Knowledge Diffusion, Markups, and Cohorts of Firms*

Umberto Muratori†

This version: October 23, 2020

Click here for the latest version

Abstract

This paper investigates the differences in markups between and within cohorts of US firms. I document substantial between-cohort differences and a relatively flat profile over time within cohorts. The paper uses administrative patent data to provide suggestive evidence that knowledge creation and diffusion explain these patterns. Namely, the between-cohorts pattern is associated with improvements in the innovation quality, and the within-cohort pattern is the result between the interaction of innovation and knowledge diffusion. Motivated by this new empirical evidence, I develop a general equilibrium endogenous growth model of creative destruction augmented with knowledge diffusion. I build the model for two purposes. First, I estimate the changes in the intensity of knowledge diffusion. I find knowledge diffuses 38% faster in 2010 than it did in 1980. Second, I quantify the effect of changes in the innovation step size and intensity of knowledge diffusion on growth and welfare. The quantitative exercises show the consumption-equivalent welfare increases by 0.29% if the innovation step size and intensity of knowledge diffusion increase from their 1980 to their 2010 values. By contrast, the counterfactual experiments highlight the increase in these parameters has no substantial effect on growth. Changes in the innovation quality and the speed of knowledge spillover can be achieved by reforming the patent system, such as changing the *non-obviousness requirements* or the *patent term*. The findings suggest policymakers, who aim to maximize the welfare, should choose short patent terms and low innovative-step requirements.

JEL Codes: D4, E2, L12, O31, O33, O34

Keywords: Markup, Cohort of firms, Innovation, Knowledge Diffusion, Patent Policy

---

*I am deeply indebted to my advisor Toshihiko Mukoyama for his continuous support, guidance, and encouragement. I have benefited from the comments and the teaching of Adriana Kugler, Dan Cao, Mark Huggett, Minsu Chang, and Margit Reischer, among others. I am grateful to my colleagues Carolina Concha, Maria Hernandez-de-Benito, Pedro Juarros, Juan Margitic, Arturo Romero Yanez, Mayra Sáenz, Yangfan Sun, and Daniel Valderrama for their feedback. I acknowledge the generous financial support from the Washington Center for Equitable Growth. All errors are my own.

†Georgetown University; Economics Department; 3700 O St NW, Washington, DC 20057; Email: um22@georgetown.edu

1 Introduction

Market power is commonly measured by markups, defined as the wedges between the selling prices of a good and the costs of variable inputs used in its production.[1] Markups have exponentially increased in the US economy starting from the 1980s.[2] Most regions of the world, although with different timings, have experienced similar trends.[3] Market power is linked to several long-term macroeconomic patterns, such as declining labor and capital shares (Barkai, 2020; De Loecker et al., 2020), widening income inequality (Ennis et al., 2019), lowering investment and productivity (Gutiérrez and Philippon, 2017), increasing input misallocation (Peters, 2013; Baqaee and Farhi, 2020), and a slowdown in business dynamism (Edmond et al., 2018).[4] Markups also affect the magnitude of a recession and the speed of the recovery.[5]

Knowledge creation and diffusion are important determinants of markups (Andrews et al., 2015; Akcigit and Ates, 2019). Knowledge creation is the production, through innovation, of new technologies that improve productivity. Knowledge diffusion consists of the flow of existing technologies from the technological leaders to their competitors. Knowledge creation and diffusion affect markups through changing the technological gap between the leader and the followers. One important mission of the patent system is to guarantee a balance between stimulating invention and promoting knowledge diffusion.[6] Patent-reform legislation has been central to the policy debate of the US Congress since the beginning of the 2000s,[7] and legislative changes have sought to find a better balance between knowledge creation and diffusion.

This paper builds on these two pillars, market power and knowledge creation/diffusion, and investigates how the size of the technological gap and the speed of catching up with the frontier affect markups, growth, and welfare. To achieve this goal, the paper addresses three research questions. First, what are the markup patterns by cohorts of firms in the US? Second, what are the mechanisms that account for these patterns? Third, what are the welfare and growth implications of changes in the innovation step size and intensity of knowledge diffusion? The answers to these questions will contribute to the recent academic and policy debate on markups, innovation, and knowledge diffusion.

To answer these questions, I proceed in three steps. In the first step, I document new facts about the markup heterogeneity using US firm-level data. Namely, I explore the differences in markups between and within cohort of firms. A cohort includes all firms born in a specific year. The analysis highlights two

[1]Some studies, for example, Azar et al. (2017), Grullon et al. (2019), and Gutierrez and Philippon (2018), use market concentration to measure market power.

[2]De Loecker and Eeckhout (2017), De Loecker et al. (2020), and Hall (2018) show markups have increased. The literature has not reached a unanimous consensus. Anderson et al. (2018) proxy markups with gross margins and find markups in the retail sector are relatively stable over time. Rossi-Hansberg et al. (2020) focus on product-market concentration rather than markups, and document diverging trends with an increase in concentration at the national level and a decrease in concentration at the local level.

[3]In addition to North America and Europe, the increase is also common to Latin America and Asia. See De Loecker and Eeckhout (2018) and Diez et al. (2018) for the description of the different geographical trends.

[4]See Syverson (2019) for an overview of the macro market literature.

[5]Nekarda and Ramey (2020) argue markups are pro-cyclical, and pushing demand up during downturns speeds up recovery. Other authors argue markups are counter-cyclical (Rotemberg and Woodford, 1999; Galí et al., 2007) or acyclical (Anderson et al., 2018; Nekarda and Ramey, 2011). Heise et al. (2020) argue that an increase in the market concentration that affects pass-through from labor costs to prices and explains the lack of inflation after the most recent recessions.

[6]See OECD (2004) for an extensive review of the evolution and challenges of patent systems in the OECD countries.

[7]Hall (2009) and Ouellette and Williams (2020) discuss some legislative proposals.

new findings. First, at the time of their entry, new firms charge higher markups than existing firms, thus generating substantial between-cohorts differences. I denote this pattern as the "cohort effect." Second, the average markup within cohorts remains relatively flat over time.

I quantify the effect of the between- and within-cohort patterns on aggregate markup for the US economy by implementing two thought-experiments. In the first experiment, I only remove the differences between cohorts, whereas in the second experiment, I abstract from the growth in markups within firms. I find the persistence in markups jointly with the increase in market shares makes the initial differences across cohorts of firms quantitatively sizeable for the aggregate. My estimates suggest the weighted-average markup in the US economy would have been about 5%-10% lower if the differences at entry were removed. For example, the aggregate markup in 2010 would have been 40% over the marginal cost instead of 48%, implying a decline of 8 percentage points. By contrast, removing the markup growth over the firm life cycle shows no clear pattern. Whereas the simulated series of aggregate markups would have been lower than the observed one in the 1960s, the opposite would have been true for the 1970s. Since the mid-1980s, removing the markup growth over the life cycle has no substantial effect on the aggregate markup.

I provide some suggestive evidence that links the between- and within-cohort patterns to knowledge creation and diffusion. I show the "cohort effect" is linked to improvements in the innovation quality over time. A higher quality of entrants' innovation generates greater technological gaps over the competitors and higher markups. Because incumbents also innovate, innovation becoming more radical is not sufficient to explain the flatness of the markup profile within cohorts. The interaction of two forces explains the within-cohort markup profile: innovation and knowledge diffusion. I provide support for this mechanism by implementing an indirect test.[8] Innovations widen the technological gap over the competitors and increase markups. In the absence of innovations, competitors learn about the technological frontier through a process of knowledge diffusion. Thus, the technological gap narrows and markups decrease. Motivated by this intuition, I explore the markup changes after the occurrence of innovations, controlling also for other potential determinants of markups. The results show markups first significantly increase, and then decline. I interpret this result as support for knowledge creation and diffusion as important determinants of markups over the life cycle of a firm.

In the second part of the paper, I build a parsimonious general equilibrium Schumpeterian growth model of creative destruction augmented with knowledge diffusion to account for the empirical findings. The model is in the tradition of the pioneering works of Grossman and Helpman (1993) and Aghion and Howitt (1992).[9] The framework assumes imperfect product markets. Firms engage in a Bertrand competition in which the efficiency leader drives all lower-productivity competitors out of the market by practicing "limit pricing." This pricing rule implies the leader optimally sets the price of a product equal to the marginal cost of production of the follower, defined as the second-most-efficient producer. Because the leader has a

[8]Due to data limitations, I do not observe knowledge diffusion.

[9]Different versions of this theoretical framework have been recently used in the growth literature, among others, Acemoglu and Akcigit (2012), Akcigit and Kerr (2018), Atkeson and Burstein (2019), Akcigit and Ates (2019), and Acemoglu et al. (2020).

technological advantage over the follower, markups are positive and proportional to the technological gap.

Innovation decisions are endogenous and widen the technological gap. The model features two sources of knowledge creation: innovation by entrants and innovation by incumbents. The first case consists of the process of creative destruction in which entrants displace incumbents by improving the production efficiency of an intermediate good. The second source of growth comes from incumbent firms that successfully innovate in their product line and increase the technological gap with the follower. Thus, successful innovations by the leader not only improve productivity but also increase markups.

I augment this theoretical framework by introducing a process of knowledge diffusion. The follower gradually learns to replicate the frontier technology via an exogenous technological catching-up. As a consequence, if the leader does not innovate, the follower narrows the technological gap and erodes the monopolistic rents of the leader.

Two model parameters, the innovation step size and the intensity of knowledge diffusion, discipline the evolution of markups and account for the between- and within-cohort patterns. Changes in the innovation step size proxy for changes in the innovation quality. Higher innovation step sizes mean more radical innovations, wider technological gaps, and greater markups. Changes in the intensity of knowledge diffusion affect the speed at which followers catch up with the technology frontier, and consequently, the markups that the leader can charge. Changes in these parameters affect welfare through three separate channels: growth, labor reallocation, and input misallocation.

In the third part of the paper, I first estimate the two model parameters of interest: the innovation step size and the intensity of knowledge diffusion. Based on the model intuition, I can compute the innovation step size directly from the data. The results show the innovation step size increased between 1980 and 2010. The intensity of knowledge diffusion is not directly observable, and the theoretical structure helps in estimating this parameter. I derive a meaningful interpretation of the knowledge diffusion process by calculating the half-life. The half-life measures the periods needed for the follower to catch up with half of her technological gap. The estimates suggest that in 1980, the half-life for a follower whose leader was one innovation step ahead was eight years. Estimates in 2010 highlight that the half-life decreased by 38%, from eight to five years.[10] This result pairs with the view that the information and communication technology (ICT) revolution has opened new perspectives for accessing a broader set of external information and speeding up the diffusion of knowledge across firms.

I then use the estimated model to run counterfactual experiments and examine the effect of changes in the innovation step size and intensity of knowledge diffusion on growth and welfare. Although the two parameters are not modeled as endogenous to specific policy changes, their changes may be achieved by reforming the patent system. The *non-obviousness requirement* is the height of the inventive step required for a patent application to be granted.[11] The higher the required contribution, the more radical innovation

[10] The literature has not reached a consensus on the evolution of knowledge diffusion in recent decades. Whereas Baslandze (2016) documents an increase in knowledge diffusion, Akcigit and Ates (2019) find the opposite result.

[11] See OECD (2004) for further details.

efforts become. Similarly, changes in the *patent term*, the maximum time during which a patent can be maintained in force, or the *timing of codified information disclosure* affect the speed of knowledge diffusion. Indeed, shorter patent terms or faster disclosure imply higher intensity of knowledge diffusion.[12]

The first set of counterfactual experiments changes one of these two parameters at a time. The scope of these experiments is to understand the mechanisms through which the innovation step size and intensity of knowledge diffusion affect welfare. The quantitative exercises identify three channels: growth, labor reallocation, and input misallocation. Changes in the innovation step size and intensity of knowledge diffusion move these channels in opposite directions. Increases in the innovation step size boost growth and consequently increase the welfare. Increases in the innovation step size also imply greater monopolistic rents that incentivize investment in innovation reallocating labor away from production. Lower production is associated with lower consumption and welfare losses. Finally, changes in the innovation step size affect the shape of the distribution of markups. Markup dispersion affects input misallocation and production efficiency. By contrast to increases in innovation step size, increases in the intensity of knowledge diffusion lower growth, boost production through labor reallocation, and have ambiguous effects on production efficiency.

The last counterfactual exercise quantifies the changes in welfare and growth by jointly moving the innovation step size and the intensity of knowledge diffusion from their 1980 to their 2010 values. The results point out that the representative household would require an increase of 0.29% in consumption to remain in the 1980 baseline economy rather than moving to the counterfactual economy with the 2010 parameter values. By contrast, the counterfactual experiments highlight that the increase in these parameters has no substantial effect on growth. Namely, the growth rate increases by 0.07 percentage points, which corresponds to 1.6% of the 1980 growth rate. The quantitative results also suggest policymakers, who aim to maximize the welfare, should target a high intensity of knowledge diffusion and a small innovation step size. In terms of policy, they should choose short patent terms and low innovative-step requirements. By contrast, policymakers can achieve a high growth rate by choosing the opposite combination of policies.

This paper contributes to four branches of the literature. First, I build on the fast-growing literature on markups and their potential consequences on macroeconomic outcomes. This paper does not contribute to the methodological debate on how to best estimate markups,[13] but rather uses state-of-the-art methodology and explores sources of heterogeneity in markups that, to my best knowledge, previous studies have not investigated. This paper contributes to both the empirical and the quantitative literature on markups. Empirically, I document the differences in markups between and within cohorts of firms and the determinants of these patterns. Quantitatively, similarly to Edmond et al. (2018) and Boar and Midrigan (2019), I quantify the welfare implications connected to markup changes.

[12] OECD (2004) provides some examples of the connection between patent terms and technology diffusion. Baruffaldi and Simeth (2020) study the empirical linkages between the timing of codified information disclosure and knowledge diffusion.

[13] Edmond et al. (2018), Traina (2018), Flynn et al. (2019), and Bond et al. (2020) raise methodological concerns on the estimation of markups. Edmond et al. (2018) argue the individual markups should be aggregated using the variable-cost shares rather than the sale shares. Traina (2018) argues that the rise in markups is due to the increase in marketing and management costs. Flynn et al. (2019) show standard methods to estimate production functions do not identify markups and propose an alternative identification strategy.

The second contribution is to the literature on knowledge diffusion and imitation.[14] Most of these papers are theoretical studies. Recently, Baslandze (2016) and Akcigit and Ates (2019) estimated the evolution of the intensity of knowledge diffusion and find opposite trends. This paper contributes to this debate by using a different approach to estimate the evolution of knowledge diffusion. My results point in the same direction as Baslandze (2016), emphasizing an increase in knowledge diffusion over the last four decades.

The third contribution is to the literature on innovation policy. The main focus of this literature has been the investigation of the effect of innovation taxes and subsidies on R&D efforts and innovation occurrences (Bloom et al., 2002; Atkeson and Burstein, 2019; Akcigit et al., 2020; Bloom et al., 2019).[15] Unlike the previous literature that studies the frequency of innovation, this paper examines the changes in the innovation quality and its effect on growth and welfare.

Finally, this paper is related to the literature on misallocation. Misallocation is an important determinant of productivity and growth.[16] Peters (2013) develops a growth model and quantifies the effect of markups on misallocation. Differently from Peters's (2013) work that focuses on the welfare effect of relaxing entry barriers, I study the effect of changes in the innovation step size and the intensity of knowledge diffusion on misallocation through changes in the endogenous markup distribution.

The remainder of the paper is structured as follows. Section 2 presents the empirical strategy to estimate markups. Section 3 describes the data. Section 4 presents the new empirical findings. In section 5, I introduce the structural model. In section 6, I estimate the model parameters and assess the fit of the model. Section 7 presents the results of the numerical exercises. Finally, I conclude in section 8.

2 Markups Computation: Empirical Strategy

The computation of markups relies on the supply-side approach proposed by De Loecker and Warzynski (2012) based on the work of Hall (1988). This approach does not require any explicit assumptions on the demand system and the competition structure between firms. The production-based approach relies on the cost minimization of variable inputs of production. The key assumption is that within one period, firms adjust variable inputs, whereas capital is subject to adjustment costs and does not adjust within one period.

The main theoretical takeaway from the work of Hall (1988) is that markups can be expressed as

$$\mu_{it} = \beta_{it}^{V} \frac{P_{it}^{Q} Q_{it}}{P_{it}^{V} V_{it}}, \tag{1}$$

where β_{it}^{V} is the output elasticity of variable input, $P_{it}^{Q} Q_{it}$ is the value of output, and $P_{it}^{V} V_{it}$ is the total cost

[14] Among others, Luttmer (2007), Luttmer (2012), Alvarez et al. (2013), Perla et al. (2014), Perla and Tonetti (2014), Lucas and Moll (2014), Lucas and Moll (2014), Sampson (2016), and Lashkari (2018). Buera and Lucas (2018) provide an extensive review of the literature on knowledge diffusion.

[15] Another branch of the literature, for example, Gilbert and Shapiro (1990) and Acemoglu and Akcigit (2012), has studied the theoretical implications of intellectual-property-rights protection on innovation.

[16] A vast literature measures the impact of misallocation on productivity. A non-exhaustive list includes Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Restuccia and Rogerson (2013), and Midrigan and Xu (2014). Restuccia and Rogerson (2017) provide an excellent overview of state of the art in this topic.

for variable inputs. The derivations of this theoretical result are reported in Appendix A.

I use industry-specific Cobb-Douglas production functions that employ variable input and capital:

$$q_{it} = \beta_s^v v_{it} + \beta_s^k k_{it} + \omega_{it} + \varepsilon_{it} \qquad \forall s,$$

where q_{it} measures log sales, v_{it} is the log variable costs, k_{it} is the log capital stock, ω_{it} is the log productivity, and ε_{it} is the error term. The estimated output elasticity, β_s^v, differs in two respects from the theoretical elasticity derived from the cost minimization problem, β_{it}^V. First, the theoretical elasticities vary at the firm level. Due to estimation limitations, I assume a production function is common to all firms in an industry. Thus, the estimated elasticities are at the industry level rather than the firm level. Second, the theoretical elasticities in equation (1) vary over time. In the empirical specification, I assume these elasticities remain constant over time.[17]

The identification of the input elasticity, β_s^v, is challenging because the input choice and firm productivity may be correlated. Because productivity ω_{it} is unobservable, we cannot separate the contribution of the elasticity from the contribution of productivity. As a consequence, the elasticity estimates may be biased.

To solve this transmission bias, I implement a control function approach that consists of two-stages. In the first stage, I draw from the work of Olley and Pakes (1996) that shows unobserved productivity is a function of the firm inputs. I remove the idiosyncratic measurement error from the production process by estimating the productivity term as a non-parametric function of the firm inputs:

$$q_{it} = \beta_s^v v_{it} + \beta_s^k k_{it} + h(v_{it}, k_{it}) + \varepsilon_{it} \qquad \forall s,$$

where $h(v_{it}, k_{it})$ is the non-parametric function.[18]

In the second stage, I use the predicted sales to derive the implied productivity as a function of elasticity parameters. Conditional on the assumption that the productivity follows an AR(1) process, $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$. I project the productivity on its lag to recover the productivity shock as a function of the elasticity. Then, I solve the following moment condition to recover the flexible input elasticity:

$$\mathbb{E}\left[\xi_{it}(\beta) v_{it-1}\right] = 0.$$

Finally, I compute the markups using the formula in equation (1):

$$\mu_{it} = \beta_s^v \, \frac{P_{it}^Q Q_{it}}{P_{it}^V V_{it}},$$

[17] One could estimate time-varying elasticities but would face a trade-off between time variation in the elasticities and the level of industrial aggregation. De Loecker and Eeckhout (2017) and De Loecker et al. (2020) show estimating time-varying elasticities instead of constant elasticities does not significantly affect the results.

[18] I approximate the $h(v_{it}, k_{it})$ function by second-order polynomials of the input variables and their cross products. In the empirical model, I also include time fixed effects to capture time-invariant unobservable heterogeneity.

where β_s^v is the time-invariant elasticity for the variable input in sector s, $P_{it}^Q Q_{it}$ measures the total value of sales, and $P_{it}^V V_{it}$ measures the expenditure in flexible inputs.

The identification strategy uses the lag of flexible input expenditure as an instrument. The lag of variable input use is a valid instrument for the current use of flexible input if two crucial identification assumptions are satisfied. First, the variable input use responds to contemporaneous productivity shocks, whereas the lag variable input use does not. The first assumption guarantees the exogeneity of the instrument. The second assumption requires that the lag and current use of variable input are correlated. This second assumption guarantees the relevance of the instrument. In subsection 4.3, I discuss an alternative identification strategy of the elasticity for variable inputs.

3 Data

The primary dataset used in this paper is the Fundamental Annual Compustat file from Wharton Research Data Services. The collected data cover 1950-2015 and include only publicly traded firms.[19] For the scope of this research, the availability of the longest possible period and broad coverage of economic activity drive the choice of the data. The Compustat file classifies firms according to the North American Industry Classification System (NAICS).[20] The dataset contains information on sales values (item SALE), variable input "Cost of Goods Sold" (item COGS), and the physical capital from the annual financial statements.[21] I derive the capital stock by the recursive use of a perpetual-inventory method similar to Stein and Stone (2013). The method consists of fixing the level of physical capital stock for the first period of activity of a firm at the reported value of net plants, property, and equipment (item PPENT). The levels of capital stock for the following periods are updated as the previous capital stock discounted by price-level changes and depreciation plus the capital expenditure (item CAPX).[22] For consistency, all items are converted from nominal to real terms by using price deflators. I use the nonresidential fixed-investment good deflator (line 9 of the NIPA Table 1.1.9) to transform the physical capital stock and the GDP deflator (line 1 of the NIPA Table 1.1.9) to convert the remaining variables.

The sample of firms includes only US domestic firms. I select those firms by using the standard industry format observations in USD with Foreign Incorporation Codes (item FIC) in the US. I exclude two groups of firms from the sample: (1) Public Utilities firms (SIC codes between 4900 - 4999), because heavily regulated on prices; and (2) Finance, Insurance, Real Estate firms (SIC codes between 6000 - 6999), because their balance sheets are not comparable with non-financial firms.[23] To guarantee data quality, I remove observations with negative or missing total assets (item AT), sales (item SALE), cost of goods sold (item

[19] All items are recorded at the end of the fiscal year.

[20] For the estimation of markups, I define an industry based on the 3-digit North American Industry Classification System (NAICS). I also compare the benchmark results with the estimation obtained by defined an industry based on the Standard Industrial Classification (SIC).

[21] The assumptions of the production approach imply one should recover the same markup from any variable input. Raval (2019) compares markups estimated using different measures of variable inputs and finds some differences in the estimation of markups. In this paper, I use variable costs as a measure of variable inputs.

[22] The capital stock is winsorized to be non-negative in each period. The annual depreciation rate is assumed to be 10%.

[23] The exclusion of these two industry classification groups accounts for a loss of approximately $3,000$ distinct firms.

COGS), or net plants, property, and equipment (item PPENT). To avoid picking up merger and acquisition distortions, I also exclude observations in which acquisitions are larger than 5% of the value of total assets. I also remove all firms that do not follow the Income Statement Model 1 (item ISMOD), because the operating-expense items are reported in a different format from the one described above.[24]

Finally, the Compustat file identifies records by GVKEYs that refer to securities, not to firms. A single organization may correspond to multiple entries within the Compustat data. To uniquely identify firms, I construct a new firm identifier. In most cases, the unique firm identifier is the same as the Compustat GVKEY. However, in few cases, multiple GVKEYs are associated with a single identifier.[25] The final sample is an unbalanced panel between 1950 and 2015 with approximately $20,000$ distinct firms and with about $265,000$ firm-year observations.

4 Empirical Facts

This section contains three subsections. In the first subsection, I present new empirical facts about markups focusing on the between- and within-cohorts differences. In the second subsection, I provide some suggestive evidence that links these empirical patterns with knowledge creation and diffusion. In the last subsection, I implement robustness checks on the definition and computation of markups.

4.1 Markups by Cohorts of Firms

The first contribution of this paper is to document new facts about the heterogeneity in markups. For ease of the exposition, in the remainder of the paper, I denote a firm as an "entrant" in the current year if that firm went public this year. Similarly, I denote a firm as "incumbent" in the current year if that firm went public any year before the current. Finally, a "cohort" includes all entrants from a specific year.

Figure 1 shows the average markup by selected cohorts of firms, and it highlights two main empirical results.[26] First, a between-cohorts comparison shows entrants have higher markups than incumbents.[27] In the remainder of the paper, I refer to this pattern as the "cohort effect." Second, in addition to a between-cohorts comparison, Figure 1 also shows the evolution of markups within cohorts of firms. The results point out that the average markup profile within cohorts is relatively flat over time for most of the cohorts. To my best knowledge, this study is the first to document these empirical regularities.[28] Figure B.1 in Appendix

[24] The excluded firms, because of a different structure of the Income Statement, are approximately $6,500$ firms.

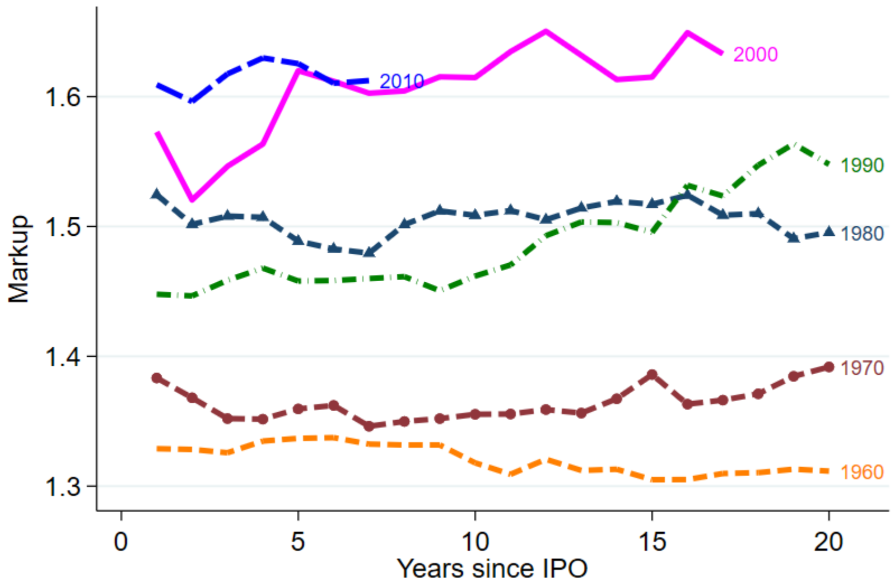
[25] For $1,259$ GVKEYs, no one-to-one relationship exists with the new firm identifier. These $1,259$ GVKEYs are grouped into 595 new firm identifiers.

[26] The macro market literature has debated the correct weighting measure to evaluate the aggregate effect. De Loecker et al. (2020) use sales-weighted markups to study the increase in aggregate markups. Edmond et al. (2018) show costs-weighted markups are the proper weighting to evaluate the aggregate welfare effects. The results in Figure 1 report the unweighted markups. The use of weighted rather than unweighted markups would simultaneously combine two effects: the change in markups and the change in weights. As Figure 1 aims to isolate the changes in markups by cohorts of firms from other factors, without investigating welfare effects, the unweighted markups seem to be more reasonable metrics.

[27] For graphical purposes, I compute the unweighted average markups as the average of three years around the reported date. I also trim the distribution of markups by the top and bottom 1%. The relaxation of these restrictions produces qualitatively similar results.

[28] The findings from this paper do not contradict the conclusions of Foster et al. (2008) that entering businesses have lower prices than incumbents. Still, they are more productive, and consequently lower marginal costs. One could simultaneously

Figure 1: Markups by Cohorts of Firms



B presents two robustness checks. The first uses the SIC instead of the NAICS. The industrial classification matters in computing the elasticities of output to variable inputs. The second robustness check restricts the analysis to a balanced sample of firms. Both robustness checks substantially confirm the previous findings.

A brief discussion of the age-period-cohort (APC) identification problem is worthwhile. Meaningfully partitioning change into these three dimensions with statistical methods is not straightforward, because the three dimensions are collinear. In principle, different combinations of these three effects can produce the same result. The "cohort effect" could result from a combination of the period and age effects operating in opposite directions. Although an alternative combination is a possibility, it seems unlikely, and it is more natural to think of the results above as "cohort effects."

I formally test the differences between cohorts of firms by using firm-level data to compare the markups of entrants with the markups of incumbents. I implement the following empirical specification:

$$\log(\mu_{its}) = \beta Ent_{its} + \delta X_{its} + \nu_s + \nu_t + \varepsilon_{its}, \tag{2}$$

where μ_{its} is the markup for firm i at time t operating in industry s. The main variable of interest is Ent_{its}, which takes the value of 1 if firm i is an entrant in time t, and 0 otherwise. I include a set of industry and time fixed effects. The inclusion of these fixed effects implies the comparison between entrants and incumbents is within industry s and time t. Finally, I include a set of time-varying controls X_{its} that account for other potential determinants of markups such as the years after the IPO, the market share by industry, the asset value, and the number of employees. I also control for the unemployment rate and the cyclical deviation of the real gross domestic product from its trend at the time of entry for firm i. These variables capture the

observe entrants having lower prices and higher markups than incumbents.

long-lasting effect of initial conditions on firm performance.

The results of the specification (2) are reported in Panel A of Table 1. The inclusion of the fixed effects and the time-varying controls decreases the magnitude of the difference in markups between entrants and incumbents, but it does not affect the direction and significance of the estimate. The estimate from column 2 implies markups of entrants are, on average, 1% higher than the markups of incumbents. These results reinforce the conclusion that entrants charge significantly higher markups than incumbents and that differences exist between cohorts of firms.

Table 1: Between and Within Cohorts Effects

	(1)	(2)		(3)	(4)
Panel A: Cohort Effect			*Panel B: Markup Growth*		
Entrants	0.03***	0.01**	Years of Activity	-0.00***	0.00
	(0.00)	(0.00)		(0.00)	(0.00)
Observations	210,823	190,556	Observations	210,823	190,556
R^2	0.00	0.32	R^2	0.00	0.53
Industry FE	No	Yes	Industry FE	No	Yes
Time FE	No	Yes	Cohort FE	No	Yes
Controls	No	Yes	Controls	No	Yes

I use the specification in equation (3) to test the evolution of markups over the life cycle:

$$\log(\mu_{itcs}) = \beta \, \textit{Years after IPO}_{itcs} + \delta X_{itcs} + \nu_s + \nu_c + \varepsilon_{itcs}, \tag{3}$$

where μ_{itcs} is the markup for firm i belonging to cohort c at time t operating in industry s. The main regressor is $\textit{Years after IPO}_{itcs}$ counts for firm i the number of years passed between the IPO year and time t. I include industry and cohort fixed effects. The set of time-varying controls, in addition to the ones previously mentioned for specification (2), includes the firm's initial markup and proxies for the business cycle.

The results reported in Panel B of Table 1 highlight that once we control for other sources of markup variation, markups, on average, do not change significantly over time. Furthermore, in addition to not being significant, the estimates also suggest the economic magnitude of the effect is relatively small. After 10 years, the average markup increases by only 0.06% due to the firm aging. The dispersion of markups increases over the years of business activities. The average standard deviation of markups in the first five years after the IPO is 0.40, whereas the average dispersion within 10 to 15 years after the IPO is 0.49.[29] I interpret the increase in dispersion jointly with the estimates as evidence that some firms increase their markups over the life cycle, whereas other firms lower their markups over time.[30] This results suggests markups over a

[29]For comparability, the standard deviations are computed restricting the sample to a balanced panel.

[30]Most of the standard theories of markup formation predict increasing markups over the life cycle. For example, customer base models (Foster et al., 2016; Gourio and Rudanko, 2014) state that firms charge lower markups to accumulate customer

firm's life cycle may either rise or fall. In the next subsection, I investigate a potential mechanism that may generate these differences in firm behavior.

Before investigating the determinants of these patterns, I quantify their contribution to the changes in the aggregate markup from 1950 until recently. To achieve this goal, I implement two thought-experiments. In the first experiment, I remove the differences across cohorts. I proceed in three steps. I first compute the markup growth rates by firm and the percentage deviation of each entrant from the cohort average. I then calculate the average markup for incumbents starting from the first year containing both entrants and incumbents. Using the entrants' percentage deviations from their cohort average, I generate an adjusted distribution for entrants that has the same shape of the original distribution but an average shift to match the mean of the incumbents' distribution. I then use the firm-specific growth rates to update the adjusted series of markups from t to $t+1$. I sequentially repeat the same procedure for all years.

The second experiment aims to remove the growth in markups within firms. I construct an adjusted series in which the markup of a firm is set equal to its initial markup for all periods in which the firm operates. Both adjusted series are weighted by the firm-specific market shares. This weighting strategy presents an important caveat. If the sale share in a period depends on the level of markup in that period, adjusting the markups without changing the share may create some bias in the results.[31] In these experiments, I abstract from this potential measurement issue.

Figure 2: Thought-Experiments Comparison

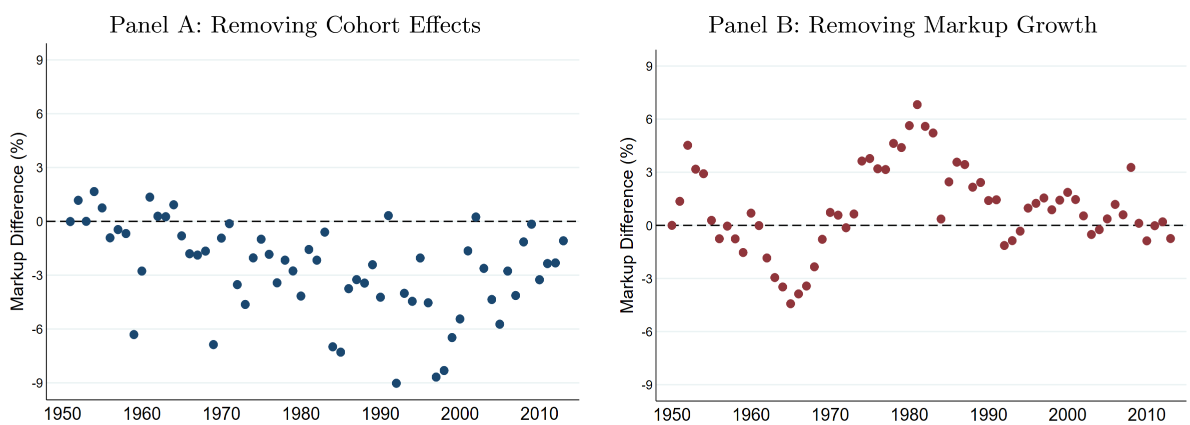


Figure 2 shows the percentage difference between the adjusted and the actual series of markups. Panel A refers to the first experiment in which the cohort effect is removed, and Panel B reports the result for the second experiment in which the markup growth within firms is removed.[32] The black dashed line is the zero

capital. Once the customers are attached to a product, and in the presence of transition costs, firms can raise prices without losing customers. A different family of models relates markups to firms' market shares (Atkeson and Burstein, 2008; Edmond et al., 2018). In these models, when market shares increase, as shown empirically, goods within a sector are more substitutable than goods across sectors, and therefore markups also increase. The mechanisms proposed in these theories face some challenges in explaining declines in markup over the life cycle.

[31] Autor et al. (2017) and Van Reenen (2018) argue that "superstar" firms have larger market shares but also higher profit margins and markups. Markups and market shares may be related, so changing the markups but keeping the market shares unchanged may create some bias.

[32] Figure B.2 in Appendix B replicates this exercise using the variable inputs weights, as argued by Edmond et al. (2018).

line in which no differences exist between the adjusted and observed series. The two panels show different patterns.

Panel A highlights that the weighted-average markup in the US economy would have been 5%-10% lower if the differences at entry were removed. For example, the aggregate markup in 2010 would have been 40% over the marginal cost instead of 48%, implying an 8-percentage-point decline. The differences in markups between cohorts of firms produce a substantial aggregate effect because firms' market shares increase over time. Indeed, approximately 70% of firms have larger market shares five years after the IPO date. Although the contribution of entrants at the time of the IPO is quantitatively small, the markup persistence and the increase of market shares over the life of a firm imply the initial differences become gradually more important for the aggregate. De Loecker et al. (2020) find the increase in markups is mostly due to a *reallocation* of market shares from low- to high-markup firms. My findings are consistent with this result and provide some additional insights. Indeed, the reallocation occurs in part toward firms belonging to more recent cohorts that start with a higher initial markup.

Panel B shows no clear effect on the aggregate markup from removing the markup growth over the life cycle. Although the simulated series of aggregate markup would have been lower than the observed one in the 1960s, the opposite would have been true for the 1970s. Since the mid-1980s removing the markup growth over the life cycle has no substantial effect on the aggregate markup. In fact, the aggregate markup would have been only slightly higher by removing the within-firm markup growth.

In sum, these results emphasize that differences exist in markups between firms belonging to different cohorts and that the differences at entry contribute significantly to the documented change in markups. The results also highlight that the average markup within cohorts remains substantially unchanged over time. This trend is the result of firm-specific markups changing in opposite directions and offsetting each other.

4.2 Mechanism: Innovation and Knowledge Diffusion

Given the importance of these patterns for the aggregate effect, in this subsection, I dig into the mechanisms that may explain these empirical facts. This paper argues the between- and within-cohorts patterns are linked to innovation and knowledge diffusion. Other mechanisms may generate the same patterns of markups. This paper focuses on these specific mechanisms and provides some suggestive evidence in support of this explanation.

I proxy innovation using patent administrative data. I collected the patent data from PatentsView, which contains the universe of granted patents from the US Patent and Trademark Office (USPTO). The patent microdata contain a rich set of information such as the assignee identifier, application year, type of patent, citations' network, and technology class to which a patent belongs.[33] I restrict the analysis only to utility patents that account for approximately 98% of the universe of patents granted by the USPTO. Following the common practice in the patent literature, the application year identifies the year when innovation occurs.

[33] See Hall et al. (2001) and Lerner and Seru (2017) for an overview on the patent data.

I match the patent-assignee identifiers to the Compustat file by using the crosswalk constructed by Autor et al. (2019).[34] The overall sample includes 963,597 patents between 1976 and 2013 granted to 6,910 distinct firms. Because the administrative files include the universe of all granted patents by the USPTO, Compustat firms with no match in the patent dataset have zero granted patents.

The first hypothesis is that the "cohort effect" is linked to improvements in the quality of innovation. If the quality of entrants' innovation increases, entrants have a greater technological gap over the competitors and they charge higher markups.[35] An increase in the quality of innovation implies successful innovations have become more radical over time. To test this hypothesis, I measure innovation quality by following Acemoglu et al. (2020). The first indicator of innovation quality is the average number of citations per patent within a five-year window after the application year. I use the "quasi-structural" approach proposed by Hall et al. (2001) to adjust for changes over time in the propensity to cite. The second statistic is the tail innovation index and measures the share of patents that a firm owns with citations above the 95^{th} percentile. The last metric is the generality index devised by Hall et al. (2001). The generality index is computed as 1 minus the sum of the squared shares of citing patents in a technology class, and it measures the dispersion of the citations received by a patent in terms of the technology classes of citing patents. These three measures capture complementary aspects of innovation quality.

Table 2: Measures of Innovation Quality

	1980	1990	2000
Citations per Patent	1.414	2.316	3.804
Tail Innovation Index	0.167	0.124	0.138
Generality Index	0.348	0.394	0.412
Share of Innovators	5.1%	9.6%	9.0%

Table 2 shows the evolution of the innovation-quality statistics for entrants in the last three decades.[36] The statistics suggest that the innovation quality has increased over time. The average citations per patent in 1980 were 1.4, and were 3.8 in 2000. Similarly, patents have also become more general. The share of patents in the top 5^{th} percentile in terms of citations is the only indicator that shows a small decline from 17% in 1980 to 14% in 2000.[37] In addition to improvements in the innovation quality, the average markup of a cohort may also increase because a larger share of entrants are innovators. In this respect, the last row

[34] Although patent data are available until the end of 2019, the crosswalk only links patents until 2013. Therefore, I restrict the end year of this part of analysis to 2013.

[35] This mechanism is a standard building block of endogenous growth models with a quality ladder.

[36] The estimates are not reported for 2010, because they suffer from a well-known "truncation bias." Patents are released by USPTO only once they are granted. The approval process may be lengthy, implying that at the end of the sample, the number of patents is miscounted. Because the crosswalk only reports patents granted until 2013, the "truncation bias" may be severe. Although not reported, the estimates for 2000 show an increase in the innovation quality.

[37] Note the quality of innovation for incumbents has also increased similarly to entrants, which implies entrants are not the only ones in more recent years to be making more radical innovation.

of Table 2 reports the share of innovators among entrants. This share doubled between the 1980 and 2000 cohort, moving from 5% to about 10%.[38] The increase in the innovation quality is not sufficient to explain the markup profile within cohorts. Indeed, because incumbents also innovate, one would expect a much steeper within-cohort markup profile.

The second hypothesis is that the evolution of markups over the life cycle of a firm is explained by the interaction of two forces: innovation and knowledge diffusion. Due to data limitations, I do not observe the process of knowledge diffusion. I provide supportive evidence for these mechanisms by implementing an indirect test through the investigation of markups. The main intuition of the test is the following. Innovations widen the technological gap over the competitors and increase markups. In the absence of innovations, competitors learn about the technological frontier through a process of knowledge diffusion. Knowledge diffusion narrows the technological gap and markups decrease. In the empirical test, I explore the markup changes after the occurrence of innovations.[39] To avoid spurious changes in markups due to compositional changes, I restrict the analysis to a balanced panel of firms. I estimate the following empirical specification:

$$\log(\mu_{ist}) = \sum_{k=0}^{5} \beta_k D_{kit} + \delta X_{ist} + \nu_t + \nu_s + \varepsilon_{ist},$$

where D_{kit} is a dummy that takes the value of 1 if the index variable is equal to k for firm i in year t and 0 otherwise. The index variable measures the number of years between t and the year of the innovation. I also include time and industry fixed effects to absorb the heterogeneity along these dimensions. Finally, I also control for individual characteristics, such as the years after IPO and the firm's market share and size, to capture other factors that determine the markup level. Figure 3 reports the results. The red solid line represents the percentage changes relative to the markup in the year before the innovation occurs. The dashed lines delimit the 90% confidence interval.

Figure 3 is consistent with the above-mentioned intuition. After an innovation occurs, markups first significantly increase and become about 2% higher than their level in the year before innovation. After this initial increase, markups start to decline. I interpret this result as support for knowledge creation and diffusion as important determinants of markups over the life cycle of a firm.

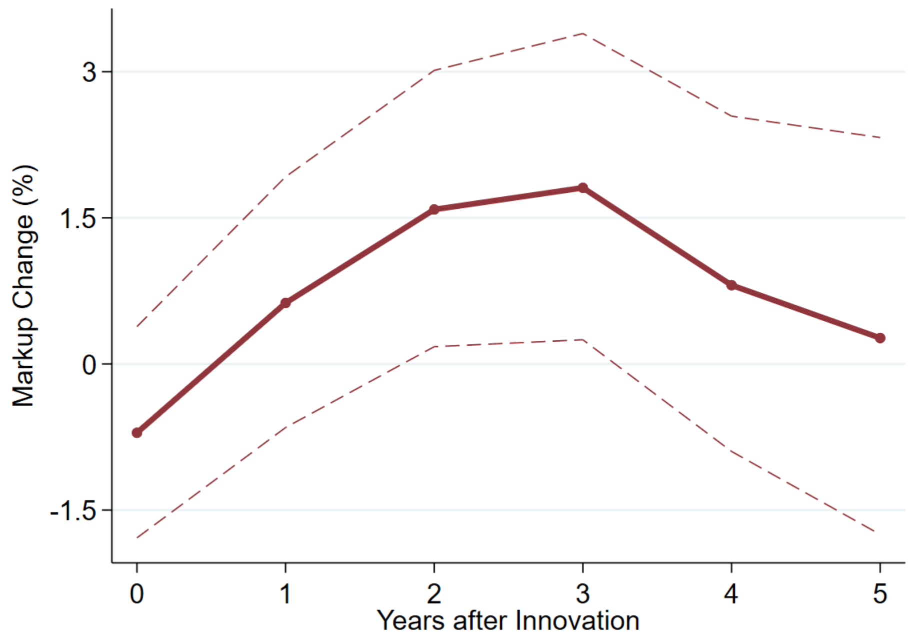
These results are consistent with previous studies that have also highlighted the importance of innovation on the markup increase. IMF (2019) argues the increase in markups has been widespread across advanced economies and industries, but within them, the increase has been more marked for more productive and innovative firms. De Ridder (2019) shows the increase in markups can be explained by the increase in fixed costs, and that firms with higher fixed costs spend more on R&D. In this paper, I do not focus on innovation inputs such as expenditures in R&D, but on innovation outcomes such as successful patenting.[40]

[38] An alternative explanation could be that entrants enter new markets where they face low competition. To address this possibility, I compute, but do not report, the market shares for entrants by industry. The results show these shares are very small and do not vary over time making entrants unlikely to face lower competition in the markets they enter.

[39] I exclude from the sample firms that innovate in multiple years. Including firms that repeatedly innovate may bias the results, because the frontier technology keeps moving, and one cannot isolate the effect of knowledge diffusion.

[40] Higher R&D expenses do not necessary turn into better innovation. Innovation inputs like R&D and innovation outputs

Figure 3: Markup Changes



Furthermore, De Loecker et al. (2020) and Diez et al. (2018) show the increase in markups is concentrated in the top 10% of the markups distribution. I find approximately 46% of the firms in the highest decile of the markup distribution by industry have innovated at least once within the previous five years. The results from this paper confirm the increase in the aggregate markup is connected to innovation activities, and substantial heterogeneity exists among firms.

4.3 Robustness Checks

Recent studies have argued the results from De Loecker and Eeckhout (2017) may be subject to some methodological and measurement drawbacks. For a robustness check, I implement alternative identification strategies and measurement refinements to show the consistency of the results from the previous subsections.

First, Flynn et al. (2019) argue elasticities are not identified under the standard assumptions proposed in De Loecker and Warzynski (2012). The authors argue the lag of variable inputs has no power as an instrument for the variable input in the proxy model. Conditional on the inputs chosen before period t, the only variation in the contemporaneous flexible inputs, v_t, comes from the productivity shock, ξ_t. Because ξ_t is assumed to be uncorrelated with the inputs chosen in the period before t, the lagged flexible inputs, v_{t-1}, are not valid instruments for the contemporaneous flexible inputs, v_t. Flynn et al. (2019) propose estimating the elasticity by assuming constant returns to scale (CRS) in the production function, specifically $\beta_v + \beta_k = 1$. The CRS assumption removes a parameter from the problem, and it allows a number of parameters to be equal to the number of instruments with capital as the only instrument with power. In

like patents are complimentary measures.

their specification, the following moment condition is solved to recover the elasticities:

$$\mathbb{E}\left[\xi_{it}(\beta)k_{it}\right] = 0.$$

Second, an important challenge in the markups estimation is to accurately measure the input costs.[41] Traina (2018) shows marketing and management costs have increased from roughly 12% in 1950 to 22% in 2015. As firms have increasingly devoted more of their inputs to marketing and management costs in the last few decades, the exclusion of this item from the production function may bias the elasticity estimates. The Compustat file also contains a measure of overhead, booked under "Selling, General and Administrative Expenses" (item XSGA). This item includes selling expenses, general operating expenses, and administration expenses. I follow De Loecker et al. (2020) by considering an augmented production technology whereby the overhead expenses are included as a separate factor of production in addition to the variable inputs.[42] Precisely, I estimate the following industry-specific production function:

$$q_{it} = \beta_v^s v_{it} + \beta_k^s k_{it} + \beta_{ov}^s ov_{it} + h(v_{it}, k_{it}) + \varepsilon_{it},$$

where the production function from section 2 also includes the overhead costs, ov_{it}.

Appendix B replicates the previous empirical analysis for these two alternative computations of markups. This robustness analysis substantially confirms the results presented above.

5 Model

The previous facts emphasize the importance of changes in the innovation quality and knowledge diffusion for markups. Motivated by these facts, I construct a model of firm dynamics. I use this theoretical framework to estimate the effect of changes in the two parameters, innovation step size and intensity of knowledge diffusion, on welfare and growth. In the model, these two parameters proxy for the innovation quality and the intensity of knowledge diffusion, respectively.

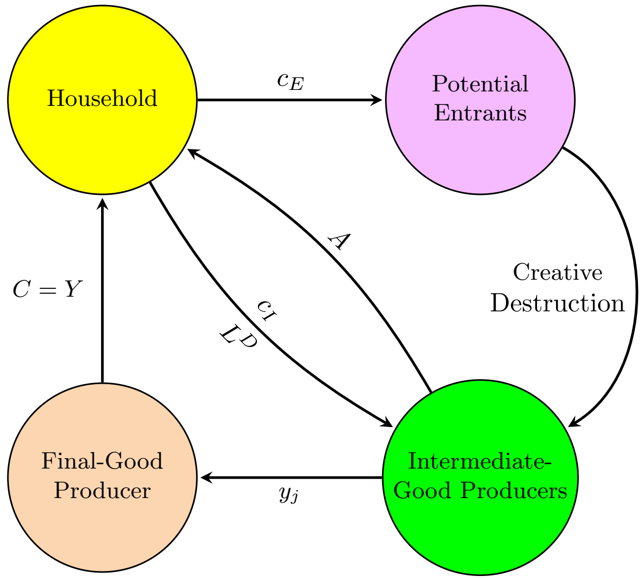
The model consists of a parsimonious growth model with endogenous innovation decisions along a quality ladder in the tradition of Grossman and Helpman (1993) and Aghion and Howitt (1992). I augment this standard model by introducing an exogenous knowledge diffusion process. The main intuition of the model is that markups charged by incumbent firms are the result of two opposite forces: innovation and knowledge diffusion. Whereas innovation improvements by incumbents increase markups, knowledge diffusion and technological catch-up by the second-most-efficient producer decline markups.

The economy is inhabited by four types of agents: the household, the final-good producer, intermediate-

[41] See De Loecker et al. (2020), Gutiérrez and Philippon (2017), and Traina (2018).

[42] The main difference in the estimation of markups between De Loecker et al. (2020) and Traina (2018) is in the accounting of the overhead expenses. Traina (2018) attributes them to variable costs, whereas De Loecker et al. (2020) include them as separate inputs in the production function. As argued in the corporate finance literature, the item XSGA includes in part variable costs but also investment costs, such as training expenses. Because the separation between the two is not straightforward, I consider them a different input in production.

Figure 4: Model Structure



good producers, and potential entrants. To simplify the exposure of the economy, Figure 4 illustrates a stylized structure of the model and highlights the interactions between the different types of agents. The representative household consumes the final good constrained by asset holdings and labor earnings that come from the household supplying labor for production and innovation efforts. The asset holdings come from owning the intermediate-good producers that generate profits, because they operate in a non-competitive environment. Differentiated product lines produced by intermediate-good producers are bundled together into the final good. Intermediate-good producers exit the market if a potential entrant successfully innovates in their product line. The remainder of this section describes the features and the optimization problems of each of these types of agents in the economy.

5.1 Household

The economy is populated by a representative household who receives utility only from an aggregate bundle of consumption C. The model is in discrete time and goes to infinity. The household supplies inelastically one unit of labor and maximizes the present value of the future stream of consumption:

$$U = \sum_{t=0}^{\infty} \beta^t \log(C_t), \tag{4}$$

subject to the budget constraint

$$C_t + A_{t+1} = (1 + r_t)A_t + w_t L, \tag{5}$$

where β represents the discount rate, A are the total asset holdings at the beginning of the period, r is the interest rate, w is the wage rate, and L is the labor supply. Because the household owns the firms, the asset market clearing condition states that in equilibrium, the total asset holdings are equal to the sum of the intermediate-good producers' values V_j:

$$A_t = \int_0^1 V_{jt} dj.$$

The logarithmic preference implies a simple relationship between the growth rate, the interest rate, and the discount factor. This relationship is highlighted by the standard Euler equation

$$\frac{C_{t+1}}{C_t} = \beta(1 + r_{t+1}). \tag{6}$$

5.2 Final-Good Producer

The final good Y_t is consumed by the representative household. The final good is a composite consumption bundle of a continuum of differentiated intermediate products of measure one. The final good is produced in a perfectly competitive sector by using a Cobb-Douglas technology:

$$Y_t = \exp\left[\int_0^1 \log(y_{jt}) dj\right], \tag{7}$$

where y_{jt} is the differentiated intermediate product j. I assume the final good Y_t to be the numeraire; thus, the aggregate price index defined as

$$P_t = \exp\left[\int_0^1 \log(p_{jt}) dj\right]$$

is normalized to one in each period. For simplicity, I also assume the final-good producer aggregates the different intermediate products without needing to use any input.

The problem of the final producer is essentially a static problem; thus, the demands for each intermediate good j are derived as solutions of the following maximization problem:

$$\max_{y_{jt}} \left\{ P_t Y_t - \int_0^1 p_{jt} y_{jt} dj \right\}. \tag{8}$$

The zero-profit condition implies the demand function for each intermediate product j is the ratio between the total value of the final good and the price of the intermediate good:

$$y_{jt} = \frac{Y_t}{p_{jt}}.$$

The final-good producer eases the exposition of the model and acts as a veil. One could reformulate the model by removing the final-good producer and by allowing consumers and firms to combine the intermediate goods on their own.

5.3 Intermediate-Good Producer

An intermediate-good producer f takes aggregate prices as given. An intermediate-good producer is defined by the intermediate good j that it produces. I assume an intermediate producer can produce at most one product line.[43] Thais assumption implies a measure one of intermediate producer. For ease of exposition, I drop the firm subscript when no risk of confusion exists.

Intermediate-good producers use a linear production function to produce the product line j with labor as the only input in the production. Specifically,

$$y_{jft} = q_{jft}\, l_{jft},$$

where l_{jft} is firm f's labor demand at time t, and q_{jft} is the productivity that is assumed to be strictly positive. The production function implies one unit of labor produces q_{jft} units of the intermediate good. Because labor is the only input in production, it follows that the marginal cost of producing one unit of intermediate good is equal to

$$mc_{jft} = \frac{w_t}{q_{jft}}.$$

Different firms producing the same good j differ in production efficiency, but the goods they produce are perfect substitutes. In each period t and each market j, intermediate producers engage in a Bertrand competition. I apply the standard tie-breaking rule. In the Bertrand-Nash equilibrium, the efficiency leader in market j practices "limit pricing." In other words, the leader optimally sets the price of good j equal to the productivity-adjusted marginal cost of production of the second-most-efficient potential producer. The leader drives out all lower-productivity competitors from that market and captures all demand and production for good j. I also assume labor is fully mobile, such that wage rate w equalizes across firms and product lines. Denote f' as the second-most-efficient producer in market j. The "limit pricing" rule implies

$$p_{jt} = mc_{jf't}.$$

This model structure generates in equilibrium positive markups that are linked to the productivity differentials between firms. In each period t, markup μ for product j is computed as

$$\mu_{jt} = \frac{p_{jt}}{mc_{jft}} = \frac{q_{jft}}{q_{jf't}}. \tag{9}$$

Positive markups exist in equilibrium as the consequence of the efficiency differential between the leader and the second-most-efficient producer. To provide an intuition of this result, the markup is defined as

[43] The model can be easily extended by considering multi-products firms as in Klette and Kortum (2004). This extension does not change the main mechanism and intuition of the model. Quantitatively, the results would be different because incumbents entering new product lines would charge markups as high as new entrants. The Compustat data files do not include product-level information. The markups, as explained in section 2, are computed at the firm-year level using the supply-side approach. As a consequence, I can observe only one markup per firm and per year. Therefore, modelling the intermediate-good producers as single-product firms produces a better mapping between the available data and the model structure.

the ratio of the selling price over the marginal cost of production. The marginal cost is linked to the efficiency. Because the efficiency leader charges as a price that is the marginal cost of the second-most-efficient producer, the implication is that its marginal cost of production is below the price it charges for the good. This mechanism generates positive markups in equilibrium.

The profits for the market leader in product line j are computed by combining the pricing rule with the product demand and the expression for markups. Formally,

$$\Pi_{jt} = Y_t\left(1 - \mu_{jt}^{-1}\right).$$

The labor demand for the market leader can be expressed in terms of the markups by using the production technology, the pricing rule, and the formula for markups. Specifically,

$$l_{jt} = \frac{Y_t}{w_t}\mu_{jt}^{-1}.$$

By combining the demand functions for intermediate goods and the price rule, we get that

$$y_{jt} = \frac{q_{jf't}}{w_t}Y_t.$$

This formulation emphasizes that labor demands and profits are both functions of markups. The evolution of the distribution of markups is a function of the productivity of the different lines across firms. In turn, the productivity distribution changes endogenously due to innovation and exogenously due to knowledge diffusion. Therefore, the evolution of the technological gaps is essential to understand the evolution of markups.

5.4 Technological Gaps

The evolution of the relative efficiencies determines markups and all other intermediate-good producers' equilibrium quantities. I define z_{jt} as the innovation gap in time t between the leader and the second-most-efficient producer in product line j, $\frac{q_{jft}}{q_{jf't}}$. The model features innovation and technological catch-up. Innovators choose the frequency at which their next product innovation arrives. The cost of innovation is in terms of labor units used in production.

Leaders: The leader in product line j at time t is defined as the firm that produces the intermediate good j at time t. Innovation decisions for incumbents, or leaders, aim to improve the relative quality of the product they own. I assume an incumbent can innovate only in the product line that it owns. An incumbent firm f generates a $0 \leq x \leq 1$ flow Poisson rate of innovation. The cost function is in terms of units of labor and is represented by the following function:

$$\mathcal{C}_I(x, z) = c_I(x, z)w = \left(\psi_I \frac{x^\gamma}{z}\right)w, \tag{10}$$

where $\psi_I > 0$ is a scale parameter that regulates the efficiency of innovation for incumbents, $\gamma > 1$ ensures the convexity of the cost function, and $c_I(x, z)$ is the labor demand for innovation activities. Successful innovations improve efficiency by factor $\lambda > 1$. In other words, the efficiency of an incumbent that successfully innovates at the beginning of $t+1$ is $q_{jft+1} = \lambda q_{jft}$. As noted from equation (10), the cost function is decreasing in the efficiency advantage accumulated by an incumbent up to the current period.

Followers: The follower in product line j at time t is defined as the firm with the second-highest efficiency in producing intermediate good j at time t. I assume followers exogenously catch-up with the frontier technology. I model the catch up step as an exponential process. The efficiency of the follower improves by a factor e^{α} for each period in which the leader does not innovate, $q_{jf't+1} = e^{\alpha} q_{jf't}$. In other words, the innovation gap between leaders and followers shrinks over time if the leader does not successfully innovate. I also assume a follower cannot replace a leader by catching up with the frontier. Thus, the follower's efficiency in period t is bounded by the efficiency of the leader, q_{jft}.

Potential Entrants: A potential entrant is defined as a firm that attempts to replace the current producer in a product line. I denote this process as *creative destruction*. Therefore, in this model, firms endogenously exit. Each period has a mass of size one of potential entrants. I assume entrants do not know ex ante in which product line they will innovate, and the resulting R&D efforts may realize in any product line with equal probability. I refer to this assumption as innovation efforts to be undirected across all product lines. One important implication of this assumption is that no space for strategic interactions is available among firms in choosing which product line to innovate. In the remainder of the paper, I focus on the case of positive entry. A potential entrant uses units of labor to generate a $0 < \tau \leq 1$ flow Poisson rate of innovation according to the following innovation cost function:

$$\mathcal{C}_E(\tau) = c_E(\tau) w = \left(\psi_E \tau^{\gamma}\right) w, \tag{11}$$

where $\psi_E > 0$ accounts for the efficiency in innovation for entrants.

A potential entrant, who successfully innovates, replaces the current intermediate-good producer in market j. Entrants build on the state-of-the-art technology. In other words, they improve the frontier quality by a factor λ. I finally assume the displaced incumbent becomes the second-most-efficient producer. Note the model predicts the innovation gap for entrants is always equal to λ. Furthermore, in equilibrium, the markup charged by firms is equal to the innovation gap. Thus, the markups charged by entrants is also equal to λ. I use this prediction from the model to estimate the parameter λ, as explained in section 6.

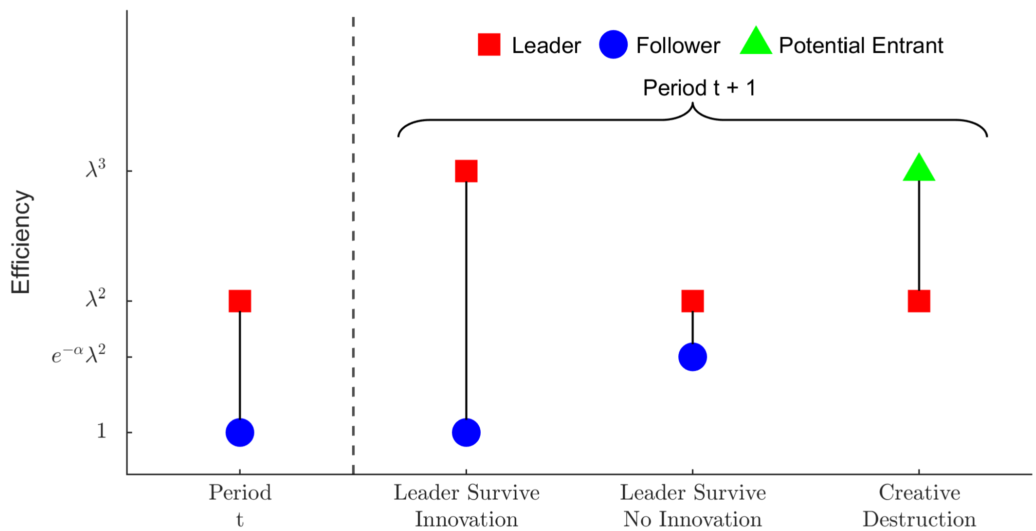
In sum, given the current relative technology gap in product line j, z_{jt} evolves according to the following law of motion:

$$z_{jt+1} = \begin{cases} \lambda z_{jt} & \text{with prob. } (1-\tau_t)x_{jt} \\ \max\{1, e^{-\alpha} z_{jt}\} & \text{with prob. } (1-\tau_t)(1-x_{jt}) \\ \lambda & \text{with prob. } \tau_t, \end{cases} \tag{12}$$

where τ and x_j are endogenous, and they are the decision rules for potential entrants and incumbents in product line j, respectively. The two parameters, λ and α, are the main object of interest in this paper. Section 6 estimates these parameters and section 7 investigates how changes in these two parameters affect growth and welfare.

Figure 5 shows the possible evolution between period t and period $t+1$ of the leader's and follower's quality and the innovation gaps generated by these changes.

Figure 5: Possible Innovation Gaps in a Product Line



At time t in Figure 5, the leader has a technological advantage over the follower equal to λ^2. There possible cases may occur in $t+1$. The first scenario corresponds to the case in which the leader survives in $t+1$ and successfully innovates. The efficiency of the leader increases by a factor λ. Because the efficiency of the follower does not change, the technological gap widens from λ^2 to λ^3, implying higher markups as well. The second scenario occurs when the leader survives but fails to innovate. In this case, the quality of the leader remains unchanged, but the follower learns about the technology frontier via a process of knowledge diffusion. The improvement in the follower's efficiency leads to a shrinking of the technological gap between the leader and the follower and an erosion of the leader's markups. Finally, the third case represents the process of creative destruction. In this scenario, a potential entrant successfully innovates in this product line and displaces the leader, who becomes the follower. By assumption, the technological gap between entrants and previous leaders is set to be equal to λ.

5.5 Dynamic Programming

The firm optimization problems are expressed as standard dynamic-programming problem. I first describe the timing of events and then the individual firm maximization problem. For ease of exposition, I omit the time subscript t if not necessary. The events in the model occur in the following order:

1. At the beginning of each period t, potential entrants' innovation efforts from the previous period realize. Successful innovators enter the market and become entrants. Displaced incumbents exit and become the followers.

2. The incumbents' innovation efforts from the previous period realize. The distribution of innovation gaps is determined based on the realization of the previous period's innovation activities.

3. Existing firms and potential entrants choose their innovation efforts that will affect the innovation gap distribution in $t+1$. At the same time, existing firms choose their labor demands. The labor market clears, and the intermediate goods are produced. The final-good producer bundles the intermediate goods, and they are sold to the household.

4. The representative household consumes and makes the asset decisions subject to her resource constraint. All markets clear. Period $t+1$ starts.

Incumbents: In each period t, an incumbent maximizes its value function by choosing the investment in innovation, x. The state space of the maximization problem consists only of the innovation gaps that an incumbent has relative to the second-most-efficient producer in its market:

$$V_t(z) = \max_{0\leq x\leq 1} \left\{ \begin{array}{c} \Pi_t(z) \ - \mathcal{C}_{It}(x,z) + \\ \dfrac{1}{1+r_{t+1}} \left[(1-\tau_t) \left[x_t V_{t+1}(\lambda z) + (1-x_t) V_{t+1}(\max\{1, e^{-\alpha} z\}) \right] \right] \end{array} \right\}. \tag{13}$$

The first term accounts for the profits that a firm earns. The second term is the innovation cost that incumbents undertake to improve their current innovation gaps. The third term represents the capital gains following successful innovation, and the last term accounts for capital losses due to followers' catching up.

Potential Entrants: A potential entrant decides its innovation intensity before learning the product line in which it enters if it successfully innovates. As all potential entrants are equal in all the dimensions, they solve the same maximization problem. A potential entrant chooses innovation rate τ. The optimal innovation decision is equal to

$$\tau_t^* \equiv arg \max_{0<\tau\leq 1} \left\{ -\mathcal{C}_{Et}(\tau) + \frac{1}{1+r_{t+1}} \tau_t V_{t+1}(\lambda)] \right\}. \tag{14}$$

The FOC of the potential entrants' optimization problem implies

$$\tau_t^* = \left(\frac{V_{t+1}(\lambda)}{(1+r_{t+1})\psi_E \gamma w_t} \right)^{\dfrac{1}{\gamma - 1}}. \tag{15}$$

Equation (15) implies the entry rate is a function of the value of a potential entrant if entry occurs. A decrease in firm value at time $t+1$ will disincentivize potential entrants from investing in innovation and entering the market, because the profits from operating in the market are lower.

5.6 Market Clearing Conditions

Because the economy is closed and no government exists, the structure of the economy implies the household consumes the final good:

$$Y_t = C_t. \tag{16}$$

The labor supply L is fixed, time invariant, and normalized to one. The labor market condition implies

$$L = L_t^D + I_t, \tag{17}$$

where L_t^D is the aggregate labor demand used for production by the existing firms,

$$L_t^D = \int_0^1 l_{jt} dj = \frac{Y_t}{w_t} \left(\int_0^1 \mu_{jt}^{-1} dj \right),$$

and I_t are the units of labor used for innovation activities,

$$I_t = c_{Et}(\tau) + \int_0^1 c_{Ijt}(x, z) dj.$$

The labor market clearing condition implies the units of labor hired for production and the units of labor used for innovation activities have to be equal to the household labor supply.

Finally, the asset market clearing condition states that the household asset holdings equalize the value of intermediate-good producers:

$$A_t = \int_0^1 V_{jt} dj. \tag{18}$$

5.7 Stationary Equilibrium

Definition 1. A stationary equilibrium, or *Balanced-Growth Path* (BGP), consists of a set of aggregate variables $\{Y, C, A\}$, a set of prices $\{r, w\}$, a set of individual allocations $\{x_j, l_j, y_j, p_j, \tau\}_j$, and an initial distribution of innovation gaps f_0 such that

- All aggregate variables $\{Y, C, A, r, w\}$ grow at a constant rate g.
- The representative household chooses a flow of consumption C to maximize the utility function (4) subject to the budget constraint from Equation (5) and the no-Ponzi condition.
- The final-good producer chooses the intermediate good demands y_j to maximize the profit function in equation (8) subject to the technology from equation (7) and the price normalization for the final good.
- Intermediate-good producers choose an allocation of x_j to solve the dynamic problem in equation (13) subject to the law of motion of efficiency.

- Potential entrants choose an innovation effort τ to solve the maximization problem in equation (14).
- The product, labor, and asset markets clear according to equations (16), (17), and (18), respectively.
- The cross-sectional distribution of markups $f(\cdot)$ is stationary and consistent with the law of motion of the state variable from equation (12) and the optimal policy decisions.

Along the balanced-growth path, consumption grows at the same rate as output, the Euler equation in (6) implies

$$1 + g = \beta(1 + r),$$

where the constant growth rate g is a function of the innovation step size λ, and the optimal innovation decisions of incumbents and entry rate

$$g = \log(\lambda)\left[\tau + \int_0^1 x_j dj\right]. \tag{19}$$

Along the balanced-growth path, the distribution of markups is stationary. Thus, the labor utilization that is a function of the markups is also stationary. This result implies the economy-wide growth comes only from changes in the efficiency index. In turn, efficiency improves upon successful innovations generated by either entrants or incumbents. Note a decline in the innovation gap between the leader and the follower does not affect the growth rate. As the economy exhibits perpetual growth, to solve the stationary equilibrium, I first normalize the economy to make it stationary, by dividing the aggregate variables by the efficiency index Q. I denote the normalized variables with a hat ($\hat{}$).

All equilibrium quantities for the intermediate-good producers can be expressed in terms of technology gaps. By the definition of z, the expression for markups for product j from equation (9) turns to be

$$\mu_{jt} = z_{jt}.$$

It follows that labor demand for product j is given by

$$l_{jt} = \frac{Y_t}{w_t} z_{jt}^{-1},$$

and, finally, the profit function for product line j is computed as

$$\Pi_{jt} = \left(1 - z_{jt}^{-1}\right) Y_t.$$

The previous results imply that to track the distribution of markups and its moments, we need to derive the distribution of innovation gaps. Because the innovation decisions are endogenous, the solution of the intermediate-good producers' optimization problem tracks the evolution of the innovation gaps.

The characterization of an analytical solution for the incumbent optimal innovation decisions is not straightforward, because the innovation gap is bounded below by one. The innovation decisions depend on the current level of innovation gap that an incumbent has relative to the second-most-efficient follower. I derive the incumbents' optimal innovation decisions by applying standard dynamic-programming techniques. Although an analytical solution cannot be derived, the characterization of the economy in terms of innovation gaps greatly facilitates the computation of the equilibrium.

Finally, to close the model, I compute the stationary distribution $f(z)$ by solving the fixed point of the contraction mapping $f \to \mathbf{T}f$, where $\mathbf{T}f$ gives the probability mass in any state for the next period given the current probability function. In the stationary distribution, the mass of firms with any innovation gap z remains unchanged. The model contains three sources of movement in the innovation gaps. Creative destruction is the first source of changes. The remaining two are linked to the incumbents' innovation process. Successful innovations increase the individual firm innovation gap, whereas failures in innovating shrink the gap. The mass of firms in any state z in the next period is given by

$$\int_1^{z'} \mathbf{T}f(z)dz = (1-\tau)\left[\int_{\lambda z \leq z'} x_I(z)f(z)dz + \int_{e^{-\alpha}z \leq z'} \big(1 - x_I(z)\big)f(z)dz\right] + \tau \mathcal{I}_{\{z'=\lambda\}},$$

where the first term accounts for the contribution of surviving incumbents, and the second term measures the mass of entrant firms. The indicator function $\mathcal{I}_{\{z'=\lambda\}}$ takes the values of 1 if the innovation gap is equal to λ that is the gap the entrants start with, and 0 otherwise.

5.8 Welfare Analysis

Changes in the innovation step size and intensity of knowledge diffusion have an ambiguous effect on the household's welfare. The model highlights three mechanisms through which these parameters influence welfare. These channels operate in the opposite direction, leading to an ex-ante uncertainty on the overall effect. Furthermore, these three channels operate in different directions for changes in either the innovation step size or the intensity of knowledge diffusion.

The first channel operates through the growth rate. Equation (19) highlights that the growth rate depends on the innovation step size and the R&D efforts. Increases in the intensity of knowledge diffusion slow down growth though a decline in the innovation efforts. Indeed, increases in knowledge diffusion cause a decline in the monopolistic rents that firms enjoy. As a consequence, the value of successful innovation decreases and discourages firms from investing in innovation. Increases in the innovation step size have a direct and an indirect effect on growth leading to higher growth rates. The direct effect is represented by an increase in the step size, and the indirect effect is represented by higher innovation rates, due to the fact that successful innovation is more profitable. Changes in the growth rate also affect welfare. Because all aggregate variables grow at the same rate, lower (higher) growth rates lead to lower (higher) equilibrium consumption, leading to welfare losses (gains).

The second channel acts through labor-force reallocation. Whereas increases in the intensity of knowledge diffusion cause drops in R&D efforts and a decline in labor usage for R&D activities, increases in the innovation-step size move these variable in the opposite direction. The labor market clearing condition induces labor reallocation between R&D activities and production. An increase (a decline) in labor demand for production also boosts (depresses) production. The resource constraint links changes in output with changes in consumption. Higher (lower) levels of consumption are associated with welfare gains (losses).

Finally, the last channel operates through changes in production efficiency. Changes in both parameters alter the distribution of markups and its dispersion. The dispersion of markups across firms negatively affects aggregate productivity. This result dates back to Lerner (1934), who argues that in the absence of cross-sectional differences in markups, relative prices would correctly signal relative marginal costs. Therefore, changes in the distribution of markups play an important role in quantifying the input misallocation. Following Peters (2013), the distortionary effect can be summarized by the *efficiency wedge*:

$$\mathcal{M} = \frac{\exp\left(\displaystyle\int_0^1 \log(\mu_{jt}^{-1})dj\right)}{\displaystyle\int_0^1 \mu_{jt}^{-1}dj}. \tag{20}$$

Increases in either of the two parameters have ambiguous effects on the production efficiency.

Because the three channels may result in overall welfare gains or losses, in section 7, I quantify the aggregate effect of changes in these two parameters on welfare. To derive a meaningful interpretation of welfare changes, I compute the consumption-equivalent welfare. This measure corresponds to the consumption change that the representative household would require to be indifferent between staying in the baseline versus the counterfactual economy. Mathematically,

$$\mathcal{W}^b = \sum_{t=0}^{\infty} \beta^t \log(C_t^b(1+\xi)) = \sum_{t=0}^{\infty} \beta^t \log(C_t^c) = \mathcal{W}^c,$$

where $\mathcal{W}$ is the welfare measure, the superscript b refers to the baseline economy, the superscript c refers to the counterfactual economy, and ξ is the share of additional consumption required by the household to be indifferent between the two economies. ξ is derived from the previous expression as

$$\xi = e^{(1-\beta)(\mathcal{W}^c - \mathcal{W}^b)} - 1.$$

Positive values of ξ imply the household would be better off in the counterfactual economy rather than the baseline. Negative values have the opposite interpretation.

6 Calibration

6.1 Parameters Identification

Table 3 summarizes the baseline calibration of the model. The model is in discrete time, and the model period is set to one year. The parameters to calibrate consist of $\{\beta, \gamma, \lambda, \psi_I, \psi_E, \alpha\}$.

The first two parameters $\{\beta, \gamma\}$ are exogenously calibrated, and they are relatively standard in the literature. The annual discount factor β is set to the value of 0.947, which is consistent with the findings of Cooley and Prescott (1995). The curvature of the R&D cost function γ is equal to 2. This value is taken by Blundell et al. (2002) and Hall and Ziedonis (2001), who estimate the elasticity of patents to R&D expenditures to be equal to 0.5. This elasticity corresponds to a quadratic R&D cost function.

The remaining four parameters are calibrated based on data moments from 1980.[44] The model states a one-to-one mapping exists between the entrants' markup and the size of the innovation step λ. The model assumes the innovation gap for entrants is always equal to λ. Furthermore, in equilibrium, the markup charged by firms is equal to the innovation gap. Thus, the markups charged by entrants is also equal to λ. I calibrate the innovation step to be equal to the average markup of new publicly traded firms. The latter is directly observed in the data.

Table 3: Parameters Calibration, 1980

Parameter	Parameter Description	Parameter Value	Target Moment Description	Data Moment	Model Moment
β	Discount Rate	0.947	–	–	–
γ	Curvature of R&D Cost Function	2	–	–	–
λ	Innovation Step	1.527	Entrants' Average Markup	–	–
ψ_I	Innovation Efficiency: Incumbent	49.323	R&D-Expenditure-to-Output Ratio	0.018	0.018
ψ_E	Innovation Efficiency: Entrant	11.900	Entry Rate	0.076	0.076
α	Intensity Knowledge Diffusion	0.025	Incumbents' Average Markup	1.368	1.353

The last three parameters are calibrated via a GMM estimation procedure in which I exactly identify the model by using three moments conditions to calibrate the three parameters. The model parameters are jointly estimated by minimizing the following distance function:

$$\min \sum_{i=0}^{3} \left\{ \frac{|\text{model}_i - \text{data}_i|}{0.5\big(|\text{model}_i| + |\text{data}_i|\big)} \right\},$$

where model_i represents the i model-generated moment, and data_i is the i data-generated moment.[45] The

[44] To alleviate the effect that business-cycle fluctuations may have on the estimated moments, I compute all the data moments as averages over the period 1978-1982.

[45] The proposed distance function is in the spirit of Akcigit and Kerr (2018) and Lentz and Mortensen (2008).

distance function implies I assign an equal weight to each of the moment conditions. The model-generated moments are computed either as solutions of the stationary equilibrium or as results of the model simulation. The detailed computational algorithm is described in Appendix C.

The R&D expenditure for incumbents' innovation depends on the incumbents' innovation efficiency ψ_I. A variation in ψ_I affects both the incumbents' innovation decisions and the cost of innovation. Given the optimal innovation decisions, I identify ψ_I by matching an R&D-to-sales ratio equal to 0.018. The model counterpart of this moment is the ratio between the cost of innovation over the output. This moment is computed from the solution of the stationary equilibrium.

The entrants' innovation efficiency ψ_E is calibrated by targeting an entry rate of 7.6%. This value is calculated as the share of new publicly traded firms relative to the number of firms that were already publicly traded. The model counterpart is part of the solution of the stationary equilibrium problem.

Finally, the markups for existing firms depend on the innovation and catching-up processes. Given the incumbents' optimal innovation decisions and λ, I discipline α by matching an average markup for existing firms of 1.368. The average markup of existing firms is computed by simulating the model. The simulated economy consists of $5,000$ firms. I simulate the model until the distribution converges. Once the economy has converged to its stationary equilibrium, I compute and collect the simulated moments.[46]

6.2 Fit of the Model

The last two column of Table 3 assess the extent to which the calibration matches the targeted moments. The model generates the entry rate and the ratio of R&D expenditure to output that perfectly match the data-generated moments at the third digit. The model produces an average markup for incumbents that is slightly smaller than the one observed in the data (1.353 vs. 1.368). The difference between the data- and model-generated average markup for incumbents is approximately 1%. Overall, despite the parsimonious structure, the model successfully fits all targeted moments.

The most straightforward interpretation of the parameter α is to convert this value in the half-life. The half-life measures the periods needed for the follower to catch up with half of the technological gap that she has with the leader. The exponential knowledge diffusion process implies the half-life $\mathcal{T}$ needed to catch up with a technological gap of size z is equal to

$$\mathcal{T} = -\frac{1}{2}\Big(\frac{1}{\alpha}\Big)\log\Big(\frac{1}{z}\Big). \tag{21}$$

The previous expression highlights that the catch-up time depends on the size of the technological gap. If the technological frontier is one-innovation-step-ahead technology, λ, the half-life is approximately eight years, assuming the leader does not innovate in the meantime. A technological gap of λ resembles a situation in which the leader is a new entrant. A second insightful case consists of calculating the periods needed to

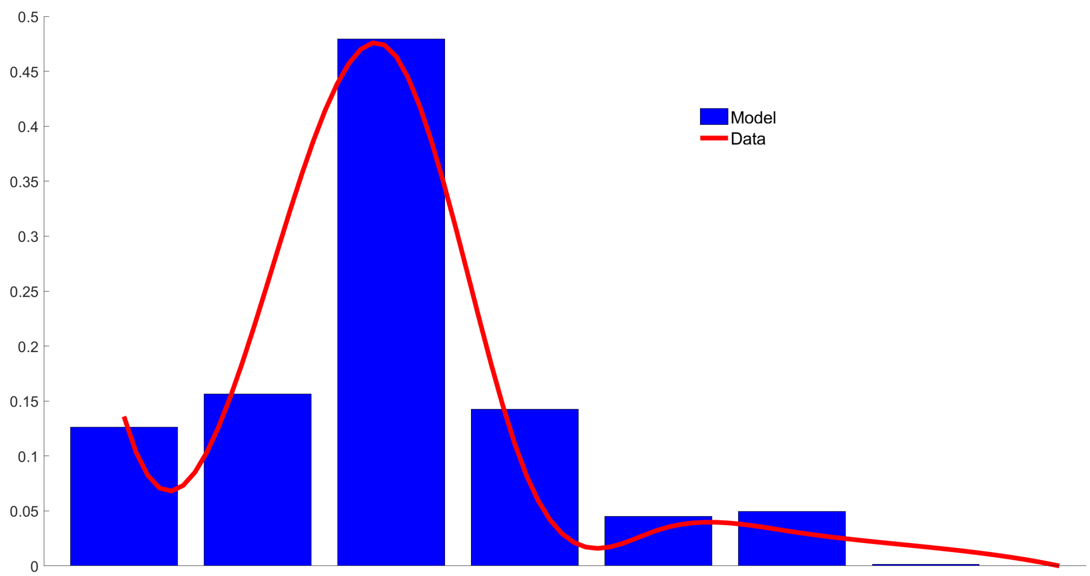
[46]Once convergence is achieved, I simulate the model for 65 periods, and I calculate the moments as averages over those periods.

catch up with a frontier that is ahead by the average technological gap in the economy equal to 1.366.[47] In this second case, the half-life is about six years.

I also investigate the fit of the model by comparing a set of untargeted moments. The calibration of the parameters does not use any cross-sectional moments. I evaluate the fit of the model by comparing the markup distribution generated by simulating the model with the one derived from the data.

Figure 6 compares the data- and the model-generated distribution of markups. The two distributions show a similar shape. Although in both distributions, most of the markups are centered around the mean, the model generates a higher share of markups in the center of the distribution and a lower variation of markups compared to the data-generated distribution.

Figure 6: Markups Distribution: Model vs. Data



To further assess the fit of the model, Table 4 reports selected percentiles of the distribution of markups. The model generates slightly thicker tails than the data. The top 10^{th} and 25^{th} percentiles of the model-generated distribution of markups are somewhat greater than the corresponding percentiles of the empirical distribution (1.737 vs. 1.685 and 1.519 vs. 1.416). By contrast, the model-generated bottom 10^{th} and 25^{th} percentiles are smaller than the data-generated moments (1.206 vs. 1.165 and 1.129 vs. 1.070). These differences likely come from the fact that the only determinant of markup in the model is innovation. In reality, other determinants left unmodeled affect the level of markups.

Thicker model-generated tails also produce larger top to bottom percentile ratios than in the data. Although both the 90^{th}-to-10^{th} percentile and the 75^{th}-to-25^{th} percentile ratios generated by the model are larger than the empirical ones, the deviation between the model-generated and data-generated ratios is in the 10% range. Finally, as already observed from Figure 6, the model generates a distribution of markups

[47] The average technological gap is a combination of the entrants' and incumbents' gaps that they are, in turn, equal to the markups of these two groups of firms.

more concentrated around the mean. Indeed, the model-generated standard deviation is 0.31, whereas the data-generated standard deviation is 0.32. These results also imply the data-generated distribution is flatter than the model-generated distribution and the markups are more spread within the interquartile range.

Table 4: Untargeted Moments: Dispersion of Markups

Untargeted Moment	Data	Model
90^{th} percentile	1.685	1.737
75^{th} percentile	1.416	1.519
25^{th} percentile	1.206	1.165
10^{th} percentile	1.129	1.070
90^{th}-to-10^{th} percentile ratio	1.492	1.624
75^{th}-to-25^{th} percentile ratio	1.175	1.304
Standard deviation	0.323	0.306

Overall, although the model is very parsimonious, these sets of results suggest it replicates a rich set of targeted and untargeted moments well.

7 Quantitative Analysis

In this section, I use the structural model for two purposes. First, I estimate the evolution of knowledge diffusion between 1980 and 2010. Second, I examine the effect of changes in the innovation-step size and intensity of knowledge diffusion on growth and welfare. Although the two parameters are not modeled as functions of specific policy changes, changes in these two parameters can be achieved by reforming the patent system. Higher innovation step size can be achieved by increasing the *non-obviousness requirements*. The non-obviousness principle states that an invention is granted a patent only if it is sufficiently beyond or above the state of the art.[48] Increasing these requirements leads to more radical innovation. Similarly, a shorter patent term and quicker disclosure of codified information can be mapped into changes in the speed of knowledge diffusion and technological catch-up.

7.1 Changes in Knowledge Diffusion

The intensity of knowledge diffusion is not directly observable in the data. My approach to identify the changes is to re-estimate the model using data from 2010. I estimate the new set of parameters by forcing the model-generated moments to be consistent with the 2010 empirical counterparts. Table 5 reports the parameters values for the 2010 calibration, and Appendix D discusses the fit of the model for the new calibration.

[48] The non-obviousness reflects a patentability requirement present in most patent laws. The term "non-obviousness" is predominantly used in US patent law. Policymakers, particularly in Europe, refer to this principle as the "inventive step" requirement. Although different names are used, the principle is the same across legislation.

Table 5: Parameters Calibration, 2010

Parameter	Parameter Description	Parameter Value	Target Moment Description	Data Moment	Model Moment
β	Discount Rate	0.947	–	–	–
γ	Curvature of R&D Cost Function	2	–	–	–
λ	Innovation Step	1.608	Entrants' Average Markup	–	–
ψ_I	Innovation Efficiency: Incumbent	23.607	R&D-Expenditure-to-Output Ratio	0.038	0.037
ψ_E	Innovation Efficiency: Entrant	14.429	Entry Rate	0.065	0.065
α	Innovation Decay Rate	0.044	Incumbents' Average Markup	1.556	1.553

Table 5 shows the estimates for both the innovation step size and the intensity of knowledge diffusion have increased. A greater estimate of α (0.044 in 2010 vs. 0.025 in 1980) does not necessary imply the knowledge diffusion process has sped up. To provide a meaningful comparison, I compute the length of time needed to catch up with the technology frontier as in equation (21). The calculations show that in 2010, the half-life for a follower who was behind by one innovation step was approximately five years.

These results suggest the catching-up process has significantly sped up. Indeed, a follower needs 38% less time to catch up with half of the one-step-ahead frontier. These findings are consistent with the work of Baslandze (2016), who uses a different empirical strategy but similarly finds knowledge diffusion has increased over time. This paper estimates the intensity of knowledge diffusion through the lens of a structural model, whereas Baslandze (2016) uses patent citations across technology classes.

The results in this paper differ from Akcigit and Ates (2019), who estimate a decline in knowledge diffusion intensity by 65% between 1980 and 2010. Although the models are similar in several respects, the sharp difference between the two estimates comes from the difference in the identification strategies of the intensity parameter. Akcigit and Ates (2019) identify the intensity of knowledge diffusion by matching the average markup in the economy with the decline in this parameter, by forcing the model to fit the downward trend in the establishment entry rate between 1980 and 2010. In this paper, I identify the intensity of the knowledge diffusion intensity by using the life-cycle patterns of markups. The trend of knowledge diffusion is still an ongoing debate in the literature. This paper aims to contribute to this debate by proposing an alternative estimation strategy not used in the literature.

The faster process of knowledge diffusion has sped up the technology adoption not only among firms but also among households. Major technological breakdowns such as the telephone and household electricity took about 25 years to reach a rate of market penetration between 40% and 75%. More recent technological advancements are spreading much faster, with television or smartphone adoption reaching a rate between 40% and 75% in less than 10 years.[49] The increase in knowledge diffusion fits well the view that the information

[49] De Gusta (2012) presents a detailed analysis of these trends and the differences in adoption lengths by different products.

and communication technology (ICT) revolution has opened new perspectives for accessing a broader set of external information and speeding up the diffusion of knowledge across firms.[50]

7.2 Counterfactual Experiments: Innovation Step Size

This subsection investigates the effect of changes in the innovation step size λ on a household's welfare and growth. To achieve this goal, I construct counterfactual economies in which I vary the innovation step size and keep all other parameters at their 1980 levels. In these experiments, the innovation-step-size parameters range between 1.445 and 1.689. This range is constructed to have the lower bound, the upper bound, and the values for 1980 and 2010 equally distant.

Figure 7: Counterfactual Experiments – Changes in λ

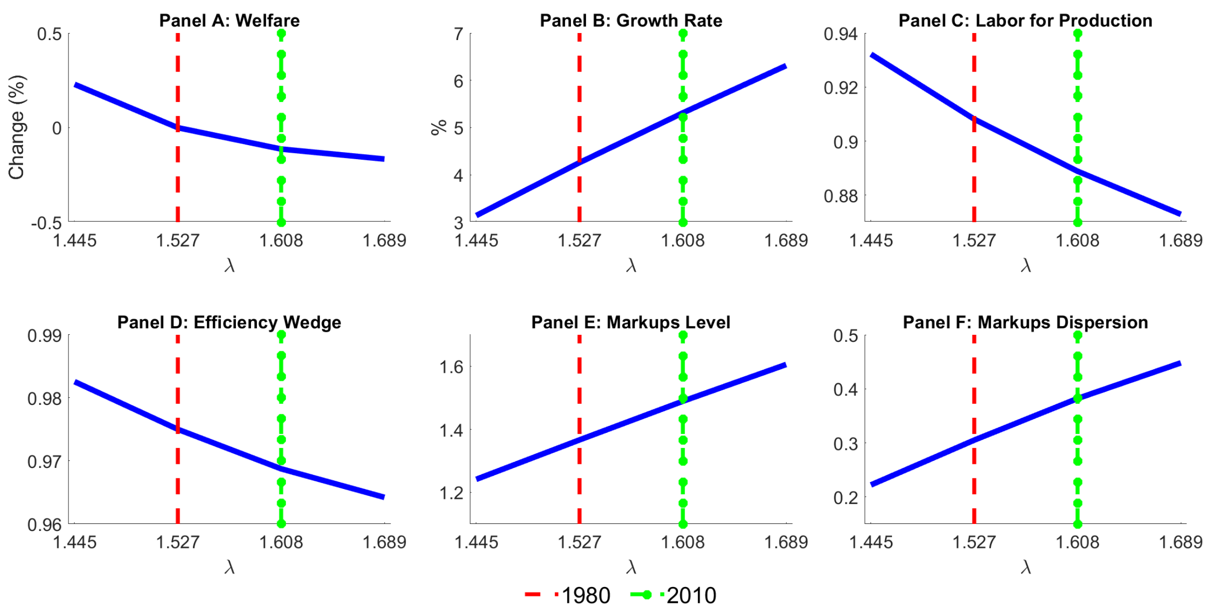


Figure 7 summarizes the effects of changes in the innovation step size on welfare and growth. Panel A highlights that the representative household is worse off as the innovation step size increases. The household would be willing to lose about 0.10% of her consumption instead of moving from the 1980 innovation-step-size level to the 2010 one. By contrast, as equation (19) emphasizes, the growth rate is an increasing function of the innovation step size. An increase in the innovation step size from 1.527 to 1.608 leads to a 1.05-percentage-point increase that corresponds to 25% of the 1980 growth rate. Increases in the innovation step size have a direct and an indirect effect on growth. The direct effect is represented by an increase of the step size, and the indirect effect is represented by higher innovation rates, due to the fact that successful innovations become more profitable. Changes in the growth rate also affect welfare. Indeed, because all aggregate variables grow at the same rate, higher growth rates lead to higher equilibrium consumption, leading to welfare gains.

[50] Baslandze (2016) studies the impact of ICT on facilitating knowledge diffusion and empirically finds knowledge diffusion is correlated with higher usage of ICT.

The decrease in welfare is due to labor reallocation (Panel C) and losses in aggregate efficiency (Panel D). With higher innovation step sizes, innovation becomes more profitable, leading to higher demand of labor for R&D effort. The fixed labor supply implies the share of labor used in production declines, causing a decrease in output. In addition to this effect, the dispersion in markups (Panel F) increases with higher innovation step sizes, causing more misallocation of labor across intermediate-good producers and a loss in production efficiency. Overall, these welfare losses dominate the gains, due to higher growth rates.

Table 6: Counterfactual Experiments: Results

	Panel A: Innovation Size λ				*Panel B: Knowledge Diffusion* α			
	1.527	1.445	1.608	1.689	0.025	0.006	0.044	0.063
Welfare	–	0.23%	−0.11%	−0.17%	–	−0.35%	0.57%	1.33%
Growth Rate	4.3%	3.1%	5.3%	6.3%	4.3%	4.9%	3.4%	2.8%
Average Markup	1.367	1.242	1.489	1.606	1.367	1.572	1.206	1.124
Dispersion Markup	0.306	0.222	0.382	0.448	0.306	0.289	0.250	0.198
Efficiency Wedge	0.975	0.983	0.969	0.964	0.975	0.985	0.979	0.985
Labor Demand	0.908	0.932	0.889	0.873	0.908	0.879	0.939	0.958

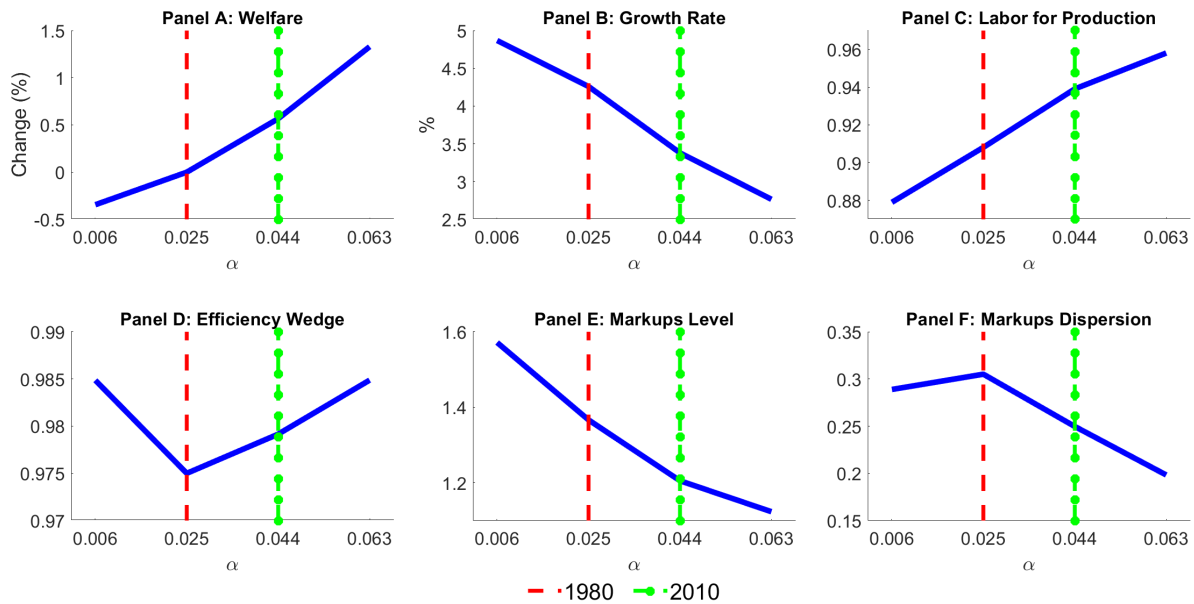
Finally, as in the model, a one-to-one correspondence exists between markups and the innovation gap. For a given intensity of knowledge diffusion, the average markup also increases. Panel E shows the increase in the innovation step size from the 1980 to the 2010 value causes markups to rise by 9%. Panel A of Table 6 summarizes the changes in the macroeconomic outcomes of interest associated with variation in the innovation step size.

7.3 Counterfactual Experiments: Knowledge Diffusion

This subsection studies the effect of increases in the intensity of knowledge diffusion on welfare. I construct counterfactual economies in which I vary only the intensity of knowledge diffusion and keep all other parameters at the 1980 levels. The knowledge diffusion parameter ranges between 0.006 and 0.063. Because the change in the estimated α between 2010 and 1980 is equal to 0.19, the upper bound of the range is set to mimic the same increase starting from the 2010 estimate. The lower bound is set by removing the difference between 2010 and 1980 from the 1980 estimate.

Figure 8 shows the main numerical result of these counterfactual experiments. Specifically, Panel A reports the consumption-equivalent welfare. The estimates show increases in knowledge diffusion generate sizeable welfare gains. Panel B of Table 6 quantifies these gains. Column 3 shows the representative household would require an increase of 0.57% in her consumption to be indifferent between two economies, one with the intensity of knowledge diffusion set at the 1980 baseline value and the other with α set at the 2010 estimate of 0.44.

Figure 8: Counterfactual Experiments – Changes in α



The remaining panels of Figure 8 explore the mechanisms behind the welfare increases. As Panel E indicates, increases in knowledge diffusion decrease markups. Lower markups make R&D efforts less profitable and slow down growth (Panel B). As Table 6 indicates, the growth rate varies significantly with changes in knowledge diffusion. As knowledge diffusion moves from 0.025 to 0.044, the growth rate declines by 21% from 4.3% to 3.4%.

Decreases in innovation efforts lead to a reallocation of the labor force toward production (Panel C). Panel D depicts an inverse U-shaped relationship between knowledge diffusion and the efficiency wedge as defined in equation (20). Increases in knowledge diffusion may have a positive or negative effect on the efficiency wedge. Whereas increases in knowledge diffusion starting from low intensities of diffusion worsen the efficiency wedge, increases starting from high intensities lead to efficiency improvements. The changes in efficiency are linked to changes in the dispersion of markups. Panel F highlights that increases in knowledge diffusion starting from low intensities generate a widening of the markups distribution and an increase in the dispersion. A higher dispersion of markups worsens the allocation of inputs and lowers the aggregate efficiency. On the other hand, increases in knowledge diffusion for high intensities make the distribution of markups more concentrated and improve the efficiency in production.

7.4 Counterfactual Experiments: Innovation Size and Knowledge Diffusion

The previous counterfactual experiments have investigated the welfare and growth effects of changes in one parameter at the time. The scope of those experiments was to understand the mechanisms through which the innovation step size and intensity of knowledge diffusion affect welfare. As argued in section 4, both the innovation step size and the intensity of knowledge diffusion have increased over time. This subsection analyzes the macroeconomic implications of jointly changing both parameters.

Figure 9: Counterfactual Experiments – Changes in λ and α

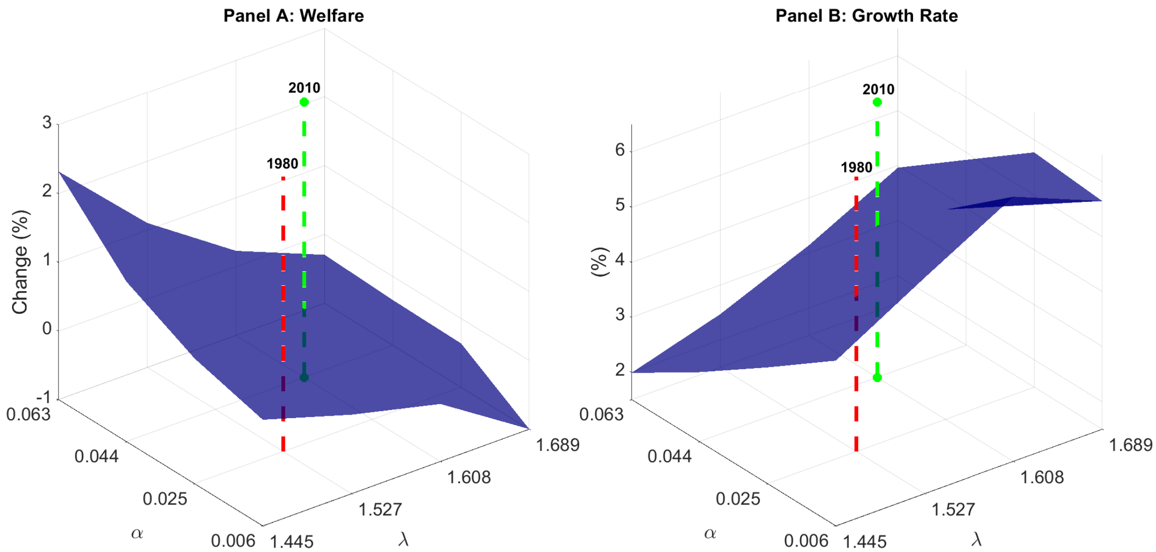


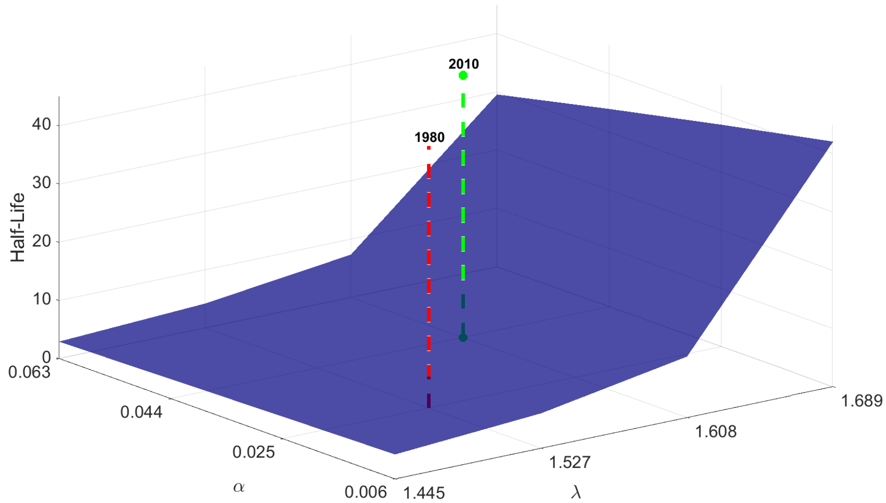
Figure 9 shows the variation in welfare and growth in response to changes in both parameters. Changes in these two parameters can be achieved by reforming the patenting system. If policymakers only aim to maximize the welfare, they should target a high intensity of knowledge diffusion and a small innovation step size. In terms of policy, this target translates to very short patent terms and very low innovative-step requirements. By contrast, if policymakers only aim to maximize the growth rate of the economy, they would choose a low intensity of knowledge diffusion and a large innovation step size. In terms of policy, these goals can be achieved by long patent terms and high non-obviousness requirements. These findings also imply that if policymakers target increases in welfare and growth, they would opt for a combination of intensity of knowledge diffusion and innovation step size that, in isolation, would maximize neither the welfare nor the growth rate.

Figure 9 shows the household would require an increase of 0.29% in consumption to remain in the 1980 baseline economy rather than moving to the counterfactual economy. By contrast, the counterfactual experiments highlight the increase in these parameters has no substantial effect on growth. Namely, the growth rate increases by 0.07 percentage points, which corresponds to roughly 1.6% of the 1980 growth rate.[51]

Figure 10 shows the half-life for different combinations of the innovation step size and the intensity of knowledge diffusion. The results highlight the half-life for small innovation step sizes vary only slightly. By contrast, for large innovation step sizes, the half-life increases drastically. Furthermore, the results also point out changes in the intensity of knowledge diffusion have a minor effects on the half-life compared with changes in the innovation step size.

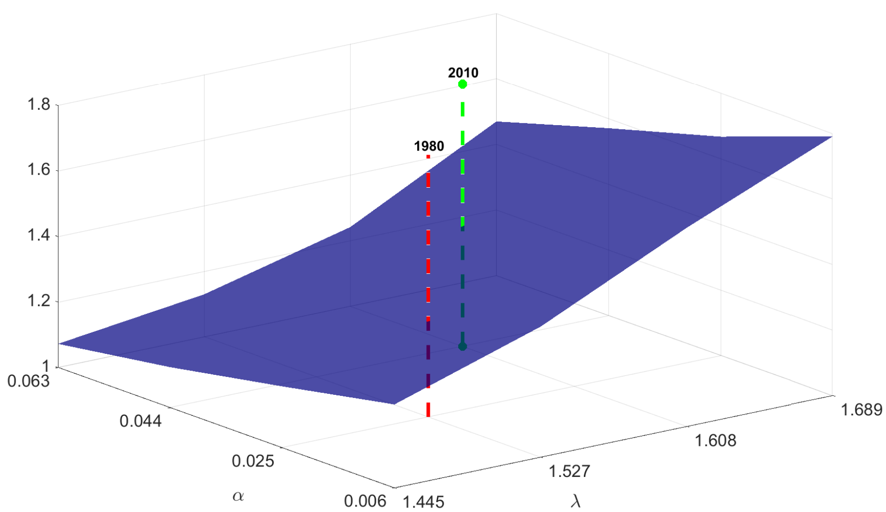
[51] The quantitative results from these counterfactual experiments are reported in Table E.1 in Appendix E.

Figure 10: Counterfactual Experiments –Half-Life



Finally, Figure 11 shows the evolution of markups due to joint changes in the innovation step size and intensity of knowledge diffusion. This result suggests average markup reacts more to changes in the innovation step size than to changes in the intensity of knowledge diffusion.

Figure 11: Counterfactual Experiments –Markups



This numerical result reinforces the empirical findings that differences in markups between cohorts of firms have played a quantitatively significant role in the increase of the aggregate markups. This paper has not investigated the determinants of the increase in the innovation step size. An avenue for future research is to disentangle the contribution of different factors to the increase in the innovation quality.

8 Conclusion

This paper builds on new empirical evidence on the evolution of markups. Namely, I explore the between and within cohorts of firms variation in markups. The analysis highlights two new findings. First, new firms, at the time of their entry, charge higher markups than existing firms, creating substantial between-cohorts differences. Second, the average markups within-cohorts remain relatively flat over time. I also find the differences between cohorts account for about $5\% - 10\%$ of the increase in the aggregate markup in the US economy.

I link the between- and within-cohort patterns to knowledge creation and diffusion. I show the differences between cohorts are linked to improvements in the innovation quality. A higher quality of entrants' innovation generates greater technological gaps over the competitors and higher markups. Because incumbents also innovate, innovation becoming more radical is not sufficient to explain the flatness of markups over the life cycle. The interaction of two opposing forces explains this second empirical finding: innovation and knowledge diffusion. Successful innovations widen the technological gap over the competitors and increase markups. In the absence of innovations, knowledge diffusion shrinks the gap between the followers and the technological leader, causing a decrease in markups. I empirically provide support for these two hypotheses, using administrative patent data.

I develop a general equilibrium Schumpeterian endogenous growth model of creative destruction augmented with a process of knowledge diffusion that accounts for the new empirical patterns and incorporates the economic mechanisms discussed above. Knowledge creation and knowledge diffusion are mapped to two model parameters: the innovation step size and intensity of knowledge diffusion. In turn, these two parameters determine the equilibrium markups and affect growth and welfare.

I estimate the model parameters using US firm-level data. Unlike the innovation step size, which is directly observable from the data, the intensity of knowledge diffusion is not observable. I use the theoretical framework to estimate this parameter and its evolution between 1980 and 2010. I find knowledge diffusion has increased by 38% over the last four decades. This finding is consistent with the view that the ICT revolution has facilitated the spread of information.

I use the estimated model to run counterfactual experiments. These experiments aim to quantify the effect of changes in innovation step size and the intensity of knowledge diffusion on growth and welfare. Although the two parameters are not modeled as functions of specific policy changes, changes in these two parameters can be achieved by reforming the patent system, such as changes in the non-obviousness requirements, or the patent term.

The first set of quantitative exercises computes welfare changes associated with changes either in the innovation step size or in the intensity of knowledge diffusion. These exercises help us understand the channels that determine welfare. Specifically, the paper identifies three mechanisms that operate in different directions: growth, labor reallocation, and input misallocation. Furthermore, the response in welfare, growth, reallocation, and efficiency to changes in the innovation step size are in the opposite direction of responses to changes in the intensity of knowledge diffusion.

The main quantitative exercise aims to quantify the effect on welfare and growth by jointly changing both the innovation step size and the intensity of knowledge diffusion from the 1980 values to the 2010 ones. The quantitative exercise points out the household would require an increase of 0.29% in consumption to remain in the 1980 baseline economy rather than moving to the counterfactual economy. By contrast, the growth rate would remain substantially unchanged between the baseline and the counterfactual economy.

This study contributes to the ongoing debate on market structure and its macroeconomic implications. In particular, the paper suggests the increase in markups is not due to a decline in the intensity of knowledge diffusion, as recently proposed in the literature. The theoretical model and the numerical exercises also provide some clear policy recommendations. If policymakers only aim to maximize the welfare, they should target a high intensity of knowledge diffusion and a small innovation step size. In terms of policy, this target translates to very short patent terms and very low innovative-step requirements. By contrast, if policymakers only aim to maximize the growth rate of the economy, they should choose a low intensity of knowledge diffusion and a large innovation step size. In terms of policy, these goals can be achieved by long patent terms and high non-obviousness requirements. These findings also imply that if policymakers target increases in welfare and growth, they would opt for a combination of intensity of knowledge diffusion and innovation step size that, in isolation, would maximize neither the welfare nor the growth rate.

References

Acemoglu, D. and Akcigit, U. (2012). Intellectual Property Rights Policy, Competition And Innovation. *Journal of the European Economic Association*, 10(1):1–42.

Acemoglu, D., Akcigit, U., and Celik, M. A. (2020). Radical and incremental innovation: The roles of firms, managers and innovators. Working Paper.

Aghion, P. and Howitt, P. (1992). A Model of Growth through Creative Destruction. *Econometrica*, 60(2):323–351.

Akcigit, U. and Ates, S. T. (2019). What Happened to U.S. Business Dynamism? NBER Working Papers 25756, National Bureau of Economic Research, Inc.

Akcigit, U., Hanley, D., and Serrano-Velarde, N. (2020). Back to basics: Basic research spillovers, innovation policy and growth. *Review of Economic Studies*. Forthcoming.

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

Alvarez, F. E., Buera, F. J., and Robert E. Lucas, J. (2013). Idea Flows, Economic Growth, and Trade. NBER Working Papers 19667, National Bureau of Economic Research, Inc.

Anderson, E., Rebelo, S., and Wong, A. (2018). Markups across space and time. NBER Working Papers 24434, National Bureau of Economic Research.

Andrews, D., Criscuolo, C., and Gal, P. N. (2015). Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries. OECD Productivity Working Papers, OECD.

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

Atkeson, A. and Burstein, A. (2019). Aggregate implications of innovation policy. *Journal of Political Economy*, 127(6):2625–2683.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Reenen, J. V. (2017). The Fall of the Labor Share and the Rise of Superstar Firms. NBER Working Papers 23396, National Bureau of Economic Research, Inc.

Autor, D. H., Dorn, D., Hanson, G. H., Pisano, G., and Shu, P. (2019). Foreign Competition and Domestic Innovation: Evidence from U.S. Patents. CESifo Working Paper Series 7865, CESifo Group Munich.

Azar, J., Marinescu, I., and Steinbaum, M. I. (2017). Labor Market Concentration. NBER Working Papers 24147, National Bureau of Economic Research, Inc.

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

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance*, 75(5):2421–2463.

Baruffaldi, S. H. and Simeth, M. (2020). Patents and knowledge diffusion: The effect of early disclosure. *Research Policy*, 49(4):103927.

Baslandze, S. (2016). The Role of the IT Revolution in Knowledge Diffusion, Innovation and Reallocation. 2016 Meeting Papers 1509, Society for Economic Dynamics.

Bloom, N., Griffith, R., and Van Reenen, J. (2002). Do R&D tax credits work? Evidence from a panel of countries 1979-1997. *Journal of Public Economics*, 85(1):1–31.

Bloom, N., Reenen, J. V., and Williams, H. (2019). A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives*, 33(3):163–184.

Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108(1):113–131.

Boar, C. and Midrigan, V. (2019). Markups and inequality. NBER Working Papers 25952, National Bureau of Economic Research, Inc.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2020). Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data. NBER Working Papers 27002, National Bureau of Economic Research, Inc.

Buera, F. J. and Lucas, R. E. (2018). Idea Flows and Economic Growth. *Annual Review of Economics*, 10(1):315–345.

Cooley, T. F. and Prescott, E. C. (1995). *Frontiers of business cycle research*. Princeton University Press.

De Gusta, M. (2012). Are smart phones spreading faster than any technology in human history. *Massachusetts Institute of Technology Review*.

De Loecker, J. and Eeckhout, J. (2017). The Rise of Market Power and the Macroeconomic Implications. NBER Working Papers 23687, National Bureau of Economic Research, Inc.

De Loecker, J. and Eeckhout, J. (2018). Global Market Power. NBER Working Papers 24768, National Bureau of Economic Research, Inc.

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

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

De Ridder, M. (2019). Market Power and Innovation in the Intangible Economy. Cambridge Working Papers in Economics 1931, Faculty of Economics, University of Cambridge.

Diez, F., Leigh, D., and Tambunlertchai, S. (2018). Global market power and its macroeconomic implications. IMF Working Paper 137, International Monetary Fund.

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). How Costly Are Markups? NBER Working Papers 24800, National Bureau of Economic Research, Inc.

Ennis, S. F., Gonzaga, P., and Pike, C. (2019). Inequality: A hidden cost of market power. *Oxford Review of Economic Policy*, 35(3):518–549.

Flynn, Z., Gandhi, A., and Traina, J. (2019). Identifying market power from production data. Technical report, Working paper.

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

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

Galí, J., Gertler, M., and López-Salido, J. D. (2007). Markups, Gaps, and the Welfare Costs of Business Fluctuations. *The Review of Economics and Statistics*, 89(1):44–59.

Gilbert, R. and Shapiro, C. (1990). Optimal Patent Length and Breadth. *RAND Journal of Economics*, 21(1):106–112.

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

Grossman, G. M. and Helpman, E. (1993). *Innovation and Growth in the Global Economy*, volume 1 of *MIT Press Books*. The MIT Press.

Grullon, G., Larkin, Y., and Michaely, R. (2019). Are US Industries Becoming More Concentrated? *Review of Finance*, 23(4):697–743.

Gutierrez, G. and Philippon, T. (2018). How EU Markets Became More Competitive Than US Markets: A Study of Institutional Drift. CEPR Discussion Papers 12983, C.E.P.R. Discussion Papers.

Gutiérrez, G. and Philippon, T. (2017). Declining Competition and Investment in the U.S. NBER Working Papers 23583, National Bureau of Economic Research, Inc.

Hall, B. H. (2009). Issues in and possible reforms of the us patent system. In *conference on 21st Century Innovation Systems for the United States and Japan: Lessons from a Decade of Change, sponsored by the National Academies, NISTEP of Japan, and the Institute of Innovation Research of Hitotsubashi University*.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Papers 8498, National Bureau of Economic Research, Inc.

Hall, B. H. and Ziedonis, R. H. (2001). The patent paradox revisited: an empirical study of patenting in the us semiconductor industry, 1979-1995. *Rand Journal of Economics*, pages 101–128.

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

Hall, R. E. (2018). Using Empirical Marginal Cost to Measure Market Power in the US Economy. NBER Working Papers 25251, National Bureau of Economic Research, Inc.

Heise, S., Karahan, F., and Şahin, A. (2020). The Missing Inflation Puzzle: The Role of the Wage-Price Pass-Through. NBER Working Papers 27663, National Bureau of Economic Research, Inc.

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

IMF (2019). World Economic Outlook: Growth Slowdown, Precarious Recovery - The Rise of Corporate market power and its macroeconomic effect. World Economic Outlook Reports Chapter 2, International Monetary Fund.

Klette, T. J. and Kortum, S. (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy*, 112(5):986–1018.

Lashkari, D. (2018). Innovation, knowledge diffusion, and selection. Ssrn working papers 3124486, SSRN.

Lentz, R. and Mortensen, D. T. (2008). An Empirical Model of Growth Through Product Innovation. *Econometrica*, 76(6):1317–1373.

Lerner, A. (1934). The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies*, pages 157–175.

Lerner, J. and Seru, A. (2017). The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond. NBER Working Papers 24053, National Bureau of Economic Research, Inc.

Lucas, R. E. and Moll, B. (2014). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, 122(1):1–51.

Luttmer, E. G. (2012). Technology diffusion and growth. *Journal of Economic Theory*, 147(2):602–622.

Luttmer, E. G. J. (2007). Selection, Growth, and the Size Distribution of Firms. *The Quarterly Journal of Economics*, 122(3):1103–1144.

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

Nekarda, C. J. and Ramey, V. A. (2011). Industry Evidence on the Effects of Government Spending. *American Economic Journal: Macroeconomics*, 3(1):36–59.

Nekarda, C. J. and Ramey, V. A. (2020). Supplemental appendix to the cyclical behavior of the price-cost markup. Working Paper.

OECD (2004). Innovation: Trends and policy challenges.

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

Ouellette, L. L. and Williams, H. (2020). Reforming the patent system. Technical report, Policy Proposal, The Hamilton Project, Brookings Institution, Washington, DC.

Perla, J. and Tonetti, C. (2014). Equilibrium Imitation and Growth. *Journal of Political Economy*, 122(1):52–76.

Perla, J., Tonetti, C., and Benhabib, J. (2014). The Growth Dynamics of Innovation, Diffusion, and the Technology Frontier. 2014 Meeting Papers 818, Society for Economic Dynamics.

Peters, M. (2013). Heterogeneous mark-ups, growth and endogenous misallocation. Working Paper.

Raval, D. (2019). Testing the production approach to markup estimation. Technical Report 3324849, SSRN Working Papers.

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

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10.

Restuccia, D. and Rogerson, R. (2017). The Causes and Costs of Misallocation. *Journal of Economic Perspectives*, 31(3):151–174.

Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2020). Diverging Trends in National and Local Concentration. In *NBER Macroeconomics Annual 2020, volume 35*, NBER Chapters. National Bureau of Economic Research, Inc.

Rotemberg, J. J. and Woodford, M. (1999). The cyclical behavior of prices and costs. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 16, pages 1051–1135. Elsevier.

Sampson, T. (2016). Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth. *The Quarterly Journal of Economics*, 131(1):315–380.

Stein, L. and Stone, E. (2013). The Effect of Uncertainty on Investment, Hiring, and R&D: Causal Evidence from Equity Options. Working paper.

Syverson, C. (2019). Macroeconomics and market power: Facts, potential explanations and open questions. *Brookings Economic Studies Report.*

Traina, J. (2018). Is Aggregate Market Power Increasing? Production Trends Using Financial Statements. Working Papers Series 17, Stigler Center for the study of the economy and the state.

Van Reenen, J. (2018). Increasing Differences between firms: Market Power and the Macro-Economy. Working paper.

Appendix A. Markups: Theoretical Framework

The computation of markups relies on the supply-side approach proposed by De Loecker and Warzynski (2012) based on the work of Hall (1988). This approach does not require any explicit assumptions on the demand system and the competition structure between firms. The production-based approach relies on the cost minimization of variable inputs of production. The key assumption is that within one period firms adjust variable inputs, while capital is subject to adjustment costs, and it cannot be adjust within one period.

The Lagrangian function associated with the cost minimization is as it follows:

$$\mathcal{L}(V_{it}, K_{it}, \lambda_{it}) = P_{it}^V V_{it} + r_{it} K_{it} + F_{it} - \lambda_{it}[Q_{it}(V_{it}, K_{it}, \Omega_{it}) - \bar{Q}_{it}],$$

where V_{it} is variable input of production, P_{it}^V is the price of variable inputs, K_{it} is the capital stock, r_{it} is the user cost of capital, F_{it} is the fixed cost, $\bar{Q}_{it}$ is a scalar, λ_{it} is the Lagrangian multiplier, and $Q_{it}(V_{it}, K_{it}, \Omega_{it})$ is the production technology that is a function of the variable input, the capital stock and the firm-specific productivity Ω_{it}.

The first order condition with respect to the variable input V_{it} is given by:

$$\frac{\partial \mathcal{L}(\cdot)}{\partial V_{it}} = P_{it}^V - \lambda_{it} \frac{\partial Q_{it}(\cdot)}{\partial V_{it}} = 0.$$

By arranging the terms, the output elasticity of variable input V_{it} is derived as:

$$\beta_{it}^V = \frac{\partial Q_{it}}{\partial V_{it}} \frac{V_{it}}{Q_{it}} = \frac{1}{\lambda_{it}} \frac{P_{it}^V V_{it}}{Q_{it}}.$$

The Lagrange multiplier λ_{it} is a direct measure of marginal cost, indeed it represents the value of the objective function as output constraint is relaxed. The markup is defined as $\mu = \frac{P^Q}{\lambda}$, where P^Q is the price of the output good. Replacing the marginal cost for the markup to price ratio, we obtain the expression for the markup:

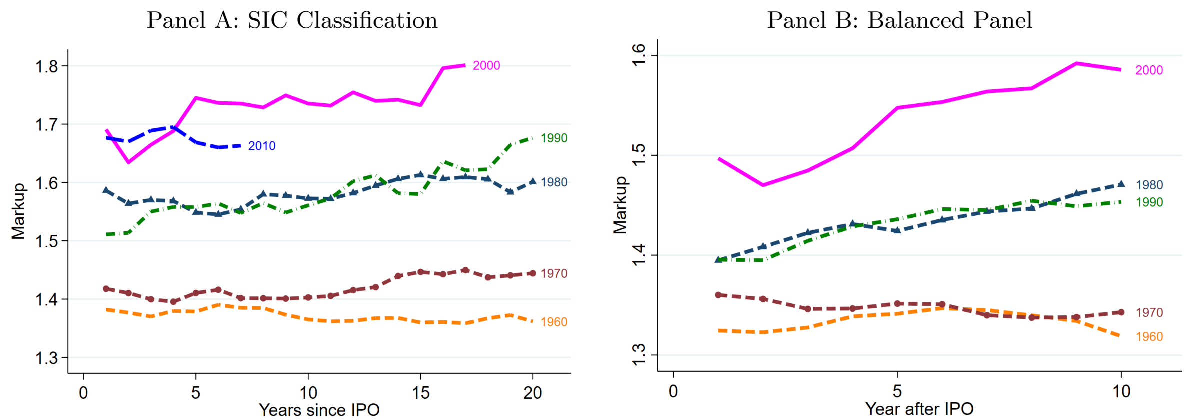
$$\mu_{it} = \beta_{it}^V \frac{P_{it}^Q Q_{it}}{P_{it}^V V_{it}}.$$

This expression will be used to implement the empirical strategy.

Appendix B. Empirical Robustness Analysis

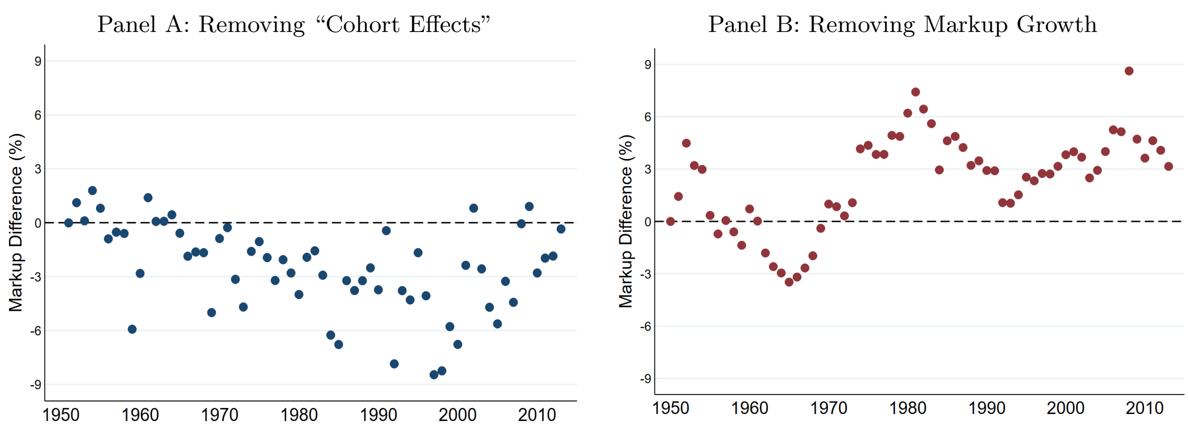
Figure B.1 shows the evolution of markups by cohorts of firms as in Figure 1. I present two robustness checks. The first one, in Panel A, uses the Standard Industrial Classification (SIC) instead of the North America Industry Classification System (NAICS). In computing the markups, I assume an industry-specific production function. Panel A tests whether a different industrial classification affects the empirical patterns discussed in the body of the paper. Results remain substantially unchanged in both classifications.

Figure B.1: Markups by Cohorts of Firms, Robustness



Panel B of Figure B.1 restricts the analysis to a balanced sample of firms. One may expect that firms that exit the market are less profitable and charge lower markups. That would imply the average markup over the life-cycle is pushed down by these firms. Panel B confirms the previous intuition. Indeed, we can observe that the markup growth within cohort is more pronounced than in Figure 1. Although the within-cohort markup growth is higher, we still observe significant differences between cohorts.

Figure B.2: Thought Experiments Comparison, Variable Inputs Weights



Edmond et al. (2018) argue that the markups have to be weighted by the variable costs shares rather than the sales shares. Figure B.2 computes the aggregate effects using the variables costs weights. Overall, the results are similar to the ones discussed in sub-section 4.1 with the only exception that markups have significantly declined in the last twenty-five years. This result is still consistent with the story described in the body of the paper.

Figure B.3: Markups by Cohorts of Firms, Alternative Markup Computation

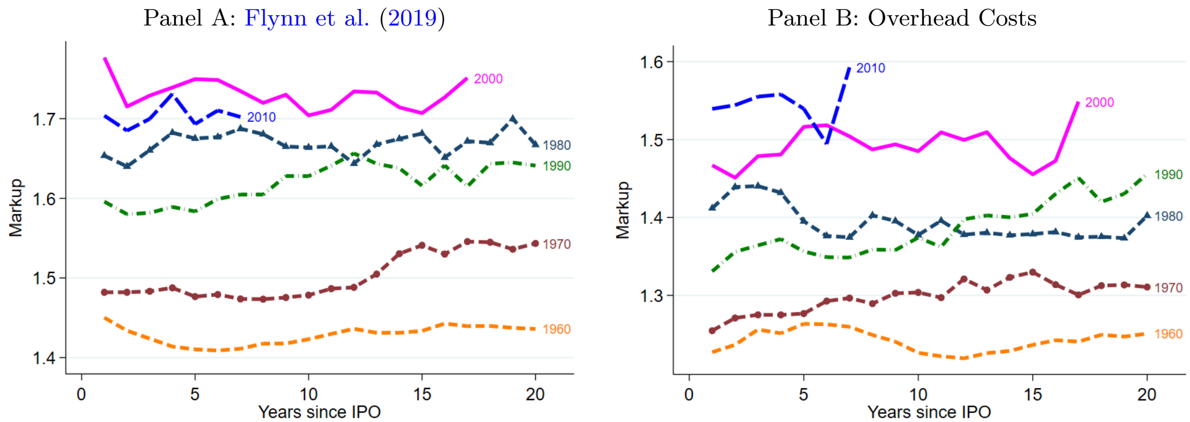


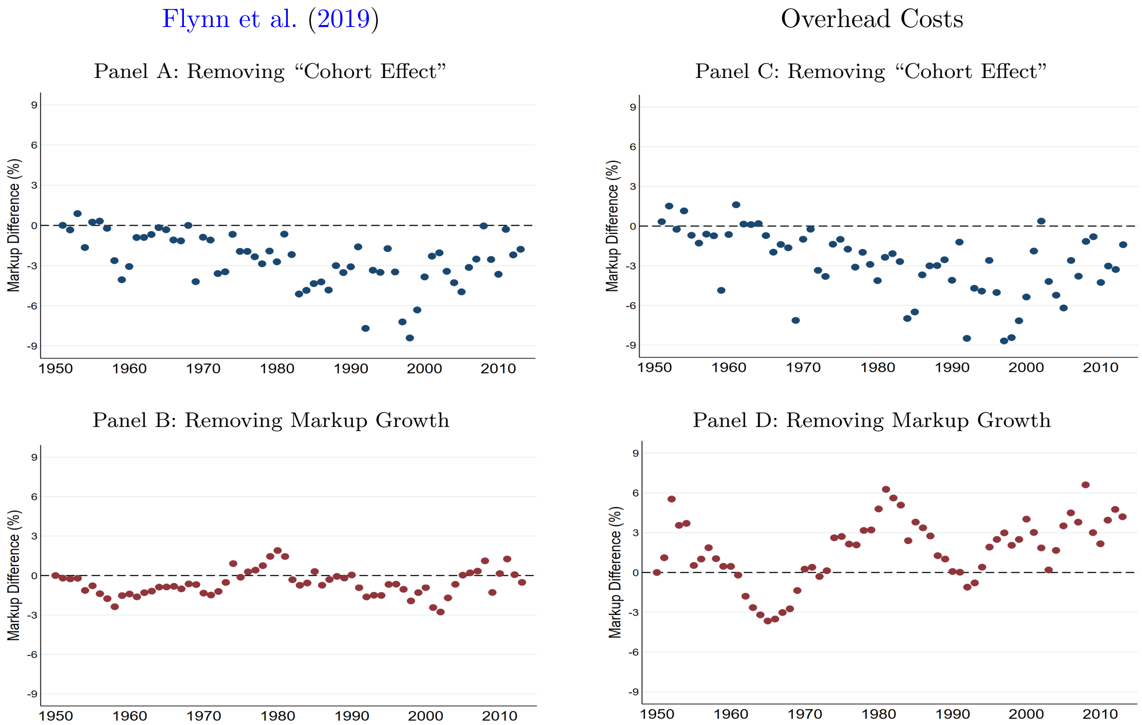
Figure B.3 shows the average markups by selected cohorts using the alternative identification strategy of elasticities proposed by Flynn et al. (2019) in Panel A, and the augmented production function with overhead costs as in De Loecker et al. (2020) in Panel B. Figure B.3 shows the same between- and within-cohort patterns as Figure 1. In fact, we observe both between-cohorts differences and within-cohorts flatness in the markup profile over the years of activity. The statistical tests also confirm that the entrants charge higher markups than incumbents.

Figure B.4 disentangles the contribution in the rise of the aggregate markup of the differences across cohorts of firms and the growth of markups over the life-cycle using the alternative identification strategy of elasticities proposed by Flynn et al. (2019) in right two panels, and the augmented production function with overhead costs as in De Loecker et al. (2020) in the left two panels. Similarly to the benchmark estimation strategy, the results show a strong "cohort effect." In both alternative specifications removing the "cohort effect" would have decreased the aggregate markup between $5-10\%$. The magnitude of this effect is comparable to the one from the benchmark specification.

The size of the effect by removing the growth of markups over the life-cycle of a firm varies between the two alternative specification. Using the alternative instrumental variable identification proposed by Flynn et al. (2019), we do not observe any significant differences in the aggregate markup. In contrast, by adding the overhead costs to the production function generates implies that the markups have declined over the life-cycle of a firm. All these findings point to the same direction as the results from the benchmark estimation strategy.

Overall, the robustness analysis confirms, both qualitatively and quantitatively, the main findings discussed in the body of the paper. The differences in markups between cohorts of firms play an quantitatively important role in determining markups. The within-cohort markup profile is relatively flat and implies

Figure B.4: Thought Experiments Comparison, Robustness



markups do not change too much over the life-cycle. The few differences are not quantitatively significant.

Appendix C. Computation Algorithm

I define the aggregate variables as ratio with respect to Q, $\hat{Y} = \frac{Y}{Q}$, $\hat{w} = \frac{w}{Q}$. The algorithm proceeds as it follows.

1. Construct a grid for innovation gaps with the sizes of gap belonging to Ξ. I use a log grid of 100 points with a gap of 0.001 between 1 and 20.
2. Calibrate the parameters β, γ, and λ and guess the parameters ψ_I, ψ_E, and α.
3. Guess $\hat{Y}$, $\hat{w}$, τ, and g.
4. Given the guesses and the Euler equation, solve for the incumbents' innovation decision x_I by iterating on the value function and using linear interpolation between grid points.
5. Given the guesses and the incumbents' innovation decision x_I, solve for the stationary distribution of quality gaps by iterating over the density and by using the histogram method.
6. Given the guesses and the incumbents' innovation decision x_I, calculate the incumbents' innovation cost, and the labor demand for a given quality gap. Calculate the aggregate variables by weighting any given quality gap by the probability density.
7. Given the guesses, calculate the entrants' innovation cost.
8. Given the guesses, compute the aggregate consumption by using the household budget constraint, the asset market clearing condition, and the Euler equation.
9. Verify that the equilibrium conditions are satisfied. The equilibrium conditions used to update the guesses are: the resource constraint to update $\hat{Y}$; the labor market clearing condition to update $\hat{w}$; The free-entry condition to update τ; and the computed growth rate to update g. Once all equilibrium conditions are verified, convergence is achieved and move to step 10. If at least one of the conditions is not verified, update the guesses for the equilibrium quantities and repeat the algorithm from step 4.
10. Given the stationary equilibrium, simulate the model until the distribution of markups converges to its stationary distribution. Once convergence is achieved, compute and collect the simulated moments.
11. Estimate ψ_I, ψ_E, and α via GMM. The parameters are chosen to minimize the distance between the model-generated and the data-generated moments from the following objective function:

$$\min \sum_{i=0}^{3} \left\{ \frac{|\text{model}_i - \text{data}_i|}{0.5\big(|\text{model}_i| + |\text{data}_i|\big)} \right\}$$

    The minimization uses the Nelder-Mead simplex algorithm. If the minimization is not achieved, then a new set of parameters is used and the algorithm is repeated starting from step 3.

Appendix D. Additional Results: Stationary Equilibrium

Table 5 shows the targeted data-generated moments, the model-generated moments, and the parameter values that minimize the distance between the two set of moments. As for the calibration for the 1980, the model matches quite well the data-generated moments. The innovation step size λ has increased to 1.608, and the intensity of knowledge diffusion α has also rise to 0.044.

Table D.1: Parameters Calibration, 2010

Parameter	Parameter Description	Parameter Value	Method	Target Moment Description	Data Moment	Model Moment
β	Discount Rate	0.947	External Calibration	–	–	–
γ	Curvature of R&D Cost Function	2	External Calibration	–	–	–
λ	Innovation Step	1.608	Direct Inference	Entrants' Average Markup	–	–
ψ_I	Innovation Efficiency: Incumbent	23.607	Indirect Inference	R&D Expenditure to Output Ratio	0.038	0.037
ψ_E	Innovation Efficiency: Entrant	14.429	Indirect Inference	Entry Rate	0.065	0.065
α	Innovation Decay Rate	0.044	Indirect Inference	Incumbents' Average Markup	1.556	1.553

To verify the fit of the model for the 2010, Figure D.1 and Table D.2 report the comparison for selected untargeted moments.

Figure D.1: Markups Distribution - Model vs. Data, 2010

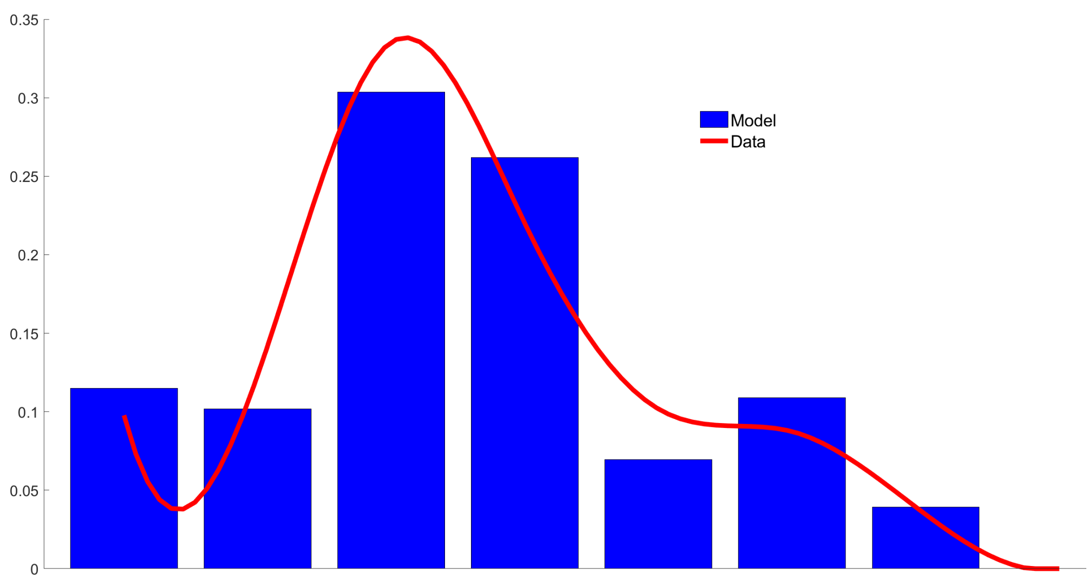


Figure D.1 shows the cross-sectional distribution of markups. Relative to the distribution in the eighties in Figure 6, the markups are less centered around the mean. Indeed, one can observe a much thicker right tail. Similarly to the data-generated distribution, the model also produces a much sparse markups distribution relative to Figure 6.

Table D.2: Untargeted Moments - Markups Distribution, 2010

Untargeted Moment Description	Data	Model
10^{th} Percentile	2.116	2.316
25^{th} Percentile	1.721	1.822
75^{th} Percentile	1.272	1.170
90^{th} Percentile	1.153	1.060
10^{th} to 90^{th} Percentile Ratio	1.836	2.185
25^{th} to 75^{th} Percentile Ratio	1.353	1.558
Standard Deviation	0.410	0.655

Table D.2 confirms the insights from the previous figure. The empirical standard deviation for the 2010 is 0.66 compared with 0.31 in the 1980. The model generates an higher standard deviation than the data. Indeed, the model-generated standard deviation is 0.66 relative to 0.41 from the data. Although the model matches the selected percentiles of the distribution quite well, it is worth to notice from the comparison between Table D.2 and Table 4 that the model overestimates the thickness of both tails, in particular the right tail, in the calibration for 2010, while it underestimates the tails in the baseline calibration for 1980.

Appendix E. Additional Results of the Counterfactual Experiments

Table E.1 reports a summary of the quantitative results for the counterfactual experiments presented in sub-section 7.4 in which both parameters λ and α vary.

Table E.1: Results of the Counterfactual Experiments

	λ	α			
		0.006	0.025	0.044	0.063
Welfare	1.445	-0.30%	0.23%	1.10%	2.32%
	1.527	-0.35%	–	0.57%	1.33%
	1.608	-0.36%	-0.11%	0.29%	0.82%
	1.689	-1.00%	-0.17%	0.15%	0.54%
Growth Rate	1.445	4.0%	3.1%	2.5%	2.0%
	1.527	4.9%	4.3%	3.4%	2.8%
	1.608	5.8%	5.3%	4.3%	3.6%
	1.689	5.7%	6.3%	5.4%	4.5%
Average Markup	1.445	1.470	1.241	1.130	1.073
	1.527	1.572	1.367	1.205	1.123
	1.608	1.668	1.489	1.293	1.187
	1.689	1.791	1.606	1.398	1.254
Dispersion Markup	1.445	0.239	0.222	0.184	0.138
	1.527	0.289	0.305	0.250	0.198
	1.608	0.331	0.382	0.317	0.266
	1.689	0.455	0.448	0.385	0.325
Efficiency Wedge	1.445	0.987	0.983	0.987	0.991
	1.527	0.985	0.975	0.979	0.985
	1.608	0.983	0.969	0.971	0.977
	1.689	0.975	0.964	0.964	0.969
Labor Demand	1.445	0.894	0.932	0.956	0.971
	1.527	0.879	0.908	0.939	0.958
	1.608	0.865	0.889	0.923	0.944
	1.689	0.657	0.873	0.903	0.931