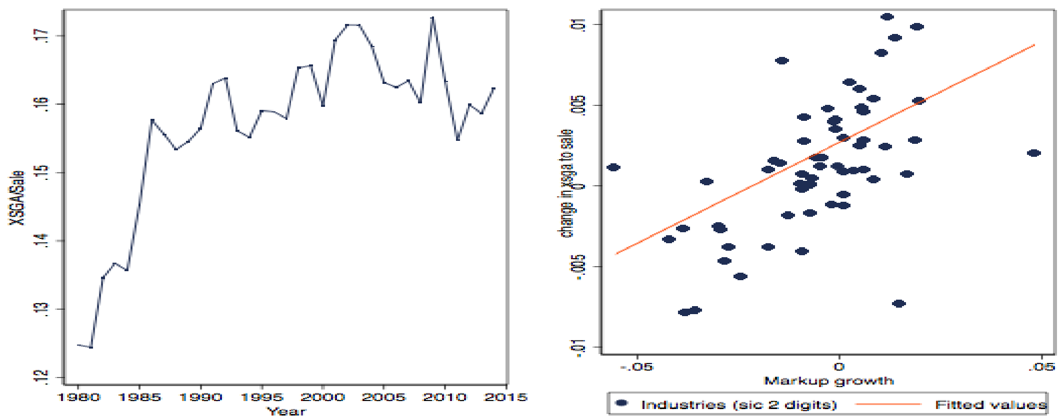

Figure 1.13: SG&A to sale ratio growth over the time and across industries across industries

Chapter 2

The Life Cycle Dynamics of Plant-level Markups*

2.1 Introduction

On average, young firms are less productive than old firms and that productivity gap partly explains the differences in firms' size and growth along the life cycle (Hopenhayn, 1992; Foster et al., 2008) but also the difference in aggregate productivity across countries (Hsieh and Klenow, 2014). Such difference in productivity along the life cycle has been explained not only by firms' characteristics such as efficiency in management practice and demand factors but also in the misallocation literature by various exogenous wedges that distort resource allocation across firms.[1] Markup's variation is an important source of friction to the input allocation because the distortion endogenously results from a firm's pricing strategy and has some implications for the average productivity. This paper analyses the implications from an endogenous variation in plants' markup for the aggregate productivity and productivity growth over the life cycle.

I start by identifying the within-age variation in markups over the firm's age. I use the De Loecker et al. (2020) framework to estimate a firm's level markup as the wedge between the input elasticity and the input share. Using the Colombia Annual Manufacturing Survey (AMS) which contains detailed information about the production process (price, quantities, value-added, materials, etc.), I estimate the firm's markup using both labor and materials inputs. The estimation strategy follows Raval (2020) where I control for labor augmented productivity shock.[2] Using both labor and materials markup estimations, I document a decline in markup dispersion over the plant's age. This

*I am grateful to Immo Schott, Baris Kaymak for helpful comments and invaluable guidance. Thanks to Raval for the dataset availability.

[1]See Hsieh and Klenow (2009, 2014); Hottman et al. (2016); Eslava and Haltiwanger (2020).

[2]Raval (2020) shows that the non-neutral labor augmented productivity explains the negative and low correlation between markup estimations using various inputs.

pattern is not driven by some selection or measurement errors correlated with age but represents a convergence of all plant's markup over the age conditional of survival.[3]

The micro foundation of the convergence in plants' markup is in the spirit of Hopenhayn (1992) where I assume a demand learning process by plants. In this environment, heterogeneous plants in both idiosyncratic productivity and demand level at entry used their current sales to build their demand (e.g. customer base accumulation). Thus, markup endogenously results from both the goods differentiation and the demand adjustment, reflecting the learning process. As a result, plants charge a low markup to make more sales and raise their future demand and profit, growing in both size and markups along the life cycle. Thus, plants grow passively through exogenous productivity shock and actively through demand accumulation. However, the increasing markup following the demand adjustment reduces the incentive to produce and harms the plant's growth. In addition, as incumbents are growing, they exogenously exit from the market and are replaced by new plants.

Markups are highly dispersed at entry due to the dispersion in both idiosyncratic productivity and idiosyncratic demand shifter. The convergence of all plants' markup results from a mismatch between the current level plants operate and the optimal level they would desire to operate given their efficiency. Over the life cycle, new plants with high idiosyncratic demand levels charge a high markup because they have fewer opportunities to build demand. As they grow, they find it costly to maintain their demand and reduce their markup with time. On the other hand, new plants with a low idiosyncratic demand level charge a low markup due to the high growth opportunities and increase their markup over their life cycle. Because the average plant at entry is less productive with low demand, it starts small and grows over the life cycle.

Markup is the only source of heterogeneity in the average revenue productivity across plants and distorts the labor allocation both within and between age. Plant's markup is positively correlated with demand shifter but negatively correlated with productivity. The variation in markup distorts the labor allocation so that plants with a high demand shifter have less incentive to produce due to the high markup and then use less labor than those with a low demand shifter. However, more efficient plants charge a low markup than less efficient plants, contributing to raising their labor input. Although the variation in markups induces an inefficiency in labor allocation, especially at a given age, the average increase in markup with age induces an additional inefficiency to labor allocation along the life cycle.

Following the misallocation literature, I analyze the implications from a within and between age variation of plant's markup to the aggregate productivity and the age-productivity growth. The life cycle dynamic in both average markups and markups dispersion shows two mitigates results. First, the age pattern from markups dispersion

[3] I exploit the covariance between those two estimations to control for a potential bias in the estimations.

suggests high friction to the labor allocation at entry, which reduces the average plant's productivity at entry and fosters the plant's productivity growth over the life cycle.[4] A fast convergence in markups results from a low variation in the within-age markups, increasing both the age-productivity growth and the aggregate productivity. Second, the age pattern in average markups increases the frictions to labor allocation along with age, which lowers the aggregate productivity.

I calibrate the model to match specific moments from Columbia manufacturing plants. I then used the model to quantify the learning effect on aggregate productivity. I show that even if the within-age variation in markups reduces the average productivity at entry and contributes to fostering productivity growth, its influence remains marginal. In addition, the absence of a learning mechanism lowers the productivity growth of the life cycle as plants don't build demand. However, the aggregate effect is higher and positive due to the absence of markup variation with age.

I used the model to test some policies implications related to a decline in the plant's exit rate. An exogenous exit rate implies that all plants have the same probability of exiting. Therefore, a reduction in exit rate increases the survival likelihood and then the incentive to build demand. As a result, plants markup converges faster, reducing the markups dispersion and increasing the productivity growth along the life cycle. However, the more the plants have the incentive to build demand, the faster is the growth in average markups with age, increasing the age dispersion in markups and reducing the aggregate productivity. Simulations show that going from our benchmark economy, the age dispersion in markups effect is stronger and increases aggregate productivity.

This paper is related to the recent literature on heterogeneity in markup. De Loecker (2011) developed a framework to estimate a firm's markup based on production information without necessarily introducing constraints and specifications on the demand side. Kaplan and al. (2020) discuss the bias and limits from such a framework in the context of data limitations and the inputs used for the estimation. Raval (2020) emphasizes the specification of the production function in the presence of labor augmenting productivity to overcome the lack of negative correlation in estimated markup when using different inputs. The current paper exploits those various strategies and the richness from the data to identify the dispersion of plant markups over the life cycle. I overcome the potential bias in the markup's estimation by assuming that the estimation errors are orthogonal for two estimated markups strategies.

This paper is also related to the literature on firm dynamics, emphasizing the size and markup dynamics as they evolve along the life cycle. Foster et al. (2008) found that, on average, new plants have lower productivity and markup than incumbent firms. Peters (2020) has written one of the first papers to build a growth model with a consistent, endogenous markup dynamic. In particular, markup results from an innovation process

[4]The fast productivity growth is driven by an increase in demand with age and the decline in markup dispersion.

and new products. He derives an endogenous dynamic markups and size distribution over the life cycle. This paper considers the behavior of the average plant along the life cycle and the heterogeneity in the markup at the entry to observe the life cycle dynamic.

This literature on firm growth (Hopenhayn, 1992; Luttmer, 2010; Eslava and Haltiwanger, 2020; Hottman et al., 2016) emphasizes the role of productivity efficiency and demand factors as the main source of plant growth along with minor contributions from wedges through a static framework. This paper builds a theory based on those main sources of growth with an emphasis on the dynamic. The main interesting feature is studying the interaction between the firm's idiosyncratic characteristic and wedge in cross-sectional and dynamic.

The misallocation literature has emphasized the role of exogenous firm-specific wedges to input allocation such as size-dependent taxes (Guner et al., 2008; Hsieh and Klenow, 2014) imperfect capital market (Midrigan and Xu, 2014) which lower aggregate productivity. Hsieh and Klenow (2014) have analyzed the contribution of those wedges to differences in product life-cycles across countries. Bento and Restuccia (2017) have shown that policy distortions that are correlated with productivity induce a misallocation of resources that lower the aggregate productivity and firm size but also distort life-cycle productivity growth. A close related paper is Peters (2020) who built a model where endogenous markups variation non correlated with a firm's productivity induce a misallocation of resources. This paper takes a different approach on two points. First, I propose a model of endogenous markup correlated with the firm's characteristics. Second, I show that friction to inputs allocation may contribute to productivity growth, especially over the life cycle.

The next sections are organized as follows. Section 2.2 presents the data and the markup estimation strategy. Section 2.2.3 discusses stylized facts on markup and size dynamic over the life cycle. Section 2.3 introduces a model of the life-cycle dynamics of markup. Finally, section 2.4 draws the implications from the markup dynamic for average productivity along the life cycle.

2.2 Markup estimation

2.2.1 Data

This paper uses the Columbia Annual Manufacturing Survey (AMS) collected by the Colombian official statistical bureau DANE.[5] AMS is a census of Colombian manufacturing establishments (plants) with more than ten employees which reports information on the plants' production process. The data reports the plant's sales, value-added, price, output and inputs (labor, capital and materials). Those data allow me to estimate plants'

[5]The Colombia AMS dataset has been recently used by Eslava and Haltiwanger (2020) to understand the source of plant's growth and by Raval (2020) to test the production approach of estimating markups

markup with less bias using the production approach (see De Loecker et al. (2020)) and follow plants over their life cycle. The overall sample covers small and large plants from 1977 to 1991.

The data report the year when plants start to operate and I use it to measure the plant's age and define plants' cohort.[6] I construct capital stock for each type of capital (lands, equipment, structures, and transportations) and measure the capital cost as the sum of each type of capital stock multiplied by the rental rate cost (user cost of capital). I measure the rental rate as the lending rate adjusted for inflation plus the depreciation rate as in Raval (2020) and construct the plant's capital stock using the perpetual inventory method. I measure raw materials spending as the total spending on raw materials adjusted for inventories by measuring the difference between the ending year and beginning year value of inventories of raw materials. Finally, I drop all observations with a negative or null value on sales, labor, and capital and observations at the top and bottom 1% of the industry's labor share and material share distribution to remove outliers in markups distribution. An industry is defined at the 2 digits SIC classification to match the Colombia industry classification (ISIC). Inputs are deflated with an appropriate input deflator to have both real and nominal values.[7] The final sample displays 611.66 plants on average per year distributed over 30 industries, with an average entry rate of 10.02 % each year while the average exit rate per year is 12.72%.

2.2.2 Plant's markup estimation

I estimate the plant's markup following the production approach (De Loecker et al., 2020). This framework is based on a specification of the production function and allows me to estimate the plant's markup without any assumptions on-demand or price-setting strategy. However, it remains sensitive to measurement errors and misspecification (Blanchard, 2020). For robustness, I estimate markup using both labor and material inputs.

Let consider a plant i from industry j who used capital (k_{it}), labor (l_{it}) and material (m_{it}) to produce a good y_{it}. Each of those inputs are respectively priced at r_t, ω_t and p^m_t while the final output is priced at p_{it}. Plants produce an output using a technology $q_{it} = y_{it}(a_{it}, k_{it}, l_{it}, m_{it})$ where a_{it} refers to the plant's efficiency. I assume that all inputs market to be competitive and there is no friction on inputs market.[8] Let assume that the plant's objective is to minimize its total spending in order to produce a level of output.

[6]That variable refers to the year a given plant starts to operate. It allows us to measure the plant's age without proxy. However, measurement errors represent around 10 % sample. I treat those errors by taking the minimum year holds by plant with the same size as is Eslava and Haltiwanger (2020).

[7]I follow the similar approach used by Raval (2020) in order to estimate plant's markup.

[8]Any additional wedges that change with the level of inputs (financial friction, capital adjustment cost, firing cost, etc.) will bias the markup estimation using from the production approach.

$$
\begin{aligned}
&\min_{k_{it},l_{it},m_{it}} \quad r_t k_{it} + \omega_t l_{it} + p_t^m m_{it} \\
&s/c \qquad y_{it}(a_{it}, k_{it}, l_{it}, m_{it}) \leq q_{it}
\end{aligned}
\tag{2.1}
$$

The constraint allows me to derive the shadow price (λ_{it}) as the cost of an additional unit of output. Plants optimally choose an input such that the cost from an additional unit of input equals the marginal gain from that additional unit. With the marginal cost expression from the optimal choice, I estimate the plant's markup as the wedge between the input share (α_{it}^x) and the input elasticity (θ_{it}^x) for a given input x.[9] Under those previous assumptions, the plant's markup estimation is independent of the input used for the estimate. However, wedges from input markets or production function misspecification may bias estimation.

$$
\mu_{it} = \frac{p_{it}}{\lambda_{it}} = \theta_{it}^x \frac{p_{it} q_{it}}{\omega_t x_{it}} = \theta_{it}^x (\alpha_{it}^x)^{-1} \qquad \forall \qquad x = l, m \tag{2.2}
$$

Raval (2020) shows that the Cobb-Douglas production function is misspecified and generates a negative correlation between markup estimations when using two different inputs (labor and materials). This misspecification is due to the absence of non-neutral labor augmented productivity.[10] I then estimate the inputs elasticities (θ_{it}^x) by considering a CES production function with non-neutral labor augmented productivity. Equation (2.3) represents the production function specification where σ is the elasticity of substitution, (b_{it}) the non-neutral productivity and (e_{it}) the efficient parameter. The share of labor and material are respectively β_l and β_m.

$$
q_{it} = e_{it} \left((1 - \beta_l - \beta_m) k_{it}^{\frac{(\sigma-1)}{\sigma}} + \beta_l (b_{it} l_{it})^{\frac{(\sigma-1)}{\sigma}} + \beta_m m_{it}^{\frac{(\sigma-1)}{\sigma}} \right)^{\frac{\sigma}{(\sigma-1)}} \tag{2.3}
$$

From this production function, the labor augmented productivity distort both the labor to material elasticities ratio and the labor to material cost ratio (equivalent to the ratio between labor share and material share) across plants. In addition, it is the source of dispersion in the labor to material cost ratio across plants (see equation 2.5).[11] Indeed, with an elasticity of substitution ($\sigma > 1$), plants with high labor augmenting productivity (b_{it}) have high labor to material cost ratio. High (b_{it}) induce plants to use more labor relative to material by reducing the cost efficiency of labor (ω_t / b_{it}). In absence of labor

[9]The plant's markup estimation is similar to the price-cost margin proxy used in the literature.

[10] Raval (2020) verify others factors such as measurement errors or cost wedges that are also consistent with a negative correlation between markup's estimations based on labor and material inputs. He shows that misspecification is more a consistent story than other potential factors.

[11]Both labor and material elasticities are defined as follows :

$$
\theta_{it}^l = \beta_l b_{it} e_{it}^{(1-1/\sigma)} \left(\frac{q_{it}}{b_{it} l_{it}} \right)^{1/\sigma} \qquad \theta_{it}^m = \beta_m e_{it}^{(1-1/\sigma)} \left(\frac{q_{it}}{m_{it}} \right)^{1/\sigma} \tag{2.4}
$$

.

augmenting productivity, plants with high (b_{it}) will have a high labor share relative to material share, leading to a negative correlation between labor and material markup.

$$\frac{\omega_t l_{it}}{p_t^m m_{it}} = \left(\frac{p_t^m}{\omega_t}\right)^{\sigma-1} \left(\frac{\beta_m}{\beta_l}\right)^{-\sigma} (b_{it})^{\sigma-1} \tag{2.5}$$

To estimate the input elasticity, I use the flexible cost shares approach proposed by Raval (2020). The approach consists of estimating both material and labor elasticities conditional on the labor augmented productivity. Because the labor augmenting productivity is the only source of variation in the labor to materials cost ratio, I split the cross-sectional distribution of labor to materials cost ratio into quantile bins (5 quantiles) and estimate inputs elasticities in each bin as the cost share.[12]

Figure (2.7) in appendix represents the cross-sectional distribution of markups using labor input (μ^l) and materials input (μ^m). Markup based on material input is more left-skewed than labor input. The two estimations are positively correlated with a low coefficient of correlation (0.3). The low correlation suggests some measurement errors or estimation bias that create noises and lead to a non-perfect correlation.

2.2.3 Stylized facts

This section presents some stylized facts on markups dispersion over the plants' age. I begin by discussing the relative importance of the within-age dispersion to the overall dispersion in markups. I then identify the pattern from such dispersion over the plant's age. Finally, I analyze the correlation from a within-age variation in markups to the size growth. All the results are presented within an industry and cohort.

I begin by documenting the relative importance of the within-age dispersion in markups. The figure (2.8) in the appendix represents the average markup (in the log) by age group. On average, the plant's markup is low at entry and increases with age. A similar dynamic in markup over the age has been documented by Peters (2020) using India manufacturing survey.

Table 2.1: Variance decomposition of markup

	l-markup	m-markup
Between age	36.825	111.387
	(0.25%)	(0.31%)
Within age	14552.685	35479.93
	(99.75%)	(99.69%)
Total	14589.51	35591.317

[12]That variation in the input elasticity arises because plants with the same labor to materials cost ratio have the same labor augmenting productivity and input elasticity

However, such variation in markup between ages only accounts for 0.3 % of the overall markups dispersion (see table 2.1). Most of the variations in markup are within age, showing that age is less informative about the markup's variation. However, factors driving the within-age dispersion in markups may also be correlated with age.

The figure (2.1) represents the dispersion in markups over the age based on the inter-quantile range regression within an industry and cohort. I used the inter-quantile range (IQR) to measure the dispersion and deal with potential outliers from the markups distribution. I use a range between the 10th and 90th deciles. Figure (2.1) shows a high dispersion in markups at the entry which declines with the plant's age for both the two measures of markups. However, the declining rate is more important during the five consecutive years following the entry and above 10 years markups' dispersion is almost constant.

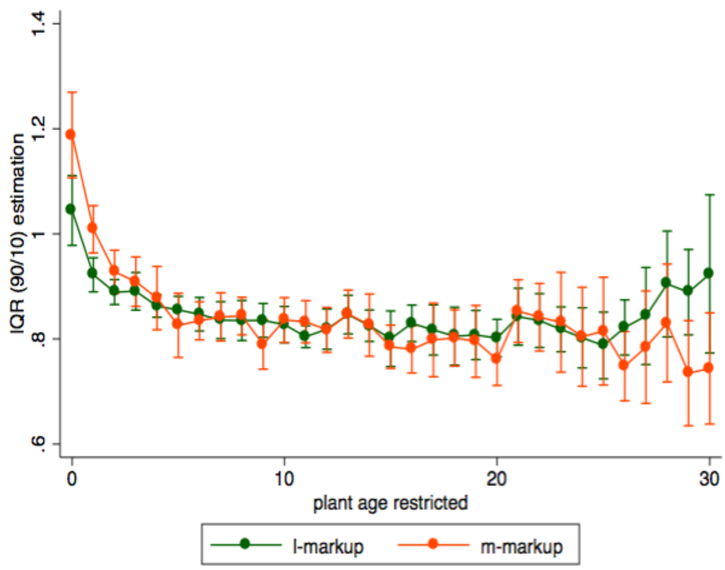

Figure 2.1: Markup dispersion over the life cycle

The decline in markups' dispersion with age is persistent even after adding control variables. Table (2.8) in the appendix shows the estimated declining rate in markup IQR per year after controlling for survival, capital to labor ratio and plant productivities.[13] Columns (1) and (3) show the declining rate in markups' dispersion between 0.386 - 0.713 % per year within a given cohort and industry. Controlling for other factors, Column (2) and (3) shows that the IQR is still declining but at a rate between 0.421-0.589 % per year. The exit probability has a mitigating effect on markup dispersion which depends on the markup we used as an estimation. While an increase in the exit probability reduces the dispersion in labor markup, it increases the dispersion in materials markup. Those results

[13] I estimate plant's productivities using the control function approach. A mean reversion of productivity over life can explain the pattern.

suggest an estimation bias in markup that may be correlated with survival. In addition, an increase in capital to labor ratio is associated with an increase in markup dispersion while an increase in plant productivity is related to a decline in markup dispersion.

Both markup estimations may also be biased by the measurement errors or estimations bias ($\mu_t^l = \mu_t \epsilon_t^l$ and $\mu_t^m = \mu_t \epsilon_t^m$) that may be correlated with age and then drive the previous pattern. Under the assumption of orthogonality between those measurement errors, the age-covariance between both labor and material markups identifies the age pattern in markups dispersion. Figure (2.9) represents the age-covariance between the labor and material markup estimations and similarly suggests that markups dispersion decreases with the plant's age. The decreasing rate in markups dispersion is high after the entry and decreases with the plant's age.

$$\sigma_{\mu,t}^2 = cov(\ln \mu_t^l, \ln \mu_t^m) \qquad \epsilon_t^m \perp \epsilon_t^l \quad \forall t \tag{2.6}$$

To understand the decline in markups dispersion considering the markups distribution, I estimate the plant's markup over the age conditional on the markup at the entry and the survival. To do so, I split the cross-sectional markups distribution at entry by industry and cohort into deciles and I follow the plants within each decile with age. Figure (2.10) in the appendix represents the average age pattern from a markup dynamic based on the markup at entry. Conditional on the markup at entry and the survival, new plants with a low markup (resp. a high markup) increase (resp. decrease) their markup over their life cycle - leading to a convergence of plant's over their life cycle. The convergence region is not a fixed point but represents a persistence in the plant's profitability in the long run. The adjustment is such that at each age, low markup plants grow faster their markup than high markup plants and that growth in markup is higher at the entry and declines with the plant's age. Thus, new plants entering the convergence region remain over their life cycle while new plants with high markup remain on the top over their life cycle.

Such decline in markups dispersion with age can be explained by various mechanisms representing a decline in the plant's profitability gap within a cohort of plants with age. Following the misallocation literature, such a pattern in markups dispersion may lower the productivity at the entry and have some implications on the productivity growth. Plants with high markup may have less incentive to produce than low markup plants. Such incentive distorts the labor allocation, reducing both the plant's size and productivity at that given age.

Table 2.2: Size and age dependance with plant's growth

	Δln(labor)				
	(1)	(2)	(3)	(4)	(5)
ln(labor)(-1)	-0.0228***	-0.0213***	-0.0297***	-0.0219***	-0.0366***
	(0.000954)	(0.000943)	(0.00160)	(0.000962)	(0.00166)
Age	0.000833**	0.00138***	0.00174***	0.000936***	0.00102***
	(0.000262)	(0.000259)	(0.000291)	(0.000265)	(0.000297)
Δ ln(μ_l)(-1)		-0.0617***	-0.0632***		
		(0.00254)	(0.00268)		
Δ ln(μ_m)(-1)				-0.0189***	-0.0204***
				(0.00172)	(0.00185)
Pr(exit=1 | u_i=0)(-1)			-0.159***		-0.302***
			(0.0254)		(0.0267)
ln(TFPQ)(-1)			-0.0000205		-0.00000641
			(0.0000929)		(0.0000942)
Constant	0.0674***	0.0566***	0.0944***	0.0634***	0.139***
	(0.00450)	(0.00446)	(0.0103)	(0.00455)	(0.0106)
Observations	82335	78512	72771	78666	71919
R^2	0.0124	0.0192	0.0220	0.0132	0.0154
Indus FE	√	√	√	√	√
Cohort FE	√	√	√	√	√

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Because the convergence is driven by the dynamic in the plant's markup within a given cohort, I then analyze how such variation in markups within an age is correlated with the expected size growth. The table (2.2) reports the correlation between plant growth with plant size and age after controlling for the change in markup. Column (1) shows the size and age dependence with the plant's growth. Small plants grow faster than large plants such as 1 % increases in plant size reduce the plant's growth by 2.28 %. Furthermore, plant's growth increase by 0.0833 % by year within a given cohort and industry.

Furthermore, conditional on age, a 1 % increase in labor-markup (resp. material-markup) growth reduces the plant's growth by 6.17 % (resp. 1.89 %). On the other hand, conditional on markup growth, plants grow on average by 0.138 % per year while conditional on age, plant's growth decreases by 2.28 % following an increase in firm's size by 1%. As the average markup increase with the plant's age, it contributes to lower the plant's growth with age. In addition, a 1 % increase in exit probability reduces the plant's growth by 15.9 %. This result is consistent with the selection as a source of growth. As the exit rate increases, low productive firms exit and are replaced by high productive firms. The result is consistent even if I consider materials markup.

2.3 The model

In this section, I build a model to explain the life cycle pattern of a plant's markup conditional on markup at entry. I use the model to assess implications from that markup dynamic on the life cycle of plants productivity. The model is built from the Hopenhayn (1992) industry dynamic framework. Heterogeneous plants and a representative consumer populate the economy. Plants value the demand as an asset and learn from their demand over the life. However, the learning is costly because it requires lower markup and profit loss but generates a return from demand adjustment. Finally, plants are heterogeneous in both idiosyncratic productivity and idiosyncratic demand level.[14]

2.3.1 Set-up

Preferences

There is a representative consumer at a given period with a CES utility who consumes a set of differentiated goods i. The relative importance of each differentiated goods in the consumer basket (b_i) is externally formed by plants. The relative importance refers a demand shifter (such as customer base or appeal) formed by plants. θ measure the sensitivity of the demand to the shifters and ($\theta = 0$) imply no effect from demand shifter. The consumer supply inelastic labor at wage ω, get dividends from plants. There is no saving problem. The consumer problem consist to allocate his income to buy a set of differentiated goods and it is describe as follows :

$$\begin{aligned}
&\max_{(c_i)_i} C = \left[\int_i (c_i b_i^{-\theta})^{1-1/\eta} di\right]^{1/(1-1/\eta)} \\
&s/c \\
&\int_i p_i c_i di = \omega + D = pY
\end{aligned} \tag{2.7}$$

where c_i is the consumption of the differentiated goods i and p_i its price. ω is the wage and D is the dividends get from plants. η is the elasticity of substitution across the differentiated goods and we assume ($\eta > 1; \theta < 0$ and $\theta(1-\eta) < 1$) to insure a decrease in the return to scale from demand formation. The consumer problem provides the demand for each differentiated goods and the price index (p) is normalized to 1.

$$c_i = p_i^{-\eta} b_i^{\theta(1-\eta)} \frac{Y}{p} \qquad\qquad p = \left(\int (p_i b_i^{\theta})^{1-\eta} di\right)^{1/(1-\eta)} \tag{2.8}$$

[14]Recent papers on firm's growth have pointed both idiosyncratic productivity and demand shifter as a source of growth (Hottman et al., 2016; Eslava and Haltiwanger, 2020).

Production

At each period t, incumbent plants produce differentiated goods y_{it} and sell at the relative price p_{it}. A plant i holds a demand shifter (b_{it}) as a stock (for example customer base) and draws a productivity shock e_{it} from an $AR(1)$ process. The demand shifter follows a law of motion given by the equation (2.9).[15] At each period, plants lose a fraction $(1-\delta)$ of their current demand but use their sales to build their future demand shifter.

$$b_{it+1} = (1-\delta)b_{it} + \delta p_{it} y_{it} \tag{2.9}$$

An Incumbent plant i at the time t chooses the level of labor n_{it}, output y_{it}, price p_{it} and next level of demand shifter b_{it+1} that would maximise his discounted expected profit. Labor is hired at the cost ω_t. At the end of the period, the plant decides whether to exit or not from the market based on its current value. Exit arises exogenously with a probability s. I assume a symmetric equilibrium which allows me to write the problem over two states (e, b). Given the states variables (e, b), an incumbent plant maximises all his expected dividends stream that is recursively formulate as follows:

$$\begin{aligned} v(e,b) &= \max_{d,p,y,b'} \{py - \omega n + (1-s)\beta E_e(v(e',b'))\} \\ y &= e n^{\alpha} \\ p &= y^{-1/\eta} b^{-\theta(1-1/\eta)} Y^{1/\eta} \\ b' &= (1-\delta)b + \delta py \end{aligned} \tag{2.10}$$

where $v(e, b)$ is the plant value. Plants have an incentive to build demand over their life because it increases their future demand and profits. However, the decrease in return to scale limits the accumulation by lowering the return from the learning. Plants' growth in this framework arises both through a change in productivity and demand, which are consistent with the main sources of plant growth describe by (Eslava and Haltiwanger, 2020).

At the optimum, the plant will charge a markup μ over its marginal cost. The markup charged by plants reflects the differentiation in goods and services and the cost from the learning process. Setting a low price (low markup) increases the sales, future demand and profits. The rise in the future demand and profits allows plants to charge a high price (high markup) in the future. Let define by $\tilde{\beta} = (1-s)\beta$ the adjusted discount factor to the survival rate and $\bar{\mu}$ the monopolistic markup (in absence of demand learning). The optimal markup internalizes cost and the expected marginal gain from increasing the demand through the learning.

$$\mu^{-1} - \bar{\mu}^{-1} = \tilde{\beta} E_e \left((1-\delta)(\mu'^{-1} - \bar{\mu}^{-1}) - (\theta\delta\bar{\mu}^{-1}) \frac{p'y'}{b'} \mu'^{-1} \right) \tag{2.11}$$

[15] The difference with the model describes in chapter 1 is the investment decision. In chapter 1, firms invest a share of their sales to build their demand while in chapter 2, the demand shifter follows a law of motion.

Equation (2.11) shows the plant trade-off is to set a low price to increase their sales and build the demand or set a high price and increase his current short-run profit. Thus, the left hand of the equation (2.11) represents the marginal cost from an additional unit of demand in terms of profit loss and the right hand of the equation represents the marginal gain from an additional unit of demand representing the long-run profit.

An increase in the long-run profit inducing by the demand shift raises the future demand and markup. Besides, the shift from the demand raises the future sales, inducing a positive co-movement between sales and markup along the life cycle. However, equation (2.11) shows that the more productive plants have a high expected return from building their demand than the less efficient because they can make more sales. Therefore, they charge a low markup relative to the fewer productivity plants to benefit more from the growth opportunities.

Entry

Entry to the market requires to pay a fixed cost (k_e) per unit of labor. Paid that fixed cost allows plants to draw an initial productivity shock (e_0) from a distribution and an initial idosyncratic demand level (b_0) respectively from two independents distribution $\tau(e_0)$ and $\chi(b_0)$. The distribution $\chi(b_0)$ is derived from the steady-state distribution. After drawing a new productivity shock and stock, the new plant becomes an incumbent plant and start to produce. The recursive formulation of the new plant problem is as follows :

$$\begin{aligned}
&v_{new} = -\omega k_e + \beta \int_{b_0,e_0} \left(\max_{d,p,y,b'} \{py - \omega n + \tilde{\beta} E_{e_0}(v(e', b'))\} \right) \chi(b_0)\tau(e_0) \\
&y = e_0 n^{\alpha} \\
&p = y^{-1/\eta} b_0^{-\theta(1-1/\eta)} Y^{1/\eta} \\
&b' = (1-\delta) b_0 + \delta p y
\end{aligned} \tag{2.12}$$

The free entry condition implies that plants will enter until the expected gain from entry equals to the cost of entry. That free entry condition allows characterizing the plant distribution over the state variables. Given the productivity transition matrix and the policy function, I define the transition probability across the current and future state variables. With a mass of new plants m_t, the law of motion of the plants distribution is defined by :

$$\Gamma_{t+1}(e', b') = \sum_{e,b} \Psi(e', b'|e, b)\Gamma_t(e, b) + m_t \chi(b_0)\tau(e_0) \tag{2.13}$$

2.3.2 Competitive equilibrium

A stationary equilibrium with entry which consists to a policy function $b'(e, b)$, $y(e, b)$, $n(e, b)$, $p(e, b)$, $d(e, b)$ and positive number Y, ω, m such that : *i)* $b'(e, b)$, $y(e, b)$, $n(e, b)$,

$p(e,b)$, $d(e,b)$ solve the plant problem given the Y and ω. *ii)* $c(e,b)$ solve the consumer problem given Y and ω. *iii)* free entry condition : $v_{new} = 0$. *iv)* labor market clear : $\int n(e,b)d\Gamma(e,b) + mk_e = 1$. *v)* output market clear : $y(e,b) = c(e,b) \quad \forall a, b$ and *vi)* The stationary distribution of plants $\Gamma(e,b)$ solves the following equation :

$$\Gamma(e', b') = \sum_{(a,b)} \Psi(e', b'|e, b)\Gamma(e,b) + m\chi(b_0)\tau(e_0)$$

2.3.3 Growth and markup's convergence

This section shows how an active demand learning mechanism explains the convergence of markup with age. I also characterize the long-run markup over and demand shifter dynamic. The decrease in the return to scale in a demand accumulation ensures a steady level in the shifter where plants want to operate in the long run. Let define by $b_{ss}(e)$ and $\mu_{ss}(e)$ respectively the steady-state level of demand shifter for a level of productivity e. The following proposition characterizes the steady-state equilibrium :

Proposition 1: *In the long run, plant's markup converge to $\mu_{ss}(e)$ such that :*

- $\forall e$ *,* $\forall t$ $b_t < b_{ss}(e)$*,* $b_t < b_{t+1} < b_{ss}(e) \Rightarrow \mu_t < \mu_{t+1} < \mu_{ss}(e)$
- $\forall e$*,* $\forall t$ $b_t > b_{ss}(e)$*,* $b_{ss}(e) < b_{t+1} < b_t \Rightarrow \mu_{ss}(e) < \mu_{t+1} < \mu_t$

where l;

$$b_{ss}(e) = \left(\frac{\alpha}{\omega} Y^{1/\eta} e^{1/\alpha} \mu_{ss}^{-1}\right)^{\frac{1}{\frac{1}{\alpha} - (1-\theta)(1-\frac{1}{\eta})}} \tag{2.14}$$

$$\mu_{ss}(e) = \mu_{ss} = \bar{\mu} \left(\frac{1 - \tilde{\beta}(1 - \delta - \theta\delta\bar{\mu}^{-1})}{1 - \tilde{\beta}(1-\delta)}\right) \tag{2.15}$$

Proposition (1) shows the convergence of the plant-level markup over the life cycle. Markup is highly dispersed at entry because of the plant's heterogeneity in demand shifter and productivity at the entry. New plants with a low (resp. high) demand shifter would have more (less) growth opportunities and charge a low (resp. high) markup to build their demand. Thus, it results in a wide dispersion in markups among new plants at the entry as the equilibrium is stationary. Besides, the plant's heterogeneity in productivity at entry induces an additional dispersion in markup as the more efficient plants charge low markup relative to the less productive plants to benefit more from growth opportunities.

The decline in markups dispersion over age results from a mismatch between the current state of a plant operator and the optimal level they desire to operate for a given level of productivity. By lowering (resp. increasing) their markup, new plants with a low (resp. high) demand shift increase (decrease) their sales - raising (reducing) their future demand, profit, and markup in such a way that is consistent with their productivity. The decline in the return to scale reduces the return from investing in habit stock, inducing a convergence in all plant's markup. The convergence in the plant's markup characterizes the decrease in markup dispersion over the plant's age. At the steady-state, there is still a dispersion in markup induced by the uncertainty shock in the plant's productivity which causes a persistence in markup dispersion in the long run.

Proposition (1) also characterizes the equilibrium at the steady-state. This steady-state markup characterizes the long-run dynamic in markup and it is lower than the monopolistic markup $\bar{\mu}$.[16] It depends on the elasticity of substitution across goods, the exit probability, the depreciation rate and the elasticity of habit stock to the respect of the demand (see equation 2.15). An increase in the substitution among goods reduces the long-run markup by lowering the monopolistic markup and return from building the demand. In addition, the high elasticity from the idiosyncratic demand shock lowers the steady-state markup.

Proposition 2: *Plant's growth is defined by:*

$$\Delta \ln(n) = (\frac{\gamma - 1}{\alpha})\Delta \ln(e) - \gamma\theta(1 - \frac{1}{\eta})\Delta \ln(b) - \gamma\Delta \ln(\mu) \tag{2.16}$$

$$\gamma = \frac{1}{\frac{1}{\alpha} - (1 - \frac{1}{\eta})} > 0$$

Two mechanisms drive the plant's growth: an exogenous growth through productivity and an endogenous growth through demand accumulation. The state variables and equation characterize the plant's size and equation (2.16) presents the dependence between size and the expected plant's growth. Small firms are less productive with low demand. They grow faster because they are more likely to be more productive in the future and have more opportunities to build demand (left hand of figure 2.2). As they grow, their incentive to build their demand declines due to the return to scale in demand and they grow less as they age. It suggests a negative correlation between size and growth along the life cycle.

However, the demand adjustment is followed by an adjustment in markup over the life cycle. By lowering their markup, plants learn about their demand and increase their future demand and markup based on their efficiency. Equation (2.16) shows that controlling for plant's size, plant's growth and markup's growth are negatively correlated. The result is consistent with table (2.2) suggesting that conditional on survival and size, markup growth is friction for plants' growth.

[16] It is low because plants at the steady-state still need to keep an amount of demand shifter maintain their growth. Therefore, they charge a low markup relative to the monopolistic one.

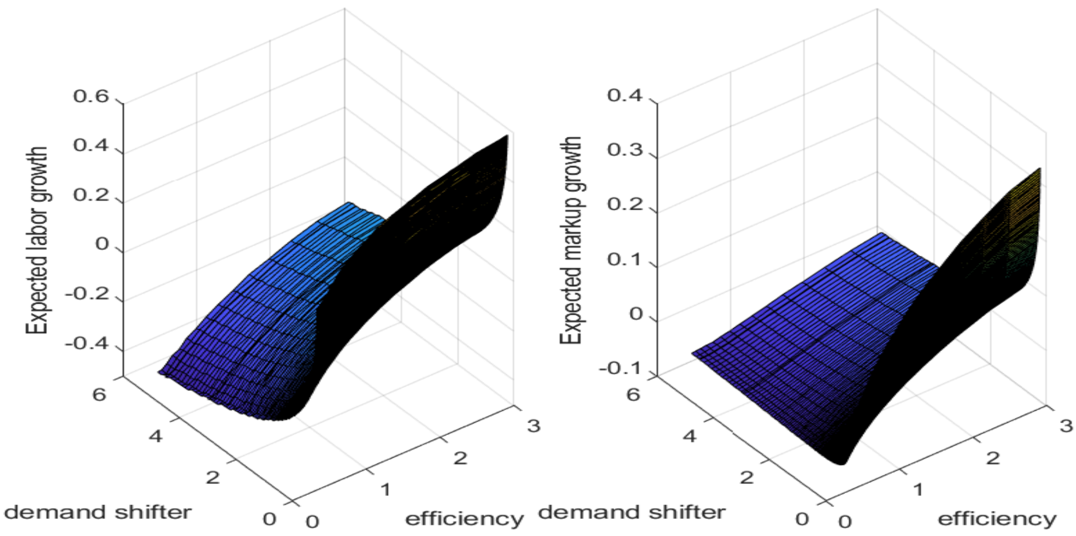

Figure 2.2: Expected growth in markup and size

2.3.4 Life cycle of average productivity

This section analyzes the implications from an age variation in markups to the average productivity over the life cycle. The misallocation literature emphasizes the role of correlated or uncorrelated distortions with the plant's productivity which are sources of heterogeneity in marginal productivity across plants. They affect the allocative efficiency of inputs and contribute to lower aggregate productivity.[17] The productivity effect is stronger with the correlation between idiosyncratic productivity and distortions.

Following Hsieh and Klenow (2009), the marginal revenue product of labor is measured using the average revenue product of labor. Equation (2.17) shows that markups variation is the only source of dispersion in the average revenue product of labor. Markup is endogenously correlated with both plants' productivity and the demand shifter with opposite sign of correlation. Plants' markup increases with demand shifter and decreases productivity as the most productive plant lowers its markup to build more demand.

$$arpl = \frac{py}{n} \propto \mu \tag{2.17}$$

To analyze the implications from such variation in markups to the average productivity growth with age, I decompose the final output with age. I denote the age by t. Equation (2.18) represents the decomposition of the final output over the age where Γ_t is the plants distribution at age t. The final output is a CES function of the aggregate

[17] Hsieh and Klenow (2014) shows that those distortions induce friction to inputs allocation and distort the average productivity over the firm's life cycle.

output at each age (y_t) with an elasticity of substitution η. As the equilibrium is stationary, the aggregate output at each age is only characterized by the plants distribution which captures the growth dynamic. I define both average markup at a given age ($\tilde{\mu_t}^{-1}$) and at the aggregate ($\tilde{\mu}^{-1}$) using the sales-weighted harmonic.

$$Y = \left(\sum_{t=0}^{\infty} \int (yb^{-\theta})^{(1-\frac{1}{\eta})} d\Gamma_t(e,b) \right)^{\frac{1}{(1-\frac{1}{\eta})}} = \left(\sum_{t=0}^{\infty} y_t^{(1-\frac{1}{\eta})} \right)^{\frac{1}{(1-\frac{1}{\eta})}} \tag{2.18}$$

$$\begin{aligned} \tilde{\mu_t}^{-1} &= \int \frac{py}{p_t y_t} \mu^{-1} d\Gamma_t(e,b) \\ \tilde{\mu}^{-1} &= \int \frac{py}{Y} \mu^{-1} d\Gamma(e,b) \end{aligned} \tag{2.19}$$

Let defines $\hat{e} = e^{(1-\frac{1}{\eta})}$; $\hat{\alpha} = \alpha(1-\frac{1}{\eta})$; $\hat{\theta} = \theta(1-\frac{1}{\eta})$ and N_t as the total labor at the age t. I characterize both the within age (γ_{wt}) and between age (γ_{bt}) allocation of labor. Those allocations are respectively distorted by the within age variation in markup and the age-variation in markup. The within-age allocation of labor depends on the distribution of productivity, demand shifter and the endogenous variation in markups.

$$\begin{aligned} \frac{n}{N_t} &= \gamma_{wt} = \frac{\left(\hat{e}b^{-\hat{\theta}}\mu^{-1}\right)^{\frac{1}{1-\hat{\alpha}}}}{\int \left(\hat{e}b^{-\hat{\theta}}\mu^{-1}\right)^{\frac{1}{1-\hat{\alpha}}} d\Gamma_t(e,b)} = \left(\frac{N_t^{\alpha}}{y_t} e b^{-\theta}\right)^{\rho} \left(\frac{\mu^{-1}}{\tilde{\mu_t}^{-1}}\right)^{\frac{1}{1-\hat{\alpha}}} \qquad \rho = \frac{(1-\frac{1}{\eta})}{1-\hat{\alpha}} \\ \frac{N_t}{N} &= \gamma_{bt} = \frac{\int \left(\hat{e}b^{-\hat{\theta}}\mu^{-1}\right)^{\frac{1}{1-\hat{\alpha}}} d\Gamma_t(e,b)}{\int \left(\hat{e}b^{-\hat{\theta}}\mu^{-1}\right)^{\frac{1}{1-\hat{\alpha}}} d\Gamma(e,b)} = \left(\frac{y_t}{N_t^{\alpha}} \frac{N^{\alpha}}{y}\right)^{\rho} \left(\frac{\tilde{\mu_t}^{-1}}{\tilde{\mu}^{-1}}\right)^{\frac{1}{1-\hat{\alpha}}} \end{aligned} \tag{2.20}$$

The within-age share of labor (γ_{wt}) increases with both productivity and demand shifter but decreases with markup. In addition, the correlations between markup, productivity and demand shifter contribute to distorting the allocative efficiency of labor input. As markup decrease with plant productivity, conditional on demand shifter, less productive plants have less incentive to produce and become smaller than high productive plants. However, markup increases with the demand shifter inducing plants with a high demand to have less incentive to produce. Therefore, the plant's distribution over the state variables characterizes the age-markup distribution and then allocative inefficiency in labor at a given age.

The labor allocation depends on both the average productivity and average markup along with plant age. Because the average plant at entry is less productive with a low demand shifter, new plants are small even if they charge a low markup to build their demand. As they grow, they become more productive, learn about their demand and increases their markup. Therefore, their size increases despite the rise in their markup, which reduces their size.

Using the labor allocation and the final output both at each age and at the aggregate level, I characterize both the average productivity over the age (A_t) and the aggregate productivity (A).[18] The average productivity at the age t is a geometric average of plant's efficiency, the demand shifter and the variation in plant's markup. The aggregate productivity is characterized by the life cycle of plant productivity and the between age dispersion observed in markup.

$$
\begin{aligned}
A_t &= \frac{y_t}{N_t^{\alpha}} = \left[\int \left(eb^{-\theta} \left(\frac{\mu^{-1}}{\tilde{\mu}_t^{-1}} \right)^{\alpha} \right)^{\rho} d\Gamma_t(e,b) \right]^{\frac{1}{\rho}} \\
A &= \frac{y}{N^{\alpha}} = \left[\sum_{t=0}^{\infty} \left(A_t \left(\frac{\tilde{\mu}_t^{-1}}{\tilde{\mu}^{-1}} \right)^{\alpha} \right)^{\rho} \right]^{\frac{1}{\rho}}
\end{aligned}
\tag{2.21}
$$

On the dynamic side, the change in average productivity over age is driven by the dynamic in the plants distribution. Change in plants distribution characterizes the dynamic in markups distribution along the life cycle, driven by the selection, an exogenous growth in plan's productivity and the accumulation in demand shifter. The dynamic in demand shifter is the main source of convergence in markups and endogenously shapes the plant markups distribution.

The log-linear approximation of productivity at the aggregate and over the life cycle are given by equation (2.22). Dispersion in markups contributes to lower the average productivity at a given age. Thus, at the entry, the high dispersion in markups induces an allocative inefficiency in labor input and lowers the average productivity at entry. Over the age, the decrease in markups dispersion contributes to increasing productivity growth. The result suggests that new plants face high friction to inputs allocation induced by the heterogeneity in markups. In addition, the friction to input allocation diminishes with age and contributes to increasing the average productivity over the life cycle.

$$
\begin{aligned}
a_t &= log(A_t) \simeq \bar{e}_t - \theta \bar{b}_t + \frac{\rho}{2} \left(\sigma_{e,t} + \theta^2 \sigma_{b,t} - \alpha \sigma_{\mu,t} \right) \\
a &= log(A) \simeq \overline{a_t} + \frac{\rho}{2} \left(\sigma_{a_t} - \alpha \sigma_{\mu_t} \right)
\end{aligned}
\tag{2.22}
$$

where $\bar{e}_t$,$\bar{b}_t$ are respectively the average log productivity and the average log demand stock across all plants. The aggregate productivity is characterized by both productivity and markup over the life cycle. However, the age variation in markups reduces the aggregate productivity. Those two results show a mitigate effect from the overall variation in

[18] Note that both average productivity at the age t and the aggregate productivity can be represented in terms of within and between age allocation of labor input. I used that allocation expression to capture the effect from the within-age dispersion of aggregate markup when efficiently allocating a markup over the age.

markups on aggregate productivity. Indeed, while the within age dispersion in markups improves the productivity growth over the life cycle and then the aggregate productivity, the age variation in markups lowers the aggregate productivity growth. The net effect on aggregate productivity is given by the elasticities ρ and α.

2.4 Quantitative analysis

2.4.1 Calibration

The model period is a year. I pre-set some parameters related to preferences, and I calibrated other parameters to match the key characteristics from the firm markups and size distribution. I set the discount factor to $\beta = 0.96$ which corresponds to a 4% yearly interest rate. I set both the labor elasticity and the demand shifter depreciation respectively to $\alpha = 0.7$ and $\delta = 0.67$ following Foster et al. (2016). The exit rate is exogenously set to $s = 10\%$ and the remaining parameters are jointly calibrated to match some selected moments in data.[19]

Table 2.3: Jointly identify parameters

	Values	Moments	Target	Model
ke	3.6788	Average log markup	0.2092	0.2091
η	2.1899	AR(1) log sales	0.9925	0.9906
ρ_a	0.8893	SD markup at entry	0.1083	0.1128
σ_a	0.1907	Correlation of markup and sales	0.1030	0.1039
θ	-0.7242	Dispersion at age 20	0.0686	0.0652

The estimated parameters are shown in the Table (2.3). The productive efficiency parameters (ρ, σ) are chosen to match both the autocorrelation of sales and the dispersion in markups at entry. Indeed, for a fixed value of δ, ρ tracks the dependence between a plant's size and growth. σ is the source of heterogeneity in both productivity and demand shifter. Finally, I choose the price demand elasticity (η) to match the long-run dispersion in markups. Indeed, the price demand elasticity drives the upper bound of the markups distribution and the return from building the demand. I finally choose θ to match the correlation between markups and sales. Indeed, θ drives the return from the demand learning and then the markup growth. A high value of θ suggests a high return from a demand adjustment, induces a fast growth in markup over time and increases the correlation between markups and sales.

[19]The exit rate is set to be consistent with the average exit rate in the sample.

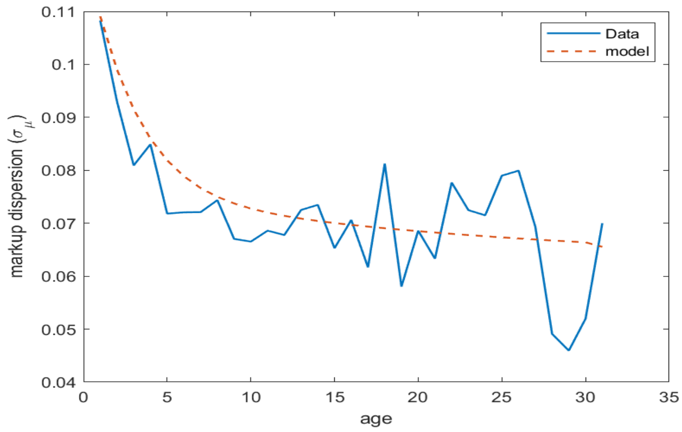

Figure 2.3: Matching between the model and the data

To pin down the entry cost, I target the average log markup. The average log markup is the sales-weighted harmonic mean of firms' markup (in log). The value is about 0.2092, which represents an average markup of 1.232. A fall in the wages increases the firm's value and then the entry cost from the entry condition. Change in the wages is more likely to increase the distortion in the firm's marginal cost and the outside option. Therefore, I use the change in the wage to target the log average markup and then identify the entry cost. Figure (2.3) shows both the age pattern of markup dispersion from both the model and the data. The model fits the markup dispersion pattern of data.

2.4.2 Productivity decomposition

I use the model to quantify the productivity losses over the age induces by the within-age variation in markups and the demand accumulation. To quantify the productivity losses over the life cycle induces by the markup variation, I simply compute the average productivity at each age and omitting the variation in markup.[20] I complement this analysis by simulating an economy with no demand learning ($\theta = 0$). In this case, there is no demand accumulation and markup variation.

The left-hand panel of the figure (2.4) shows the cumulative losses in productivity induced by the variation in markups. As we can observe, the cumulative losses resulting from markups' variation is low (less than 1%).[21] The right-hand panel of the figure (2.4) decomposes the source of markup variation. The green region represents the variation induced by the demand shifter and the white region the variation induced by productivity.

[20] This approach can be seen as a back of the envelop because we keep the keep the effect from demand shift but remove the markups variation.

[21] As reported on the figure (2.3), the within-age dispersion in markups remains low with a labor elasticity of 0.67, the log-linear decomposition show a low contribution from markup dispersion to the aggregate productivity. With a high within-age dispersion in markups, the productivity losses would be high.

Most of the variations in markups are driven by the demand shifter which lowers the incentive to produce and generates an inefficiency.

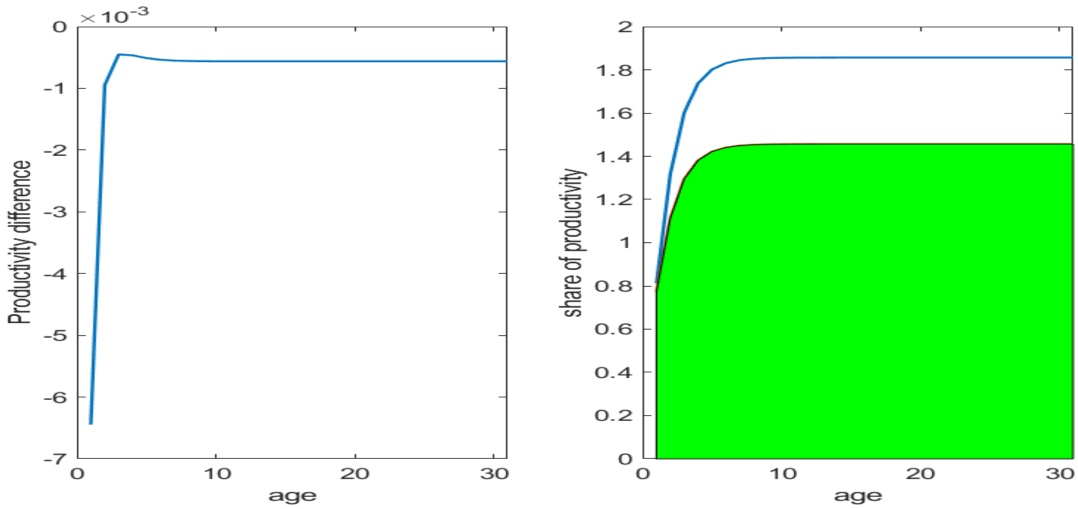

Figure 2.4: Productivity decomposition

With $\theta = 0$, there is no dynamic markup because there is no return from demand learning. The only source of a plant's markup is the differentiation in goods and services driven by the degree of substitution. With a constant elasticity of substitution, all plants charge the same markup. In this case, both markup and the idiosyncratic demand don't contribute to the change in the average productivity over the life cycle. Plant productivity growth is driven exogenously.

In this case, the plants distribution over productivity matters. The average age productivity declines due to the absence of demand accumulation which fosters productivity growth. Thus plants would grow less over their life. In addition, the absence of markups variation over the age would improve the labor allocation and increased productivity growth. Table (2.4) presents the productivity gain at the aggregate level in the absence of demand learning.

As θ is declining, the return from adjusting the demand falls (see equation 2.11), reducing the plant's incentive to learn from its demand. Therefore, the dispersion in markups at the entry declines and the rate at which markups decline also falls. Those two variations of markup reduce the within-age dispersion in markups and the age variation of markups, reducing the overall dispersion in markups (see table 2.4). Although the markups dispersion is declining, the average markup increases because plants have less incentive to learn about their demand and become close to the monopoly markup.

With the decline in markups dispersion driven by a fall in θ, the within-age input allocation becomes more specific to the plant's productivity. As a result, there is less friction to inputs allocation. The decrease in markup dispersion at entry and the low convergence rate of markups dispersion with age increase the average productivity at entry and reduce the average age-productivity growth over the life cycle. In addition, the low increase in average markup due to the low return from the learning reduces the markup variation along with age and then the allocative inefficiency of labor of the life cycle. Thus, it contributes to increasing aggregate productivity. Those two mechanisms increase aggregate productivity, especially through the decline in markup variation, which reduces the friction to labor allocation.

Table 2.4: Change in θ

		Change in θ		
Value	$\theta = -0.7242$	-1%	-2%	$\theta = 0$
VAR	0.0090	0.0086	0.0081	0.0000
Within VAR	0.0068	0.0065	0.0063	0.0000
Between VAR	0.0022	0.0020	0.0018	0.0000
Avr markup	1.2331	1.2405	1.2467	1.6394
TFP	1.5858	2.3461	3.2458	57.2930
No Within	1.5891	2.3514	3.2537	57.2930
No Between	1.5861	2.3466	3.2466	57.2930

2.4.3 Fall in exit rate

In this section, I study policy implications resulting from a change in the exit rate. I study those implications on markups dispersion and productivity both over the life cycle and the aggregate level. Exit rate is assumed exogenous and independent of the firm's characteristics. This allows us to not consider the creative destruction process following the shock.

Figure (2.5) shows both the average markup and the markup dispersion following a fall in exit rate. A fall in the exit rate increases the survival rate and then the growth opportunities. Plants with high growth opportunities build more demand shifters at the cost of markup adjustment. Thus, the fast accumulation is followed by a high adjustment in markup, inducing a fast convergence in markup. As described by the left-hand panel of the figure (2.5), the fast convergence is followed by a fast decline in markup dispersion. Furthermore, because the rapid growth by firms is associated with an increase in markup, the average markup grows faster over the life cycle (right-hand panel of figure 2.5). Therefore, a decline in the exit rate foster growth through learning and leads to a fast convergence in markup. The fast convergence increases the average age-productivity growth through demand accumulation and allocative inefficiency reduction, increasing aggregate productivity.

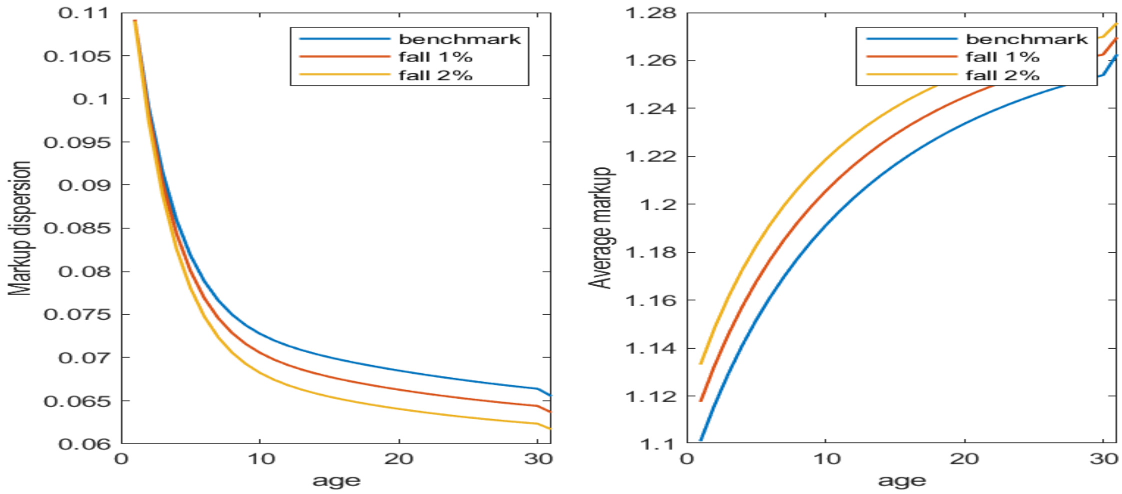

Figure 2.5: Change in s

Table (2.5) quantifies the change in productivity following a decline in the exit rate. As the exit rate declines, markup dispersion declines following the fast convergence in the age dispersion in markup. However, the fast growth in markup over the life cycle increases the age dispersion in the markup, contributing to increasing the overall markup dispersion. However, the change in the within-age dispersion in markup is strong enough to lower the overall dispersion in markup.[22]

Table 2.5: Change in s

		Change in s		
	$s = 0.10$	- 1%	-2%	- 3%
VAR	0.0090	0.0090	0.0089	0.0088
Within VAR	0.0068	0.0067	0.0065	0.0064
Between VAR	0.0022	0.0023	0.0024	0.0024
Avr markup	1.2331	1.2332	1.2332	1.2337
TFP	1.5858	1.8957	2.2832	2.7874
No Within	1.5891	1.8995	2.2877	2.7928
No Between	1.5861	1.8960	2.2836	2.7879

The fast decline in markup dispersion and the accumulation of customer base increase the productivity growth and markup growth over the life cycle. While that fast productivity growth contributes to increasing the aggregate productivity, the fast growth in

[22] An endogenous exit model would provide a different result on the within age dispersion of markup because exogenous exit would introduce a cut-off in the markups distribution. A fall in the exit rate would shift the cut-off, increasing the markup dispersion at a given age and then induce a low declining rate in markup dispersion over the age.

markup over the life cycle contributes to lower the aggregate productivity (see equation 2.22). Table (2.5) shows an increase in the aggregate productivity following a fall in the exit rate, suggesting a strong effect on the productivity over the life cycle driven by the within age dispersion in markup and the demand formation over the life cycle.

2.4.4 Test of the learning mechanism

In this section, I test the learning mechanism from the model. From the model, demand shifter arises as a residual factor from the demand. To estimate the demand shifter, I jointly estimate the production function and the demand to recover both productivity and habit stock as the residual factor after controlling for industry and time-specific effects. I then exploit information on the plant's price to recover the demand shifter.

Table 2.6: Test of learning

	(1) ln(b)	(2) ln(sale)	(3) ln(sale)
ln(b)(-1)	0.656***		
	(0.00650)		
ln(sale) (-1)	0.342***	0.984***	0.990***
	(0.00658)	(0.000810)	(0.00143)
Pr(pexit=1 | u_i=0)(-1)	0.386***		0.0962**
	(0.0375)		(0.0331)
Δ ln(TFPQ)			0.000310**
			(0.000108)
Δ ln(labor)			0.382***
			(0.00430)
Δ ln(capital)			0.180***
			(0.00364)
Constant	-3.724***	0.403***	0.335***
	(0.0714)	(0.00879)	(0.0180)
Observations	74474	82345	73608
R^2	0.9450	0.9659	0.9727
Industry FE	√	√	√
Cohort FE	√	√	√

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To solve the learning mechanism, I run two regressions. The first regression matches the law of demand formation, which is the source of the learning mechanism. The second one is the autocorrelation in sales which is a consequence of the learning mechanism. Table (2.6) shows the estimates result from those regressions.

Column(1) shows the autocorrelation between the demand residual and the sales. The autocorrelation in demand residual is significant and shows a negative correlation between demand shifter growth and the size of the demand shifter. Plants with a low shifter grow faster such as a 10 % increase in the demand shifter increases the demand shifter's growth by 3.44 %. Furthermore, column(1) shows the significant effect from sales of the shift in the future demand, a 10 pp increase in sales increase the demand shifter increases the demand residual by 3.42 %.

Table (2.6) also reports the autocorrelation in sales after controlling for the plant's productivity growth and inputs growth. I control for productivity input growth to remove the effect of a change in the production side on the sales growth. Therefore, the autocorrelation in sales would be driven by the demand side. Column (2) shows a significant autocorrelation between sales after controlling for a change in inputs and plants productivity. This result suggests that the demand side plays a significant role in the prediction of future sales.

2.5 Conclusion

This paper studies the implications of plant's markup variation for productivity both over the life cycle and at the aggregate level. I found that markup dispersion is high at entry and decreases with the plant's age. I explain this pattern by an active learning process of the demand over the life cycle. New plants get higher (resp. lower) idiosyncratic demand shock than what is required for their efficiency and optimally learn from their demand through markup adjustment. In the long run, the dynamic results in a convergence in markups.

The life cycle pattern of markups dispersion implies high friction in inputs allocation at the entry, which lowers the average productivity of new firms. In addition, it contributes to foster productivity growth over the life cycle. In a model with endogenous exit, the high variation in markups at entry would improve the creative destructive process as the less productive plants reduce their size and exit.

Thus, this paper presents an alternative view on distortions that induces friction to the input allocation. An emphasis is made on the type of firms that bear the distortions. Depending on the group of firms where the distortion is the most important, friction to the resource allocation may improve productivity growth. In addition, this result may explain why the net effect of distortions found in Eslava and Haltiwanger (2020) remains low.

2.6 Appendix

2.7 Figures

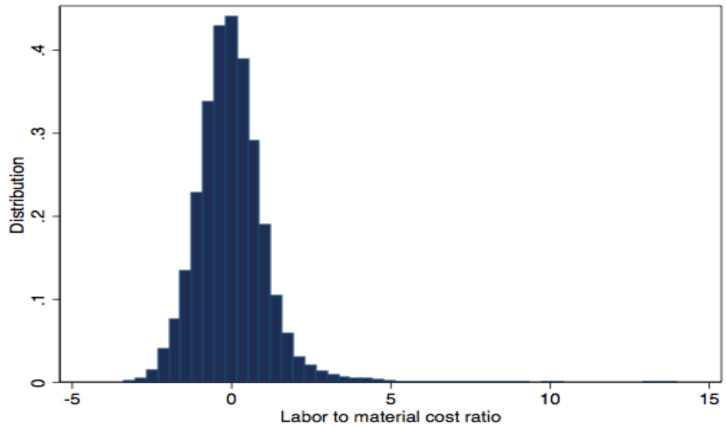

Figure 2.6: Labor to material cost ratio distribution

Note: The figure represents the labor to materials cost ratio distribution in log after controlling for industry cohort and time-specific effects. The dispersion in that ratio suggests the presence of some wedges, measurement errors, non-neutral productivity shock or heterogeneity in the return to scale.

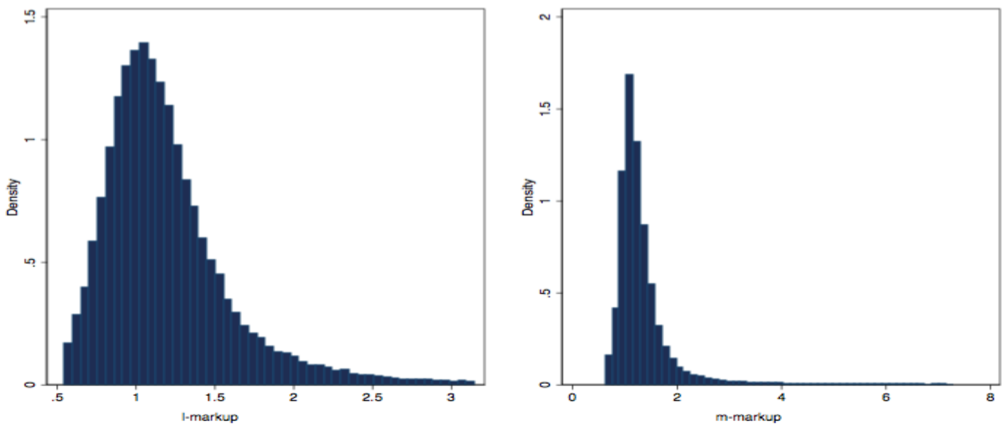

Figure 2.7: Markups distribution

Note.- The figure represents markup estimation using both labor (l-markup) and materials (m-markup). The estimate distribution doesn't control for industry time and cohort effect. Markup is estimated using the De Loecker et al. (2020) framework where the input elasticity estimated using the flexible cost-share approach proposed by Raval (2020).

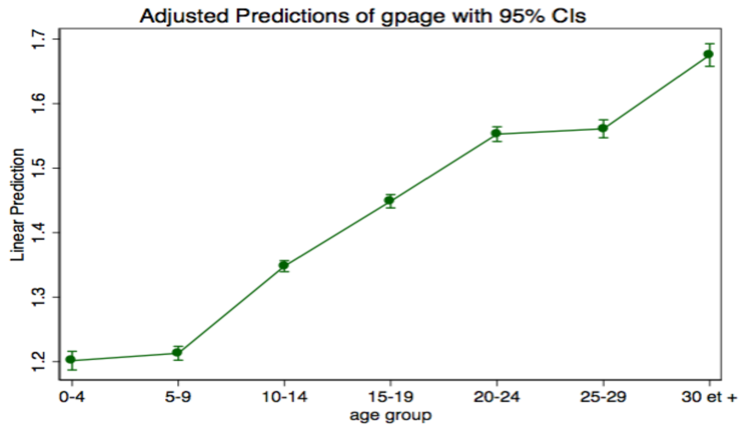

Figure 2.8: Average (l-markup) over the age

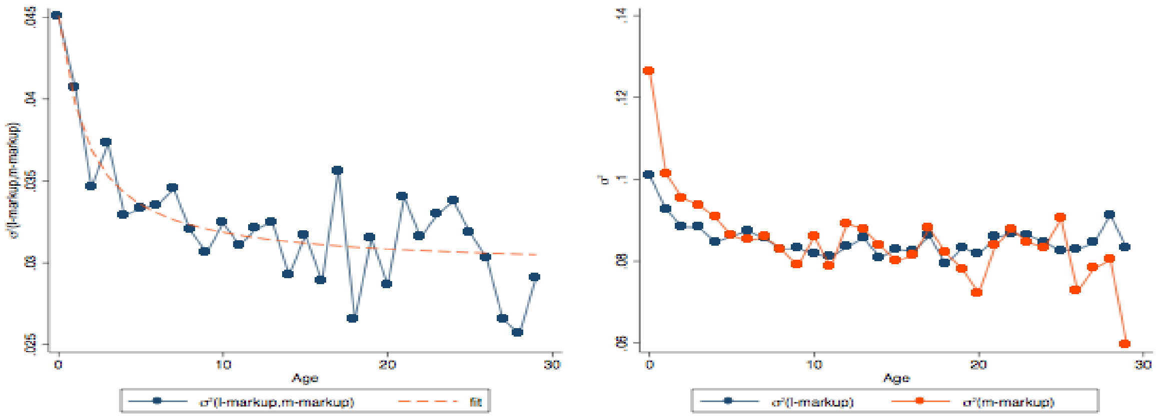

Figure 2.9: Variance and Covariance of l-markup and m-markup over the age

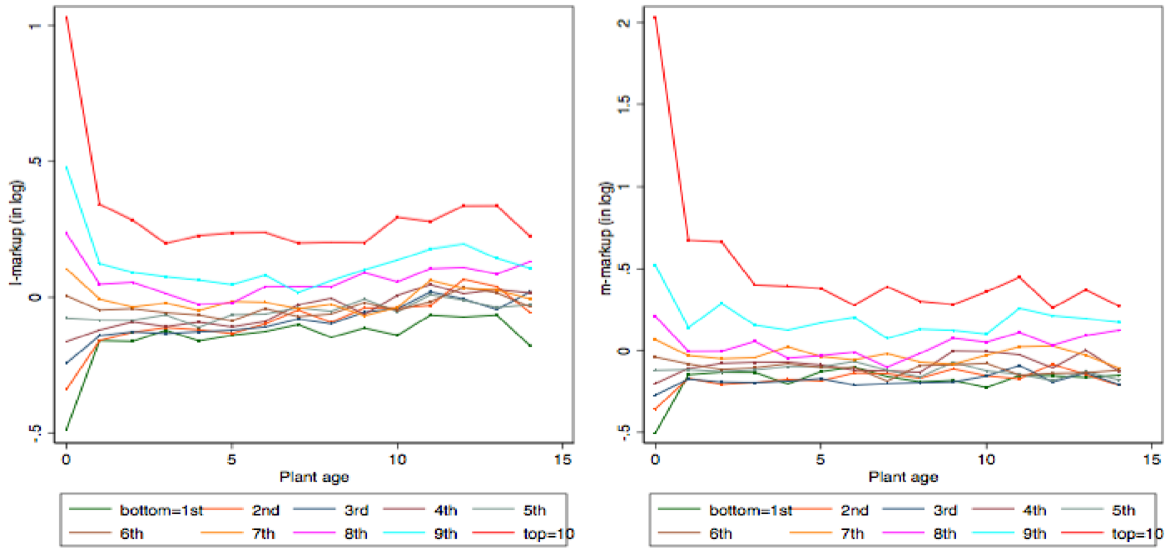

Figure 2.10: Life cycle pattern of markup conditional on markup at entry and survival

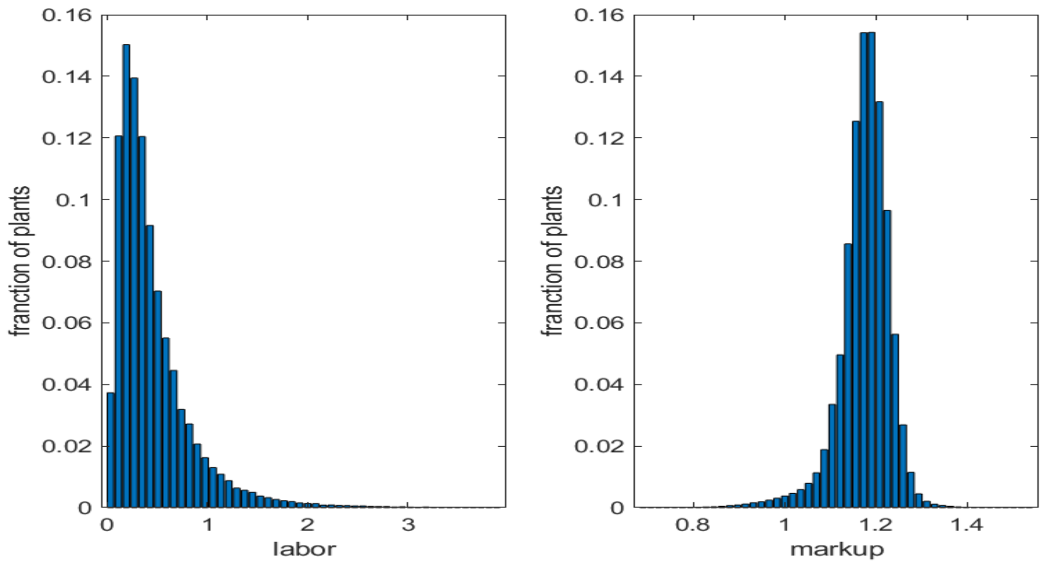

Figure 2.11: Firms size and markups distribution from the model

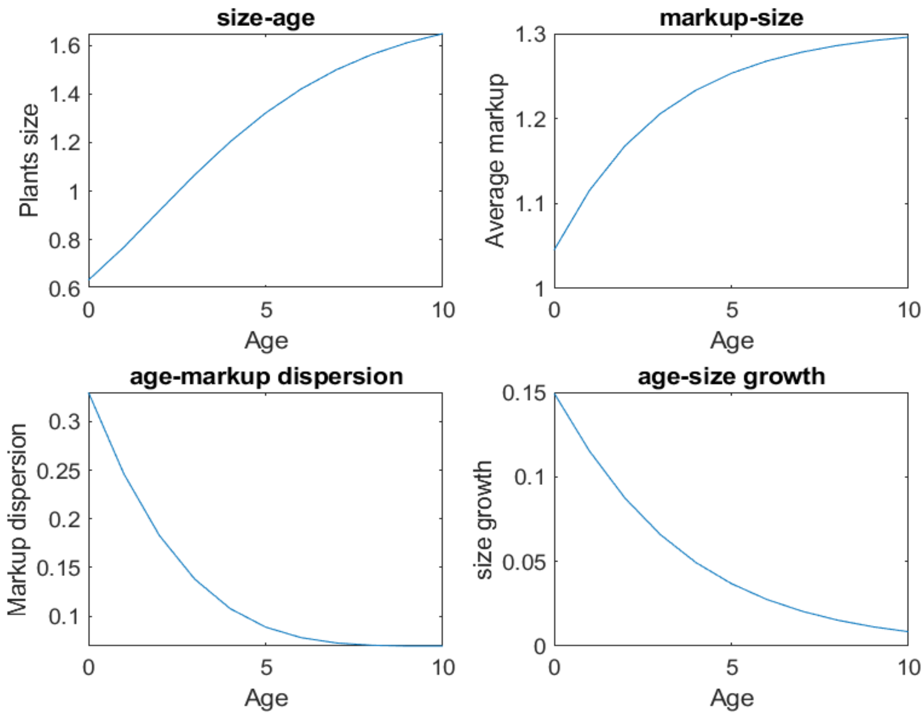

Figure 2.12: Life cycle dynamic

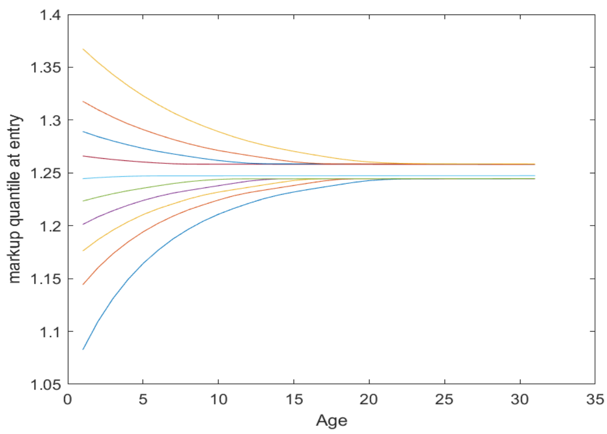

Figure 2.13: Age pattern of markups conditional on markups at entry and survival

2.8 Tables

Table 2.7: Flexible cost share markup estimation

	Statistics		Cross-sectional correlation	
	Mean	Std	l-markup	m-markup
l-markup	1.1953	0.3936	1	-
m-markup	1.3455	0.6094	0.3578	1

Table 2.8: Inter-quantile range regression (10-90)

	(1)	(2)	(3)	(4)
	ln(μ_l)	ln(μ_l)	ln(μ_m)	ln(μ_m)
age	-0.00386***	-0.00589***	-0.00713***	-0.00421***
	(0.000482)	(0.000405)	(0.000603)	(0.000642)
Pr(pexit=1 | u_i=0)		-0.855***		1.814***
		(0.0358)		(0.0738)
$ln(k/l)$		0.107***		0.0887***
		(0.00250)		(0.00368)
ln(tfpq)		-0.000172***		0.00000197
		(0.0000491)		(0.0000854)
Constant	0.888***	0.546***	0.931***	0.376***
	(0.00479)	(0.0107)	(0.00848)	(0.0215)
Observations	95739	88612	95749	87255
Industry FE	√	√	√	√
Cohort FE	√	√	√	√

Standard errors in parentheses

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

Table 2.9: Correlation between markup growth and plant's size growth

	Δln(sale)			
	(1)	(2)	(3)	(4)
$\Delta\ln(\mu)$	0.162***	0.0977***	0.0959***	0.0832***
	(0.00313)	(0.00291)	(0.00302)	(0.00536)
$\ln(\mu)$(-1)		-0.135***	-0.131***	-0.131***
		(0.00124)	(0.00127)	(0.00127)
age		0.0250***	0.0252***	0.0252***
		(0.000394)	(0.000416)	(0.000416)
Pr(pexit=1 | u_i=0)		-3.457***	-3.433***	-3.434***
		(0.0276)	(0.0286)	(0.0286)
Δln(tfpq)			0.0000993	0.0000999
			(0.0000651)	(0.0000651)
$\Delta\ln(\mu)\times$ age				0.00113**
				(0.000392)
Constant	0.235***	1.684***	1.647***	1.648***
	(0.00119)	(0.0128)	(0.0135)	(0.0135)
Observations	78522	78218	72621	72621
R^2	0.0463	0.2074	0.2063	0.2064
Industry FE	√	√	√	√
Cohort FE	√	√	√	√

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 3

The Welfare Gain from New Infrastructure Investments*

3.1 Introduction

Shipping costs remain one significant trade barrier, especially in developing countries with a lack of infrastructure. They are driven by both the supply and demand of transport services (Asturias, 2020). Although investment in infrastructures drives the supply side, the welfare effect resulting from a change in the demand for transport services following infrastructure shocks is not wide study. That effect is especially important given the lack of competition in the transport sector (Hummels et al., 2009; Asturias, 2020). This paper study the welfare gain from new infrastructure investments when considering an endogenous transport sector.

We build a multi-country, multi-sector Ricardian model as in Caliendo and Parro (2015) where we add a multi-modal endogenous transport sector. The endogenous transport sector considers that exporters of goods have shipping preferences for specific transportation modes and transporters don't perfectly compete in shipping goods. A multi-modality and the lack of competition allow us to study the endogenous change in transport cost along three margins following an infrastructure shock: a change in the supply through infrastructure shock, a change in demand through markup adjustment, and a reallocation of sales shares across transportation modes. The interaction between those margins defines the magnitude of shipping costs and then the welfare gain. Thus, the model has the advantage of studying sector-countries interaction as in Caliendo and Parro (2015), but it also introduces shipping modes-sector interaction.

In the model, countries import intermediate goods from the lowest supplier subject to tariff and shipping costs. Intermediate goods are used to produce goods and services used as intermediate goods for other sectors and final goods for tradable and non-tradable

*I am greatly indebted to Mathilde Lebrand for her invaluable guidance.

sectors. The transport sector is divided into sub-sectors relative to the mode of transportation used. An endogenous number of identical transporters compete in Bertrand per route and ship all goods for each transport mode. Transporters belong to the origin country and pay a fixed cost per unit of labor to enter on a route. Their production costs are valued per unit of goods produced and they charge a markup on their price as they compete in Bertrand. Markup decreases with the number of transporters and generates an incomplete pass-through from the cost to the shipping price.

Infrastructure shocks are mode-specific and contribute to lower the cost of producing shipping services. Shipping cost falls following an infrastructure investment in sub-sector m, especially in the sub-sector where the investment occurs, increasing the imports. Thus, transporters' profit increases in the sub-sector m, increasing the number of transporters in that sub-sector and lowering the shipping cost through markup adjustment. Although consumers benefit from shipping cost reduction through an adjustment of the supply and demand for transport services, the shipping cost reduction makes attractive the shipping mode m, inducing a sales shares reallocation across transportation sub-sectors which contribute to amplify the decrease in the shipping cost. The relative importance of this gain depends on the mode-specific shipping preferences for a good.

In sub-sector/mode where there is no shock in infrastructures (m'), the sales share falls even if the overall imports. Therefore, change in the number of transporters is essentially driven by the net effect from sales shares loss and imports gains. Shipping preferences drive this net effect. When a good is strongly related to a specific shipping mode, fewer sales shares are reallocated across sub-sectors and the number of shippers increases in the sub-sector m'. However, when a good is related to any particular shipping mode, a large share of sales are reallocated, reducing the transporter profits in the sub-sector m', reducing the number of shippers, and increasing the shipping cost of the sub-sector m'.

In addition, competition is an important channel by which changes in tariffs adjust the transport cost and then the welfare (Hummels et al., 2009). Indeed, an increase in tariffs raises the overall trade cost and lowers the imports. On the other hand, low substances reduce the transporters' profit, reduce the number of shippers, and increase the shipping cost through increased markup. Therefore, the overall trade cost through transport cost and contributes to lower consumer welfare.

Therefore, I used a simulation to assess the implications of new infrastructure investments. The simulation is based on two identical countries with different preferences in shipping modes. The structural modeling allows me to endogenously estimate the shipping mode preferences per goods using the ad-valorem shipping costs distribution and sales shares distribution across the transport sub-sectors. Using a set of parameters chosen in the literature for the simulation, I found a spillover effect from new infrastructure investments across the transport sub-sectors, goods and the bilateral trade between countries. The welfare gain is more important in the country and transport sub-sector where the investments are realized.

This paper is related to the international trade literature where structural models are used to quantify the welfare effects from different policies such as new infrastructure investments. This paper follows Caliendo and Parro (2015) framework which is an extension of Eaton and Kortum (2002) with multi-countries, to study the interaction between sectors and countries following a tariffs change. In addition, the paper introduces an endogenous non-competitive and multi-modal transport sector. Thus, the framework allows studying how the exporters shift in shipping modes following an infrastructure shock and the implications of that reallocation on shipping cost and welfare.[1]

The paper contributes to the new literature on endogenous transport cost. Asturias (2020) develops a model of trade with an oligopolistic competitive transport industry where shippers choose their transport technology. Both oligopolistic competition and technology choice induces markups in the transport sector. He shows that half of the welfare changes following tariffs changes result from a transport cost adjustment. Wong (2018) develops a framework where trade imbalance affects transportation costs. The channel from the imbalance to the transport cost is the round trip effect associated with the shippers' capacities on the two directions of a route. Behrens and Picard (2011) study the role of a competitive transport market in shaping the economic activity and the trade pattern. They show that the spatial distribution of firms and the concentration of the economic activity prevails, especially when the freight rate is endogenously determined. Finally, this paper considers a Bertrand competition across transporters as the source of the transporters' margins. Tariffs changes have a similar effect on transport cost as described by Asturias (2020). The main difference is the sales shares reallocation across transport mode, which endogenously drive the shipper's competition within each mode. In addition, this paper focus on infrastructures policies and not tariff changes.

The next sections are organized as follows. Section 2 introduces the economic model and studies the implications. Section 3 presents the equilibrium used to study the counterfactual. Section 4 discusses simulation impact from the model and finally, section 5 concludes.

3.2 Model

This section presents the multi-country, multi-sector model from Caliendo and Parro (2015) where we introduce a non-competitive and multi-modal transportation sector. We use the framework later to quantify the welfare gain from new infrastructure investments. Both the non-competitiveness and multi-modality in the transport sector characterize the general equilibrium effect in the transport sector. Infrastructures are modeled as public goods that lower the transporters' marginal cost.

A set of N country represents the economic environment and each country is populated by a consumer and a continuum of firms by sectors j. Between each pair of countries, an endogenous number of transporters produce a shipping service in a non-

[1]Others references include de Soyres et al. (2019); de Soyres et al. (2020); etc.

competitive manner, using various shipping modes (e.g. airplane, car, truck, cargo, etc.). Investment in infrastructures is transport mode-specific and is financed by taxes.

3.2.1 Household

Let us consider N countries and a representative household from country n who consumes goods from all sectors (j) in the economy. The household supply labor (L_n) at the wage w_n, get dividends from the transport sector (Π_n), pays a lump-sum taxes (T_n) and receives a transfer from the government (R_n). The transfer is the import tariffs revenue. Consumer buys goods on both domestic and foreign markets. The household problem is to maximise his utility subject to his budget constraint :

$$\max c_n = \prod_{j=1}^{J} (c_n^j)^{\alpha_n^j} \qquad with \qquad (\sum_{j=1}^{n} \alpha_n^j = 1)$$

$$\begin{aligned} \sum_{j=1}^{J} p_n^j c_n^j &= w_n L_n + R_n + \Pi_n - T_n = I_n \\ R_n &= \sum_{j=1}^{J} \sum_{i=1}^{N} \nu_{in}^j M_{in}^j \end{aligned} \tag{3.1}$$

where c_n^j is the consumption of goods j in country n and α_n^j is the consumption share in the consumer basket. ν_{in}^j is the add-valorem tariffs cost that is collected by the government from country n on importations (M_{in}^j) and rebate as a lump-sum to the household.[2] The household n chose the quantities of each goods (j) such as his total spending on that good j represents a share α_n^j of his total income (equation 3.2). The aggregate price index in country n given by equation (3.3) is the weighted average of all priced across goods in that country.

$$p_n^j c_n^j = \alpha_n^j I_n = \alpha_n^j p_n c_n \tag{3.2}$$

$$p_n = \prod_{j=1}^{J} (\frac{p_n^j}{\alpha_n^j})^{\alpha_n^j} \tag{3.3}$$

3.2.2 Production

Within each sector j and country n, there is a continuum of producers $z \in (0, 1)$ producing one variety of the good j. Producers are heterogeneous in productivity a_n^j draw from a Frechet distribution $F_n^j(a) = \exp\{-T_n^j a^{\theta^j}\}$. As in Parro and Calliendo (2012), T_n^j captures the absolute advantage of country n to produce the goods j and θ^j the comparative

[2] We denote by M_{in}^j and E_{in}^j respectively the importations and exportations of goods j from country i to country n.

advantage of country n to produce the same goods j. Each producer hires labor (l_n^j) and buys intermediate inputs (m_n^{kj}) from other sectors k to produce the good j. β_n^j and β_n^{kj} respectively represent the return of both labor and intermediate inputs. The producer problem is given by :

$$\min \omega_n l_n^j(z) + \sum_{k=1}^{J} p_n^k m_n^{kj}(z) \qquad (\beta_n^j + \sum_{k=1}^{J} \beta_n^{kj} = 1)$$

$$q_n^j(a_n^j(z)) \geq a_n^j (l_n^j(z))^{\beta_n^j} \prod_{k=1}^{J} (m_n^{kj}(z))^{\beta_n^{kj}}$$

The cost minimisation solution shows that producer's marginal cost $\lambda_n^j(z)$ has two components. An industry specific component (c_n^j) and a firm specific component (a_n^j) (see equation 3.4). The industry specific component is a weighted average of input cost weighted by the inputs elasticities. This component is the source of correlation between price across sectors. Producers productivity is the only source of heterogeneity in the firm's marginal cost and they are negatively correlated.[3]

$$\lambda_n^j(a_n^j(z)) = \frac{1}{a_n^j(z)} \left(\frac{\omega_n}{\beta_n^j}\right)^{\beta_n^j} \prod_{k=1}^{J} \left(\frac{p_n^k}{\beta_n^{kj}}\right)^{\beta_n^{kj}} = \frac{c_n^j}{a_n^j(z)} \tag{3.4}$$

Producers sell goods on both domestic and foreign markets. Selling goods abroad requires two additional costs. An ad-valorem transport cost (τ_{in}^j) that is paid for shipping goods from country i to country n and an ad-valorem tariffs (ν_{in}^j) that is charged on the consumer n on imported goods j from i. Let define by $\kappa_{in}^j = \tau_{in}^j + \nu_{in}^j$ the overall ad-valorem trade cost. Given the set of imported goods, the consumer n will buy the good j at the minimum price p_n^j. This feature imply that only the most productive firm in country i will ship his good to country n.

$$p_n^j = \min_i \left\{(1 + \kappa_{in}^j)\lambda_i^j\right\} \tag{3.5}$$

3.2.3 Trade flows

Following Eaton and Kortum (2002), we derive the trade flows properties based on the Frechet productivity distribution. Because consumer chooses to buy goods at the lowest price after trade cost, the probability $F_n^j(p)$ that the consumer n buys the goods j at a price below p is given by equation (3.6). The final price for the good j also follows a Frechet distribution with a scale parameter θ^j and a dispersion ϕ_n^j. The dispersion

[3] The producer choice is $\omega_n = \beta_n^j \frac{q_n^j}{l_n^j} \lambda_n^j$ and $p_n^k = \beta_n^{kj} \frac{q_n^j}{m_n^{kj}} \lambda_n^j$

reflects the price dispersion across countries after adding the trade cost and those trade costs increase the price dispersion. Therefore, the probability (π_{in}^j) that the country i supplies a particular good j to country n at the cheapest price is :

$$F_n^j(p) = 1 - e^{-\phi_n^j p^{\theta^j}} \qquad\qquad \pi_{in}^j = \frac{\phi_{in}^j}{\phi_n^j} \tag{3.6}$$

where $\phi_n^j = \sum_{i=1}^N \phi_{in}^j = \sum_{i=1}^N T_i^j \left(c_i^j \kappa_{in}^j\right)^{-\theta^j}$. Using the price distribution I compute the consumer price in country n for the good j based on expectation. Equation (3.7) shows that the consumer price is a CES aggregation of price industry specific component of the marginal cost of the good j around the world, added with all trade cost measured ad-valorem and weighted by the absolute advantage of each countries in good j.

$$p_n^j = \Gamma\left(\frac{1+\theta^j}{\theta^j}\right)\left(\sum_{i=1}^N T_i^j \left(c_i^j \kappa_{in}^j\right)^{-\theta^j}\right)^{-1/\theta^j} \tag{3.7}$$

By defining X_n^j the total expenditure in good j by the country n, the share of expenditure by the country n in good j from the country i is equal to the probability that the country i supply the good j to the country n. Both equations (3.7) and (3.8) show that country n are more likely to trade with countries with the highest absolute advantage of good j, the lowest trade cost and the lowest industry-specific marginal cost in good j.

$$\frac{X_{in}^j}{X_n^j} = \pi_{in}^j = \frac{T_i^j \left(c_i^j \kappa_{in}^j\right)^{-\theta^j}}{\sum_{i=1}^N T_i^j \left(c_i^j \kappa_{in}^j\right)^{-\theta^j}} \tag{3.8}$$

3.2.4 Endogenous transport cost

The transport sector is divided into M sub-sectors, where each sub-sector is specific to a mode of transportation. Between two countries, importers have preferences on using a particular mode of transport (b_{in}^{jm}) to ship a specific good j. A route/line (i, n) is defined by an origin country i and a destination country n and is used for shipping goods. An endogenous number of transporters use a specific mode of transportation m and compete in price (Bertrand competition).

Let define by q_{in}^j the quantity of good j that is imported by country n from country i. Each good j is shipped using the various mode of transportation and consumers have preferences on shipping using a specific mode of transportation (b_{in}^{jm}). The shipping preference varies with to the route, the shipping mode and the good that is shipped. We assume that all shipping services per sub-sector are bundle to match the quantities shipped through a CES function. σ is the elasticity of substitution across the m sub-sector mode of transportation and t_{in}^{jm} is the transport price index for shipping the good

j in through the route (i, n). For a given route, the demand for transport service per sub-sector and the transport price index per good are :

$$x_{in}^{jm} = \left(\frac{t_{in}^{jm}}{t_{in}^{j}}\right)^{-\sigma} (b_{in}^{jm})^{\sigma} q_{in}^{j} \qquad\qquad t_{in}^{j} = \left(\sum_{m=1}^{M} (b_{in}^{jm})^{\sigma} (t_{in}^{jm})^{(1-\sigma)}\right)^{1/(1-\sigma)} \tag{3.9}$$

Within each sub-sector, there is an endogenous number of identical transporters h_{in}^{m} per route. Transporter provide their shipping service one a route for all goods based on the demand received. However, on the same route, transporters using the same shipping mode offer a differentiated service. η is the elasticity of substitution of shipping services within each sub-sector. Because transporters on the same route and sub-sector compete in Bertrand, they internalize the effect from their price change in the price index when their set price. Lets consider the transporter u. The price $(t_{in,u}^{jm})$ and the demand $(x_{in,u}^{jm})$ face by this transporter are:

$$x_{in,u}^{jm} = \left(\frac{t_{in,u}^{jm}}{t_{in}^{jm}}\right)^{-\eta} x_{in}^{jm} \qquad\qquad t_{ni}^{jm} = \left(\sum_{u=1}^{h_{in}^{m}} (t_{in,u}^{jm})^{(1-\eta)}\right)^{1/(1-\eta)} \tag{3.10}$$

We denote by $mc_{in,u}^{jm}$, the marginal cost of producing an additional unit of shipping service for the good j by the transporter u. The cost of shipping services is valued per unit of goods shipped, which allows us to obtain an add-valorem expression of shipping cost. Equation (3.11) represents the marginal cost expression for a given transporter. The transporter marginal cost has an idiosyncratic component $e_{in,u}^{j}$. It increases with the distance on the route and decrease with the level of infrastructures in both the origin and the destination countries.[4] Given that a transporter ships all goods j through a route (i, n) using a specific mode m, the transporter u maximize his profit across all the goods shipped given the shipping demand received. The profit maximisation problem is given by:

$$\begin{aligned}
&\max \sum_{j=1}^{J} \left\{ \tau_{in,u}^{jm} x_{in,u}^{jm} - mc_{ni,u}^{jm} x_{in,u}^{jm} \right\} \\
&mc_{in,u}^{jm} = \frac{e_{in,u}^{j} (d_{in})^{\delta_d}}{(g_i^m)^{\delta_i^{jm}} (g_n^m)^{\delta_n^{jm}}} \lambda_i^j \\
&x_{in,u}^{jm} = \left(\frac{t_{in,u}^{jm}}{t_{in}^{jm}}\right)^{-\eta} x_{in}^{jm} \\
&x_{in}^{jm} = \left(\frac{t_{in}^{jm}}{t_{in}^{j}}\right)^{-\sigma} (b_{in}^{jm})^{\sigma} q_{in}^{j}
\end{aligned} \tag{3.11}$$

[4]We use that specification to compute transport cost in ad-valorem.

Because transporters compete in Bertrand, they internalize the effect of their price change on the sub-sector price index. The price demand elasticity for a single transporter is $\eta_{in,u}^{jm} = \eta - (\eta - \sigma)s_{in,u}^{jm}$ where $s_{in,u}^{jm}$ is the share of sales holds by the transporters u in the sub-sector m. Transporters charge a markup for two reasons. First, they provide differentiated services, which allows them to charge a markup. Second, they compete in Bertrand, which generates heterogeneity in markups that are correlated with sales. Large transporters charge a high markup.

We assume that all transporters are symmetric per route and mode and have the same the sales share $s_{in,u}^{jm} = 1/h_{in}^{m}$. Therefore, the price demand elasticity for a single transporter is $\eta_{in,u}^{jm} = \eta - (\eta - \sigma)(1/h_{in}^{m})$. A decrease in the number of transporters reduces the substitution across the shipping services, allowing transporters to increase their markup. Let define by τ_{in}^{jm} the ad-valorem shipping cost. The price charges by the transporter is :

$$t_{in}^{jm} = \mu_{in}^{m} mc_{in}^{jm} = \left(\frac{\eta_{in}^{m} - 1}{\eta_{in}^{m}}\right) \frac{e_{in}^{j}(d_{in})^{\delta_d}}{(g_i^m)^{\delta_i^{jm}}(g_n^m)^{\delta_n^{jm}}} \lambda_i^j = \tau_{in}^{jm} \lambda_i^{1j} \tag{3.12}$$

The ad-valorem shipping cost index for the good j is defined by:

$$\tau_{ni}^{j} = \left(\sum_{m=1}^{M} (b_{in}^{jm})^{\sigma} (\tau_{in}^{jm})^{(1-\sigma)}\right)^{1/(1-\sigma)} \tag{3.13}$$

The number of shippers is endogenously determined by the free entry condition on each route. Entry through each line is costly and there is a fixed cost (ke_{in}^{m}) per unit of labor from the origin country. The entry condition within each route implies that transporters will enter within each route until all sources of profit disappear. The cost of the entry acts as a barrier at the entry, limiting the number of transporters that operate on a line. An increase in the fixed cost for a fixed amount of goods traded reduces transporters' number on a given route as the profitability margin falls.

$$w_i ke_{in}^{m} = \frac{(1 - \mu_{in}^{m^{-1}})}{h_{in}^{m}} \sum_{j=1}^{J} s_{in}^{jm} \tau_{in}^{j} M_{in}^{j} \tag{3.14}$$

Taxes (T_n) are collected by the government to finance infrastructures goods in each sub-sector m. Infrastructures goods are assuming non-excludable and non-rivalrous and they are paid at the marginal cost r_n^m. The government problem consists to allocate his resources into the various sub-sector infrastructures goods $(g_n^1, ..., g_n^M)$.

$$\sum_{m=1}^{M} r_n^m g_n^m = T_n \tag{3.15}$$

3.3 Equilibrium and welfare analysis

3.3.1 Equilibrium

Equilibrium in this economy is defined by a vector of input prices, sector-country prices and a route-number of transporters that satisfy the following equations:

$$D_n = \left(\sum_{j=1}^{J}\sum_{i=1}^{N} \frac{\pi_{ni}^j}{1+\kappa_{ni}^j} X_i^j\right) - \left(\sum_{j=1}^{J}\sum_{i=1}^{N} \frac{\pi_{in}^j}{1+\kappa_{ni}^j} X_n^j\right) \tag{3.16}$$

$$X_n^j = \alpha_n^j I_n + \sum_{k=1}^{J} \beta_n^{kj} \left(\sum_{i=1}^{N} \frac{\pi_{ni}^k}{1+\kappa_{ni}^k} X_i^k\right) \tag{3.17}$$

$$I_n = \omega_n L_n + \sum_{j=1}^{J}\sum_{i=1}^{N} \nu_{in}^j \frac{\pi_{in}^j}{1+\kappa_{in}^j} X_n^j \tag{3.18}$$

$$\omega_i k e_{in}^m = \frac{(1-\mu_{in}^{m^{-1}})}{h_{in}^m} \sum_{j=1}^{J} s_{in}^{jm} \tau_{in}^j \left(\frac{\pi_{in}^j}{1+\kappa_{in}^j} X_n^j\right) \tag{3.19}$$

Equation (3.16)-(3.19) refers to the market-clearing conditions. Equation (3.16) represents the trade deficits D_n which is the difference between the sum of exports and the sum of imports across goods. Equation (3.17) shows how the total production for each sector equates to the total absorption, while equation (3.18) decomposes the total household income into labor income and tariffs revenue. Finally, equation (3.19) refers to the entry condition per route (i, n) and a specific mode of transportation m. It shows how the number of transporters balances the cost paid to be transporters and the expected profit on a route.[5]

To asset the welfare impact from an infrastructure shocks, we define equilibrium in relative change by looking at the relative change in the policies function. Let us define by $\hat{x} = \frac{x'}{x}$ the growth in x. An equilibrium in relative change is a vector change in input price, sector-country prices and a route-number of transporters $(\hat{\omega}_n, \hat{p}_n, \hat{h}_{in}^m)$ relative to the change from infrastructures g_n^m to $g_n'^m$ and that satisfy :

$$\hat{\tau}_{in}^{jm} = \hat{\mu}_{in}^m (\hat{g}_i^m)^{-\delta_i^{jm}} (\hat{g}_n^m)^{-\delta_n^{jm}} \tag{3.20}$$

$$\hat{\mu}_{in}^m = \hat{\eta}_{in}^m \frac{\eta_{in}^m - 1}{\eta_{in}^m \hat{\eta}_{in}^m - 1} \qquad where \qquad \hat{\eta}_{in}^m = \frac{\eta - (\eta - \sigma)(1/(\hat{h}_{in}^m h_{in}^m)}{\eta_{in}^m} \tag{3.21}$$

$$\hat{\tau}_{in}^j = \left(\sum_{m=1}^{M} s_{in}^{jm} (\hat{\tau}_{in}^{jm})^{(1-\sigma)}\right)^{1/(1-\sigma)} \tag{3.22}$$

[5]Importations from country i to country n in good j and exportations from country n to country i in good j are respectively given by $M_{in}^j = \frac{\pi_{in}^j}{(1+\kappa_{in}^j)} X_n^j$ and $E_{ni}^j = \frac{\pi_{ni}^j}{(1+\kappa_{ni}^j)} X_i^j$.

$$\hat{s}_{in}^{jm} = \left(\frac{\hat{\tau}_{in}^{jm}}{\hat{\tau}_{in}^{j}}\right)^{(1-\sigma)} \tag{3.23}$$

$$\hat{\kappa}_{in}^{j} = \frac{\tau_{in}^{j}\hat{\tau}_{in}^{j} + \nu_{in}^{j}}{\kappa_{in}^{j}} \tag{3.24}$$

The optimal policies function in the transport sector are given by equations (3.20)-(3.23). An increase in infrastructure investments in a sub-sector m reduces the ad-valorem shipping cost for a good j in that sub-sector and changes the transporter's profit margin (equation (3.20)). By lowering the ad-valorem shipping cost in the sub-sector m, the average shipping cost for the good j falls (equation (3.22)). In addition, the share of shipping goods using the shipping mode m increases relative to other sub-sector (sales shares reallocation effect). Therefore, the relative size of the sub-sector m increases driving by both the reallocation effect and the shipping cost reduction (equation 3.23).[6]

The reduction in the average shipping cost for the good j and the increase of size for the sub-sector m determine the change in the number of transporters. The reduction in the average shipping cost for the good j increases the imports and raises the overall transport profits, especially in the sub-sector m. In addition, the sales shares reallocation toward the sub-sector m increases the profits in that sub-sector. Because profits have increased in the sub-sector m, the sub-sector becomes more attractive and the number of transporters within that sub-sector increases. With many transporters in the sub-sector m, there is more substitution across shipping services which contribute to lower transporter markup within the sub-sector m, amplifying the reduction in the ad-valorem shipping cost in the sub-sector m.

For the other sub-sectors m' where there is no new infrastructure investment, the sales shares reallocation across the sub-sector lowers the size of the sub-sector m'. Although the rise of imports increases the overall profits in the transport sector, the size reduction in the sub-sector m' mitigates the expected profits from shipping services in the sub-sector m'. The net effect from the sales shares reallocation effect and the profits gained through imports depends on the mode shipping preference for the good j. If importers don't use the mode m' for shipping the good j, there will be no reallocation effect and the profits gain from imports would increase the number of transporters and reduces the markup in the sub-sector m'. Therefore, the shipping cost would decline for those using the mode m'.

However, if the shipping mode m' is used for the good j and importers highly prefer it, the reallocation effect would be important and even dominate the profits gain from imports. In this case, the sales shares reallocation toward the sub-sector m would reduce the relative size of the sub-sector m'. Because the relative size effect is more important than the profits gain from imports, the expected profit for transporters in the sub-sector m' declines, reducing the number of transporters in that sub-sector and

[6]The relative size of a sub-sector is measured by the share of sales of that sub-sector.

increasing the transporters markup as there is less substitution across transporters. The rise in the markup within the sub-sector m' increases the ad-valorem shipping cost in that sub-sector. However, if the profit gain effect dominates the reallocation effect, the transporter's profit would increase, reducing the shipping cost in the sub-sector m'.

In addition, the framework allows us to study the effect of tariffs reduction on shipping costs. Hummels et al. (2009) shows a positive correlation between change in the shipping costs and change in tariffs. We show that this positive correlation may result from the lack of competition in the transport sector. Indeed, a decrease in tariffs costs would increase the imports and then the transporter's profit. Therefore, more transporters would enter, reducing the markup and then shipping cost.

$$\hat{c}_n^j = (\hat{\omega}_n)^{\beta_n^j} \prod_{k=1}^{J} (\hat{p}_n^k)^{\beta_n^{jk}} \tag{3.25}$$

$$\hat{p}_n^j = \left(\sum_{i=1}^{N} \pi_{in}^j \left(\hat{c}_i^j \hat{\kappa}_{in}^j \right)^{-\theta^j} \right)^{-1/\theta^j} \tag{3.26}$$

$$\hat{\pi}_{in}^j = \left(\frac{\hat{c}_i^j \hat{\kappa}_{in}^j}{\hat{p}_n^j} \right)^{-\theta^j} \tag{3.27}$$

The optimal policy function in good and service sectors is given by equations (3.25)-(3.27). As the ad-valorem transport cost fall, the ad-valorem trade cost $\hat{\kappa}_{in}^j$ also fall. A decline in the ad-valorem trade cost reduces the price of goods and services through a direct effect from the shipping cost and an additional decline in the marginal cost of producing goods worldwide. As the shipping cost has declined, the imported goods become cheaper, making cheaper inputs, lowering the marginal cost for producing goods and services. This additional effect from a decline in the marginal cost of goods and services reduces imported goods' final price. Therefore, origin countries where it becomes cheaper to ship goods and cheaper to produce goods increase their trade flow with their partners.

3.3.2 The welfare decomposition

Following Caliendo and Parro (2015), I decompose the consumer welfare following an infrastructures shocks into terms of trade and volume of trade effects. This decomposition allows us to understand the effects from a change in transport cost following infrastructure shocks across countries and sectors. The trade deficit and the profit condition from the transport sector imply ($\Pi_n = 0$) and (D_n). The consumer welfare is defined by $W_n = \frac{I_n}{p_n}$ and the change in the consumer welfare is :

$$d\ln W_n = \frac{1}{I_n}\sum_{j=1}^{J}\sum_{i=1}^{N}\kappa_{in}^{j}M_{in}^{j}\left(d\ln M_{in} - (d\ln c_i^j + d\ln\kappa_{in}^{j})\right) + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N}\left(E_{ni}^{k}d\ln c_n^k - M_{in}^{j}d\ln c_i^j\right) \tag{3.28}$$

As in Caliendo and Parro (2015), the first term measures the multi-lateral volume of trade effect and multi-sectoral and the second term the multi-lateral and multi-sectoral terms of trade effect. The terms of trade effect quantify the gain from an increase in exporters prices relative to importer prices, weighted by the sectoral-bilateral exports and imports. The first term allows us to recover each sector's contribution to the terms of trade effect, which depends on the bilateral trade deficit between countries and the change in imports and exports prices. The volume of trade terms measures the change in imports volume following an infrastructures shock. Infrastructure shock reduces trade cost ($d\ln\kappa_{in}^{j}$) and then the imports price ($d\ln c_i^j$), increasing the volume of trade ($d\ln M_{in}$).

3.3.3 Quantitative analysis

This section studies the change in shipping cost and the welfare gains from new infrastructure investments. The goal of this quantitative analysis is to identify the mode shipping preferences and analyses the implications in terms of sales shares reallocation and welfare gains. Due to the current lack of data availability, we simulate an economy of two countries to quantify the welfare gains.[7] For the simulation, we select some realistic parameters from the data and the literature.

We choose two identical countries that produce two goods. As in the model, consumers buy the two goods in both domestic and foreign markets. Shipping goods on those two markets is costly (tariffs and shipping costs) and consumers can use two shipping modes (for example, road (1) or airplane (2)). The two countries are only different in terms of shipping preferences. In country (1), consumers buy goods on domestic and foreign markets using the road as a shipping mode. However, in the country (2), consumers buy goods on the domestic market using the road but use both road and airplane to ship goods from the foreign market. We assume that transporters sales are equally distributed between the two shipping modes. Each country has a stock of infrastructures allocated across transportation sub-sectors.

To solve the model, we set some moments to characterize the initial equilibrium. Each country has a Gross Domestic Product (GDP) of 20. Within each country and transportation sub-sector, we set the infrastructure stock g_i^m =30 (m=road, airplane). The elasticity of substitution within each transport sub-sector is $\eta = 10$, while the elasticity of substitution between the transport sub-sector is $\sigma = 1.5$.[8] The return to scale from

[7]We are currently building data to quantify the welfare gain from new infrastructure investments in Africa using the model.

[8]There is less substitution between transport sub-sector (or mode) than within each transport sub-sector.

infrastructure investment is set to $\gamma = 0.2$ for both shipping modes. We assume that tariffs costs represent 0.2 % of each imported goods price. Finally, the labor share is identical for the two sectors of production j and set to $\beta_i^j = 0.3$. The Input-Output table is described in table (3.1).

Table 3.1: Input-Output table

Input-Output		value	
		Good 1	Good 2
	Good 1	0.6	0.4
	Good 2	0.4	0.6

Using the sales shares distribution across shipping modes, I estimate shipping mode preferences per good. The model shows that for a given shipping costs distribution across shipping modes, there is a unique distribution of preference b_{in}^{jm} that is consistent with the sales shares distribution across shipping modes. We solve for the shipping preferences by solving recursively the following equation:

$$s_{in}^{jm} = \frac{(\tau_{in}^{jm})^{(1-\sigma)}}{\sum_{m=1}^{M} (b_{in}^{jm})^{\sigma} (\tau_{in}^{jm})^{(1-\sigma)}} \tag{3.29}$$

To solve the counterfactual, I only introduce new infrastructure investments on the country's (2) road and airplane sub-sectors. Table (3.2) presents the result from the counterfactual experiment. Because it is a simulation, we will give less interpretation to the level of the change. The first panel of the table (3.2) shows the change in both the number of transporters and the ad-valorem shipping cost. Following the new investments in country (2), columns (2) and (3) show that the number of transporters increases in each bilateral route due to a spillover effect. Indeed, the new infrastructures in the country (2) increases both imports and exports and therefore the transporter's profits. There is no reallocation effect in the country (1) where there is a strong preference for using the road as a shipping mode. Therefore, changes in imports and exports increase the expected profit and then the number of road transporters. In country (2), where shipping from the foreign market use both road and airplane, the new infrastructure investments increase the imports, exports and the domestic trade. Given that the magnitude of shocks is identical for both sub-sectors and we have a similar return to scale, the sales shares reallocation effect is less important. Therefore, transporter's profits increase, inducing an increase in the number of transporters in both sub-sectors.

The table's (3.2) columns (3) and (4) represent the changes in the ad-valorem shipping costs following the new infrastructure investments. Because there is no change in the domestic transport sector in the country (1) and no investment has been made, the domestic shipping cost remains constant for both goods. In addition, on the bilateral route going from country (1) to country (2), the rise in the number of transporters reduces the shipping cost for both goods. Given that the only way of shipping in the country (1) is to use the road, there is no reallocation of sales shares. Therefore, the fall in shipping costs reflects an improvement in the tightening of competition on that route/line.

Table 3.2: Counterfactual analysis

	Change in h_{in}^m		Change in τ_{in}^j	
Route/Line	Road	Airplane	Good 1	Good 2
Country 1 - Country 1	1.0000	-	1.0000	1.0000
Country 1 - Country 2	1.0516	-	0.7579	0.7579
Country 2 - Country 1	1.0516	1.0516	1.0000	0.7579
Country 2 - Country 2	1.0149	-	0.7579	0.5743
Welfare gain				
Country 1	1.2664			
Country 2	1.5907			

In country (2) where the investments are made, the domestic gain features a change in the transporter's marginal cost and improved competition in the transport sector. The reallocation sales shares are less important because the two shipping modes have the same sales share and return to scale. However, on the bilateral trade between countries (1) and (2), the decline in shipping cost is driven by all three margins: reduction in transporter's marginal cost, a decrease in transporter's profit margins and the sales shares reallocation. Finally, the second panel shows that both countries benefit from the new infrastructure investments through the spillover with a much larger gain in welfare for the country (2).

3.4 Conclusion

In sum, we propose a new theoretical framework to quantify the welfare gain following new investments in infrastructures. The novelty is to introduce both the competition between nodes and the sales shares reallocation. In addition, introducing an endogenous transport sector provides additional margins where new infrastructure investments affect the shipping costs and welfare. However, those margins are strongly determined by the preference in shipping. Thus, the joint effect from the competition and the sales shares reallocation mitigate the benefit of new infrastructure investment.

This paper emphasizes the role of the transport sector in the policies analysis in international trade. The source of transporters' profit margins is essential to assess the benefit of new investment in infrastructure. In addition, shipping preferences are important to determine the magnitude of the shipping cost and welfare changes. This is essentially driven by the consumer shift to the most preferred shipping mode.

The framework will be interesting in the case of developing countries where there are few transport infrastructures. In addition, the framework would be useful to study the optimal allocation of infrastructures across modes, especially for the World Bank.

3.5 Appendix

Solving for the welfare change, we have :

$$d\ln W_n = \frac{\omega_n L_n}{I_n} d\ln\omega_n + \sum_{j=1}^{J}\sum_{i=1}^{N}\frac{\nu_{in}^j M_{in}^j}{I_n} d\ln M_{in} - d\ln p_n \tag{3.30}$$

For each of those component we have :

$$\begin{aligned}\frac{\omega_n L_n}{I_n} d\ln\omega_n &= \frac{\omega_n L_n}{I_n}\left(\frac{d\ln\kappa_n^j}{\beta_n^j} - \sum_{k=1}^{J}\frac{\beta_n^{jk}}{\beta_n^j} d\ln p_n^k\right)\\ &= \frac{1}{I_n}\sum_{j=1}^{J}\sum_{i=1}^{N} E_{ni}^j\left(d\ln\kappa_n^j - \sum_{k=1}^{J}\beta_n^{jk} d\ln p_n^k\right)\end{aligned}$$

$$\begin{aligned}d\ln p_n &= \sum_{j=1}^{J}\alpha_n^j d\ln p_n^j\\ &= \sum_{j=1}^{J}\left(\frac{X_n^j}{I_n} - \sum_{k=1}^{J}\frac{\beta_n^{kj}}{\beta_n^k}\frac{\omega_n l_n^k}{I_n}\right) d\ln p_n^j = \frac{1}{I_n}\sum_{j=1}^{J} X_n^j d\ln p_n^j - \frac{1}{I_n}\sum_{j=1}^{J}\sum_{k=1}^{J}\beta_n^{kj}\sum_{i=1}^{N} E_{ni}^k d\ln p_n^j\\ &= \frac{1}{I_n}\sum_{j=1}^{J} X_n^j d\ln p_n^j - \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k - \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k\left(\sum_{j=1}^{J}\beta_n^{kj} d\ln p_n^j\right)\\ &= \frac{1}{I_n}\sum_{j=1}^{J} X_n^j d\ln p_n^j - \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k\left(d\ln\kappa_n^k - \sum_{j=1}^{J}\beta_n^{kj} d\ln p_n^j\right)\\ &= \frac{1}{I_n}\sum_{j=1}^{J} X_n^j d\ln p_n^j - \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k + \frac{\omega_n L_n}{I_n} d\ln\omega_n\end{aligned} \tag{3.31}$$

Therefore :

$$\begin{aligned}d\ln W_n &= \frac{1}{I_n}\sum_{j=1}^{J}\sum_{i=1}^{N}\nu_{in}^j M_{in}^j d\ln M_{in} + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k - \frac{1}{I_n}\sum_{j=1}^{J} X_n^j d\ln p_n^j\\ &= \frac{1}{I_n}\sum_{j=1}^{J}\sum_{i=1}^{N}\nu_{in}^j M_{in}^j d\ln M_{in} + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k\\ &- \frac{1}{I_n}\sum_{j=1}^{J} X_n^j\sum_{i=1}^{N}(1+\psi_{in}^j)\frac{\pi_{in}^j}{(1+\psi_{in}^j)}(d\ln\kappa_n^j + d\ln\tilde{\psi}_{in}^j)\\ &= \frac{1}{I_n}\sum_{j=1}^{J}\sum_{i=1}^{N}\nu_{in}^j M_{in}^j d\ln M_{in} + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N} E_{ni}^k d\ln\kappa_n^k - \frac{1}{I_n}\sum_{j=1}^{J}(1+\psi_{in}^j) M_{in}^j(d\ln\kappa_i^j + d\ln\tilde{\psi}_{in}^j)\\ &= \frac{1}{I_n}\sum_{j=1}^{J}\sum_{i=1}^{N} M_{in}^j\left(\nu_{in}^j d\ln M_{in} - \psi_{in}^j(d\ln\kappa_i^j + d\ln\psi_{in}^j)\right) + \frac{1}{I_n}\sum_{k=1}^{J}\sum_{i=1}^{N}\left(E_{ni}^k d\ln\kappa_n^k - M_{in}^j d\ln\kappa_i^j\right)\end{aligned} \tag{3.32}$$

Conclusion Générale

Dans cette thèse divisée en trois chapitres nous analysons les causes et les implications des marges bénéficiaires sur le bien-être et la productivité. Dans le premier chapitre, nous étudions les causes de la hausse des marges bénéficiaires des entreprises, measurant le pouvoir de marché des entreprises dans un environnement où celles-ci accumulent des actifs intangibles. Nous proposons une nouvelle méthode d'aggrégation des marges bénéficiaires et nous montrons qu'elles ont augmenté ces 40 dernières années aux Etats-Unis. Cette hausse résulte principalement d'une réallocation des parts de marché des entreprises à faibles marges bénéficiaires vers celle à forte marges bénéficiares. Toutefois cette réallocation est associée à une hausse des marges bénéficiaires des entreprises. L'étude des causes de la hausse des marges bénéficiaires montrent que la hausse des investments en actifs intangibles ces quatre dernières décennies, particulièrement les bases de clientèle, ont contribué à 50%, à la hausse de la marge bénéficiaire au niveau aggrégée. Cette hausse des investments en actifs intangibles ne peut toutefois expliquer la recente de la concentration des parts de marché par une minorité d'entreprises appélées 'superstars'. Nous montrons que la hausse de la concentration est probablement due la hausse de la différence de productivité entre les entreprises dites superstars et les non-superstars.

Dans le deuxième chapitre, nous étudions les causes d'une variation des marges bénéficiaires sur la productivité moyenne des entreprises. Nous montrons que les marges bénéficiaires sont dispersées à l'entrée du marché et diminue avec l'âge des usines, représentant une convergence des marges bénéficiaires des entreprises d'une même cohorte durant leur vie. Nous montrons que cette convergence s'explique par un apprentissage de la demande par les entreprises. Elle est due à une différence entre l'échelle de production cible d'une entreprise et son niveau courant de production étant donné sa productivité. La variation des marges bénéficiaires entre entreprises induit une friction à l'allocation du travail qui contribue à réduire la productivité moyenne de usines à leur création. Par ailleurs, la diminution de la dispersion des marges bénéficiaires reflète une diminution des frictions à l'allocation et augmente la croissance de la productivité moyenne avec l'âge.

Dans le troisième chapitre, nous proposons un cadre d'analyse des gains économiques suite à des investissements en infrastructures. La particularité de ce cadre d'analyse est de ténir compte de l'équilibre et caracteristiques du secteur de transport. Nous montrons qu'en tenant compte de l'offre et la demande des services de transport, la faible de compétition du secteur du transport et la diversité des moyens de transport, une hausse des infrastructures de transport a pour éffet réduire les coûts de transport suivant trois marges. Prémièrement une diminution des coûts marginaux des transporters, ensuite une hausse de la compétition qui contribue à réduire les marges bénéficiaires et finalement une réallocation des parts de marché vers des moyens de transport plus bénéfiques. Toutefois, selon les préférences des agents économiques sur l'usage des moyens de transport, la réallocation des parts de marché a des éffets mitigés. Ce cadre d'analyse offre une possibilité d'analyse d'une allocation optimale des infrastructures de transport selon les modes et les secteurs de l'économie.

Bibliography

D. A. Ackerberg, K. Caves, and G. Frazer. Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451, 2015. ISSN 1468-0262.

P. Aghion, A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li. A Theory of Falling Growth and Rising Rents. Technical Report 2019-11, Federal Reserve Bank of San Francisco, Mar. 2019.

J. Asker, J. Farre-Mensa, and A. Ljungqvist. Corporate Investment and Stock Market Listing: A Puzzle? *The Review of Financial Studies*, 28(2):342–390, Feb. 2015. ISSN 0893-9454.

J. Asturias. Endogenous transportation costs. *European Economic Review*, 123:103366, Apr. 2020. ISSN 0014-2921.

D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen. Concentrating on the Fall of the Labor Share. *American Economic Review*, 107(5):180–185, May 2017. ISSN 0002-8282. doi: 10.1257/aer.p20171102.

D. Baqaee and E. Farhi. Productivity and Misallocation in General Equilibrium., 2018.

S. Barkai. Declining Labor and Capital Shares Simcha Barkai. 2016.

S. Basu. Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence. *Journal of Economic Perspectives*, 33(3):3–22, Aug. 2019. ISSN 0895-3309.

K. Behrens and P. Picard. Transportation, freight rates, and economic geography. DEM Discussion Paper Series, Department of Economics at the University of Luxembourg, 2011.

P. Bento and D. Restuccia. Misallocation, Establishment Size, and Productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303, July 2017. ISSN 1945-7707.

S. Bond, A. Hashemi, G. Kaplan, and P. Zoch. Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data. Technical Report w27002, National Bureau of Economic Research, Apr. 2020.

A. Burstein, B. Grassi, and V. Carvalho. Bottom-Up Markup Fluctuations. Technical Report 505, Society for Economic Dynamics, 2019.

L. Caliendo and F. Parro. Estimates of the Trade and Welfare Effects of NAFTA. *The Review of Economic Studies*, 82(1):1–44, Jan. 2015. ISSN 0034-6527. Publisher: Oxford Academic.

N. Crouzet and J. Eberly. Intangibles, Investment, and Efficiency. *AEA Papers and Proceedings*, 108:426–431, May 2018. ISSN 2574-0768.

S. J. Davis, J. Haltiwanger, R. Jarmin, and J. Miranda. Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms. Working Paper 12354, National Bureau of Economic Research, July 2006.

J. De Loecker. Recovering markups from production data. *International Journal of Industrial Organization*, 29(3):350–355, May 2011. ISSN 0167-7187.

J. De Loecker and F. Warzynski. Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–2471, May 2012. ISSN 0002-8282.

J. De Loecker, J. Eeckhout, and G. Unger. The Rise of Market Power and the Macroeconomic Implications. *The Quarterly Journal of Economics*, 135(2):561–644, May 2020. ISSN 0033-5533. Publisher: Oxford Academic.

M. De Ridder. Market Power and Innovation in the Intangible Economy. Technical Report 1931, Faculty of Economics, University of Cambridge, Mar. 2019. Publication Title: Cambridge Working Papers in Economics.

F. de Soyres, A. Mulabdic, S. Murray, N. Rocha, and M. Ruta. How much will the Belt and Road Initiative reduce trade costs? *International Economics*, 159:151–164, Oct. 2019. ISSN 2110-7017.

F. de Soyres, A. Mulabdic, and M. Ruta. Common transport infrastructure: A quantitative model and estimates from the Belt and Road Initiative. *Journal of Development Economics*, 143:102415, Mar. 2020. ISSN 0304-3878.

R. A. Decker, J. C. Haltiwanger, R. S. Jarmin, and J. Miranda. Changing Business Dynamism and Productivity: Shocks vs. Responsiveness. Working Paper 24236, National Bureau of Economic Research, Jan. 2018.

J. Eaton and S. Kortum. Technology, Geography, and Trade. *Econometrica*, 70(5): 1741–1779, 2002. ISSN 0012-9682. Publisher: [Wiley, Econometric Society].

C. Edmond, V. Midrigan, and D. Y. Xu. How Costly Are Markups? Working Paper 24800, National Bureau of Economic Research, July 2018.

G. Eggertsson, J. Robbins, and E. G. Wold. Kaldor and PikettyâĂŹs Facts: The Rise of Monopoly Power in the United States. Technical Report w24287, National Bureau of Economic Research, Cambridge, MA, Feb. 2018.

M. Eslava and J. Haltiwanger. The Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges. Technical Report w27184, National Bureau of Economic Research, Cambridge, MA, May 2020.

Z. Flynn, A. Gandhi, and J. Traina. Measuring Markups with Production Data. SSRN Scholarly Paper ID 3358472, Social Science Research Network, Rochester, NY, June 2019.

L. Foster, J. C. Haltiwanger, and C. J. Krizan. Aggregate Productivity Growth: Lessons from Microeconomic Evidence. *New Developments in Productivity Analysis*, pages 303–372, Jan. 2001.

L. Foster, J. Haltiwanger, and C. Syverson. Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425, Mar. 2008. ISSN 0002-8282.

L. Foster, J. Haltiwanger, and C. Syverson. The Slow Growth of New Plants: Learning about Demand? *Economica*, 83(329):91–129, 2016. ISSN 1468-0335.

L. Foster, C. Grim, J. C. Haltiwanger, and Z. Wolf. Innovation, Productivity Dispersion, and Productivity Growth. Technical Report w24420, National Bureau of Economic Research, Mar. 2018.

A. Gandhi, S. Navarro, and D. A. Rivers. On the Identification of Gross Output Production Functions. *Journal of Political Economy*, June 2020. ISSN 0022-3808. Publisher: The University of Chicago PressChicago, IL.

F. Gourio and L. Rudanko. Customer Capital. *The Review of Economic Studies*, 81(3): 1102–1136, July 2014. ISSN 0034-6527.

G. Grullon, Y. Larkin, and R. Michaely. Are U.S. Industries Becoming More Concentrated? SSRN Scholarly Paper ID 2612047, Social Science Research Network, Rochester, NY, Oct. 2018.

N. Guner, G. Ventura, and Y. Xu. Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4):721–744, Oct. 2008. ISSN 1094-2025.

G. Gutierrez and T. Philippon. Declining Competition and Investment in the U.S. Working Paper 23583, National Bureau of Economic Research, July 2017.

G. Gutierrez and T. Philippon. Fading Stars. *AEA Papers and Proceedings*, 109:312–316, May 2019. ISSN 2574-0768.

G. Gutierrez, C. Jones, and T. Philippon. Entry Costs and the Macroeconomy. Working Paper 25609, National Bureau of Economic Research, Feb. 2019.

R. E. Hall. The Relation between Price and Marginal Cost in U.S. Industry. *Journal of Political Economy*, 96(5):921–947, 1988. ISSN 0022-3808.

R. E. Hall. New Evidence on the Markup of Prices over Marginal Costs and the Role of Mega-Firms in the US Economy. Working Paper 24574, National Bureau of Economic Research, May 2018.

J. Hastings, A. HortaÃğsu, and C. Syverson. Sales Force and Competition in Financial Product Markets: The Case of Mexico's Social Security Privatization. *Econometrica*, 85(6):1723–1761, 2017. ISSN 1468-0262.

H. A. Hopenhayn. Entry, Exit, and firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5):1127–1150, 1992. ISSN 0012-9682.

C. J. Hottman, S. J. Redding, and D. E. Weinstein. Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364, Aug. 2016. ISSN 0033-5533. Publisher: Oxford Academic.

C.-T. Hsieh and P. J. Klenow. Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009. ISSN 0033-5533. Publisher: Oxford University Press.

C.-T. Hsieh and P. J. Klenow. The Life Cycle of Plants in India and Mexico. *The Quarterly Journal of Economics*, 129(3):1035–1084, Aug. 2014. ISSN 0033-5533. Publisher: Oxford Academic.

D. Hummels, V. Lugovskyy, and A. Skiba. The trade reducing effects of market power in international shipping. *Journal of Development Economics*, 89(1):84–97, May 2009. ISSN 0304-3878.

L. Karabarbounis and B. Neiman. Accounting for Factorless Income. Working Paper 24404, National Bureau of Economic Research, Mar. 2018.

J. Levinsohn and A. Petrin. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2):317–341, 2003.

E. G. Luttmer. Models of Growth and Firm Heterogeneity. *Annual Review of Economics*, 2(1):547–576, 2010.

V. Midrigan and D. Y. Xu. Finance and Misallocation: Evidence from Plant-Level Data. *American Economic Review*, 104(2):422–458, Feb. 2014. ISSN 0002-8282.

C. J. Nekarda and V. A. Ramey. The Cyclical Behavior of the Price-Cost Markup. SSRN Scholarly Paper ID 2276363, Social Science Research Network, Rochester, NY, June 2013.

G. S. Olley and A. Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297, 1996. ISSN 0012-9682.

S. Peltzman. Industrial Concentration under the Rule of Reason. *The Journal of Law & Economics*, 57(S3):S101–S120, 2014. ISSN 0022-2186.

M. Peters. Heterogeneous Markups, Growth, and Endogenous Misallocation. *Econometrica*, 88(5):2037–2073, 2020. ISSN 1468-0262.

D. Raval. Testing the Production Approach to Markup Estimation. SSRN Scholarly Paper ID 3324849, Social Science Research Network, Rochester, NY, Feb. 2020.

M. Ravn, S. Schmitt-Grohe, and M. Uribe. Deep Habits. Working Paper 10261, National Bureau of Economic Research, Feb. 2004.

J. V. Reenen. Increasing Differences between firms: Market Power and the Macro-Economy. page 46, 2018.

E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter. Diverging Trends in National and Local Concentration. Working Paper 25066, National Bureau of Economic Research, Sept. 2018.

J. J. Rotemberg and M. Woodford. Chapter 16 The cyclical behavior of prices and costs. In *Handbook of Macroeconomics*, volume 1, pages 1051–1135. Elsevier, Jan. 1999.

C. Syverson. Macroeconomics and Market Power: Context, Implications, and Open Questions. *Journal of Economic Perspectives*, 33(3):23–43, Aug. 2019. ISSN 0895-3309.

J. Traina. Is Aggregate Market Power Increasing? Production Trends Using Financial Statements. SSRN Scholarly Paper ID 3120849, Social Science Research Network, Rochester, NY, Feb. 2018.

J. Weiss. Intangible investment and market concentration. Technical report, Job Market Paper, New York University, 2020.

W. F. Wong. The Round Trip EïňĂect: Endogenous Transport Costs and International Trade. page 73, 2018.

J. M. Wooldridge. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114, Sept. 2009. ISSN 0165-1765.