ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Structural Change and the Rise in Markups

Authors	Ricardo Marto
Working Paper Number	2024-002F
Revision Date	September 2025
Citable Link	https://doi.org/10.20955/wp.2024.002
Suggested Citation	Marto, R., 2025; Structural Change and the Rise in Markups, Federal Reserve Bank of St. Louis Working Paper 2024-002. URL https://doi.org/10.20955/wp.2024.002

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Structural Change and the Rise in Markups*

Ricardo Marto†

Federal Reserve Bank of St. Louis

November, 2023

Abstract

Is the recent rise in markups caused by increased monopoly power or is it a natural consequence of structural change? I show that the rise in aggregate markups has been driven by a reallocation of market share away from non-services to services-producing firms and a faster increase of services' markups. I develop a two-sector model to assess the sources of the rise in markups, in which the two forces of structural change play opposing roles. On one hand, an increase in the relative productivity of manufacturing leads to a decline of the relative price of manufactured goods and to an increase of the goods markups. On the other hand, the increase in incomes that triggers the rise of the services sector leads to higher markups for firms in services. I show that the rise in markups is in line with the rise of the services sector and the fall of the relative price of manufactured goods, and may not necessarily reflect a decline of competition. I provide novel experimental evidence supporting the notion that the price elasticity of demand decreases with income.

Keywords: Endogenous markups, income elasticity of demand, manufacturing, non-homothetic preferences, online experiment, price elasticity of demand, services, skill premium, structural change, survey, technological progress

JEL Classifications: D11, D12, D22, D43, E21, E23, L11, L16, O41, O47

*I am deeply indebted to my advisors Jeremy Greenwood, Aviv Nevo, and Joachim Hubmer for their guidance and continued support. I also thank Harun Alp, Yoshiki Ando, Gadi Barlevy, Harold Cole, Alessandro Dovis, Nezih Guner, Dohan Kim, Dirk Krueger, Guillermo Ordoñez, Jordan Peeples, Michael Peters, Víctor Ríos-Rull, Raul Santaeulalia-Llopis, Ina Simonovska, Akos Valentinyi, and Jon Willis for their insightful comments and suggestions. I am also grateful to seminar participants at Arizona State University, Asian and European Meetings of the Econometric Society, Católica Lisbon, Computing in Economics and Finance conference, ECB, ESADE, Federal Reserve Bank of Atlanta Firm Dynamics and the Macroeconomy conference, Federal Reserve Bank of Chicago Rookie conference, Federal Reserve Bank of St. Louis, Federal Reserve Board, FGV/EPGE, Insper, IMF, Inter-American Development Bank, Lisbon Macro conference, Penn State, PUC-Rio, SED, University of Oxford, University of Pennsylvania, University of Toronto, and Virtual Family Macro seminar for helpful suggestions. I thank the EEA for the 2023 Young Economist Award.

†The views expressed here are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System. Email address: ricardo.marto@stls.frb.org.

1 Introduction

What has caused the recent rise in markups? Is this the result of an increase in monopoly power? Does it call for a strengthening of competition policy? Those questions have sparked the interest of researchers and policymakers around the world.[1] Three hypotheses have been advanced to explain the rise of markups. The first hypothesis relates to the rise of superstar firms. Highly innovative firms have reaped the benefits of technological advancement by creating new markets and acting like monopolists in those markets, which has led to an increase in markups. The second hypothesis suggests that firms have charged higher markups in response to increased barriers to entry, which is also potentially related to the increased spending on advertising and managerial expenses. The third hypothesis relates to the potential mismeasurement of markups.[2]

This paper takes a different view by asking: Does structural change—the reallocation of economic activity away from manufacturing toward the services sector—explain the rise in markups? Backed by empirical evidence and a quantitative model matched to the trends observed in the U.S. economy over the past forty years, this paper shows that the rise in markups is intimately related to the process of structural transformation economies go through. Why structural change? The increase in the services share in advanced economies is accompanied by an increase in the relative price of services (44% between 1980 and 2015 in the U.S.). By definition, prices equal markups times marginal costs. Therefore, markups in service industries grew faster than in non-service industries and/or marginal costs in non-service industries declined faster than in service industries. I argue both happened.

In the empirical analysis (Section 2), I show that the rise in markups has been driven by a reallocation of market share away from non-services-producing to services-producing firms along with a larger increase in the average markup of services. The contribution of the services sector to the aggregate markup increased from 46% in 1980 to 72% in 2015. About two-thirds of the increase in the aggregate markup is driven by the rise of the average markup in the services sector and 7% due to the sectoral reallocation.

By relying on Roy's (1947) identity to derive the Marshallian demand function, I show that in theory three channels can affect firms' markups through the price elasticity of demand: (i) changes in prices, (ii) changes in incomes, and (iii) changes in the demand composition. I present two key propositions to establish their effects on the price elasticity of demand and therefore on markups (Section A of the Online Appendix). The first provides conditions for the price elasticity of demand to be decreasing in the consumer's income. This relates to Harrod's (1936) *Law of Diminishing Elasticity of Demand*, which Bretherton (1937) summarizes as follows

> *"...as people's incomes become larger, the ratio between the trouble involved in finding the cheapest market, and the real gain in utility which will result in so doing, increases."*

[1]For instance, a high-level Competition Council was established in the United States in the summer of 2021 with the responsibility of overseeing policy measures related to competition. Several other countries, including emerging and developing economies, have recently strengthened their competition laws to curb anticompetitive practices. Syverson (2019) highlights the open questions in the macroeconomics of market power that need to be addressed.

[2]See, for instance, Autor, Dorn, Katz, Patterson, and Van Reenen (2020), Bond, Hashemi, Kaplan, and Zoch (2021), De Ridder, Grassi, and Morzenti (2022), Raval (2022), and Traina (2018).

This statement is interpreted here as implying that a consumer's price elasticity of demand for goods and services is lower the wealthier they are. The second proposition establishes the conditions for the price elasticity of demand to be increasing in the product's price. This relates to Marshall's (1890) *Second Law of Demand* and implies that the cost pass-through is smaller than one.

The insights from these two propositions are then embedded into a model of structural change that rationalizes the observed trends in markups, relative price of services, and services share. In particular, I propose a class of non-homothetic preferences that allow these propositions to hold. The economy features two sectors with monopolistically competitive firms offering differentiated varieties of goods and services. Firms are retailers and, hence, sell directly to consumers. There are two types of consumers, high-skilled (or wealthy) and low-skilled (or poor), who have preferences over goods and services. Since households have different incomes, their price elasticities of demand for goods and services differ (Section 3).

The economic intuition behind the mechanisms that give rise to increasing markups in the model is simple and that is where the two forces of structural change are handy—namely, income effects and relative price changes caused by differential rates of technological progress across sectors.

As productivity grows faster in the non-services sector, marginal costs decline at a faster rate for these firms. This in turn allows them to reduce their prices. The cost pass-through is less than one, however. Hence, some of the efficiency gains will be retained by the firm in the form of higher markups. This driver of structural change explains the increase in the average markup of firms in the non-services sector. There is a caveat, however. As households buy more goods, the share of the non-services sector in the aggregate output could increase. This is not observed in the data, so a second driver of structural change is necessary.

Income effects play the countervailing role. As households' income grows, commodities that were luxuries become more accessible and consumption starts flowing toward the sector providing them—i.e., the services sector. Hence, the services share increases. There is a second effect resulting from the introduction of non-homothetic preferences. As households' income increases, their price elasticity of demand decreases. Therefore, firms catering to wealthier consumers, who are now more willing to buy their services, are able to command higher markups. This driver of structural change explains both the rise of the services share and why the average markup in the services sector has increased.

The paper quantifies the role of three exogenous drivers on the rise of markups between 1980 and 2015: (i) technological change (both neutral and skill-biased), (ii) the rise in entry costs, and (iii) the increase in the fraction of wealthier consumers (Section 4). Skill-biased technological progress is the main driver of the rise of markups. This results from two effects. The first is that it reduces marginal costs. Second, and more importantly, it creates wealthier consumers whose willingness to buy services grows over time.

Four counterfactual experiments are performed (Section 5). In the first experiment, the increase in skill-biased technological change is shut down to keep incomes of high- and low-skilled households at their 1980 levels. This experiment implies a much lower level of income inequality than in the data. The resulting aggregate markup would have been lower than its level in 1980. As households are poorer, their willingness to buy luxuries is now reduced. Although there are fewer firms entering the market, consumers are still better off in this economy.

In the second experiment, the neutral technological changes are shut down to keep the prices of goods and services at their 1980 levels. Although prices of goods and services would be higher, the resulting aggregate markup is half as high as it was in 2015 relative to the baseline economy. Because goods and services are now more expensive, households are less willing to buy them, reducing the room for firms to increase their markups. This is particularly noticeable in the goods sector because the decline in prices in the baseline economy was much larger. This in turn discourages potential entrants, leading to fewer firms operating in the market.

In the third experiment, entry costs are shut down. Although these are estimated to have increased about 3 percentage points of sectoral output between 1980 and 2015, reducing them has little bearing on markups. Reducing entry barriers has, however, a direct impact on the number of active firms, which increased significantly.

In the fourth experiment, the increase in the share of wealthy consumers is shut down. The economy here features a larger share of poorer households than in the data. Firms are now more likely to sell their goods and services to poorer consumers than in the baseline economy. Because the demand share of low-skilled households increases, firms put more weight on their price elasticity of demand, which decreases the markup of firms in both sectors. Nonetheless, changes in the composition of firms' customer base has little effect on the aggregate markup.

I also present additional extensions to the quantitative framework (Section 6). First, the market structure now features Cournot competition. This allows the markup of a firm to be determined by both changes in consumers' price elasticities of demand as well as the number of competing firms in the sector. In counterfactual experiments, the rise of entry costs now gains more prominence. In another exercise, I show how other commonly used preferences (e.g. Kimball, non-homothetic CES) have to be modified to study this particular problem.

The last part of the paper assesses the extent to which price elasticities of demand vary across the income distribution (Section B of the Online Appendix). This issue is addressed by conducting a new online survey eliciting consumers' price elasticities of demand for broad categories of goods and services. The survey questions are designed as experiments to capture individuals' perceptions of the impact of changes in prices on their purchase of different goods and services. In the main empirical exercise, individuals' perceived price elasticities of demand for a product are regressed on their incomes and a set of demographic controls. The key finding is that wealthier households are less willing to adjust their demand when prices increase, while less well-off consumers are more likely to reduce their demand. This holds for different categories of goods and services. For instance, in response to a 20% price increase of child care, 87% of households earning between \$50,000 and \$60,000 would reduce their demand for such services as opposed to less than 38% for households earning between \$150,000 and \$200,000.

1.1 Related literature

This paper builds on two different strands of literature. The first relates to the recent wave of studies on the macroeconomics of market power that seek to understand the causes and consequences of the rise in markups. The second relates to the literature on structural change that seeks to understand the drivers of the rebalancing of economic activity toward the services sector. Each of these is discussed in greater detail below.

First, this paper builds on the empirical studies on market power. De Loecker, Eeckhout, and Unger (2020) documented a generalized increase in market power in the United States over the last forty years. Following a similar methodology, Díez, Fan, and Villegas-Sánchez (2021) use data on both listed and private firms across many advanced economies to confirm the rise of market power.[3] This paper contributes to this literature by showing that, using their empirical strategy, the services sector has been the main driver of the rise in markups. Other papers have studied the rise in markups and its relation with income and changes in preferences. Döpper, MacKay, Miller, and Stiebale (2022) find that consumers have become less price sensitive over time. Sangani (2023) provides empirical evidence that the price elasticity of demand for retail goods declines with income. Both papers provide support for the theory proposed in this paper that increases in income have led to a fall in price elasticities of demand. This paper also complements the quantitative work on markups. Edmond, Midrigan, and Xu (2021) study the welfare costs of markup distortions. In line with this paper, they cast doubts on the possibility of increasing barriers to entry as the main force behind the rise in markups. De Loecker, Eeckhout, and Mongey (2021) study the relationship between business dynamism and the rise of market power in a framework with rich heterogeneity. They show that technological change is an important driver of the increase in markups, which is also the case in this paper.[4]

Second, this paper is also related to the literature on structural change. The interest in the services sector follows the work of Buera and Kaboski (2012), who study the rise of high-skilled labor in driving the growth of the services sector, and Hsieh and Rossi-Hansberg (2021), who show that the services sector underwent its own industrial revolution. The differential rates of neutral technological progress across services and non-services was discussed by Ngai and Pissarides (2007).[5] The process of skill-biased structural change draws on Buera, Kaboski, Rogerson, and Vizcaino (2021),

[3] Hall (2018) shows that markups across industries have increased. Other papers have discussed the potential pitfalls of the production function estimation strategy. Raval (2022) shows that using other variable inputs to recover firms' markups can deliver a different distribution of markups. Traina (2018) also argues that including administrative expenses would display a smaller increase in markups in the United States. Bond, Hashemi, Kaplan, and Zoch (2021) show that relying on firms' revenue to estimate output elasticities might distort the level of markups. In contrast, De Ridder, Grassi, and Morzenti (2022) assess the biases in markup estimates from using revenue and show that these estimates are highly correlated with true markups.

[4] This paper also complements the alternative hypotheses for the rise in markups. Namely, De Ridder (2021) proposes the increasing use of intangible inputs as an important driver of the rise in market power, and Afrouzi, Drenik, and Kim (2021) show that market power is associated with the size of firms' customer base and propose a framework in which firms grow through customer acquisition.

[5] Fan, Peters, and Zilibotti (2022) show that productivity growth in consumer services is an important driver of structural change.

who use it to explain the rise of the skill premium. Bridgman and Herrendorf (2022) propose a model of structural change with input-output linkages to study the decline of the labor share, which is intimately related to markups. This paper adds imperfect competition to that class of models and puts on the map the rise of markups as a consequence of structural change.

Non-homothetic preferences play an important role in the structural transformation literature, which is also the case in this paper. I build on the work of Kongsamut, Rebelo, and Xie (2001), who use Stone-Geary preferences to generate the reallocation of economic activity away from agriculture toward manufacturing and services; Boppart (2014), who proposes preferences suited to analyze the joint role of changes in relative prices and income as drivers of structural change; Comin, Lashkari, and Mestieri (2021) and Matsuyama (2019), who use generalized CES preferences to study structural transformation; Bertoletti and Etro (2017) and Bertoletti, Etro, and Simonovska (2018), who use indirectly additive preferences with monopolistically competitive firms to study the gains from trade liberalization. I contribute to this strand of the literature by proposing preferences that extend our understanding of the relationship between structural transformation and variable markups.

2 Empirical Evidence

The goal of this section is to document three facts. First, the process of structural transformation is accompanied by an increase in the relative price of services. Second, the rise of services is the main driving force of the rise of market power, connecting structural change to the rise of markups. Third, the competing explanations for the rise of market power, namely the rise of fixed costs and the emergence of superstar firms, are not able to jointly explain the rise of markups together with the differential role of the services sector.

2.1 Data description

Several datasets at the industry and firm levels for the United States and other advanced economies are used. These datasets are briefly described below and the key variables used in the analysis are highlighted.

Industry-level data. Industry-level data is taken from the EUKLEMS dataset ("Basic File"), which covers the U.S. and several European countries for the period 1970-2015 (see van Ark and Jäger (2017)). This dataset allows me to compute sectoral value added shares and cost shares as well as relative prices. The goods or non-services sector is comprised of: Agriculture, Forestry, and Fishing; Mining and Quarrying; Manufacturing; Electricity, Gas, and Water Supply; and Construction. The services sector corresponds to the remaining industries.[6]

Goods- and services-producing industries. In line with the BLS, service-producing industries are defined using two-digit NAICS sector codes and comprise all industries with code 42 and higher. The goods sector encompasses all the other remaining primary and secondary sectors. At the European

[6] Activities of households as employers and extraterritorial organizations are excluded from the analysis.

level, I use concordance tables between NAICS and NACE codes to define the goods and services sectors.[7]

U.S. firm-level data. Firm-level data for the United States relies on Compustat, which is widely used in the estimation of production functions and the computation of markups.[8] The dataset provides financial information on listed firms starting in the 1950s and includes measures of sales, input expenditures, and capital. It also includes the firms' main industry classification, which allows me to group firms into non-services and services sectors. Despite being widely used, the dataset also poses some limitations. First, it only includes publicly traded firms. Second, only sales are recorded and therefore prices cannot be distinguished from quantities. Bond, Hashemi, Kaplan, and Zoch (2021) discuss the issue with the latter in greater detail. Whenever appropriate, the analysis is complemented with the Census Bureau's Business Dynamics Statistics (BDS). This dataset offers detailed statistics on the firm-size distribution and the sectoral composition of firms.

European firm-level data. The analysis at the European-level is based on Orbis Historical provided by Bureau van Dijk. Orbis provides harmonized cross-country financial information for both private and public firms since the mid-90s for many European countries. This dataset allows me to compute firm-level statistics, including markups, for millions of firms across Europe. Díez, Fan, and Villegas-Sánchez (2021) also use it to document the rise of market power across many countries. The cleaning steps closely follow Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, and Yesiltas (2019) and Gopinath, Kalemli-Ozcan, Karabarbounis, and Villegas-Sanchez (2017). This dataset has greater industry coverage but has a shorter time span.

2.2 Structural change and the increase of the relative price of services

The reallocation of economic activity and employment from agriculture and manufacturing toward the services sector—structural change—is accompanied in the United States and several other advanced countries by an increase in the relative price of services. Figure 2.1a shows that the relative (value added) price of service industries (over non-service industries) increased about 44% since 1980. Figure 2.1b presents the rise of the services sector, measured both as valued added and as a final consumption share of households' incomes. Both measures show that the shares of the services sector increased by more than 13 percentage points between 1980 and 2015, hovering around 79% of the economy (using value added shares).[9] Figure 2.1c displays the cumulated inflation rate for selected

[7] I further break each sector into broad economic categories (BEC) following the United Nations BEC's classification. Following their System of National Accounts end-use dimension, I define three types of goods: (i) consumption goods, (ii) intermediate goods, and (iii) capital goods; as well as two type of services: (i) consumer services and (ii) producer services. The mapping between NAICS and BEC is done at the 6-digit level. Since there is no direct concordance table, a few steps are required to make that mapping feasible. The first step is to use the conversion table between the Standard International Trade Classification (SITC) and BEC, the second is to use the concordance table between the International Standard of Industrial Classification (ISIC) codes and the SITC, and then finally map ISIC codes to NAICS codes. Expert judgment is then needed to ensure there are no weird outliers.

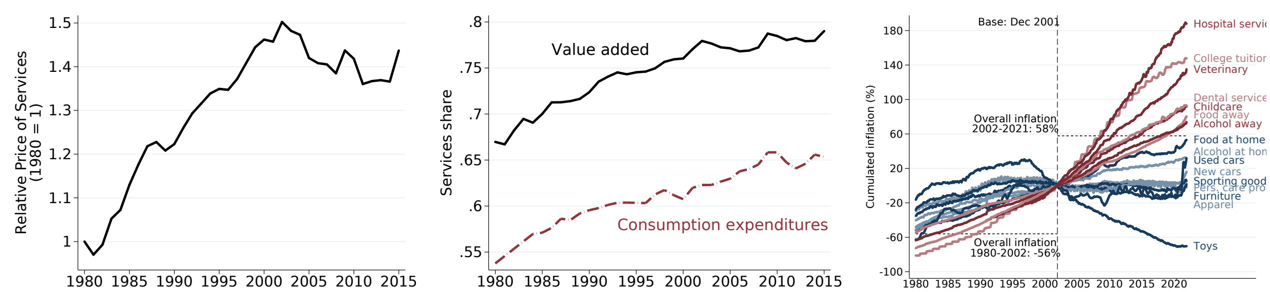
[8] See De Loecker and Warzynski (2012) for the empirical methodology used to estimate firm-level output elasticities and markups.

[9] See Figure C.1.1 for the services share and relative price of services in the U.S. starting in 1947 and Figure C.1.2 in Appendix C.1 for these trends for a few other European countries.

goods (in blue) and services (in red) relative to December 2001. There has been a rapid increase in the average prices of hospital services, college tuition, dental services, food and alcoholic beverages consumed away from home over the last forty years, while the prices of traditional goods have risen at a much slower pace.

Why are these trends important? The increase of the relative price of services is intimately related with the evolution of both markups and marginal costs across sectors. On the one hand, this hints that the services sector had potentially larger markups than the non-services sector and/or that the latter has experienced faster productivity growth than the former. On the other hand, the rise of the services share help explains the reallocation of economic activity towards higher markup firms. The next subsection documents that, indeed, markups in the services sector have experienced faster growth relative to non-services firms.

Figure 2.1: Structural change in the U.S.



(a) Relative price of services

(b) Services shares

(c) Inflation of goods and services

Note: Panel (a) shows the relative price of service industries, measured as the chain-weighted Fisher price index of the value added price indices of individual industries, using data from EUKLEMS. Panel (b) shows the value added share of the services sector (black) and the share of income spent on the final consumption of services (red), using data from EUKLEMS and the BLS's Consumption Expenditure Survey. Panel (c) shows the cumulated inflation of selected goods (blue) and services (red) relative to December 2001 and contrasts these trends with the overall inflation in December 2021 (58%) and January 1980 (-56%), using data from the BLS's Consumer Price Index.

2.3 Markups and the rising importance of services

A markup is defined as the ratio of a firm's output price to its marginal cost. From a firm's cost minimization problem, this can be shown to equal the ratio of the firm's output elasticity to a variable input and its sales share. The numerator is usually obtained by estimating firms' production functions. The denominator can be read off directly from balance sheet data. For the United States, the estimated output elasticities are taken from De Loecker, Eeckhout, and Unger (2020) to ensure the results are not driven by differences in estimation routines and the sales shares of variable inputs are taken from Compustat (here *cost of goods sold* or `Cogs`).[10] The aggregate markup is then the weighted sum of firm-level markups, where a firm's weight is the ratio of its variable costs to total variable

[10] The challenges related with the estimation of production functions is discussed in Bond, Hashemi, Kaplan, and Zoch (2021), Basu (2019), and Syverson (2019).

costs.[11]

The aggregate markup, M_t, can also be written as the sum of sectoral markups where each sector's markup is simply the product of its sectoral cost share in the aggregate economy, ω_{j_t}, and the average markup within that sector, $\overline{m}_{j_t}$ for $j = \{G, S\}$; i.e.,

$$M_t = \underbrace{(1-\omega_{S_t})\,\overline{m}_{G_t}}_{M_{G_t}} + \underbrace{\omega_{S_t}\overline{m}_{S_t}}_{M_{S_t}}.$$

The services (cost) share is measured using industry-level data from EUKLEMS, which accounts for the entire industrial production of the economy, and refers to the *compensation of employees* and *intermediate inputs*.[12] The average markup within each sector is based on firm-level data from Compustat, with the underlying assumption that the estimated markup of listed firms is a good proxy for the markups of nonlisted firms.[13]

Figure 2.2a shows the aggregate markup, M_t, over time using both sectoral cost and gross output shares (from EUKLEMS). The aggregate markup (using cost shares) increased 11% between 1980 and 2015. The increase in markups is stronger if output shares are used to measure average markups and sectoral shares, displaying an increase of 46% between 1980 and 2015. Figure 2.2b displays the contribution of each sector to the aggregate markup, (M_{j_t}/M_t). The increase in the aggregate markup is entirely driven by the services sector. Between 1980 and 2015, the contribution of services grew by more than 26 percentage points, from 46% to 72% of the aggregate markup. This increase is driven by a stronger growth rate of the average markup of services and by an increases in the services share as displayed in Table 2.1.

Table 2.1: Average markups and sectoral shares

	Non-services			Services		
	1980	2015	Δ	1980	2015	Δ
Average markups (cogs)	**1.13**	**1.21**	**7.3%**	**1.14**	**1.27**	**11.9%**
Average markups (cogs + sga)	1.18	1.44	22.3%	1.19	1.65	37.8%
Average markups (sales)	1.17	1.47	25.5%	1.22	1.86	52.2%
Sectoral shares (comp + II)	**54.0**	**28.6**	**-47.0%**	**46.0**	**71.4**	**55.1%**
Sectoral shares (gross output)	47.4	27.6	-41.8%	52.6	72.4	37.7%

Note: The average markups are computed using Compustat data and sectoral shares using EUKLEMS data.

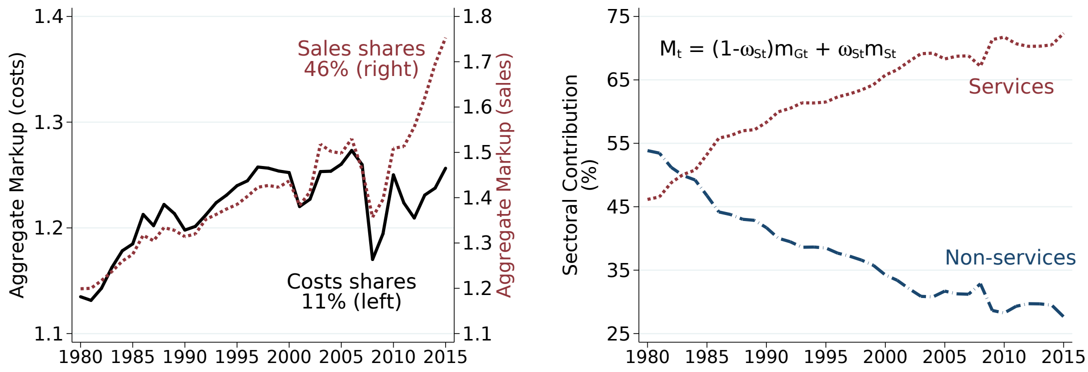
The rise of the aggregate markup can be decomposed into three sources: (i) the rise of the average markup of non-service industries; (ii) the rise of the average markup of service industries; and (iii)

[11] An alternative specification is to compute the aggregate markup as the harmonic mean of firm-level markups weighted by each firm's sales share in total sales.

[12] Compustat firms' cost of goods sold is also used to measure the services (cost) share. That share increased from 35% of total costs in 1980 to 52% in 2015, while industry-level data display an increase from 46% to 71% over the same period. Although the services cost share from Compustat is smaller than what is implied at the aggregate level from industry data, the aggregate markup computed with Compustat's sectoral shares shows the same pattern as the aggregate markup from the industry-level data.

[13] To be precise, the average markup within sector $j = \{G, S\}$ is given by $\overline{m}_{j_t} = \sum_{i=1}^{N_{j_t}} \tilde{\omega}^j_{i_t} m^j_{i_t}$ where $\tilde{\omega}^j_{i_t} = \frac{\mathrm{Cogs}^j_{i_t}}{\sum_{\kappa=1}^{N_{j_t}} \mathrm{Cogs}^j_{\kappa_t}}$ is firm i's cost share in sector $j = \{G, S\}$ variable costs.

Figure 2.2: Aggregate markups in the U.S.



(a) Aggregate markup

(b) Sectoral contribution

Note: Panel (a) shows the aggregate markup measured as the cost-weighted average of firms' markups (black, left axis) and as the sales-weighted average of firms' markups (red, right axis), using Compustat and EUKLEMS data. Panel (b) shows the sectoral contribution to the aggregate markup (non-services in blue, services in red).

the increase of the services share. Specifically,[14]

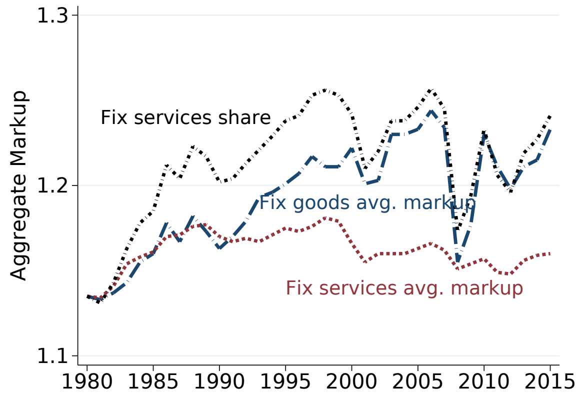
$$
\begin{aligned}
M_{2015} - M_{1980} \quad = \quad & \left(\frac{\omega_{G_{1980}} + \omega_{G_{2015}}}{2}\right)\left(\overline{m}_{G_{2015}} - \overline{m}_{G_{1980}}\right) && \text{Non-services avg. markup (28\% / 20\%)} \\
& + \left(\frac{\omega_{S_{1980}} + \omega_{S_{2015}}}{2}\right)\left(\overline{m}_{S_{2015}} - \overline{m}_{S_{1980}}\right) && \text{Services avg. markup (65\% / 72\%)} \\
& + \left(\frac{\overline{m}_{S_{2015}} - \overline{m}_{G_{2015}} + \overline{m}_{S_{1980}} - \overline{m}_{G_{1980}}}{2}\right)\left(\omega_{S_{2015}} - \omega_{S_{1980}}\right) && \text{Services share (7\% / 8\%).}
\end{aligned}
$$

The increase in average markups within the goods sector contributed 28% to rise of the aggregate markup (and 20% if markups are aggregated using sales shares), while the increase in average markups within the services sector contributed 65% (and 72% if markups are aggregated using sales share). The rise of the services share contributed 7% (and 8% if markups are aggregated using sales share).

Figure 2.3 shows the aggregate markup when (i) the average markup in the goods sector is fixed at its 1980 level, (ii) the average markup in the services sector is fixed at its 1980 level, and (iii) the services share is fixed at its 1980 level. The rise of the average markup within the services sector was the strongest driver of the rise in markups. Shutting down that increase leads to the aggregate markup barely changing over the last forty years. Instead, when the average markup in the non-services sector is shut down, the rise in the average markup of services is still strong enough to drive the aggregate markup up.

[14]The following expression is rescaled by M_{1980} so that the three components sum up to the change in the aggregate markup.

Figure 2.3: Aggregate markups across scenarios



Note: The figure shows the aggregate markup when the average markup within each sector is fixed at its 1980 level (non-services in blue, services in red) and when the sectoral cost shares are fixed at their 1980 level (black), using Compustat and EUKLEMS data.

2.4 Not the superstars, nor the fixed costs

The rise in markups has been linked to a rise in monopoly power, potentially related with the emergence of superstar firms and the rapid increase in fixed costs and barriers to entry. To address these arguments, I now focus on superstar firms—here defined as firms in the right tail of the markup distribution—and fixed costs. Table 2.2 shows the correlation coefficients between the growth rate of a firm's markup between 1980 and 2015 and the growth rates of its sales share in total sales and its fixed costs share in total fixed costs. Among the surviving firms, we should expect that firms that experienced significant increases in markups would also have gained market share through an increase in sales. As the table shows, the rise in a firm's markup is not positively correlated with an increase of its market share (first column). We should also expect firms with higher increases in fixed costs to have increased their markups. As the table shows, that is not the case (second column).

Table 2.2: Correlation coefficients for the 1980-2015 change in markups, sales, and cost shares

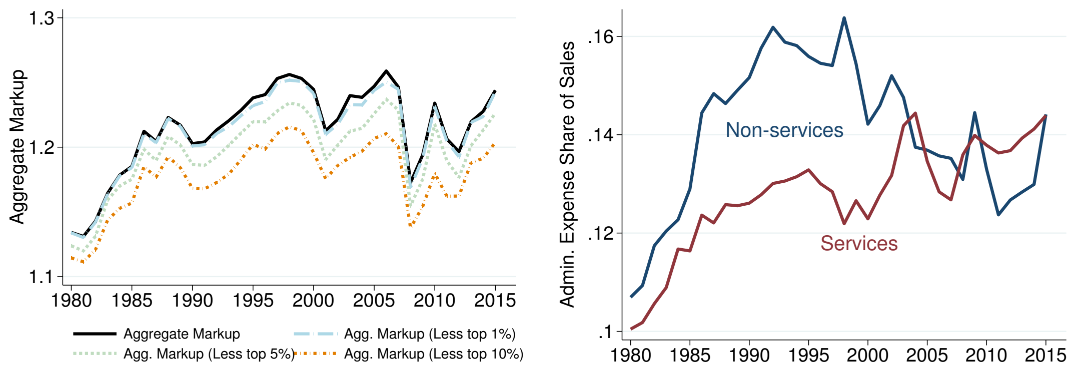
	Δ Markups, Δ Sales share	**Δ Markups, Δ Fixed costs share**
Aggregate	-0.0790**	-0.0212
Non-services	-0.0465	0.0343
Services	-0.1028*	-0.0464

Note: Each variable corresponds to the three-year average centered around 1980 and 2015. Correlation coefficients are weighted by firms' cost shares. ** $p<0.05$,* $p<0.1$. Data is taken from Compustat.

When dropping firms that are in the top 1%, 5%, and 10% of the markup distribution within each sector and year, the increase in aggregate markups over the last four decades is still noticeable—albeit to a smaller extent as displayed in Figure 2.4a. In particular, the rapid increase between 1980 and the 2000's does not appear to be driven by the superstars, nor the uptick experienced after the

Great Financial Crisis.[15] Figure 2.4b shows total general and administrative expenses over total sales for each sector. Starting in 1980, fixed costs increased from about 10% of revenue to less than 15% in 2015. Yet, the non-services sector saw a stronger increase in fixed costs relative to the services sector in the 1980s and 1990s, despite having lower average markups. The rise of superstar firms and fixed costs are certainly part of the explanation of why markups have increased but they are not sufficient to capture the differences in the rise of markups across sectors.

Figure 2.4: Ruling out competing drivers



(a) Dropping the right tail

(b) Fixed costs/Sales

Note: Panel (a) shows the aggregate markup when firms in the top 1%, 5%, and 10% of the markup distribution within each sector and year are removed, using Compustat data. Panel (b) shows the sum of Selling, General, and Administrative Expenses divided by the sum of Sales for the non-services (blue) and the services (red) sectors, using Compustat data.

I decompose the services and the non-services sector into subsectors according to their purpose. Firms producing consumption goods and in the consumer service industries have had on average higher markups than their peers. Table 2.3 summarizes the changes in average markups for different categories of goods and services-producing firms. The upshot of the table is that the firms in consumer services industries have been a key contributor to the rise of markups.

Table 2.3: Average markups within sectors

	Non-services			**Services**		
	1980	**2015**	**Δ**	**1980**	**2015**	**Δ**
Average markups (cogs)	**1.13**	**1.21**	**7.3%**	**1.14**	**1.27**	**11.9%**
Capital goods	1.12	1.24	10.9%			
Consumption goods	1.19	1.67	40.5%			
Intermediate goods	1.12	1.13	0.4%			
Consumer services				1.19	1.29	8.6%
Producer services				1.06	1.20	12.6%

Note: The average (cost-weighted) markups are computed using Compustat data.

Do changes in capital intensity or capital shares help explain the sectoral differences in markups?

[15]Figure C.1.3 in Appendix C.1 replicates the decomposition of the aggregate markups into goods and services. As before, the services sector drives most of the increase in aggregate market power in the absence of the superstars.

The answer is no. Capital intensity, measured as the ratio of capital to costs of goods sold, has been fairly stable over the past forty years for the services sector at a factor of 1, while the ratio is about 1.5 for the non-services sector. The capital share of sales is also smaller in the services sector than in the non-services sector. Additionally, the share of sales spent on variable inputs has been steadily declining across both sectors.[16] These trends are confirmed by the regression results presented in Table C.2.1, which shows the correlation between firms' markups and their capital share (tangible and intangible), variable costs share, and fixed costs share of sales. An increase in the capital share, both tangible and intangible, of non-services firms tends to have a larger impact on their markups than for services-producing firms (e.g. a one percentage point increase in a firm's capital share is associated with a 0.028% increase in its markups for non-services and 0.017% for services). A one percentage point decline in a services firm's variable costs tends to have a stronger impact on its markup (a 1.228% increase as opposed to 1.072% for non-services firms).

2.5 Markups and the rising importance of services across Europe

The importance of services is not unique to the United States. A similar pattern emerges when looking at several European countries. Using Orbis data to estimate production functions at the three-digit NACE code for both private and listed firms in each country separately, markups are computed as the ratio of the output elasticity of labor and materials over their sales shares. In each country, markups are then aggregated using variable cost shares. Figure 2.5 shows the difference in average markups in the services sector versus the non-services sector for several European countries, in which the average markup is the mean over each country's sample period. With the exception of Austria, Latvia, and Slovakia, services firms have higher markups than their peers (ranging from 4% for Spain to twice as large for the Netherlands and Portugal).

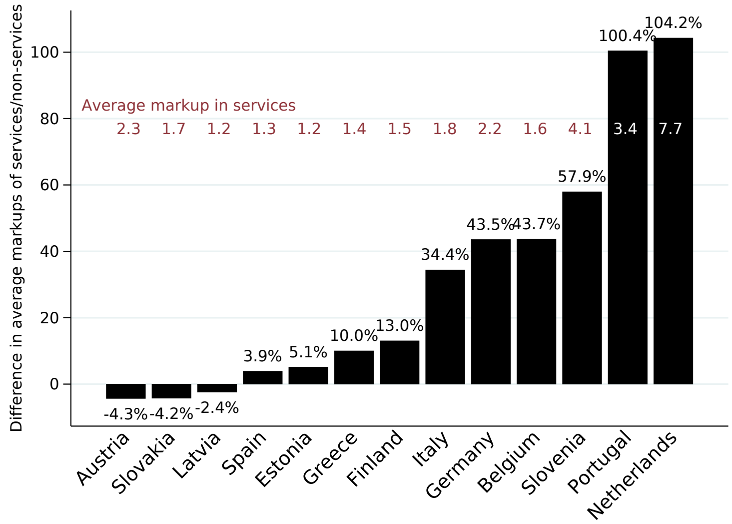
3 A Model of Structural Change and Rising Markups

The empirical evidence presented in Section 2 shows that markups increased faster in services than in the non-services sector and that usual supply-side metrics are insufficient to explain why that is the case. Drawing on the literature on structural transformation, I present a general equilibrium model where demand forces gain more prominence. In particular, I add two ingredients to study the rise of markups in an otherwise standard model of structural change: (i) monopolistic competition, and (ii) non-homothetic preferences. The first ingredient introduces firm-level markups, while the second makes them endogenous and responsive to changes in productivity, consumers' incomes, and firms' demand composition.

The Online Appendix A presents the theoretical underpinning for this avenue. In a nutshell, when preferences are non-homothetic, a consumer's price elasticity of demand can vary endogenously along a variety's price and the consumer's income. As a result, the standard forces of structural change—namely, differential rates of technological progress across sectors and income effects—not

[16] See Figure C.1.4 in Appendix C.1 for the trends in the gross profit margin, capital intensity, capital share of sales, the costs of goods share of sales across sectors.

Figure 2.5: Difference in average markups of services vs. non-services



Note: The figure shows the difference in average markups of the services sector relative to the non-services sector, using data from Orbis, aggregated using cost shares and averaged over the sample period (from 1993 (Belgium and Netherlands), 1994 (Spain), 1995 (Greece), 1996 (Italy), 1997 (Finland and Slovenia), 1998 (Estonia), 1999 (Portugal), 2000 (Latvia), 2002 (Germany), 2004 (Austria and Slovakia) to 2015 (all countries)). The plot also shows the average markup in services in red over the sample period.

only account for the rise of the services share and the increase in the relative price of services, but will also drive the increase in markups within both sectors.

The forces of structural change will however play opposing roles. On the one hand, faster technological progress in the non-services sector that brings the average price of goods down relative to services will also lead to an increase in the average markup of goods. The reason is that consumers' price elasticities of demand are increasing with prices. On the other hand, as households become richer and start shifting their consumption toward services, firms will adjust their prices upward leading to an increase in the average markup of services. The reason is that consumers' price elasticities of demand are decreasing with income.

Environment. Time is discrete and indexed by t. The economy is populated by a unit mass of heterogeneous households that differ in their skill level, which can either be high or low, $i \in \{H, L\}$. A fraction μ_H are high-skilled, while $\mu_L \equiv (1 - \mu_H)$ are low-skilled. Households are endowed with one unit of productive time that is supplied inelastically in the labor market in exchange for the wage w_i. High-skilled workers receive a skill premium in the labor market, i.e. $w_H / w_L > 1$. Households also receive nonlabor earnings Λ_i from owning firms.

There are two sectors in this economy, one that produces goods and another that produces services, $j \in \{G, S\}$. Within each sector, there is a continuum of firms producing a differentiated variety of commodity j and behaving as monopolistic competitors. Labor is freely mobile across sectors and firms take factor prices as given. A variety of commodity j differs in terms of its price and quality, both chosen by the firm. Firms are retailers and therefore sell directly to consumers. A variety can be purchased by both types of households for the same price and quality.

3.1 Households

Preferences. Preferences play an important role in determining the sources of market power as discussed in the Online Appendix A. In what follows, preferences will be non-homothetic (in prices) and admit an analytic representation for both the direct and indirect utilities.

Households have preferences over consumption of different varieties of goods and services, denoted $\mathbf{c}_{G_t}$ and $\mathbf{c}_{S_t}$, and their respective quality, $\mathbf{q}_{G_t}$ and $\mathbf{q}_{S_t}$, where the bold variables correspond to vectors of the different varieties of goods and services. Preferences are represented by the direct utility function $u(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})$. Each variety ω of goods and services is indexed by its price $p_{j_t}(\omega)$ and quality $q_{j_t}(\omega)$ taken as given by the household.

I proceed with the indirect utility function, i.e., the household's maximal attainable utility given her income, e_t, the vector of prices of goods and services, $\mathbf{p}_{G_t}$ and $\mathbf{p}_{S_t}$, and their respective quality, $\mathbf{q}_{G_t}$ and $\mathbf{q}_{S_t}$. Let the indirect utility be a composite of two sectoral indirect utilities, one for goods and one for services, aggregated in a Cobb-Douglas fashion according to

$$v(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t}) \quad = \quad v_G(e_t, \mathbf{p}_{G_t}, \mathbf{q}_{G_t})^{\lambda} \, v_S(e_t, \mathbf{p}_{S_t}, \mathbf{q}_{S_t})^{1-\lambda}, \tag{3.1}$$

where $\lambda \in (0, 1)$ is the weight on the indirect utility from goods. Each sectoral indirect utility is in turn additively separable across the differentiated varieties of commodity j, implying

$$v_j(e_t, \mathbf{p}_{j_t}, \mathbf{q}_{j_t}) \quad = \quad \int_{\mathcal{N}_{j_t}} \widehat{v}_j \left(e_t, p_{j_t}(\omega), q_{j_t}(\omega)\right) d\omega, \tag{3.2}$$

where the sector-specific indirect subutility satisfies the standard properties of indirect utility functions as defined in Assumption D.1 in Appendix D. The sectoral indirect subutility for each variety ω of commodity j is taken to be

$$\widehat{v}_j \left(e_t, p_{j_t}(\omega), q_{j_t}(\omega)\right) \quad = \quad \frac{1}{1+\gamma} \left[\frac{\left(\phi_j e_t - p_{j_t}(\omega)\right) q_{j_t}(\omega)^{\delta}}{e_t} \right]^{1+\gamma} \qquad \text{for } \; p_{j_t}(\omega) \leqslant \phi_j e_t \tag{3.3}$$

and zero otherwise. Here, $\phi_j e_t > 0$ is the sectoral choke price of any variety of commodity $j \in \{G, S\}$, i.e., the maximum price the household is willing to pay in order to consume a positive amount of that variety. A price above the consumer's choke price is not purchased and therefore yields a utility of zero. The higher the value of $\phi_j > 0$, the higher is the consumer's choke price. Similarly, the higher the household's income e_t, the higher is her choke price. Each variety is weighted by its quality $q_{j_t}(\omega)$. Varieties of higher quality are valued more than low-quality varieties. The parameter $\delta > 0$ is a quality-specific weight and $\gamma > 0$ ensures demand satisfies the law of demand. These parameters are common for both goods and services.

Proposition 3.1 shows that there is an analytic representation of the direct utility when the indirect utility has the above form (equations (3.1), (3.2), and (3.3)). Proposition 3.2 further demonstrates that the indirect utility collapses to the two-sector CES direct utility with quality. To proceed, let $\widehat{C}_t$ denote an aggregator of total consumption, given by

$$\widehat{C}_t = \phi_G \int_{\mathcal{N}_{G_t}} c_{G_t}(\omega) d\omega + \phi_S \int_{\mathcal{N}_{S_t}} c_{S_t}(\omega) d\omega,$$

and $\widetilde{C}_{j_t}$ denote a quality-adjusted composite of the different varieties of commodity $j = \{G, S\}$, given by

$$\widetilde{C}_{j_t} = \left(\int_{\mathcal{N}_{j_t}} \left[\frac{c_{j_t}(\omega)}{q_{j_t}(\omega)^{\delta}} \right]^{\frac{1+\gamma}{\gamma}} d\omega \right)^{\frac{\gamma}{1+\gamma}}.$$

Proposition 3.1. (Direct utility) *The indirect utility (equation (3.1)) admits an analytic representation of the direct utility given by*

$$u(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t}) = \psi \left[\frac{\widehat{C}_t - 1}{\widetilde{C}_{G_t}^{\lambda} \widetilde{C}_{S_t}^{(1-\lambda)}} \right]^{(1+\gamma)},$$

where $\psi = (1+\gamma)^{-1} \lambda^{\lambda(1+\gamma)} (1-\lambda)^{(1-\lambda)(1+\gamma)} > 0$.

Proof. See Appendix D.2. □

Proposition 3.2. (Two-sector CES) *Assume $\phi_j = 0$ for $j = \{G, S\}$, $\gamma < -1$, and $\delta < 0$. Then, these preferences collapse to the two-sector CES preferences with quality and $(-\gamma)$ as the elasticity of substitution within each sector.*

Proof. See Appendix D.3. □

Budget constraint. The budget constraint the household faces requires that total spending on goods and services, e_t, be paid for with labor income w_t and nonlabor earnings Λ_t. A household of skill i faces the following budget constraint

$$e_t \equiv \sum_{j=G,S} \int_{\mathcal{N}_{j_t}} p_{j_t}(\omega) c_{j_t}(\omega) d\omega = w_t + \Lambda_t. \tag{3.4}$$

Demand for varieties. The household's demand for each variety of goods and services can be recovered using Roy's identity.[17] Demand for variety ω of commodity $j \in \{G, S\}$ can then be expressed

[17]Household i's demand for variety ω of commodity j can be expressed using Roy's identity as (abstracting from functions' arguments)

$$c_{j_t}(\omega) = -\frac{\left(\partial \widehat{v}_{j_t}(\omega)/\partial p_{j_t}(\omega)\right) e_t}{\left(v_{j_t}/\lambda_j\right) \Phi_t},$$

where $\left(\partial \widehat{v}_{j_t}(\omega)/\partial p_{j_t}(\omega)\right)$ is the derivative of the sectoral indirect subutility with respect to the price of the variety, v_{j_t} is sectoral indirect utility defined in equation (3.2), $\Phi_t = (\partial v_t/\partial e_t)(e_t/v_t) > 0$ is the total utility's income elasticity, $\lambda_G = \lambda$, and $\lambda_S = (1-\lambda)$.

as

$$c_{j_t}(\omega) \;=\; \Bigg[\underbrace{\phi_j\, e_t}_{\text{choke price}} - p_{j_t}(\omega)\Bigg]^{\gamma} \underbrace{q_{j_t}(\omega)^{\delta(1+\gamma)}}_{\text{variety quality}} \underbrace{A_{j_t}}_{\text{sectoral composite}} \quad \text{for } \; p_{j_t}(\omega) \leqslant \phi_j e_t$$

and zero otherwise. Here, $A_{j_t} = \widetilde{C}_{j_t}^{(1+\gamma)}\left[\psi_j\, e_t\left(\widehat{C}_t - 1\right)\right]^{-\gamma}$ is a household-specific sectoral composite.[18] Note that firms will not set a price above households' choke prices, nor choose a quality level of zero, as that would make the demand for their variety be zero.

As $\gamma > 0$, the household's demand satisfies the law of demand and therefore her quantity demanded varies inversely with the variety's price. In particular, it increases with the distance of the variety's price to the maximum amount the household is willing to pay to consume it—the commodity's choke price. Hence, all else equal, lower-priced varieties are associated with more consumption. Similarly, the higher the quality of the variety, the larger is the household's demand for that variety. The consumption demand for different varieties of the same commodity $j = \{G, S\}$ only varies because of differences in prices and quality.

Consumption spending shares. The household's spending share on services is given by

$$\varpi_{S_t} \;\equiv\; \frac{\int_{\mathcal{N}_{S_t}} p_{S_t}(\omega) c_{S_t}(\omega)\, d\omega}{e_t} \;=\; \frac{(1-\lambda)\Phi_{S_t}}{\lambda\Phi_{G_t} + (1-\lambda)\Phi_{S_t}}, \tag{3.5}$$

where the income elasticity of the sectoral utility, $\Phi_{j_t} \equiv \int_{\mathcal{N}_{j_t}} \frac{\partial \widehat{v}_j(e_t, p_{j_t}(\omega), q_{j_t}(\omega))}{\partial e_t} \frac{e_t}{v_j(e_t, \mathbf{p}_{j_t}, \mathbf{q}_{j_t})}\, d\omega > 0$.[19] Conversely, the household's spending share on all the varieties of goods is $\varpi_{G_t} = 1 - \varpi_{S_t}$.

Price elasticity of demand. The household's consumption demand yields a direct price elasticity of demand that depends on her income and the price of the particular variety demanded. Let $\xi_{j_t}(\omega)$ denote the (negative of the) percentage change in quantity demanded of variety ω of commodity j in response to a percentage change in its own price, or $\xi_{j_t}(\omega) \equiv -\frac{\partial c_{j_t}(\omega)}{\partial p_{j_t}(\omega)} \frac{p_{j_t}(\omega)}{c_{j_t}(\omega)}$. The household's price elasticity of demand is

$$\xi_{j_t}(\omega) \;=\; \frac{\gamma\, p_{j_t}(\omega)}{\phi_j e_t - p_{j_t}(\omega)}. \tag{3.6}$$

This expression satisfies the two key propositions A.3 and A.4 discussed in the Online Appendix A.[20] First, demand becomes *less* elastic when the household's income goes up, i.e., the price elasticity of demand is decreasing in the household's income. Given that households face the same price and $e_{H_t} \geqslant e_{L_t}$, the price elasticity of demand of the high-skilled consumer is lower than the elasticity of

[18]Here $\psi_G = (1+\gamma)^{-\lambda}\, \lambda^{\lambda(1+\gamma)}$ and $\psi_S = (1+\gamma)^{-(1-\lambda)}\, (1-\lambda)^{(1-\lambda)(1+\gamma)}$.

[19]The income elasticity of sectoral utility can be written as $\Phi_{j_t} = \frac{1+\gamma}{\gamma} \frac{1}{v_{j_t}} \int \xi_{j_t}(\omega) \widehat{v}_{j_t}(\omega)\, d\omega$, with $\Phi_t = \lambda\Phi_{G_t} + (1-\lambda)\Phi_{S_t}$.

[20]The price super-elasticities of demand with respect to price and income are respectively given by

$$\frac{\partial \xi_{j_t}(\omega)}{\partial p_{j_t}(\omega)} \frac{p_{j_t}(\omega)}{\xi_{j_t}(\omega)} \;=\; 1 + \frac{\xi_{j_t}(\omega)}{\gamma} \qquad \text{and} \qquad \frac{\partial \xi_{j_t}(\omega)}{\partial e_t} \frac{e_t}{\xi_{j_t}(\omega)} \;=\; -\phi_j e_t.$$

the low-skilled consumer, $\xi_{H,j_t}(\omega) \leqslant \xi_{L,j_t}(\omega)$. As a result, firms selling to wealthier customers will be able to charge higher markups than firms selling to poorer consumers.

Second, demand becomes *more* elastic when the price of the variety goes up, i.e., the price elasticity of demand is increasing in the variety's price. As a result, firms selling cheaper varieties will be able to charge higher markups as $\xi_{j_t}(\widehat{\omega}) < \xi_{j_t}(\widetilde{\omega})$ for $p_{j_t}(\widehat{\omega}) < p_{j_t}(\widetilde{\omega})$. Also, as long as the variety's price is such that $p_{j_t}(\omega) > \frac{\phi_j e_{H_t}}{1+\gamma} \geqslant \frac{\phi_j e_{L_t}}{1+\gamma}$, the price elasticity of demand is greater than one. This guarantees that a firm's markup is well defined.

The elasticity of substitution across varieties, i.e. how demand for variety $\widetilde{\omega}$ of commodity j changes in response to a change in the consumption of variety $\widehat{\omega}$ of commodity κ, is equal to the consumer's price elasticity of demand of the variety in the numerator.[21]

Quality elasticity of demand. Let $\sigma_{j_t}(\omega)$ denote the percentage change in quantity demanded of variety ω of commodity j in response to a percentage change in its own quality, or $\sigma_{j_t}(\omega) \equiv \frac{\partial c_{j_t}(\omega)}{\partial q_{j_t}(\omega)} \frac{q_{j_t}(\omega)}{c_{j_t}(\omega)}$. As preferences are homothetic in quality, the quality elasticity of demand does not depend on the household's income, nor on the price or quality of the variety. Hence, both types of households have the same quality elasticity of demand, which is given by

$$\sigma_{j_t}(\omega) \quad = \quad \delta\,(1+\gamma)\,, \tag{3.7}$$

for $j \in \{G, S\}$. As $\delta, \gamma > 0$, an increase in a variety's quality makes households increase their demand for that variety.

3.2 Incumbent firms

Technology. Varieties of goods and services are produced by firms that differ in terms of their total factor productivity (TFP), z_{j_t}. The output of a firm is produced via a constant returns to scale nested CES production function that combines high-skilled labor h_j and low-skilled labor ℓ_j according to

$$y_{j_t} \quad = \quad z_{j_t}\left[\alpha\, x_t h_{j_t}^{\iota} + (1-\alpha)\,\ell_{j_t}^{\iota}\right]^{1/\iota}. \tag{3.8}$$

The substitution parameter between high and low-skilled labor is such that $0 < \iota \leqslant 1$, which implies that high and low-skilled labor are somewhat substitutes in production with the elasticity of substitution given by $\frac{1}{1-\iota}$. This elasticity together with the weight parameter $0 < \alpha < 1$ and skill-biased productivity, x_t, are common to all firms in both sectors. Increases in z_{j_t} and x_t over time reflect neutral and skill-biased technological progress.

Costs. Firms' total costs comprise the wage bill on high and low-skilled workers, expenses associated with quality, and other fixed costs, or

$$tc_{j_t} \quad = \quad w_{H_t} h_{j_t} + w_{L_t}\ell_{j_t} + \kappa q_{j_t}^{\vartheta} + f_{j_t}, \tag{3.9}$$

[21]Let the elasticity of substitution be $E(\widetilde{\omega}_{j_t}, \widehat{\omega}_{\kappa_t}) = -\frac{\partial\left(\frac{c_{j_t}(\widetilde{\omega})}{c_{\kappa_t}(\widehat{\omega})}\right)}{\partial\left(\frac{p_{j_t}(\widetilde{\omega})}{p_{\kappa_t}(\widehat{\omega})}\right)} \frac{\frac{p_{j_t}(\widetilde{\omega})}{p_{\kappa_t}(\widehat{\omega})}}{\frac{c_{j_t}(\widetilde{\omega})}{c_{\kappa_t}(\widehat{\omega})}}$. Then, $E(\widetilde{\omega}_{j_t}, \widehat{\omega}_{\kappa_t}) = \xi_{j_t}(\widetilde{\omega})$. Note that as the price elasticity of demand is greater than one, the elasticity of substitution is also greater than one, which implies that varieties are gross substitutes. Hence, the reduction in the relative quantity demanded of a variety exceeds the increase in its relative price. This leads to a decline of the relative expenditure on that variety. For CES preferences, the elasticity of substitution is given by $E(\widetilde{\omega}_{j_t}, \widehat{\omega}_{\kappa_t}) = -\gamma$.

where $\kappa > 0$ and $\vartheta > 1$ are parameters common across sectors. I will refer to the term f_{j_t} as entry costs, which can vary over time.

The firm solves its cost minimization problem (3.9) by optimally choosing the quantity of high and low-skilled labor it needs to produce the variety of the commodity it sells subject to the technological constraint (3.8). The first-order conditions from this problem yield the factor demands for high and low-skilled labor, with the relative demand for high-skilled labor given by

$$\frac{h_{j_t}}{\ell_{j_t}} = \left[\frac{\alpha x_t}{(1-\alpha)}\frac{w_{L_t}}{w_{H_t}}\right]^{\frac{1}{1-\iota}}, \tag{3.10}$$

which depends on the skill premium and skill-biased productivity. Note that the relative demand for high-skilled labor is the same across sectors.

By replacing these factor demands in the firm's variable cost function, an expression for its marginal cost is obtained as

$$mc_{j_t} = \frac{w_{L_t}}{z_{j_t}}\left[(\alpha x_t)^{\frac{1}{1-\iota}}\left(\frac{w_{H_t}}{w_{L_t}}\right)^{\frac{\iota}{\iota-1}} + (1-\alpha)^{\frac{1}{1-\iota}}\right]^{\frac{\iota-1}{\iota}}. \tag{3.11}$$

Note that the firm's marginal cost is decreasing in neutral as well as in skill-biased technological change as high and low-skilled labor are somewhat substitutes (or $\iota > 0$). On the other hand, marginal costs are increasing in the skill premium, w_{H_t}/w_{L_t}.

Profit maximization. A firm sets a price, p_{j_t}, and a level of quality, q_{j_t}, to maximize its profits taking the aggregate demand for its variety as given. Since the firm's production technology is constant returns to scale, its marginal cost is equal to its average variable cost. A firm producing variety ω in sector j maximizes profits by solving the following problem

$$\pi_{j_t} = \max_{p_{j_t}, q_{j_t}} \left(p_{j_t} - mc_{j_t}\right) y_{j_t} - \kappa q_{j_t}^{\vartheta} - f_{j_t} \tag{3.12}$$

$$\text{s.t.} \quad y_{j_t} = \mu_{H_t} c_{H,j_t} + \mu_{L_t} c_{L,j_t}.$$

Recall that the demand of high and low-skilled households depends on the price and quality chosen by the firm. Also, the demand faced by a firm vanishes if it is above the consumers' choke prices.

Price and markup. The solution to the firm's profit maximization problem yields its variety's price as a markup m_j over the marginal cost according to

$$p_{j_t} = m_{j_t}\, mc_{j_t}. \tag{3.13}$$

In turn, the firm's markup is a function of the (endogenous) *average* price elasticity of demand of all consumers of its variety. Let $\bar{\xi}_{j_t} \equiv -\frac{\partial y_{j_t}/y_{j_t}}{\partial p_{j_t}/p_{j_t}}$ denote the variety's average price elasticity of demand. The markup is then given by the usual expression

$$m_{j_t} = \frac{\bar{\xi}_{j_t}}{\bar{\xi}_{j_t} - 1}. \tag{3.14}$$

However, the average price elasticity of demand is now a weighted average of each consumer's own price elasticity of demand, ξ_{i,j_t}, and her demand share in the firm's customer base, ω_{i,j_t}, so that

$$\overline{\xi}_{j_t} \;=\; \omega_{H,j_t}\,\xi_{H,j_t} + \omega_{L,j_t}\,\xi_{L,j_t}. \tag{3.15}$$

The consumption share of wealthy households is $\omega_{H,j_t} = \frac{\mu_{H_t}\,c_{H,j_t}}{y_{j_t}}$ and that of poor households is $\omega_{L,j_t} = \frac{\mu_{L_t}\,c_{L,j_t}}{y_{j_t}}$. Each type's price elasticity of demand is given by equation (3.6). Note that the price is a fixed point in equation (3.13) as elasticities and consumption shares depend on it.

Introducing the average price elasticity of demand highlights the three channels behind an increase in markups: (i) the price channel, (ii) the income channel, and (iii) the composition channel. First, technological progress that reduces marginal costs allows firms to reduce their prices. As prices decrease, households are more willing to buy them. This translates into a reduction in each consumer's own price elasticity of demand, which in turn allows firms to charge higher markups. The cost pass-through is therefore less than one as firms are able to capture some of the benefits of technological progress in the form of higher markups. Second, a generalized increase in incomes reduces consumers' price elasticities of demand. Firms then respond by increasing their markups. Finally, an increase of the share of wealthier consumers in the economy (even without incomes rising) increases the probability the firm meets a wealthier shopper. As these consumers have a lower price elasticity of demand, the firm will adjust its prices by increasing its markup.

Quality. The firm's optimal choice of quality is tightly linked to its markup. A firm faces a tradeoff when choosing its price: higher markups need to be accompanied by better quality, as it weights its consumers' quality elasticity of demand and its markup. In particular, the firm equates the share of quality-related costs in terms of sales to

$$\frac{\kappa q_{j_t}^{\vartheta}}{p_{j_t} y_{j_t}} \;=\; \frac{\overline{\sigma}_{j_t}}{\vartheta}\frac{(m_{j_t}-1)}{m_{j_t}}, \tag{3.16}$$

where $\overline{\sigma}_{j_t} \equiv \frac{\partial y_{j_t}/y_{j_t}}{\partial q_{j_t}/q_{j_t}}$ denotes the average quality elasticity of demand. Since the quality elasticity of demand does not differ across households nor across sectors, the average quality elasticity of demand is equivalent to each household's own quality elasticity of demand as defined in equation (3.7), i.e.,

$$\overline{\sigma}_{j_t} \;=\; \sigma.$$

3.3 Entrants

Potential entrants consider entering the market for goods or services as long as they can make profits. If a firm chooses to enter and produce a variety ω in sector $j = \{G, S\}$, it receives the profit π_{j_t}. If instead the firm chooses to not enter the market, it gets a payoff of zero. Firms will thus keep entering the market driving down profits to zero. This implies that in equilibrium the markup of a zero-profit firm is greater than one and equivalent to

$$\frac{\overline{\xi}_{j_t}}{\overline{\xi}_{j_t}-1} \;=\; 1 + \frac{\vartheta}{(\vartheta - \overline{\sigma}_{j_t})}\frac{f_{j_t}}{\left[w_{H_t} h_{j_t} + w_{L_t}\ell_{j_t}\right]}. \tag{3.17}$$

The free-entry condition determines the aggregate number of operating firms in each sector, which is denoted by N_{j_t}. The aggregate number of operating firms in the economy is $N_t = N_{G_t} + N_{S_t}$.

3.4 Equilibrium

The equilibrium requires that all markets clear, profits and fixed costs are rebated lump sum to households, and that there is no money left on the table for potential entrants in each sector. In particular, the labor market for high and low-skilled workers must clear. Recall that an individual inelastically supplies one unit of labor and that the fraction of high-skilled workers in the economy is μ_{H_t}. Hence, the labor market clearing conditions are

$$\mu_{H_t} = \int_0^{N_{G_t}} h_{G_t}(\omega)\, d\omega + \int_0^{N_{S_t}} h_{S_t}(\omega)\, d\omega \tag{3.18}$$

$$\mu_{L_t} = \int_0^{N_{G_t}} \ell_{G_t}(\omega)\, d\omega + \int_0^{N_{S_t}} \ell_{S_t}(\omega)\, d\omega. \tag{3.19}$$

Expenses with quality and fixed costs are rebated to households. Aggregate nonlabor earnings is given by $\Lambda_t \equiv \mu_{H_t}\Lambda_{H_t} + \mu_{L_t}\Lambda_{L_t}$. It then transpires that

$$\Lambda_t = \kappa\left[\int_0^{N_{G_t}} q_{G_t}(\omega)^{\vartheta}\, d\omega + \int_0^{N_{S_t}} q_{S_t}(\omega)^{\vartheta}\, d\omega\right] + N_{G_t} f_{G_t} + N_{S_t} f_{S_t}. \tag{3.20}$$

The definition of the symmetric equilibrium follows.

Definition (EQUILIBRIUM)**.** A symmetric equilibrium consists of a solution for: (1) high and low-skilled consumers' demand for goods and services, $c_{H,G}$, $c_{H,S}$, $c_{L,G}$, and $c_{L,S}$; (2) goods and services firms' price, p_G and p_S, quality, q_G and q_S, and high and low-skilled labor demands, h_G, h_S, ℓ_G, and ℓ_S; (3) the number of operating firms in each sector, N_G and N_S; (4) the economy's high and low-skilled wages, w_H and w_L; and (5) nonlabor earnings transfered to consumers, Λ_H and Λ_L. These are determined such that

1. Given prices, p_G and p_S, quality, q_G and q_S, labor and nonlabor earnings, w_H, w_L, Λ_H, and Λ_H, consumers' indirect utility satisfies (3.1), (3.2), and (3.3). The solution yields the allocations $c_{H,G}, c_{H,S}, c_{L,G}$, and $c_{L,S}$.

2. Given consumers' demand, incumbent firms maximize their profits according to (3.12), which determines a solution for prices, p_G and p_S, and quality, q_G and q_S. The labor demanded by firms, h_G, h_S, ℓ_G, and ℓ_S, solve their cost minimization problem.

3. The free-entry condition (3.17) holds in each sector. These pin down the number of operating firms in each sector N_G and N_S.

4. Labor supplied by households must equate the labor demanded by firms for high and low-skilled workers according to (3.18) and (3.19). These conditions determine the high and low-skilled wages, w_H and w_L.

5. Aggregate nonlabor earnings is the sum of operating firms' expenses associated with quality and entry costs given by (3.20).

4 Matching the Model to the U.S.

In this section the model is matched to U.S. data to be consistent with the key macroeconomic trends documented in Section 2. In particular, the estimated model will deliver the increase in aggregate markups and average markups within both sectors together with the rise of the services share and the relative price of services. The model also accounts very well for other key trends observed over the last forty years.

The calibration proceeds in two steps. In the first step, the parameters governing preferences, technology, and costs are estimated to match the main outcomes at two different points in time (namely in 1980 and 2015). In particular, a set of parameter values can be backed out from the theory to match a set of data targets exactly. The remaining parameters are then chosen to minimize the model's prediction error relative to other targets. In the second step, the transition between these points in time is computed by allowing neutral and skill-biased productivities, and entry costs to vary in order to match the trends in the aggregate markup, the relative price of services, the high-skilled income share, the skill premium, and firm entry rates across sectors, given the parameter values estimated in the first step.

4.1 Data targets

The set of targeted moments used in the calibration is described below. All but the services share and the average markups can be matched exactly from the theory using first-order and equilibrium conditions.

Services share. The services share is taken as the value added share of service industries as done in Section 2 using KLEMS data. The targeted services shares for 1980 and 2015 are $\omega_{S_t}^{\text{sales}} = \{0.670, 0.790\}$.

Relative price of services. The relative price of services is computed using KLEMS data. In particular, sectoral prices are chain-weighted Fisher price indices of the value added price indices of individual industries for the years 1980 and 2015.[22] The relative price of services is normalized to one in 1980. The targeted relative prices of services are then $\frac{\overline{p}_{S_t}}{\overline{p}_{G_t}} = \{1.0, 1.437\}$.

Markups. The aggregate markup is measured using the average markups within each sector, as computed in Section 2 with data from listed firms in Compustat, and sectoral cost share with KLEMS data. The sectoral cost shares exclude intermediate inputs, but the aggregate markup stays almost unchanged relative to the figure presented in Section 2. The targeted aggregate markups are then $M_t = \{1.136, 1.263\}$. The average (cost-weighted) markup within each sector follows the procedure presented in Section 2 and values for 2015 are used as targets, with $\overline{m}_{G_{2015}} = 1.214$ and $\overline{m}_{S_{2015}} = 1.273$.

Income share. The income share of high-skilled households corresponds to the income share of individuals with a college degree (or some college) as reported by Kuhn and Ríos-Rull (2013) and their 2019 update using data from the Survey of Consumer Finances. The earliest period available is 1989. To compute a value for $\epsilon_{H_{1980}} = \frac{\mu_{H_{1980}} e_{H_{1980}} / \mu_{L_{1980}} e_{L_{1980}}}{1+\mu_{H_{1980}} e_{H_{1980}} / \mu_{L_{1980}} e_{L_{1980}}}$, the average growth rate of $\frac{e_{H_t}}{e_{L_t}}$

[22]Following Section 2, the goods sector corresponds to industries A to F and the services sector to industries G to S in KLEMS's classification.

(2.4% per year) and the evolution of the share of high and low-skilled households in the economy (μ_{H_t} and μ_{L_t}) are used. The value of $\frac{e_{H_{2015}}}{e_{L_{2015}}}$ is obtained by linearly interpolating the available data points (2013 and 2016). The resulting targets for the income share of high-skilled households are $\epsilon_{H_t} = \{0.365, 0.603\}$.

Skill premium. The skill premium corresponds to the ratio of the median income of males with four-year college degrees vs. high school graduates from the Census' Current Population Survey. The resulting skill premiums are $\frac{w_{H_t}}{w_{L_t}} = \{1.347, 1.928\}$. These targets are similar to Buera, Kaboski, Rogerson, and Vizcaino (2021), who adjust the skill premium for differences in hourly wage rates among skill groups.

Skilled households. The shares of high-skilled households at the initial and terminal dates ($\mu_{H_{1980}}$ and $\mu_{H_{2015}}$) are measured directly from the data as the fraction of people employed in skilled jobs. The Census' 1980 Current Population Survey and the 2015 American Community Survey are used to pin down those values. Following ILO's ISCO categories, high-skilled labor corresponds to workers in the following occupations: legislators, senior officials, and managers; professionals; and technicians and associate professionals (ILO's categories 1 to 3). Low-skilled labor comprises the following occupations: clerks; service workers and shop and market sales; skilled agricultural and fishery workers; crafts and related trades workers; plant and machine operators; and elementary occupations (ILO's categories 4 to 9). The resulting shares of high-skilled households in 1980 and 2015 are $\mu_{H_t} = \{0.3253, 0.4237\}$.

Quality costs. The costs associated with quality as a share of sales in the services are targeted in the calibration. These correspond to the ratio of selling, general, and administrative expenses to sales in each sector as reported in Compustat in line with the data presented in Section 2. The targeted quality cost share in 2015 is $\frac{\kappa Q_{S_{2015}}}{PY_{S_{2015}}} = 0.144$.

Entry. The number of active firms operating in the non-services sector went from 751,565 to 844,487 firms between 1980 and 2015, while the number of active firms in the services sector grew from 3,050,428 to 4,374,412 firms over the same period. The data is taken from the Census' Business Dynamism Statistics. To compute the entry rates in each sector, the number of firms are rescaled by total population. This implies entry rates of -20.4% in the non-services sector and 1.6% in the services sector between 1980 and 2015. The number of firms operating in the non-services sector is normalized to one in 1980 and to 0.796 in 2015. The number of firms in services is then 4.059 in 1980 and 4.123 in 2015. The targeted relative numbers of firms operating in the services sector are $\frac{N_{S_t}}{N_{G_t}} = \{4.059, 5.180\}$, which matches the entry rates in each sector between 1980 and 2015.

4.2 Estimated parameters

There are five preference parameters to be calibrated, $\{\gamma, \lambda, \phi_G, \phi_S, \delta\}$, eight technology parameters, $\{\alpha, \iota, x_t, z_{G_t}, z_{S_t}\}$, six cost parameters, $\{f_{G_t}, f_{S_t}, \kappa, \vartheta\}$, as well as the fraction of high-skilled households in the economy, μ_{H_t}, for t in 1980 and 2015. Some of these parameters are exogenously imposed, while others are matched to the targets discussed above. In particular, technology and cost parameters are identified from the theory, i.e., backed out from the first-order and equilibrium conditions. Preference parameters are then recovered from a minimization routine. The low-skilled wage is the numeraire and is thus normalized to one in both periods, i.e. $w_{L_t} = \{1.0, 1.0\}$.

Externally chosen. Five parameters, $\{\iota, x_{1980}, \vartheta, \mu_{H_{1980}}, \mu_{H_{2015}}\}$, are exogenously imposed. The substitution parameter in the firms' technology, ι, guides the elasticity of substitution between high and low-skilled labor. Acemoglu and Autor (2011), Buera, Kaboski, Rogerson, and Vizcaino (2021), Katz and Murphy (1992) estimate the elasticity of substitution for different periods and find values ranging from -2.9 to -1.4, which corresponds to a value of $\iota \in [0.291, 0.661]$. This range is consistent with skill-biased technological change decreasing marginal costs as high and low-skilled labor are substitutes. A value of $\iota = 0.4$ is chosen. Since α and x_t cannot be separately identified, skill-biased productivity is normalized to 1 in 1980, i.e., $x_{1980} = 1$. I assume that fixed costs are convex in quality and hence set $\vartheta = 2$. The shares of high-skilled households at the initial and terminal dates ($\mu_{H_{1980}}$ and $\mu_{H_{2015}}$) have a direct counterpart in the data and can therefore be set exogenously.

Technology and cost parameters. The vector of parameters $\Theta^\star \equiv \{\alpha, x_{2015}, z_{G_{1980}}, z_{G_{2015}}, z_{S_{1980}}, z_{S_{2015}}, f_{G_{1980}}, f_{G_{2015}}, f_{S_{1980}}, f_{S_{2015}}, \kappa\}$ is calibrated to match exactly the eleven data targets discussed above. This procedure uses the model's first-order and equilibrium conditions evaluated at the data targets to back out these parameter values. The solution to this system of nonlinear equations takes as given the values for the preference parameters, $\widetilde{\Theta}$. The vector $\Theta^\star$ satisfies the following condition

$$\mathfrak{M}(\widetilde{\Theta}, \Theta^\star) - \mathfrak{D} = 0,$$

where $\mathfrak{D}$ is a vector of data targets, and $\mathfrak{M}$ is the model's solution for these data targets.

In a nutshell, the share of high-skilled labor used in production, α, and skill-biased productivity, x_{2015}, help match the skill premium $\left(\frac{w_{H_t}}{w_{L_t}}\right)$ in 1980 and 2015. The sectoral neutral productivity parameters, z_{G_t} and z_{S_t}, help discipline the relative price of services $\left(\frac{\overline{p}_{S_t}}{\overline{p}_{G_t}}\right)$ as well as the aggregate markup (M_t) in 1980 and 2015. The entry costs, f_{G_t} and f_{S_t}, help match the relative number of firms operating in the services sector $\left(\frac{N_{S_t}}{N_{G_t}}\right)$ and the income share of high-skilled households (ϵ_{H_t}) in 1980 and 2015. The linear quality cost parameter, κ, is used to match the fixed costs as a share of sales in the services sector in 2015 $\left(\frac{\kappa Q_{S_{2015}}}{PY_{S_{2015}}}\right)$.

Preference parameters. The vector of preference parameters $\widetilde{\Theta} \equiv \{\gamma, \lambda, \phi_G, \phi_S\}$ minimizes the model's prediction error with respect to the services share of output in 1980 and 2015 as well as the average markup of goods and services in 2015. Specifically, denote the i'th data target by $\mathfrak{d}_i$ and the model's solution for this target by $\mathfrak{m}_i(\widetilde{\Theta}, \Theta^\star)$. Weighting each observation uniformly, the preference parameters solve the following minimization problem

$$\min_{\widetilde{\Theta}} \quad \sum_i \left[\frac{\mathfrak{d}_i - \mathfrak{m}_i(\widetilde{\Theta}, \Theta^\star)}{\mathfrak{d}_i}\right]^2.$$

This procedure internalizes how the choice of preference parameters in the outer loop affects the solution for the technology and cost parameters in the inner loop. In summary, the indirect utility weight on goods, λ, and the exponent on the subutility, γ, help discipline the services share in 1980 and 2015 (ω_t^{sales}). The choke price parameters, ϕ_G and ϕ_S, are estimated to match the average markup within each sector in 2015 ($\overline{m}_{G_{2015}}$ and $\overline{m}_{S_{2015}}$). Finally, δ is used to normalize the average quality of goods in 1980 ($\overline{q}_{G_{1980}}$) to 1.

Table 4.1: Parameter values

Parameter	Description	Value	Identification
Preferences			
λ	Indirect utility's weight on goods	0.181	Services share
γ	Exponent in indirect subutility	17.359	Services share
ϕ_G	Choke price of goods	7.725	Average goods markups
ϕ_S	Choke price of services	12.780	Average services markups
δ	Exponent related with quality	0.072	Normalization
Technology			
α	High-skilled weight	0.465	Skill premium
ι	Elasticity of substitution between high and low-skilled	0.400	Exogenous
x_t	Skill-biased prod. in 1980, 2015	1.000, 1.844	Normalization, Skill premium
z_{G_t}	TFP in goods sector in 1980, 2015	0.530, 0.485	Aggregate markup
z_{S_t}	TFP in services sector in 1980, 2015	0.580, 0.355	Relative price of services
Costs			
f_{G_t}	Entry costs in goods sector in 1980, 2015	0.009, 0.027	High-skilled income share
f_{S_t}	Entry costs in services sector in 1980, 2015	0.010, 0.024	Rel. number of service firms
κ	Linear term related with quality	0.018	Quality costs/sales in services
ϑ	Exponent related with quality	2.000	Exogenous
Measure			
μ_{H_t}	Share of high-skilled households in 1980, 2015	0.325, 0.424	Empl. in high-skilled occupations

Results. Table 4.1 presents the parameter values used in the baseline exercise. Of note, the services choke price parameter is almost twice as large as the one for goods, which reflects the consumers' higher willingness to pay for services. The indirect utility has also a larger weight on services than on goods. Technological progress is entirely driven by skill-biased productivity. This helps sustain the rise in inequality observed in the data, despite the increase in the fraction of high-skilled workers in the economy. The differential growth rate of neutral productivities across sectors helps explain the strong decline of the price of goods over time. Entry costs grew threefold in the goods sector and twofold in the services sector. The larger costs in the goods sector justify the strong decline in the number of goods-producing firms observed in the data.

Table 4.2: Targeted moments: Data vs. model

Moment	Description	Model 1980, 2015	Data 1980, 2015	Source
Markups				
M_t	Aggregate markups	1.136, 1.263	1.136, 1.263	Compustat, KLEMS
$\overline{m}_{G_t}$	Average goods markups	1.215	1.214	Compustat
$\overline{m}_{S_t}$	Average services markups	1.276	1.273	Compustat
Relative prices and income				
$\overline{p}_{S_t}/\overline{p}_{G_t}$	Relative price of services	1.000, 1.437	1.000, 1.437	KLEMS
w_{H_t}/w_{L_t}	Skill premium	1.347, 1.928	1.347, 1.928	CPS
ϵ_{H_t}	High-skilled income share	0.365, 0.603	0.365, 0.603	Kuhn-Ríos-Rull (2013), CPS
$\omega_{S_t}^{\text{sales}}$	Services share	0.670, 0.790	0.670, 0.790	KLEMS
Entry				
$\kappa Q_{S_t}/PY_{S_t}$	Sales share of costs with quality in services	0.144	0.144	Compustat
N_{S_t}/N_{G_t}	Relative number of services firms	4.059, 5.180	4.059, 5.180	BDS, CPS

Table 4.2 displays the results of the calibration exercise. The resulting fit is very good. The model matches perfectly by construction the aggregate markup, the relative price of services, the income share of high-skilled households, the skill premium as well as net entry rates across sectors. The model does well in replicating the rise of the services share. The fit of the average markup of goods and services is also good.

4.3 Matching trends

Once all the parameters are estimated, the model is solved yearly from 1980 to 2015 by finding the set of productivity terms, z_{G_t}, z_{S_t}, and x_t, and entry costs, f_{G_t} and f_{S_t}, that match the time series of the aggregate markup (M_t), the relative price of services $\left(\frac{\overline{p}_{S_t}}{\overline{p}_{G_t}}\right)$, the relative income share of high-skilled households (ϵ_{H_t}), the skill premium $\left(\frac{\overline{w}_{H_t}}{\overline{w}_{L_t}}\right)$, and the relative number of firms in the services sector $\left(\frac{N_{S_t}}{N_{G_t}}\right)$. The fraction of high-skilled households in the economy μ_{H_t} is taken directly from the data. All other parameters are constant over time. The model matches these six aggregate trends perfectly. The underlying productivity and entry costs that match them are presented in Figure D.4.1 in Appendix D.4.

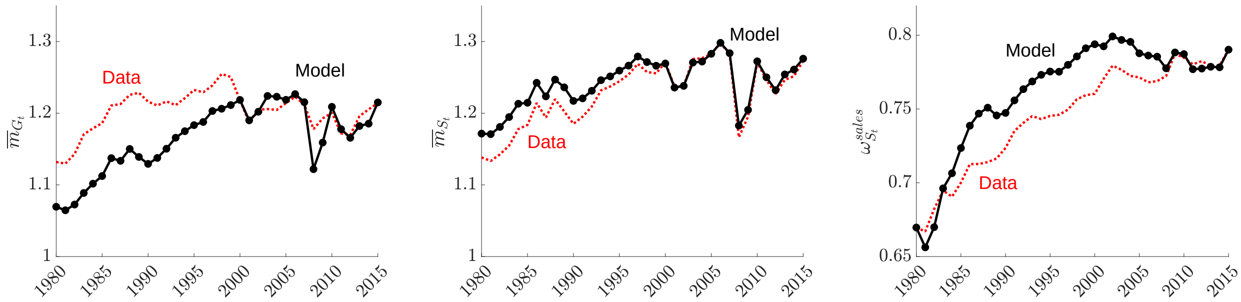
4.4 Model validation

To validate the model, time series statistics not directly targeted in the calibration procedure are compared with their data counterparts.

Average markups and demand elasticities. Although the evolution of the aggregate markup is by construction matched perfectly, the average markup within each sector as well as the associated sectoral cost shares are not. Figure 4.1 (panels (a) and (b)) displays the average markups in the services and non-services sector in the model with their data counterparts. Although only 2015 is targeted in the calilbration, the model tracks the evolution of the average markup within the services sector particularly well. What is remarkable is that the rise in average markups is achieved together with an increase in the number of firms in services and a decline in the price of goods (see panel (a) of Figure 4.2).

The increase in average markups is directly linked to the decline in the average price elasticity of demand as equation 3.14 shows. In turn, the average price elasticity responds to changes in each consumer's price elasticity of demand and their demand share (equation 3.15). The firms' sales share from high-skilled consumers increased by 69% for goods and 63% for services between 1980 and 2015 (from 35% to 60% of sales in the goods sector and from 37% to 61% of sales in the services sector) as depicted in Figure D.4.2 in Appendix D.4. Hence, firms were more likely to sell their goods and services to wealthier consumers in 2015 than they were in 1980. This increase is due both to the rise of the fraction of wealthier consumers in the economy and to the rise of inequality as high-skilled consumers experienced stronger income growth relative to poorer households. Panel (b) of Figure 4.2 shows that the average markup of services would have grown by half if the demand shares were held fixed at their 1980 values.

Figure 4.1: Average markups and services sales share



(a) Average markup of goods (b) Average markup of services (c) Services sales share

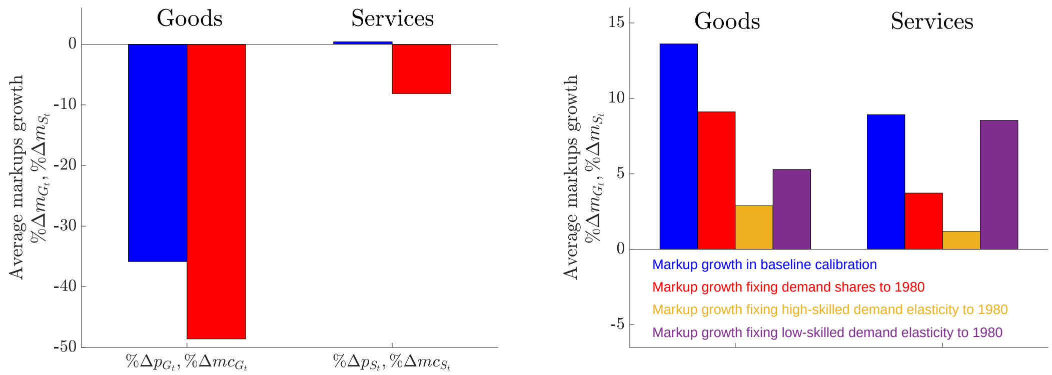
Note: Panel (a) shows the average markup within the goods sector ($\overline{m}_{G_t}$) in the model (black) and data (red). Panel (b) shows the average markup within the services sector ($\overline{m}_{S_t}$) in the model (black) and data (red). Panel (c) shows the services sales share ($\omega^{sales}_{S_t}$) in the model (black) and data (red).

In addition, both high and low-skilled consumers became less price sensitive over time. In particular, the price elasticity of demand for services declined 50% for higher-income consumers and 2% for poorer households between 1980 and 2015 as shows Figure D.4.3 (panels (a) and (b)) in Appendix D.4. The price elasticity of demand for goods declined even further (72% for high-skilled households and 48% for low-skilled households), but as households shifted their consumption toward services the impact on the average markup in the goods sector was more muted. The model also predicts that as the income of poorer consumers dropped significantly during the Great Financial Crisis, they became a lot more price sensitive during that time. This was not the case for the high-skilled consumers. Panel (b) of Figure 4.2 shows that holding fix the price elasticity of demand of wealthier consumers to its level in 1980 would have led to a much lower increase in markups than the one estimated in the baseline.

As Proposition A.2 showed, the price elasticity of demand is intimately related with the income elasticity of demand when preferences are non-homothetic. As is usually the case in models of structural change, services are luxuries and hence feature an income elasticity of demand greater than one. Aguiar and Bills (2015) estimate income elasticities for different categories of goods and services. Several services (such as food away from home, entertainment, education, childcare) have income elasticities well above one. In the model, services are luxuries for both rich and poor households (their income elasticities of demand were 1.12 and 1.19 in 1980 and declined to 1.01 and 1.02 in 2015, respectively) as shows Figure D.4.3 (panels (c) and (d)) in Appendix D.4. In contrast, goods are necessities and thus their income elasticities of demand are below one (they went from 0.74 and 0.63 in 1980 to 0.98 and 0.92 in 2015 for high and low-skilled consumers, respectively). Note that during the Great Financial Crisis, services became even more luxuries for poorer households.

Services share. The model matches well the value added share of the services sector between 1980 and 2015; see Figure 4.1 (panel (c)). The services share increases steadily from 67% of aggregate output in 1980 to 79% in 2015. Similarly, the services cost share, used in the calculation of the aggregate markup, also increased steadily over the period. The model also predicts well the shift of consump-

Figure 4.2: Decomposition of the average markups



(a) Supply-side decomposition

(b) Demand-side decomposition

Note: Panel (a) shows the decomposition of the average markup growth (in logs) into prices (in blue) and marginal costs (in red) according to equation (3.13). Panel (b) shows the decomposition of the average markup growth by fixing demand shares to their levels in 1980 (in red), by fixing the high-skilled consumers' price elasticity of demand to its 1980 value (in yellow), and by fixing the low-skilled consumers' price elasticity of demand to its 1980 value (in purple) according to equations (3.14) and (3.15).

tion spending from goods toward services for both types of households. Figure Figure D.4.4 in Appendix D.4 presents the evolution of the goods and services consumption spending shares for high and low-skilled households. The magnitude of the increase in the model between 1980 and 2015 was 11 and 12 percentage points for high and low-skilled consumers, which is close to the increase observed in the CEX data (10 and 11 percentage points, respectively).[23]

Labor shares and employment. In this framework, the aggregate labor share is the inverse of the aggregate markup. Its level is higher than in the data (88% 1980 and 79% in 2015). However, its decline of 9 percentage points between 1980 and 2015 is line with the evidence provided by Karabarbounis and Neiman (2014). This decline is entirely driven by the fall of the low-skilled labor share, which went down by 20 percentage points between 1980 and 2015. As Figure D.4.5 (panel (b)) in Appendix D.4 displays, the low-skilled labor share for workers in the non-services sector declined from 19% of aggregate output in 1980 to 7% in 2015, while the low-skilled labor share for workers in the services sector dropped from 35% of aggregate output to 26% over the same period.

In contrast, the high-skilled labor share in aggregate output increased 21 percentage points between 1980 and 2015. The high-skilled labor share for workers in the non-services sector declined slightly over the period from 12% to 10% of aggregate output. The high-skilled labor share for workers in the services sector increased markedly from 23% to 36% of aggregate output between 1980 and 2015 (see panel (a) of Figure D.4.5 in Appendix D.4). Although not targeted, the model also captures well the reallocation of workers toward the services sector in line with ILO data. The share of both

[23]High-skilled households in CEX corresponds to Managers and professionals, and Technical, sales and clerical workers. Low-skilled households in CEX corresponds to Service workers, Construction workers and mechanics, Operators, fabricators and laborers.

high and low-skilled workers employed in services increased from 65% to 78% between 1980 and 2015 (see panel (c) of Figure D.4.5 in Appendix D.4).

Fixed and entry costs. Fixed costs have risen in response to greater spending on quality. They went from 4% and 10% of sales in the non-services and services sectors in 1980 to 12% and 14% in 2015, respectively (see panel (a) of Figure D.4.6 in Appendix D.4). Although only the value of the services' fixed costs share in 2015 is targeted in the calibration, the model reproduces very well the rise of fixed costs observed in the data. Entry costs have risen to accommodate the modest changes in the number of firms observed in the data. They went from 2% and 5% of sales in the non-services and services sectors in 1980 to 6% and 7% in 2015, respectively (see panel (b) of Figure D.4.6 in Appendix D.4). The entry costs are higher in services than in the goods sector despite the higher number of operating services firms. This is due to the higher markups in the services sector.

5 Drivers of the Rise in Markups

The model is now used to decompose the forces driving the rise in markups over time. In particular, I proceed by shutting down the exogenous forces in the model, namely neutral and skill-biased technological change, the rise of entry costs, and the increase of the share of high-skilled workers in the economy, one at a time. Additionally, the model is used to study the impact of the rise of incomes, income inequality, and of the relative price of services by tracing their effect on markups, the number of operating firms in each sector, and welfare.

5.1 Technological progress

Neutral productivity. The first experiment is to set neutral productivity in each sector to its 1980 value. First, the change in neutral productivity of goods-producing firms is shut down, i.e. $z_{G_t} = z_{G_{1980}}$, while keeping the other exogenous forces at their baseline values. The results of this experiment are presented in Table 5.1 (column 2). As neutral productivity is slightly higher in 1980 relative to 2015, firms in the non-services sector are more productive in 2015 than in the baseline economy. As the productivity gains are small, there are few noticeable changes relative to the baseline economy.

More striking patterns are visible when only the change in the neutral productivity of services-producing firms is shut down, i.e. $z_{S_t} = z_{S_{1980}}$. Column 3 of Table 5.1 presents the results of this experiment. Since neutral productivity declined, services firms are more productive in 2015 than in the baseline economy. As firms are now more productive, they are able to pass on part of the decline in marginal costs to consumers by reducing their prices. Yet, their cost pass-through is smaller than one—this is the result of having consumers' price elasticities of demand increase with prices. In turn, this allows services firms to increase their markups by 25% relative to the baseline, which brings the aggregate markup up from 1.263 to 1.494. Note that the increase in markups is achieved with a notable decline in the relative price of services (down by 27% relative to the 2015 baseline).

Stronger productivity in the services sector leads to a slower reallocation of labor toward this sector, bringing down the labor share of both high and low-skilled households working in services.

This is accompanied by an income effect resulting from higher aggregate output in the counterfactual economy (18% larger). Stronger productivity encourages more firms to enter not only in the services sector, but also in the goods sector. The total number of operating firms increases by 88% relative to the 2015 baseline.

Skill-biased productivity. In the second experiment, skill-biased technological progress is shut down so that x_t is set at its 1980 value and all other exogenous forces are kept at their baseline values. Column 4 in Table 5.1 displays the results of this experiment. The decline in skill-biased productivity makes all firms about half as productive in 2015 as they are in the baseline economy.

With lower productivity, the economy shrinks and the number of firms vanishes (by 38% and 80% relative to the 2015 baseline economy, respectively). This reduces the amount of varieties available for consumption. The aggregate markup is now much lower than in 1980—close to the perfect competition benchmark—with the decline stemming from both a reduction of the average markup of goods and services. As households are now poorer, firms have to reduce their markups to attract customers. This is because consumers' price elasticities of demand are now higher as they become more price sensitive when poorer. As firms' marginal costs also increased, the prices of goods and services are now higher (35% and 39% higher than in the 2015 baseline). This further depresses consumers as they also become more price sensitive when prices are higher.

Without skill-biased technological change, firms now benefit less from employing high-skilled workers. As a result, the skill premium almost disappears and income inequality is significantly reduced. The decline in the productivity differential between high and low-skilled workers also leads to a significantly higher low-skilled labor share in the goods sector, which doubled relative to the 2015 baseline (from 20% to 40% of total output).

5.2 Rising entry costs

The third experiment shuts down the exogenous change in each sector's entry cost parameter f_{j_t}. First, only the entry cost in the goods sector is reduced to its 1980 value, i.e. $f_{G_t} = f_{G_{1980}}$. Column 5 of Table 5.1 summarizes the results. The lower entry costs encourage more firms to produce more goods. The number of active firms in the sector increases significantly in response to lower barriers to entry, with less than two services-producing firm per goods-producing firm (as opposed to five in the 2015 baseline economy). The lower entry costs have little impact on markups. The aggregate markup increases by 10% relative to 1980 (about the same as in the data).

Shutting down the increase in entry costs in the services sector has similar effects. Column 6 of Table 5.1 presents the outcome of setting $f_{S_t} = f_{S_{1980}}$. Now reducing entry barriers significantly increases the number of active firms in the services sector. There are now almost 13 services-producing firm per goods-producing firm. The aggregate markup declines slightly relative to the 2015 baseline economy.

Table 5.1: Experiments: Technological progress, entry costs, and high-skilled share

Variable	Description	(1) Baseline 1980, 2015	(2) z_{G_t} 2015	(3) z_{S_t} 2015	(4) x_t 2015	(5) f_{G_t} 2015	(6) f_{S_t} 2015	(7) μ_{H_t} 2015
Markups								
M_t	Aggregate markups	1.136, 1.263	1.278	1.494	1.073	1.253	1.247	1.215
$\overline{m}_{G_t}$	Average goods markups	1.070, 1.215	1.251	1.271	1.055	1.204	1.198	1.171
$\overline{m}_{S_t}$	Average services markups	1.172, 1.276	1.285	1.601	1.078	1.266	1.260	1.227
Relative prices and income								
$\overline{p}_{S_t}/\overline{p}_{G_t}$	Relative price of services	1.000, 1.437	1.534	1.053	1.398	1.438	1.439	1.434
w_{H_t}/w_{L_t}	Skill premium	1.347, 1.928	1.928	1.928	1.046	1.928	1.928	2.484
ϵ_{H_t}	High-skilled income share	0.365, 0.603	0.561	0.564	0.448	0.647	0.668	0.583
$\omega^{\text{sales}}_{S_t}$	Services share	0.670, 0.790	0.804	0.724	0.766	0.790	0.790	0.789
Employment and labor shares								
$\theta_{H/L,S_t}$	High/Low-skilled empl. share in services	0.649, 0.782	0.800	0.676	0.762	0.782	0.782	0.781
$VC_{H,G_t}/PY_t$	High-skilled labor share in goods	0.122, 0.118	0.092	0.127	0.097	0.102	0.103	0.098
$VC_{H,S_t}/PY_t$	High-skilled labor share in services	0.225, 0.423	0.367	0.265	0.309	0.366	0.368	0.350
$VC_{L,G_t}/PY_t$	Low-skilled labor share in goods	0.187, 0.055	0.065	0.090	0.126	0.072	0.072	0.082
$VC_{L,S_t}/PY_t$	Low-skilled labor share in services	0.347, 0.196	0.259	0.187	0.401	0.258	0.259	0.292
Entry								
$\kappa Q_{G_t}/PY_{G_t}$	Sales share of quality costs in goods	0.043, 0.118	0.133	0.142	0.035	0.113	0.110	0.097
$\kappa Q_{S_t}/PY_{S_t}$	Sales share of quality costs in services	0.097, 0.144	0.148	0.250	0.048	0.140	0.137	0.123
N_{S_t}/N_{G_t}	Relative number of services firms	4.059, 5.180	5.111	5.196	5.099	1.736	12.651	5.315

5.3 Increasing share of high-income households

In the fourth experiment, the share of high-skilled households, μ_{H_t}, is kept fixed at its 1980 value. The economy thus features more poorer households than in the baseline. Column 7 of Table 5.1 shows the results. As there are fewer wealthier households, firms are more likely to sell their goods and services to poorer consumers. This puts more weight on the price elasticity of demand of low-skilled households, which decreases the markup of firms in both sectors. If preferences were homothetic, then this effect would simply disappear and changes in the composition of firms' customer base would therefore have no bearing on the aggregate markup. Instead, the aggregate markup is now 4% lower in 2015 in this economy relative to the baseline. As there are fewer high-skilled households in the economy and firms' skilled-biased productivity are at their baseline values, the skill premium increases sharply to respond to the reduced supply of high-skilled labor.

5.4 Taming the rise of income (inequality)

Keeping incomes constant. Circling back to the discussion of Section A, this experiment studies the effect of keeping households' incomes as in 1980. To achieve this, I change the nature of the experiments by finding the values of skill-biased productivity in both sectors, x_{G_t} and x_{S_t}, that minimize the distance between the model-implied total household income of both types, e_{H_t} and e_{L_t}, and their 1980 values, $e_{H_{1980}}$ and $e_{L_{1980}}$. All the other exogenous forces are kept at their baseline values. Column 1 of Table 5.2 shows the results of this experiment and Figure D.4.7 in Appendix D.4 shows the path of skill-biased productivity and household income.

Figure 5.1 (panel (a)) contrasts the path of aggregate markups in the economy with no changes in incomes with the baseline calibration. Without the increase in incomes, the aggregate markup would have decreased between 1980 and 2015. Three forces are pushing the aggregate markup down. The services share declines abruptly. As consumers are now poorer, they spend much less on services (57% and 51% of their income for high and low-skilled households in 2015). They are also less willing to buy goods and services in general, increasing their price elasticity of demand. The significantly lower productivity in the goods sector pushes these firms' marginal costs up, which in turn leads to an increase in the price of goods. Note that the relative price of services goes down. As goods are more expensive, consumers are less willing to buy them putting additional pressures on goods-producing firms' markups.

Panel (b) of Figure 5.1 also shows what would have happened to the aggregate number of firms in this economy. The decline in business dynamism is remarkable, with the number of operating firms declining significantly over time in both sectors. For instance, entry rates between 1980 and 2015 would be -73% in the non-services sector and -64% in the services sector (as opposed to -20% and 2% in the baseline economy in the non-services and services sector, respectively).

In this experiment, income inequality is also much lower as incomes are held constant over time. The share of total income held by high-skilled households amounts to 48% of total output and the skill premium is now lower than in 1980 at 21%. To disentangle the effect of rising incomes from inequality, the next experiment lets incomes grow while keeping inequality constant. The patterns are strikingly different.

Keeping inequality constant, but economy grows. In this experiment, inequality is kept at its 1980 level but the economy grows at a similar rate as in the baseline economy. In particular, skill-biased productivity is chosen to minimize the distance between these two variables in the counterfactual economy and the targets in the baseline economy. Column 2 of Table 5.2 shows the results of this experiment and Figure D.4.8 in Appendix D.4 shows the path of skill-biased productivity, relative incomes and aggregate output.

The aggregate markup increases even if inequality is lower. This increase is driven by the services sector. Although inequality is reduced, both high and low-skilled consumers become wealthier over time. As consumers become richer, they spend more on services—the services share increases a lot more than when incomes are constant (54% vs. 70%). As consumers become richer, they become less price sensitive, allowing firms to charge higher markups. However, skill-biased productivity is much weaker in the goods sector, leading to an increase in the marginal costs of the goods-producing firms. As a result, the price of goods goes up, further reducing their attractiveness to consumers. In contrast, the stronger productivity in the services sector leads to a fall in the price of services—note that the relative price of services remains constant. This allows services firms to keep some of the benefits of productivity growth to themselves by raising their markups.

5.5 Keeping (relative) prices constant

Keeping prices constant. In this experiment, the values of the neutral productivity terms in both sectors, z_{G_t} and z_{S_t}, are chosen to minimize the distance between the model-implied price of goods

Table 5.2: Experiments: Income and prices

Variable	Description	(1) Baseline 1980, 2015	(2) e_{H_t}, e_{L_t} 2015	(3) $\Delta PY_t, \frac{e_{H_t}}{e_{L_t}}$ 2015	(4) p_{G_t}, p_{S_t} 2015	(5) $\Delta PY_t, \frac{p_{S_t}}{p_{G_t}}$ 2015
Markups						
M_t	Aggregate markups	1.136, 1.263	1.117	1.279	1.209	1.263
$\overline{m}_{G_t}$	Average goods markups	1.069, 1.215	1.041	1.150	1.102	1.133
$\overline{m}_{S_t}$	Average services markups	1.172, 1.276	1.192	1.344	1.264	1.329
Relative prices and income						
$\overline{p}_{S_t}/\overline{p}_{G_t}$	Relative price of services	1.000, 1.437	0.753	0.996	1.000	1.000
w_{H_t}/w_{L_t}	Skill premium	1.347, 1.928	1.213	1.883	1.928	1.928
ϵ_{H_t}	High-skilled income share	0.365, 0.603	0.480	0.455	0.595	0.588
$\omega^{\text{sales}}_{S_t}$	Services share	0.670, 0.790	0.537	0.699	0.690	0.697
Employment and labor shares						
θ_{H,S_t}	High-skilled empl. share in services	0.649, 0.782	0.638	0.718	0.659	0.663
θ_{L,S_t}	Low-skilled empl. share in services	0.649, 0.782	0.384	0.591	0.659	0.663
$VC_{H,G_t}/PY_t$	High-skilled labor share in goods	0.122, 0.101	0.153	0.128	0.165	0.157
$VC_{H,S_t}/PY_t$	High-skilled labor share in services	0.225, 0.363	0.269	0.326	0.320	0.308
$VC_{L,G_t}/PY_t$	Low-skilled labor share in goods	0.187, 0.071	0.292	0.134	0.116	0.110
$VC_{L,S_t}/PY_t$	Low-skilled labor share in services	0.347, 0.256	0.181	0.194	0.226	0.217
Entry						
$\kappa Q_{G_t}/PY_{G_t}$	Sales share of quality costs in goods	0.043, 0.118	0.026	0.087	0.062	0.078
$\kappa Q_{S_t}/PY_{S_t}$	Sales share of quality costs in services	0.097, 0.144	0.107	0.170	0.139	0.165
N_{S_t}/N_{G_t}	Relative number of services firms	4.059, 5.180	5.329	5.096	5.616	5.441

and services, p_{G_t} and p_{S_t}, and their 1980 values. This makes the relative price of services equal to one over the 1980-2015 period. Figures D.4.9 in Appendix D.4 show the path of these variables and column 4 of Table 5.2 shows the results of this experiment. All the other exogenous forces are kept at their baseline values.

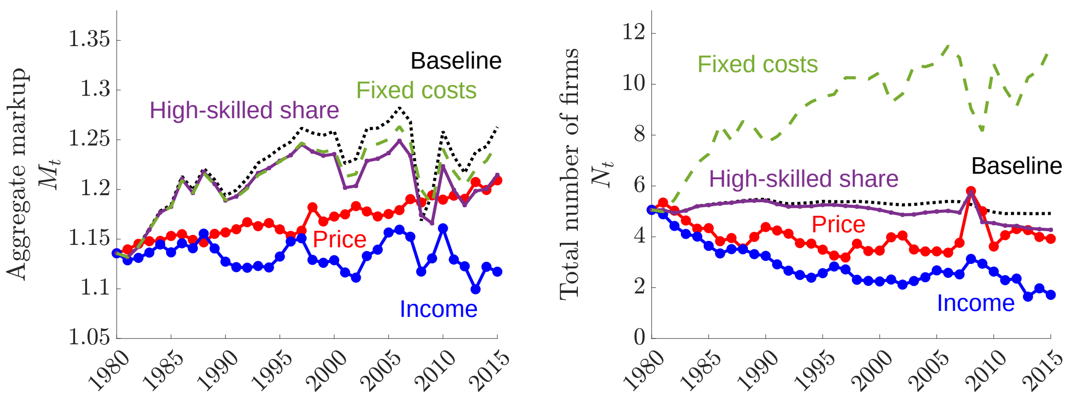
Keeping prices constant makes goods a lot more expensive in the counterfactual economy. As consumers' price elasticity of demand is increasing in prices, households are now less willing to purchase goods. This translates into much lower markups of goods-producing firms. In contrast, the price of services is fairly stable over time leading to an increase in the average markup of services relative to 1980. With a reduced weight on the services sector, the aggregate markup falls over time—but not as much as in the experiment with incomes. Panel (a) of Figure 5.1 displays the path of aggregate markups in this experiment.

Panel (b) of Figure 5.1 shows the total number of firms in the economy when prices are constant. As before, with lower aggregate markups, firms have less incentives to enter the market. However, the decline in business dynamism is smaller than under the experiment with constant incomes.

Keeping relative price of services constant, but economy grows. In the previous experiment, aggregate output fell by 4% in 2015 relative to the baseline economy. In this experiment, the relative price of services is kept constant and the path of aggregate output follows the one from the baseline economy. This is achieved by adjusting total factor productivity, z_{G_t} and z_{S_t}. Figures D.4.10 in Appendix D.4 show the path of these variables and column 5 of Table 5.2 shows the results of this experiment. All the other exogenous forces are kept at their baseline values.

Although the relative price of services is constant over time, services are now cheaper and goods more expensive relative to the baseline economy. As the economy grows at the same rate as in the baseline, consumers are as rich as before. However, as the price of goods and services followed opposite directions, the average markup of goods declined, while that of services increased.

Figure 5.1: Markups and number of firms across experiments



(a) Aggregate markup

(b) Total number of firms

Note: Panel (a) shows the evolution of the aggregate markup in the baseline economy (black) and in the experiments with constant incomes (blue), constant prices (red), constant fixed costs (green), and constant fraction of high-skilled households (pink). Panel (b) shows the evolution of the total number of active firms in the baseline economy (black) and in the experiments with constant incomes (blue), constant prices (red), constant fixed costs (green), and constant fraction of high-skilled households (pink).

5.6 Synopsis

To sum up the relative contribution of the different forces discussed above, Figure 5.2 shows the marginal effect of shutting down each exogenous term at a time on markups. In particular, each bar measures the difference between the growth rate of markups in the baseline economy and the growth rate of markups in the counterfactual economy over the 1980 and 2015.[24]

Skill-biased technological progress in the goods and services sector contributes the most to the increase in markups. This happens for two reasons. First, faster skill-biased productivity growth decreases marginal costs, which helps firms pass these gains on to consumers through reduced prices. Given the assumed form of preferences, which imply the price elasticity of demand decreases as prices fall, the cost pass-through is less than one and so firms are able to keep some of the productivity gains in the form of higher markups. Second, faster skill-biased productivity growth increases

[24]Specifically, the contribution of each experiment is measured as

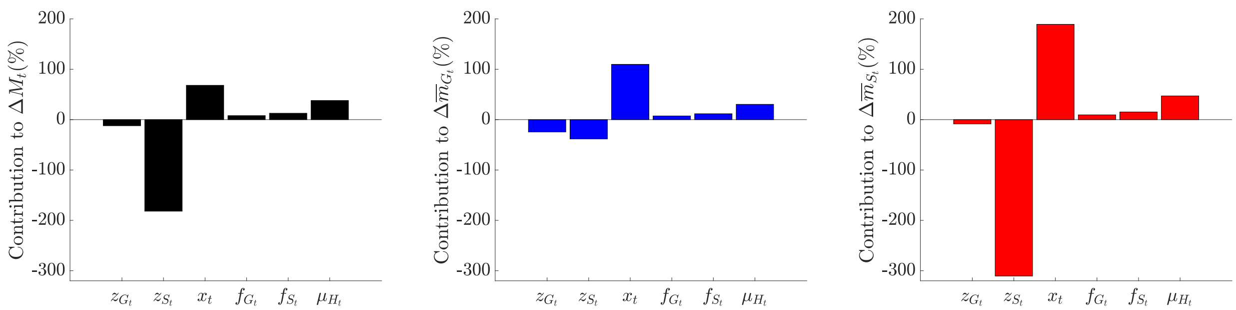
$$\text{Contribution} \quad = \quad 100 \times \left(\frac{M_{2015}^{\text{baseline}} - M_{2015}^{\text{experiment}}}{M_{2015}^{\text{baseline}} - M_{1980}^{\text{baseline}}} \right),$$

where M^{baseline} is the (aggregate or average) markup in the baseline economy and $M^{\text{experiment}}$ the markup in the counterfactual economies when z_G, z_S, x, f_G, f_S, and μ_H are fixed at their 1980 values.

income as well as income inequality by rewarding high-skilled workers in the labor market. This in turn decreases consumers' price elasticity of demand for goods and services, which allows firms to increase their markups.

In contrast, the decline in total factor productivity across services firms helps contain the rise in aggregate markups. The rise of the share of high-income consumers in the economy also plays an important role in driving markups up. The increase in entry costs plays a minor role in the rise of market power overall.

Figure 5.2: Decomposing the rise of markups



(a) Aggregate markup (b) Average goods markup (c) Average services markup

Note: The figure shows the relative importance of each exogenous force for the rise of the aggregate markup (panel (a)), the average markup of goods (panel (b)), and the average markup of services (panel (c)).

5.7 Welfare

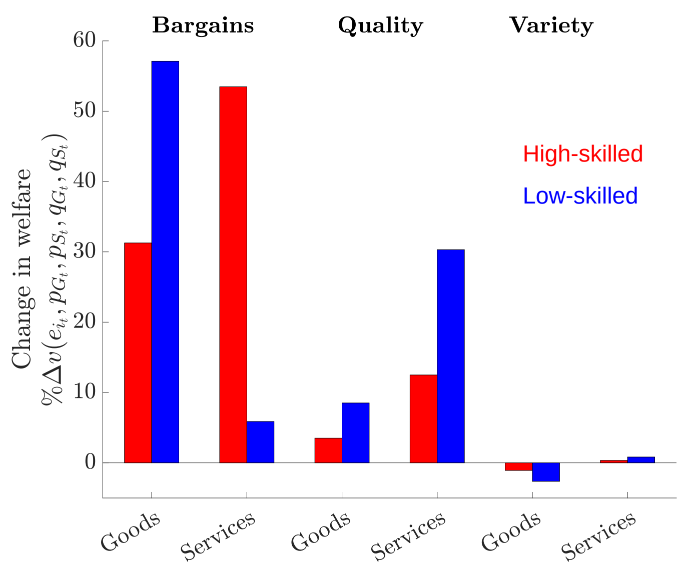
In contrast to Edmond, Midrigan, and Xu (2021) and De Loecker, Eeckhout, and Mongey (2021), this framework predicts that welfare increases along with the rise of markups. Using equations (3.1), (3.3), and (3.3), we can now decompose the changes in the log of the indirect utility of a type i household between 1980 and 2015 as stemming from (i) changes in the difference between the varieties' prices and the consumer's choke price net of changes in her income (what I henceforth call *love for bargains*); (ii) changes in the varieties' quality (capturing the *love for quality*); and (iii) changes in the number of varieties (capturing the usual *love for variety*). The change in welfare between 1980 and 2015 can thus be written as

$$\begin{aligned}\Delta v(e_{i_t}, p_{G_t}, p_{S_t}, q_{G_t}, q_{S_t}) \quad = \quad & \underbrace{(1+\gamma)\left[\lambda\Delta\left(\phi_G e_{i_t} - p_{G_t}\right) + (1-\lambda)\Delta\left(\phi_S e_{i_t} - p_{S_t}\right) - \Delta e_{i_t}\right]}_{\text{love for bargains (net of income changes)}} \\ & + \underbrace{(1+\gamma)\,\delta\left[\lambda\Delta q_{G_t} + (1-\lambda)\Delta q_{S_t}\right]}_{\text{love for quality}} \\ & + \underbrace{\lambda\Delta N_{G_t} + (1-\lambda)\Delta N_{S_t}}_{\text{love for variety}}. \end{aligned} \tag{5.1}$$

Figure 5.3 shows this decomposition as a fraction of the total change in utility between 1980 and 2015. The biggest contributor to the increase in welfare for both types of households is the *love* *for bargains* term. As consumers become richer over time (high-skilled incomes grew 77%, while low-skilled incomes grew 2%), they have more disposable income to spend on goods and services. Together with the decline of the price of goods over time, both poor and rich households see their welfare increase noticeably. As low-skilled households see their incomes increase at a much slower pace than high-skilled households, the increase in utility from consuming services is lower for the former.

Although both high and low-skilled households value the *love for quality* and the *love for variety* in the same fashion (as they do not depend on income), the contribution of these terms to the increase in welfare across the two types of households is different. The increase in quality had a bigger impact on low-skilled households than on high-skilled households. As the number of varieties of goods declined over the 1980-2015 period and low-skilled households devote a larger fraction of their income on goods relative to high-skilled households, the contribution of the *love for variety* of goods is negative and more significant for low-skilled consumers. Given the modest increase in the number of varieties of services, its contribution to the welfare increase is barely noticeable.

Figure 5.3: Decomposing welfare gains



Note: The figure shows the contribution of each term in equation (5.1) to the change in indirect utilities for high-skilled (red) and low-skilled (blue) consumers.

How much would consumers in 1980 need to receive in order to have the utility level they enjoyed in 2015? The equivalent variation measures the adjustment in income in 1980 that would make consumer i's utility equal to the level achieved in 2015, which corresponds to the value of ε_i^{ev} that solves the following equation

$$v(e_{i_{1980}}(1+\varepsilon_i^{ev}), p_{G_{1980}}, p_{S_{1980}}, q_{G_{1980}}, q_{S_{1980}}) \quad = \quad v(e_{i_{2015}}, p_{G_{2015}}, p_{S_{2015}}, q_{G_{2015}}, q_{S_{2015}}).$$

Table 5.3 shows the equivalent variation for high and low-skilled consumers. The increase in income of high-skilled households in 1980 that would give the consumer the same utility as in 2015, when markups are higher but prices are lower, is 136%. In contrast, the income of low-skilled households in 1980 would have had to be 20% higher in order to enjoy the same level of utility as in 2015.

How much would consumers need to receive in 2015 to enjoy the utility level of the counterfactual economies? The equivalent variation now delivers a value of ε_i^{ev} that solves the following

$$v\left(e_{i_{2015}}^{\text{baseline}}(1+\varepsilon_i^{ev}), \mathbf{p}_{G_{2015}}^{\text{baseline}}, \mathbf{p}_{S_{2015}}^{\text{baseline}}, \mathbf{q}_{G_{2015}}^{\text{baseline}}, \mathbf{q}_{S_{2015}}^{\text{baseline}}\right) = v\left(e_{i_{2015}}^{\text{exp}}, \mathbf{p}_{G_{2015}}^{\text{exp}}, \mathbf{p}_{S_{2015}}^{\text{exp}}, \mathbf{q}_{G_{2015}}^{\text{exp}}, \mathbf{q}_{S_{2015}}^{\text{exp}}\right).$$

Since households are worse off in the economy in which incomes are held constant at their 1980 values, both rich and poor consumers are willing to forego 9% and 13% of their 2015 income (measured in the baseline economy) to avoid the lower level of utility. In contrast, consumers are better off in the remaining counterfactual economies. Wealthier consumers stand to gain the most from the economy with 1980 prices of goods and services. Lower entry costs in the goods and services sector benefit the low-skilled households the most as they gain more from the new available goods and services—they are willing to pay half of their 2015 income to enjoy that level of utility. Poorer households are also willing to pay a similar amount to enjoy the utility level in the economy with fewer wealthier consumers.

Table 5.3: Equivalent variations

ε_i^{ev}, %	**High-skilled**	**Low-skilled**
Baseline economy, 1980 vs. 2015		
From 1980 to 2015	136.1	20.4
Baseline economy vs. Counterfactual economy, 2015		
Incomes constant at 1980 values	-8.6	-12.5
Prices constant at 1980 values	84.1	3.5
Entry costs constant at 1980 values	18.3	50.0
High-skilled share constant at 1980 values	15.5	44.9

6 Extensions and Robustness

I have presented a parsimonious model that can be extended in various directions. The first extension is to consider another market structure. Monopolistic competition delivers a markup that is solely a function of consumers' price elasticities of demand. As a result, changing the number of firms in any given market does not directly impact a firm's markup and only has an indirect effect through equilibrium conditions. Introducing oligopolistic competition allows markups to be both directly and indirectly impacted by the evolution of the number of firms in the market. The second extension is to make the model dynamic. Introducing savings does not alter the definition of a consumer's price elasticity of demand and therefore has no bearing on firms' markups. The third extension discusses the reasons why alternative preferences frequently used in the literature are not suitable to study this problem with multiple commodities and variable markups within and across sectors.

6.1 Oligopoly

Assume now that firms compete *à la* Cournot in each market. Let ω_j denote a variety of commodity $j \in \{G, S\}$. A firm chooses its level of output and quality, $y_{\omega_{j_t}}$ and $q_{\omega_{j_t}}$, to maximize profits taking the output of its competitors as fixed. The price of commodity j is now a function of all output produced in that industry, y_{j_t}. I now write the firm's problem (3.12) as

$$\begin{aligned} \pi_{\omega_{j_t}} &= \max_{y_{\omega_{j_t}}, q_{\omega_{j_t}}} \left(p(y_{j_t}) - mc_{\omega_{j_t}} \right) y_{\omega_{j_t}} - \kappa q_{\omega_{j_t}}^{\vartheta} - f_{j_t} \qquad (6.1) \\ \text{s.t.} \quad y_{j_t} &= \mu_{H_t} c_{H,j_t} + \mu_{L_t} c_{L,j_t} \\ y_{j_t} &= \sum_{\widetilde{\omega}_j = 1}^{N_{j_t}} y_{\widetilde{\omega}_{j_t}}. \end{aligned}$$

The firm's optimal pricing decision is still a markup over marginal costs as in equation (3.13). The markup is now not only a function of the average price elasticity of demand of a firm's consumers, but also depends on the firm's share of total sales of commodity j, i.e.,

$$m_{\omega_{j_t}} = \frac{\overline{\xi}_{\omega_{j_t}}}{\overline{\xi}_{\omega_{j_t}} - \frac{p(y_{j_t}) y_{\omega_{j_t}}}{p(y_{j_t}) y_{j_t}}}. \qquad (6.2)$$

Focus on a symmetric equilibrium. Then, the average price elasticity of demand is the same as the one derived in equation (3.15) and a firm's sales share of sectoral output is simply given by $\frac{p(y_{j_t}) y_{\omega_{j_t}}}{p(y_{j_t}) y_{j_t}} = \frac{1}{N_{j_t}}$. The firm's optimal choice of quality is unaltered and given by equation (3.16).

The model is calibrated following the same strategy as in Section 4. The fit is as good as before. The set of parameters that deliver these results is presented in Table 6.1. The most noticeable difference between these values and the ones estimated from the monopolistic competition model pertains to the choke price parameters, ϕ_G and ϕ_S. Now the maximum price a consumer is willing to pay for a service is seven times larger than for goods (it was a factor of two under monopolistic competition). Figure D.4.11 in Appendix D.4 depicts the average markups of goods and services as well as the services sales share. The time series from the Cournot oligopoly model trail the ones from the monopolistic competition very closely.

The experiments are more interesting to focus on. Tables 6.2 and D.5.1 in Appendix D.5 replicate the results displayed in the previous section. The upshot from the first table is that entry costs in the services sector became a stronger driver of the increase in markups. Setting the services' entry cost at its value in 1980 translates into a meager 4% increase in the aggregate markup over the 1980-2015 period (Column 6, Table 6.2) as opposed to the 11% increase observed in the data. Skill-biased technological progress also now plays a weaker role in the increase in markups (Column 4, Table 6.2). The increase in the aggregate markup over the period was 5% in the model with Cournot competition vs. a 6% decline in the model with monopolistic competition. Similarly, reducing the share of wealthy consumers in the economy has a milder impact on markups when firms compete *à la* Cournot (Column 7, Table 6.2). Keeping incomes constant at their 1980 values confirms the weaker

Table 6.1: Parameter values

Parameter	Description	Monopolistic comp.	Cournot comp.
Preferences			
λ	Indirect utility's weight on goods	0.181	0.043
γ	Exponent in indirect subutility	17.359	28.862
ϕ_G	Choke price of goods	7.725	5.317
ϕ_S	Choke price of services	12.780	38.269
δ	Exponent related with quality	0.072	0.045
Technology			
α	High-skilled weight	0.465	0.465
ι	Elasticity of substitution between high and low-skilled	0.400	0.400
x_t	Skill-biased prod. in 1980, 2015	1.000, 1.844	1.000, 1.844
z_{G_t}	TFP in goods sector in 1980, 2015	0.530, 0.485	0.924, 0.881
z_{S_t}	TFP in services sector in 1980, 2015	0.580, 0.355	1.029, 0.644
Costs			
f_{G_t}	Entry costs in goods sector in 1980, 2015	0.009, 0.027	0.008, 0.027
f_{S_t}	Entry costs in services sector in 1980, 2015	0.010, 0.024	0.011, 0.024
κ	Linear term related with quality	0.018	0.015
ϑ	Exponent related with quality	2.000	2.000
Measure			
μ_{H_t}	Share of high-skilled households in 1980, 2015	0.325, 0.424	0.325, 0.424

role of skill-biased technological progress (Column 2, Table D.5.1, Appendix D.5). Markups would still be considerably lower than in the baseline economy, but not as low as in the model with monopolistic competition. Keeping prices constant at their 1980 values also imply a smaller decline in markups relative to the model with monopolistic competition (Column 4, Table D.5.1, Appendix D.5).

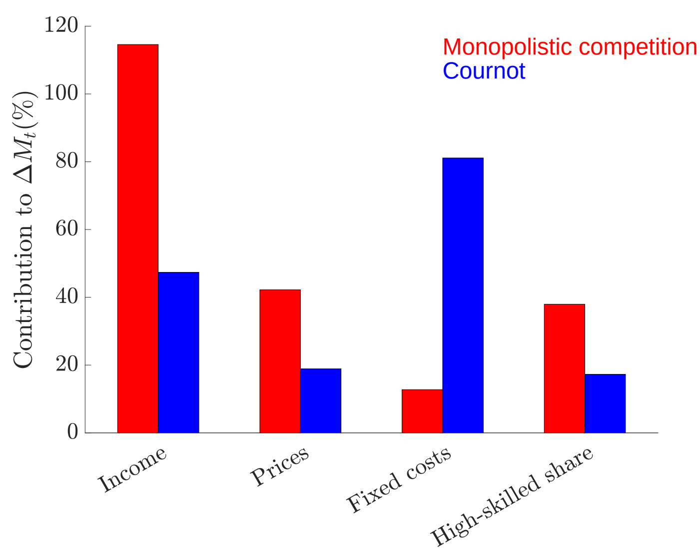
Figure 6.1 depicts the contribution of the different factors to the increase in the aggregate markup. Under monopolistic competition, the rise in markups is mostly the result of consumers becoming richer over time. In contrast, under Cournot competition both the increase in incomes and fixed costs led to the rise in markups.

In terms of welfare, quality is a much more important driver of the change in utility for both high and low-skilled consumers (see Figure D.4.12 in Appendix D.4). Although the quality of services increases less in the model with Cournot competition than in the model with monopolistic competition, its weight in consumers' utility is much larger (the value of $(1-\lambda)$ is now 0.96 vs. 0.82). Table D.5.2 in Appendix D.5 shows the coefficients of equivalent variation over time for the baseline economy as well as across scenarios. High-skilled consumers would now need a fourfold increase of their 1980 income to be as well off as in 2015 in the economy with Cournot competition. Welfare gains are also significantly higher for low-skilled households in this model.

Table 6.2: Experiments (Cournot): Technological progress, entry costs, and high-skilled share

Variable	Description	(1) Baseline (Cournot) 1980, 2015	(2) z_{G_t} 2015	(3) z_{S_t} 2015	(4) x_t 2015	(5) f_{G_t} 2015	(6) f_{S_t} 2015	(7) μ_{H_t} 2015
Markups								
M_t	Aggregate markups	1.136, 1.263	1.265	1.323	1.192	1.238	1.183	1.241
$\overline{m}_{G_t}$	Average goods markups	1.058, 1.215	1.228	1.190	1.132	1.113	1.205	1.191
$\overline{m}_{S_t}$	Average services markups	1.178, 1.276	1.275	1.382	1.214	1.273	1.176	1.255
Relative prices and income								
$\overline{p}_{S_t}/\overline{p}_{G_t}$	Relative price of services	1.000, 1.437	1.489	0.996	1.467	1.565	1.335	1.441
w_{H_t}/w_{L_t}	Skill premium	1.347, 1.928	1.928	1.928	1.046	1.928	1.928	2.484
ϵ_{H_t}	High-skilled income share	0.365, 0.603	0.593	0.567	0.464	0.635	0.601	0.548
$\omega_{S_t}^{sales}$	Services share	0.670, 0.790	0.798	0.724	0.748	0.806	0.775	0.788
Employment and labor shares								
$\theta_{H/L,S_t}$	High/Low-skilled empl. share in services	0.645, 0.782	0.792	0.693	0.735	0.785	0.780	0.779
$VC_{H,G_t}/PY_t$	High-skilled labor share in goods	0.123, 0.101	0.097	0.136	0.097	0.102	0.109	0.097
$VC_{H,S_t}/PY_t$	High-skilled labor share in services	0.224, 0.363	0.367	0.307	0.268	0.372	0.387	0.342
$VC_{L,G_t}/PY_t$	Low-skilled labor share in goods	0.189, 0.071	0.068	0.096	0.126	0.072	0.077	0.081
$VC_{L,S_t}/PY_t$	Low-skilled labor share in services	0.345, 0.256	0.259	0.217	0.348	0.262	0.273	0.286
Entry								
$\kappa Q_{G_t}/PY_{G_t}$	Sales share of quality costs in goods	0.037, 0.118	0.123	0.106	0.077	0.067	0.113	0.107
$\kappa Q_{S_t}/PY_{S_t}$	Sales share of quality costs in services	0.101, 0.144	0.143	0.184	0.117	0.143	0.100	0.135
N_{S_t}/N_{G_t}	Relative number of services firms	4.059, 5.180	5.152	5.117	5.053	2.762	7.892	5.275

Figure 6.1: Decomposing the increase in aggregate markups



Note: The figure shows the contribution of changes in income (through skill-biased technological change), prices (through neutral technological change), fixed costs (in both sectors), and the share of wealthy consumers to the aggregate markup in the economy with monopolistic competition (red) and Cournot competition (blue).

6.2 Dynamic model

Assume now that households can save a fraction of their income in exchange for a return rate R_t. Discounting the future at rate $1/\beta$, consumers now have a lifetime indirect utility given by

$$\sum_{t=0}^{\infty} \beta^t v(e_t, p_{G_t}, p_{S_t}, q_{G_t}, q_{S_t}), \tag{6.3}$$

where $v(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})$ is defined as in equations (3.1), (3.2), and (3.3). The budget constraint they now face is given by

$$e_t + a_{t+1} = w_t + R_t a_t + \Lambda_t, \tag{6.4}$$

where a_t is the amount of wealth owned by the household. A consumer now maximizes equation (6.3) subject to (6.4). The static problem of deciding how much to spend on goods and services is the same as in Section 3. Hence, the spending share on services as well as the price and quality elasticities of demand are still given by equations (3.5), (3.6), and (3.7), respectively. The consumer's optimal savings decision is in turn given by the following Euler equation[25]

$$\left(\frac{e_{t+1}}{e_t}\right)^{2+\gamma} = \beta R_{t+1}\left(\frac{A_{t+1}}{A_t}\right), \tag{6.5}$$

where

$$\begin{aligned} A_t &= \lambda \int_0^{N_{G_t}} [\phi_G e_t - p_{G_t}(\omega)]^{\gamma} p_{G_t}(\omega)\, q_{G_t}(\omega)^{\delta(1+\gamma)}\, d\omega \\ &\quad + (1-\lambda) \int_0^{N_{S_t}} [\phi_S e_t - p_{S_t}(\omega)]^{\gamma} p_{S_t}(\omega)\, q_{S_t}(\omega)^{\delta(1+\gamma)}\, d\omega. \end{aligned}$$

Firms now produce a variety of commodity $j \in \{G, S\}$ using capital and labor according to the following technology

$$y_{j_t} = z_{j_t} k_{j_t}^{\alpha_k} \left[\alpha_h x_t h_{j_t}^{\iota} + (1-\alpha_h)\, \ell_{j_t}^{\iota}\right]^{\frac{1-\alpha_k}{\iota}}. \tag{6.6}$$

Capital is mobile across sectors and in order to use it firms have to pay a rental rate r_t. Firms' total costs therefore include the expenses related to the rental of capital, which in turn affects its marginal cost. The latter is now given by

$$mc_{j_t} = \frac{1}{z_{j_t}}\left(\frac{r_t}{\alpha_k}\right)^{\alpha_k}\left(\frac{w_{L_t}}{1-\alpha_k}\right)^{(1-\alpha_k)}\left[(\alpha_h x_t)^{\frac{1}{1-\iota}}\left(\frac{w_{H_t}}{w_{L_t}}\right)^{\frac{\iota}{\iota-1}} + (1-\alpha_h)^{\frac{1}{1-\iota}}\right]^{\frac{(\iota-1)(1-\alpha_k)}{\iota}}. \tag{6.7}$$

This equation is equivalent to the marginal cost derived in the previous section for $\alpha_k = 0$ (see equation (3.11)). Marginal costs are increasing in the skill premium as well as in the rental rate of capital. In addition to the high-to-low-skilled labor ratio being the same across sectors (equation (3.10) still holds), the capital-labor ratio also is the same for goods and services-producing firms. The capital-to-low-skilled labor ratio demanded by a firm is then

$$\frac{k_{j_t}}{\ell_{j_t}} = \frac{\alpha_k (1-\alpha_h)}{1-\alpha_k}\frac{w_{L_t}}{r_t}\left[\left(\frac{\alpha_h x_t}{1-\alpha_h}\right)^{\frac{1}{1-\iota}}\left(\frac{w_{L_t}}{w_{H_t}}\right)^{\frac{\iota}{1-\iota}} + 1\right]. \tag{6.8}$$

[25] We can derive an *approximate* intertemporal elasticity of substitution of expenditure (IES) as

$$\text{IES} \equiv \frac{\partial (e_{t+1}/e_t)}{\partial R_{t+1}} \frac{R_{t+1}}{(e_{t+1}/e_t)} \approx \frac{1}{2+\gamma}.$$

As the interest rate rises, the growth rate of total consumption spending is approximately equivalent to $1/(2+\gamma)$.

Introducing capital does not alter the optimal decision of a firm with respect to the price and quality of its variety. The firm's markup is therefore still given by equation (3.14).

Finally, the asset market clearing condition requires that all households' savings, $A_t = \mu_{H_t} a_{H_t} + \mu_{L_t} a_{L_t}$, equate the total capital demanded by firms, $K_t = \left[\int_0^{N_{G_t}} k_{G_t}(\omega)\, d\omega + \int_0^{N_{S_t}} k_{S_t}(\omega)\, d\omega\right]$. This equilibrium condition determines the rental rate of capital, where $R_t = 1 + r_t - \delta$ and δ is the capital depreciation rate.

Like Greenwood, Hercowitz, and Huffman (1988) preferences, the utility function presented here is not consistent with a balanced growth path. The reason is that markups and therefore prices do not grow at a constant rate. To illustrate the point, take a firm that sells its variety to only one of the consumers. The firm's markup is now given by $m_{j_t}(\omega) = \frac{\gamma p_{j_t}(\omega)}{(1+\gamma)p_{j_t}(\omega) - \phi_j e_{i_t}}$, which evolves according to

$$\frac{m_{j_{t+1}}(\omega)}{m_{j_t}(\omega)} = \frac{(1+\gamma) - \phi_j e_{i_t}/p_{j_t}(\omega)}{(1+\gamma) - \phi_j e_{i_{t+1}}/p_{j_{t+1}}(\omega)}.$$

Here, the presence of the choke price prevents markups from growing at a constant rate.

6.3 Alternative preferences

It's worth explaining why alternative preferences are not suitable to study this problem. First, Kimball and non-homothetic CES preferences, by themselves, do not deliver both an increase in the services share and an increase in sectoral markups. To be precise, when a discrete number of commodities (e.g. goods and services) with many varieties within them are aggregated, they imply a consumer spends a constant share of income on services. In order to break that result, these preferences have to be nested, for instance, within a Stone-Geary utility function. Under some restrictions, these preferences allow the price elasticity of demand to be falling in the consumer's income and increasing in the variety's price. I explore these issues next.

Kimball. Kimball (1995) preferences have been extensively used to introduce markups that vary endogenously across firms. Although these preferences allow markups to vary over time, they do not allow the services share to increase as incomes rise. To see this, let a consumer's direct utility be represented by a Cobb-Douglas function of a goods bundle, C_G, and a services bundle, C_S. The consumer's problem is then to solve the following

$$\max_{\{c_{G_t}(\omega)\},\{c_{S_t}(\omega)\},C_{G_t},C_{S_t}} \quad C_{G_t}^{\lambda} C_{S_t}^{1-\lambda}, \tag{6.9}$$

subject to the budget constraint (3.4). Here the sector-specific consumption bundle is implicitly defined by the Kimball aggregator $\Upsilon_j(\cdot)$ for $j = \{G, S\}$ according to

$$\int_{\mathcal{N}_{j_t}} \Upsilon_j\left(\frac{c_{j_t}(\omega)}{C_{j_t}}\right) d\omega = 1,$$

where $\Upsilon_j(\cdot)$ satisfies the constraints $\Upsilon_j(1) = 1$, $\Upsilon_j'(\cdot) > 0$, and $\Upsilon_j''(\cdot) < 0$.[26] The solution to the

[26] Here I have abstracted from quality, but introducing it is straightforward.

consumer's problem delivers a demand for a variety of commodity j given by

$$c_{j_t}(\omega) \quad = \quad \Psi_j\left(\frac{p_{j_t}(\omega)D_{j_t}}{P_{j_t}}\right)C_{j_t},$$

where D_{j_t} is a sector-specific demand index defined as $D_{j_t} = \int_{\mathcal{N}_{j_t}} \Upsilon_j'\left(\frac{c_{j_t}(\omega)}{C_{j_t}}\right)\frac{c_j(\omega)}{C_j}\,d\omega$, P_{j_t} is a sector-specific price index such that $P_{G_t} = \frac{\lambda e_t}{C_{G_t}}$ and $P_{S_t} = \frac{(1-\lambda)e_t}{C_{S_t}}$, and $\Psi_j(\cdot) \equiv \Upsilon_j'^{-1}(\cdot)$ is the inverse of the derivative of the Kimball aggregator $\Upsilon_j(\cdot)$ such that $\Psi_j(\cdot) > 0$ and $\Psi_j'(\cdot) < 0$.[27]

In turn, a consumer's price elasticity of demand is given by

$$\xi_{j_t}(\omega) \quad = \quad -\frac{\frac{p_{j_t}(\omega)D_{j_t}}{P_{j_t}}\Psi_j'\left(\frac{p_{j_t}(\omega)D_{j_t}}{P_{j_t}}\right)}{\Psi_j\left(\frac{p_{j_t}(\omega)D_{j_t}}{P_{j_t}}\right)}.$$

Under the Klenow and Willis (2016) specification, this expression simplifies to

$$\xi_{j_t}(\omega) \quad = \quad \frac{\gamma}{1+\phi_j \ln\left(\frac{\gamma-1}{\gamma}\frac{P_{j_t}}{p_{j_t}(\omega)D_{j_t}}\right)}.$$

As before, a consumer's price elasticity of demand is increasing in the price of the variety, i.e. $\frac{\partial \xi_{j_t}(\omega)}{\partial p_{j_t}(\omega)} > 0$.[28] This implies that firms can charge higher markups if they sell their products at lower prices. Note, however, that these preferences annihilate any role for changes in demand, in particular through rising incomes or shifts in demand shares. As a consequence, market power arises solely from the supply side as shifts in productivity trigger changes in prices. Finally, these preferences imply that services spending is a constant fraction of total income over time, with

$$\int_{\mathcal{N}_{S_t}} p_{S_t}(\omega)c_{S_t}(\omega)d\omega \quad = \quad (1-\lambda)e_t.$$

Non-homothetic CES. Another popular utility function is the one proposed by Comin, Lashkari, and Mestieri (2021). We can write the consumer's problem as the one described above (equation (6.9)), but replace how the sector-specific consumption bundle is defined. In this particular case, let C_j be given by

$$\int_{\mathcal{N}_{j_t}}\left(\frac{c_{j_t}(\omega)}{C_{j_t}^{\phi_j}}\right)^{\frac{\gamma-1}{\gamma}} d\omega \quad = \quad 1,$$

[27] Following Klenow and Willis (2016), the Kimball aggregator is defined as $\Upsilon(x;\phi,\gamma) = 1 + (\gamma - 1)\ e^{1/\phi}\phi^{\ \gamma/\phi-1}\left[\Gamma\left(\frac{\gamma}{\phi},\frac{1}{\phi}\right)-\Gamma\left(\frac{\gamma}{\phi},\frac{x^{\phi/\gamma}}{\phi}\right)\right]$, where $\Gamma(u,z)$ is the incomplete gamma function $\Gamma(u,z) = \int_z^\infty s^{u-1}e^{-s}ds$, γ is the steady-state elasticity of substitution across varieties and ϕ is a super-elasticity that controls the strength of the strategic complementarities between varieties. For $\phi \to 0$, the Kimball aggregator reduces to the constant elasticity of substitution (CES) aggregator with $\Upsilon(x) = x^{\frac{\gamma-1}{\gamma}}$.

[28] To be more precise, the price elasticity of demand is increasing in a variety's price as long as $\xi_j(\omega) > -1 - \left(\frac{p_j(\omega)D_j}{P_j}\right)\frac{\Psi_j''\left(\frac{p_j(\omega)D_j}{P_j}\right)}{\Psi_j'\left(\frac{p_j(\omega)D_j}{P_j}\right)}$.

where γ is the price elasticity of demand and ϕ_j controls the income elasticity of demand. The demand for a particular variety of commodity j is defined by

$$c_{j_t}(\omega) \quad = \quad \left(\frac{p_{j_t}(\omega)}{D_j e_t}\right)^{-\gamma} C_{j_t}^{\phi_j(1-\gamma)},$$

where D_j is a time-invariant sector-specific demand index such that $D_G = \frac{\lambda/\phi_G}{\lambda/\phi_G+(1-\lambda)/\phi_S}$ and $D_S = \frac{(1-\lambda)/\phi_S}{\lambda/\phi_G+(1-\lambda)/\phi_S}$. Now not only the price elasticity of demand is constant over time, it is also the same for both goods and services (i.e., given by γ for both varieties of goods and services). The spending share of the many varieties of services is also constant over time and simply given by

$$\int_{\mathcal{N}_{S_t}} p_{S_t}(\omega) c_{S_t}(\omega) d\omega \quad = \quad D_S\, e_t.$$

Stone-Geary. Stone-Geary preferences are particularly popular in the structural transformation literature because they generate non-unitary income elasticities of demand. These are characterized by a subsistence point in the direct utility and can be easily combined with (homothetic or non-homothetic) CES. Now define the consumer's problem as above (equation (6.9)), but let each consumption bundle be explicitly defined by

$$C_{j_t} \quad = \quad \left[\int_{\mathcal{N}_{j_t}} \left(c_{j_t}(\omega) + \bar{c}_j\right)^{\frac{\gamma-1}{\gamma}} d\omega\right]^{\frac{\gamma}{\gamma-1}},$$

where $\bar{c}_j > 0$ is a sector-specific subsistence point. For the price elasticity of demand to vary for both goods and services, we need both $\bar{c}_G$ and $\bar{c}_S$ to be different from zero. In addition, for the price elasticity of demand to satisfy the properties defined in Section A in the Online Appendix A, we must impose that $\bar{c}_j$ is positive (more on this below). A consumer's demand for a variety of commodity j is given by

$$c_{j_t}(\omega) \quad = \quad \left(\frac{p_{j_t}(\omega)}{P_{j_t}}\right)^{-\gamma} C_{j_t} - \bar{c}_j,$$

where P_{j_t} is a sectoral price index such that

$$P_{j_t} C_{j_t} \quad = \quad \lambda_j \left[e_t + \bar{c}_G \int_{\mathcal{N}_{G_t}} p_{G_t}(\omega) d\omega + \bar{c}_S \int_{\mathcal{N}_{S_t}} p_{S_t}(\omega) d\omega\right],$$

where $\lambda_j = \lambda$ for goods and $\lambda_j = 1-\lambda$ for services. It is easy to see that the spending share of services can increase if the income share of the goods and services subsistence baskets, $\left[\bar{c}_G \int_{\mathcal{N}_{G_t}} p_{G_t}(\omega) d\omega + \bar{c}_S \int_{\mathcal{N}_{S_t}} p_{S_t}(\omega) d\omega\right] / e_t$, rises over time.

The price elasticity of demand can now vary both as a result of changes in price and income, as was the case in the baseline economy. A consumer's price elasticity of demand for a variety of commodity j can in turn be written as

$$\xi_{j_t}(\omega) \quad = \quad \gamma \frac{\left(\frac{p_{j_t}(\omega)}{P_{j_t}}\right)^{1-\gamma} P_{j_t} C_{j_t}}{\left(\frac{p_{j_t}(\omega)}{P_{j_t}}\right)^{1-\gamma} P_{j_t} C_{j_t} - p_{j_t}(\omega)\bar{c}_j}.$$

How does the price elasticity of demand varies with price and income? The super-elasticity of demand with respect to price is now given by

$$\frac{\partial \xi_{j_t}(\omega)}{\partial p_{j_t}(\omega)} \frac{p_{j_t}(\omega)}{\xi_{j_t}(\omega)} = \xi_{j_t}(\omega) - \gamma$$

and the super-elasticity of demand with respect to income is

$$\frac{\partial \xi_{j_t}(\omega)}{\partial e_t} \frac{e_t}{\xi_{j_t}(\omega)} = \frac{\lambda_j}{P_{j_t} C_{j_t}} e_t \left[1 - \frac{\xi_{j_t}(\omega)}{\gamma}\right],$$

where $\lambda_j = \lambda$ for goods and $\lambda_j = 1 - \lambda$ for services. Hence, a consumer's price elasticity of demand is increasing in the price of the variety and falling in income as long as $\xi_{j_t}(\omega) > \gamma$. For that condition to hold, it must be that $\bar{c}_j > 0$. Contrast these expressions with the ones derived in the baseline model (footnote 20). Although the super-elasticity of demand with respect to price is only a function of the level of the elasticity itself, the super-elasticity with respect to income now depends on both the income of the consumer and the elasticity itself.

7 Conclusion

This paper documents that the rise of services is the key driver of the rise in markups. This rise is observed despite similar trends in the share of fixed costs in terms of sales for firms in the services and non-services sectors and is robust to dropping superstar firms. In particular, this rise is consistent with the pattern of structural change in which economic activity and consumption shift from manufacturing toward services and the relative price of services increases.

Therefore, this suggests that the two standard drivers of structural change—namely, differential rates of technological progress and income effects through non-homotheticities—can potentially explain the rise of markups. This paper quantitatively demonstrates that the drivers of structural change lead to the rise of markups. The condition necessary for this to happen is to design preferences that make the price elasticity of demand for goods and services increasing in consumers' income—what Harrod (1936) called the *Law of Diminishing Elasticity of Demand*—as well as the pass-through of efficiency gains to consumers to be smaller than one—often referred to as Marshall's (1890) *Second Law of Demand*. This paper provides the theoretical foundation for this mechanism. Results from novel experimental data based on a representative online survey eliciting demand elasticities supports the assumed form of preferences.

A two-sector model of structural change is built and calibrated to U.S. data over the 1980-2015 period. I use the model to assess the quantitative importance of structural change for understanding the rise in markups. I find that skill-biased technological change, which reduces marginal costs but increases income inequality, was the main driver of the rise in markups. In particular, keeping incomes constant at their 1980 levels would have led to a decline in the aggregate markup. In contrast, changes in fixed costs seem to have played a minor role in the increase in markups. Keeping the price of goods and services constant over time mitigates the increase in markups.

The findings in this paper have important implications for the interpretation of the rise of markups in particular and for models of imperfect competition in general. Jointly modeling changes in demand and supply provides a new avenue for analyzing markups and market power. In addition, the increasing importance of services poses new challenges that have yet to be quantified. They allowed firms to offer more targeted and specialized products to consumers, increasing their abilities to price discriminate between them. The advent of digital advertising and big data may have facilitated this better targeting of consumers. Those considerations are left for future research.

References

D. Acemoglu and D. Autor. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4(12):1043–1171, 2011.

H. Afrouzi, A. Drenik, and R. Kim. Growing by the masses: Revisiting the link between firm size and market power. *Working Paper*, 2021.

P. Aghion, A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li. A theory of falling growth and rising rents. *Working Paper*, 2022.

M. Aguiar and M. Bills. Has consumption inequality mirrored income inequality? *American Economic Review*, 105(9):2725–2756, 2015.

U. Akcigit and S. T. Ates. Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, 13(1):257–298, 2021.

D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709, 2020.

S. Basu. Are price-cost markups rising in the united states? a discussion of the evidence. *Journal of Economic Perspectives*, 33(3):3–22, 2019.

P. Bertoletti and F. Etro. Monopolistic competition when income matters. *The Economic Journal*, 127: 1217–1243, 2017.

P. Bertoletti, F. Etro, and I. Simonovska. International trade with indirect additivity. *American Economic Journal: Microeconomics*, 10(2):1–57, 2018.

S. Bond, A. Hashemi, G. Kaplan, and P. Zoch. Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 121: 1–14, 2021.

T. Boppart. Structural change and the kaldor facts in a growth model with relative price effects and non-gorman preferences. *Econometrica*, 82(6):2167–2196, 2014.

R. F. Bretherton. A note on the law of diminishing elasticity of demand. *The Economic Journal*, 47(187): 574–577, 1937.

B. Bridgman and B. Herrendorf. Labor share, markups, and input-output linkages – evidence from the u.s. national accounts. *Working Paper*, 2022.

F. J. Buera and J. P. Kaboski. The rise of the service economy. *American Economic Review*, 102(6): 2540–2569, 2012.

F. J. Buera, J. P. Kaboski, R. Rogerson, and J. I. Vizcaino. Skill-biased structural change. *Review of Economic Studies*, 2021.

O. Coibion, D. Georgarakos, Y. Gorodnichenko, and M. van Rooij. How does consumption respond to news about inflation? field evidence from a randomized control trial. *AEJ Macroeconomics, forthcoming*, 2021.

O. Coibion, D. Georgarakos, Y. Gorodnichenko, G. Kenny, and M. Weber. The effect of macroeconomic uncertainty on household spending. *NBER Working Paper*, (28625), 2022.

D. Comin, D. Lashkari, and M. Mestieri. Structural change with long-run income and price effects. *Econometrica*, 89(1):311–374, 2021.

J. De Loecker and P. F. Warzynski. Markups and firm-level export status. *American Economic Review*, 102:2437–2471, 2012.

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

J. De Loecker, J. Eeckhout, and S. Mongey. Quantifying market power and business dynamism in the macroeconomy. *mimeo*, 2021.

M. De Ridder. Market power and innovation in the intangible economy. *mimeo*, 2021.

M. De Ridder, B. Grassi, and G. Morzenti. The hitchhiker's guide to markup estimation. *mimeo*, 2022.

A. Deaton. Estimation of own- and cross-price elasticities from survey data. *Journal of Econometrics*, 36:7–30, 1987.

A. Deaton. Quality, quantity and spatial variation of price. *American Economic Review*, 78(3):418–430, 1988.

A. Deaton. Price elasticities from survey data. *Journal of Econometrics*, 44(3):281–309, 1990.

F. J. Díez, J. Fan, and C. Villegas-Sánchez. Global declining competition? *Journal of International Economics*, 132, 2021.

H. Döpper, A. MacKay, N. H. Miller, and J. Stiebale. Rising markups and the role of consumer preferences. *mimeo*, 2022.

C. Edmond, V. Midrigan, and D. Xu. How costly are markups? *Journal of Political Economy*, 2021.

N. Fabra and M. Reguant. Pass-through of emissions costs in electricity markets. *American Economic Review*, 104(9):2872–2899, 2014.

T. Fan, M. Peters, and F. Zilibotti. Growing like india: The unequal effects of service-led growth. *Working Paper*, 2022.

G. Gopinath, S. Kalemli-Ozcan, L. Karabarbounis, and C. Villegas-Sanchez. Capital allocation and productivity in south europe. *The Quarterly Journal of Economics*, 132(4):1915–1967, 2017.

J. Greenwood, Z. Hercowitz, and G. W. Huffman. Investment, capacity utilization, and the real business cycle. *American Economic Review*, 78(3):402–417, 1988.

G. Grullon, Y. Larkin, and R. Michaely. Are us industries becoming more concentrated? *Review of Finance*, 23(4):697–743, 2019.

R. E. Hall. New evidence on the markup of prices over marginal costs and the role of mega-firms in the u.s economy. *NBER Working Paper*, (24574), 2018.

R. Harrod. *The Trade Cycle: An Essay*. Clarendon Press, Oxford, 1936.

C.-T. Hsieh and E. Rossi-Hansberg. The industrial revolution in services. *working paper*, 2021.

S. Kalemli-Ozcan, B. Sorensen, C. Villegas-Sanchez, V. Volosovych, and S. Yesiltas. How to construct nationally representative firm level data from the orbis global database: New facts and aggregate implications. *NBER Working Paper Series*, (21558), 2019.

L. Karabarbounis and B. Neiman. The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1):61–103, 2014.

L. F. Katz and K. M. Murphy. Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics*, 107:35–78, 1992.

M. S. Kimball. The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking*, pages 1241–1277, 1995.

P. J. Klenow and J. L. Willis. Real rigidities and nominal price changes. *Economica*, 83:443–472, 2016.

P. Kongsamut, S. Rebelo, and D. Xie. Beyond balanced growth. *Review of Economic Studies*, 68(4): 869–882, 2001.

M. Kuhn and J.-V. Ríos-Rull. 2013 update on the u.s. earnings, income, and wealth distributional facts: A view from macroeconomics. *Mimeo*, 2013.

A. Marshall. *The principles of economics*. London: Macmillan, 1890.

K. Matsuyama. Engel's law in the global economy: Demand-induced patterns of structural change, innovation, and trade. *Econometrica*, 87(2):497–528, 2019.

S. F. Moreira. Inside the decline of the labor share: technical change, market power, and structural change. *Working Paper*, 2022.

E. Nakamura and D. Zerom. Accounting for incomplete pass-through. *The Review of Economic Studies*, 77(3):1192–1230, 2010.

R. Ngai and C. Pissarides. Structural change in a multisector model of growth. *American Economic Review*, 97(1):429–443, 2007.

D. Raval. Testing the production approach to markup estimation. *Review of Economic Studies*, 2022.

R. Roy. La distribution du revenu entre les divers biens. *Econometrica*, 15(3):205–225, 1947.

K. Sangani. Markups across the income distribution: Measurement and implications. *Working Paper*, 2023.

S. Stantcheva. Understanding economic policies: What do people know and learn? *Harvard University Working Paper*, 2020.

S. Stantcheva. Understanding tax policy: How do people reason? *Quarterly Journal of Economics*, 136 (4):2309–2369, 2021.

S. Stantcheva. Understanding of trade. *NBER Working Paper*, (30040), 2022.

C. Syverson. Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33(3):23–43, 2019.

J. Traina. Is aggregate market power increasing? production trends using financial statements. *mimeo*, 2018.

B. van Ark and K. Jäger. Recent trends in europe's output and productivity growth performance at the sector level, 2002-2015. *International Productivity Monitor*, (33), 2017.

Online Appendix

A From Structural Change to Rising Markups: Theoretical Underpinnings

This section proposes the key ingredients needed for structural change to impact markups. It starts by offering a novel theorem linking the price elasticity of demand to the income elasticity of demand for a general class of preferences. In particular, it shows that non-homothetic preferences, which imply that for individuals of different income levels some commodities are luxuries and some are necessities, also mean that individuals of different income levels will have a different price elasticity of demand for the same commodity. This in turn has an effect on the markups firms can charge.

The theorem is used as a stepping stone for two key results. The first states the conditions for the price elasticity of demand to be increasing in a commodity's own price, which is to say that an individual's price sensitivity is lower for cheaper products. This is often referred to as Marshall's (1890) *Second Law of Demand*. The second states the conditions for the price elasticity of demand to be decreasing in a consumer's income, i.e. individuals' price sensitivity is lower the wealthier they are. This second result connects with Harrod's (1936) *Law of Diminishing Elasticity of Demand*.

These results then allow me to develop a theory of why certain firms are able to charge higher markups. In particular, three channels are highlighted. First, technological progress that reduces marginal costs translate into less expensive products. As consumers are now more willing to buy those products, the firm captures part of the gains from the lower marginal costs by increasing its markup (i.e. its cost pass-through is smaller than one because the price elasticity of demand has risen).[29] Second, the rise in customers' income decreases their price elasticity for a firm's products. In turn, the firm responds by increasing its markup. Third, changes in the composition of the firm's customer base also have an effect on the firm's markup—even if their incomes did not change. If a firm is now more likely to face wealthier consumers and cannot price discriminate them, then it will update its markup to reflect the average consumer's price elasticity of demand.

A.1 Why demand matters more than you think

Pricing with market power. A firm's markup depends on the slope of the demand curve as profit maximizing firms set their prices by equating marginal revenues to marginal costs. A firm's marginal revenue depends on the price and quantity of the product it is selling, which in turn depend on its consumers' own price elasticities of demand. If aggregate demand is composed of different consumers, all facing the same price, the price elasticity of the total demand faced by the firm can be written as the average of each individual's own price elasticity of demand weighted by their consumption share. Proposition A.1 shows that in models in which firms have market power

[29] Fabra and Reguant (2014) document that the pass-through of emissions costs to electricity prices in Spain is about 80 percent and Nakamura and Zerom (2010) show that the pass-through of exchange rates to prices in the coffee industry is also not complete.

markups can be written as a weighted average of the firm's consumers' price elasticities of demand. Let $\xi_i(q^\star)$ denote individual i's own price elasticity of demand and $\varpi_i(q^\star)$ her consumption share, where $q^\star = \sum_j y_j = \sum_i c_i$ is the aggregate quantity traded in equilibrium. Define $\epsilon_j(q^\star) \equiv \left(\frac{\partial y_j(q^\star)}{\partial q^\star}\right)^{-1} \frac{y_j(q^\star)}{q^\star}$ as firm j's output elasticity of aggregate demand.

Proposition A.1. (MARKUP) *In models of imperfect competition, in which the market structure is composed of a monopolist, monopolistic competitors or oligopolistists à la Cournot, firm j's markup, $m_j(q^\star)$, is given by*

$$m_j(q^\star) \quad = \quad \frac{\sum_i \varpi_i(q^\star)\, \xi_i(q^\star)}{\sum_i \varpi_i(q^\star)\, \xi_i(q^\star) - \epsilon_j(q^\star)}.$$

If firm j is a monopolist or a monopolistic competitor, then $\epsilon_j(q^\star) = 1$.

Proof. Assume firms have constant returns to scale technologies. If the firm is a monopolist or a monopolistic competitor, it solves the following profit maximization problem

$$\max_{c>0} \; p(c)\, c - mc\, c.$$

A solution to this problem must satisfy the first-order condition, which equates the marginal revenue to the marginal cost. Dividing both sides by $p(c)$, we have

$$\frac{\partial p(c)}{\partial c}\frac{c}{p(c)} + 1 \quad = \quad \frac{1}{m(c)},$$

where $m(c)$ is the firm's markup, and rearranging

$$m(c) \quad = \quad \frac{\xi(c)}{\xi(c) - 1},$$

where $\xi(c)$ is the price elasticity of aggregate demand or the weighted of each individual's own price elasticity of demand $\xi(c) = \sum\limits_i \varpi_i(c)\, \xi_i(c)$.

If the firm is an oligopolist competing à la Cournot, it solves the following profit maximization problem

$$\max_{c_j>0} \; p(c)\, c_j - mc_j\, c_j \quad \text{s.t. } c_j + \sum_{\kappa=1}^{N} c_\kappa = c,$$

where N is the number of oligopolistic competitors. As before, a solution to this problem must satisfy the first-order condition, which equates the marginal revenue to the marginal cost. This can be written as

$$\frac{\partial p(c)}{\partial c}\frac{c}{p(c)}\frac{\partial c}{\partial c_j}\frac{c_j}{c} + 1 \quad = \quad \frac{1}{m_j(c)},$$

and rearranging

$$m_j(c) \quad = \quad \frac{\xi(c)}{\xi(c) - \epsilon_j(c)},$$

where $\epsilon_j(c)$ firm j's output elasticity of aggregate demand, which includes the strategic interactions over the oligopolists as $\frac{\partial c}{\partial c_j} = 1 + \sum_{\kappa=1}^{N} \frac{\partial c_\kappa(c_j)}{\partial c_j}$. Firms' markup is increasing in $\epsilon_j(c)$ as $\frac{\partial m_j(c)}{\partial \epsilon_j(c)} = \frac{m_j(c)^2}{\xi(c)} > 0$. $\square$

The roots of market power are thus intertwined with how preferences are defined as they determine in equilibrium consumers' price elasticities of demand, $\xi_i(q^\star)$, and their consumption shares, $\varpi_i(q^\star)$. To proceed I resort to the indirect utility and Roy's (1947) identity. Let e_i denote individual i's expenditures (or income), $p(\omega)$ the price of variety $\omega \in [0, N]$, and $\mathbf{p}$ a vector of all prices. The identity establishes that demand for a variety, $c(e_i, p(\omega), \mathbf{p})$, can be derived using an individual's indirect utility, $\nu(e_i, \mathbf{p})$, and its derivatives with respect to the variety's price as

$$c(e_i, p(\omega), \mathbf{p}) \quad = \quad -\frac{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)}{\partial \nu(e_i, \mathbf{p})/\partial e_i}, \tag{A.1}$$

where the indirect utility satisfies the usual properties postulated in Assumption A.1 below. The results that follow require the additional Assumption A.2, which ensures all objects are well defined. In particular, Assumption A.2 (i) is needed to ensure both the price and income elasticities of demand are well defined, while (ii) ensures the pass-through between the price and income elasticities of demand is not degenerate (i.e. $\chi(e_i, p(\omega), \mathbf{p}) \neq 0$) and (iii) ensures the price elasticity of demand is positive (i.e. $\xi(e_i, p(\omega), \mathbf{p}) > 0$). These objects are defined formally below. Proceeding in this fashion will make clear the link between the price elasticity of demand and the income elasticity of demand.

Assumption A.1. **(Indirect utility)** *The indirect utility $\nu(e_i, \mathbf{p})$ is: (i) continuous on $\mathbb{R}^N \times \mathbb{R}$; (ii) decreasing in prices, $\frac{\partial \nu(e_i, \mathbf{p})}{\partial p(\omega)} \leqslant 0$ for all $p(\omega)$; (iii) strictly increasing in income, $\frac{\partial \nu(e_i, \mathbf{p})}{\partial e_i} > 0$; (iv) homogeneous of degree 0 in $(e_i, \mathbf{p})$; (v) quasiconvex in $(e_i, \mathbf{p})$.*

Assumption A.2. **(Differentiability)** *The indirect utility function $\nu(e_i, \mathbf{p})$ is at least twice continuously differentiable and satisfies (i) $\frac{\partial \nu(e_i, \mathbf{p})}{\partial p(\omega)} < 0$ for all $p(\omega)$; (ii) $\frac{\partial^2 \nu(e_i, \mathbf{p})}{\partial e_i \partial p(\omega)} \neq 0$ for all $p(\omega)$; and (iii) $\frac{\partial^2 \nu(e_i, \mathbf{p})}{\partial p(\omega)^2} > 0$.*

Price elasticity of demand. Start with an inidividual's price elasticity of demand, $\xi(e_i, p(\omega), \mathbf{p}) \equiv -\frac{\partial c(e_i, p(\omega), \mathbf{p})}{\partial p(\omega)} \frac{p(\omega)}{c(e_i, p(\omega), \mathbf{p})}$. Using the indirect utility, the price elasticity of demand can be expressed as

$$\xi(e_i, p(\omega), \mathbf{p}) \quad = \quad -p(\omega) \left[\frac{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2}{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)} - \frac{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i \partial p(\omega)}{\partial \nu(e_i, \mathbf{p})/\partial e_i} \right]. \tag{A.2}$$

This expression highlights the different channels through which changes in the price elasticity of demand materialize. Notably, changes in the variety's price, and possibly all other prices, and in the household's income can alter a consumer's price elasticity of demand. In models without strategic interactions, the dependence on competitors' prices does not affect the price elasticity of demand directly. Likewise, in models with homothetic preferences the price elasticity of demand does not depend on the consumer's income. Finally, demand for a variety is said to be inelastic when the price elasticity is less than one (i.e. $\xi(e_i, p(\omega), \mathbf{p}) < 1$): that is, changes in price have a relatively small effect on the quantity demanded (perfectly inelastic if the elasticity is zero). Demand for a variety is said to be elastic when the elasticity is greater than one (i.e. $\xi(e_i, p(\omega), \mathbf{p}) > 1$; perfectly elastic if the elasticity is infinity). Varieties conform to the law of demand as long as $\xi(e_i, p(\omega), \mathbf{p}) \geqslant 0$.

Income elasticity of demand. The income elasticity of demand, $\eta(e_i, p(\omega), \mathbf{p}) \equiv \frac{\partial c(e_i, p(\omega), \mathbf{p})}{\partial e_i} \frac{e_i}{c(e_i, p(\omega), \mathbf{p})}$, measures how demand changes in response to changes in income. Using the consumer's indirect utility, the elasticity is given by

$$\eta(e_i, p(\omega), \mathbf{p}) \quad = \quad e_i \left[\frac{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i \partial p(\omega)}{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)} - \frac{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i^2}{\partial \nu(e_i, \mathbf{p}) \partial e_i} \right]. \tag{A.3}$$

A variety is said to be a luxury for the consumer if the income elasticity is greater than one (i.e. $\eta(e_i, p(\omega), \mathbf{p}) > 1$), a necessity if the elasticity is positive but less than one (i.e. $0 < \eta(e_i, p(\omega), \mathbf{p}) < 1$), and an inferior good if the elasticity is negative (i.e. $\eta(e_i, p(\omega), \mathbf{p}) < 0$).

Income elasticity and super-elasticity of utility. The income elasticity of utility, $\Phi(e_i, \mathbf{p}) \equiv \frac{\partial \nu(e_i, \mathbf{p})}{\partial e_i} \frac{e_i}{\nu(e_i, \mathbf{p})}$, measures how the consumer's utility changes when income changes. Note that this elasticity is common to all varieties and takes into account all the possible interactions across varieties when income changes. As households tend to enjoy more utility if their income grows, $\Phi(e_i, \mathbf{p})$ is usually positive. In turn, the income super-elasticity of utility, $\varphi(e_i, \mathbf{p}) \equiv \frac{\partial \Phi(e_i, \mathbf{p})}{\partial e_i} \frac{e_i}{\Phi(e_i, \mathbf{p})}$, measures how responsive the utility's income elasticity is to changes in the household's income. This super-elasticity can also be written as $\varphi(e_i, \mathbf{p}) = (1 - \Phi(e_i, \mathbf{p})) + e_i \frac{\partial \nu^2(e_i, \mathbf{p})/\partial e_i^2}{\partial \nu(e_i, \mathbf{p})/\partial e_i}$.

Pass-through. The variety's pass-through, $\chi(e_i, p(\omega), \mathbf{p})$, measures the relative strength of the income elasticity of demand and the price elasticity of demand and it is given by

$$\chi(e_i, p(\omega), \mathbf{p}) \quad = \quad -\frac{p(\omega)}{e_i} \frac{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2}{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i \partial p(\omega)}. \tag{A.4}$$

Proposition A.2 below establishes the relationship between the price elasticity of demand and the income elasticity of demand, and is the fundamental mechanism behind the demand channel underlying markups. Often overlooked and obscured by simplifying assumptions, this relationship has important implications for the rise in markups observed in the data.

Proposition A.2. **(PRICE AND INCOME ELASTICITIES OF DEMAND)** *Given Assumptions A.1 and A.2, the price elasticity of demand of individual i for variety ω is related to their income elasticity of demand through the following expression*

$$\underbrace{\xi(e_i, p(\omega), \mathbf{p})}_{\text{price elast of demand}} \quad = \quad \underbrace{\alpha(e_i, p(\omega), \mathbf{p})}_{\text{fixed effect}} + \underbrace{\chi(e_i, p(\omega), \mathbf{p})}_{\text{pass-through}} \left[\underbrace{\eta(e_i, p(\omega), \mathbf{p})}_{\text{income elast of demand}} + \underbrace{(\Phi(e_i, \mathbf{p}) + \varphi(e_i, \mathbf{p}))}_{\substack{\text{income elast of utility} \\ \text{and super-elast}}} - 1 \right],$$

where $\alpha(e_i, p(\omega), \mathbf{p}) = p(\omega) \frac{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i \partial p(\omega)}{\partial \nu(e_i, \mathbf{p})/\partial e_i}$ is a variety-specific fixed effect (that is approximately 0).

Proof. Rearrange equation (A.2) to have

$$\frac{1}{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)} \quad = \quad \frac{1}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} \left[\frac{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i \partial p(\omega)}{\partial \nu(e_i, \mathbf{p})/\partial e_i} - \frac{\xi(e_i, p(\omega), \mathbf{p})}{p(\omega)} \right]$$

and rewrite the income super-elasticity of utility $\varphi(e_i, \mathbf{p})$ as

$$\varphi(e_i, \mathbf{p}) \quad = \quad -\left[1 + \Phi(e_i, \mathbf{p}) + e_i \frac{\partial^2 \nu(e_i, \mathbf{p})/\partial {e_i}^2}{\partial \nu(e_i, \mathbf{p})/\partial e_i}\right].$$

Next, plug these in equation (A.3) to have the income elasticity of demand as

$$\eta(e_i, p(\omega), \mathbf{p}) \quad = \quad 1 + \Phi(e_i, \mathbf{p}) + \varphi(e_i, \mathbf{p}) + \frac{[\xi(e_i, p(\omega), \mathbf{p}) - \alpha(e_i, p(\omega), \mathbf{p})]}{\chi(e_i, p(\omega), \mathbf{p})}.$$

Rearranging this equation gives the result in the proposition, i.e.

$$\xi(e_i, p(\omega), \mathbf{p}) \quad = \quad \alpha(e_i, p(\omega), \mathbf{p}) + \chi(e_i, p(\omega), \mathbf{p}) \left[\eta(e_i, p(\omega), \mathbf{p}) + (\Phi(e_i, \mathbf{p}) + \varphi(e_i, \mathbf{p})) - 1\right]$$

If the price elasticity of demand is instead defined as $\xi(e_i, p(\omega), \mathbf{p}) = -p(\omega)\frac{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2}{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)}$, we have that

$$\alpha(e_i, p(\omega), \mathbf{p}) - p(\omega)\frac{\partial^2 \nu(e_i, \mathbf{p})/\partial e_i \partial p(\omega)}{\partial \nu(\partial p(\omega))/\partial e_i} \quad = \quad 0,$$

which defines the variety-specific fixed effect. In that case, the relationship between the price and income elasticities of demand is simply

$$\xi(e_i, p(\omega), \mathbf{p}) \quad = \quad \chi(e_i, p(\omega), \mathbf{p}) \left[\eta(e_i, p(\omega), \mathbf{p}) + (\Phi(e_i, \mathbf{p}) + \varphi(e_i, \mathbf{p})) - 1\right].$$

□

Remark A.1. *It is common to drop the variety-specific fixed effect term, $\alpha(e_i, p(\omega), \mathbf{p})$, and define the price elasticity of demand as $\xi(e_i, p(\omega), \mathbf{p}) = -p(\omega)\frac{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2}{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)}$. In that case, the relationship still holds with only a minor change, i.e. the fixed effect is dropped and*

$$\xi(e_i, p(\omega), \mathbf{p}) \quad = \quad \chi(e_i, p(\omega), \mathbf{p}) \left[\eta(e_i, p(\omega), \mathbf{p}) + (\Phi(e_i, \mathbf{p}) + \varphi(e_i, \mathbf{p})) - 1\right]. \tag{A.5}$$

Without loss of generality, that is the the definition of the price elasticity of demand used hereafter.

A.2 Price elasticity of demand: Two key results

For the two forces of structural change, namely differential rates of technological progress across sectors and income effects, to alter the price elasticity of demand and therefore markups, two additional results are needed. Assumptions A.3 and A.4 provides additional conditions for the results to go through.

Assumption A.3. **(INDIRECT UTILITY AND INCOME)** *The indirect utility $\nu(e_i, \mathbf{p})$ is at least thrice continuously differentiable with $\frac{\partial^3 \nu(e_i, \mathbf{p})/\partial p(\omega)^2 \partial e_i}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} < \frac{\xi(e_i, p(\omega), \mathbf{p})}{\chi(e_i, p(\omega), \mathbf{p})\, e_i}$.*

Assumption A.4. **(INDIRECT UTILITY AND PRICE)** *The indirect utility $\nu(e_i, \mathbf{p})$ is at least thrice continuously differentiable with $\frac{\partial^3 \nu(e_i, \mathbf{p}) \partial p(\omega)^3}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} > -\frac{(1+\xi(e_i, p(\omega), \mathbf{p}))}{p(\omega)}$.*

Two results follow. First, the price elasticity of demand must be decreasing in the consumer's income, which sustains Harrod's (1936) *Law of Diminishing Elasticity of Demand*. This property has the following effect: as households' income increases over time and people start shifting their consumption basket toward luxuries, i.e. services, their price elasticity of demand decreases. This in turn allows firms to charge a higher markup. Proposition A.3 formalizes that intuition.

Proposition A.3. **(PRICE ELASTICITY OF DEMAND ACROSS INCOME)** *Under Assumptions A.1, A.2, and A.3, the price elasticity of demand for a variety ω is decreasing in the consumer's income.*

Proof. The derivative of a consumer's price elasticity of demand for a variety ω with respect to her income is given by

$$\frac{\partial \xi(e_i, p(\omega), \mathbf{p})}{\partial e_i} = \xi(e_i, p(\omega), \mathbf{p}) \left[\frac{\partial^3 \nu(e_i, \mathbf{p})/\partial p(\omega)^2 \partial e_i}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} - \frac{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega) \partial e_i}{\partial \nu(e_i, \mathbf{p})/\partial p(\omega)} \right].$$

The expression in the square brackets must be negative for the price elasticity of demand for a variety ω to be decreasing in the consumer's income, e_i. Recall that under Assumption A.2, the following holds: (i) $\frac{\partial \nu(e_i,\mathbf{p})}{\partial p(\omega)} < 0$ for all $p(\omega)$; (ii) $\frac{\partial^2 \nu(e_i,\mathbf{p})}{\partial e_i \partial p(\omega)} \neq 0$ for all $p(\omega)$; and (iii) $\frac{\partial^2 \nu(e_i,\mathbf{p})}{\partial p(\omega)^2} > 0$. Using the definition of the pass-through (equation (A.4)) and rearranging the term in the square brackets implies the result in the proposition, i.e.,

$$\frac{\partial^3 \nu(e_i, \mathbf{p})/\partial p(\omega)^2 \partial e_i}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} < \frac{\xi(e_i, p(\omega), \mathbf{p})}{\chi(e_i, p(\omega), \mathbf{p}) e_i}.$$

□

Second, the price elasticity of demand must be increasing in the variety's price in line with Marshall's (1890) *Second Law of Demand*. Stronger technological progress in the manufacturing sector will decrease the marginal costs of goods faster than that of services. This allows firms in the manufacturing sector to decrease their prices. However, if firms have some market power, the cost pass-through is not one-to-one and the firm is able to capture some of the efficiency gains. As consumers' price elasticities of demand have now increased, firms will be able to retain some of those gains and therefore charge higher markups. Proposition A.4 formalizes that intuition.

Proposition A.4. **(PRICE ELASTICITY OF DEMAND ACROSS PRICE)** *Under Assumptions A.1, A.2, and A.4, the price elasticity of demand for a variety ω is increasing in the variety's price.*

Proof. The derivative of a consumer's price elasticity of demand for a variety ω with respect to its price is given by

$$\frac{\partial \xi(e_i, p(\omega), \mathbf{p})}{\partial p(\omega)} = \frac{\xi(e_i, p(\omega), \mathbf{p})}{p(\omega)} \left[1 + p(\omega) \frac{\partial^3 \nu(e_i, \mathbf{p})/\partial p(\omega)^3}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} + \xi(e_i, p(\omega), \mathbf{p}) \right].$$

The expression in the square brackets must be positive for the price elasticity of demand for a variety ω to be increasing in its price, $p(\omega)$. Recall that under Assumption A.2, $\frac{\partial^2 \nu(e_i,\mathbf{p})}{\partial p(\omega)^2} > 0$. Rearranging the term in the square brackets implies the result in the proposition, i.e.,

$$\frac{\partial^3 \nu(e_i, \mathbf{p})/\partial p(\omega)^3}{\partial^2 \nu(e_i, \mathbf{p})/\partial p(\omega)^2} < -\frac{(1 + \xi(e_i, p(\omega), \mathbf{p}))}{p(\omega)}.$$

□

Given Propositions A.3 and A.4, firms' markups will be higher the lower the price of the variety they sell and/or the wealthier their consumers are. This has implications both for the cross-sectional distribution of markups and for the distribution of markups over time. If Proposition A.3 holds and households are heterogeneous in terms of income or wealth, then changes in the composition of demand has an effect on markups—even if each household's price elasticity of demand does not change. For instance, if the demand share from wealthier consumers increases, firms best respond by charging higher markups. In addition, if consumers become wealthier over time, firms respond by charging higher markups. Similarly, firms that are more productive will be able to sell their varieties at lower prices, which in turn will allow them to charge higher markups. These considerations will be featured in the quantitative model in the following section.

B Eliciting Demand Elasticities

Do price elasticities of demand vary across the income distribution? Section A provided the theoretical underpinnings for this to hold and Sections 4 and 5 showed that the mechanism is quantitatively important for explaining the rise of markups. The section below proposes a novel strategy to address this question by conducting a new online survey covering 607 consumers in the United States.[30] The survey questions are designed to capture individuals' perception of the impact of changes in prices on their purchase of different goods and services. The categories of goods and services follow the structure of the Bureau of Labor Statistics' Consumer Expenditure Survey (CEX).

B.1 Survey design

Recruitment of survey participants. Survey participants were selected through ResearchMatch, a platform developed by the National Institutes of Health (NIH) to help connect potential survey participants with researchers. A first message was sent in March 2022 through ResearchMatch to a pool of registered volunteers, age 18 and above who reside in the United States and can read English. Of the 37,497 individuals contacted, 1,765 volunteers showed interest in the study and agreed to receive more information about the survey. Of those who showed interest, 607 individuals completed the survey between March and May 2022. Although 812 individuals responded to the survey, 205 responses were discarded either because they were duplicates, were filled out too quickly, or only responded to the demographic questions. The survey could be completed online using a smartphone or a computer. The average time to complete the survey was 54 minutes and the median time was 14 minutes. Participants did not receive any compensation from participating in the study and were free to withdraw at any point without any consequences.

Survey questions. The questionnaire consists of three main sections. The first section collects detailed data on the socio-economic background of the individual and includes questions about gender, age, race, educational attainment, relationship status, household composition, zipcode of residence,

[30]The survey was approved by the University of Pennsylvania's Institutional Review Board (IRB) in February 2022.

yearly household income, home and vehicle ownership, employment status, occupation, and industry of employment.

The second section covers average spending per month on: (1) food at home, (2) food away from home, (3) alcoholic beverages, (4) mortgage payments and rent, (5) home insurance, (6) utilities, (7) housekeeping expenses, (8) apparel, (9) gasoline, (10) public transportation, (11) vehicle insurance, (12) medical and dental services, drugs and medical supplies, (13) health insurance, (14) child care, preschool tuition, and related expenses or care of elderly, (15) school and college tuition and related expenses, (16) personal insurance and pensions, (17) tobacco and other smoking products. It also asks about the average spending per year on: (18) house maintenance and repairs, (19) other lodging expenses out of town, (20) furniture, (21) household appliances, (22) audio and visual equipment and services, (23) entertainment fees and admissions, hobbies, pets, and toys, (24) vehicle purchases, maintenance and repairs, leases and rental charges, (25) other expenses. These spending categories follow the structure of the CEX.

The goal of the third section is to gauge each person's perceived own price elasticity of demand. The question posits the following scenario for the different categories of goods and services defined above *"Suppose you spent \$$x$ on the following items in any given y. If the same items you purchased in the past now cost \$1.2x, how much would you now be willing to spend (US\$)?"*.[31] The respondent has five options, ranging from \$0.9x to \$1.3x. These values correspond to a price elasticity of demand greater than 1, equal to 1, between 0 and 1, equal to 0, and negative, respectively.[32]

Sample. Table B.1 shows key statistics regarding the characteristics of the final sample of 607 individuals who filled out the questionnaire relative to the U.S. population. Population statistics are taken from the 2019 Current Population Survey. Each response is weighted through an iterative proportional raking procedure that minimizes the difference between the sample and the known population statistics along different demographic dimensions. The resampling procedure aims to make the survey representative of the U.S. population in terms of gender, age distribution, race, educational attainment, marital status, employment status, home ownership status, and household income. As the table shows, the sample is representative of the U.S. population.

B.2 Price elasticities of demand

Table B.2 shows the distribution of price elasticities of demand for the 24 categories of goods and services, together with the sample average spending per year.[33] For instance, for more than half of the sample, the price elasticity of demand for mortgage payments and rent is zero. This implies that most consumers would not switch homes if their rent or mortgage payments increase. Similarly, very few individuals would change their demand for health or personal insurance if their premiums

[31] Here $\{x, y\} = \{100, \text{week}\}$ for regular purchases, $\{x, y\} = \{1{,}000, \text{month}\}$ for bigger purchases, and $\{x, y\} = \{10{,}000, \text{year}\}$ for irregular big-ticket items.

[32] The survey included three additional sections. A fourth section aims to gather information about the average quality of goods and services purchased by the consumer, a fifth section focuses on the income elasticity of demand, and a last section allows respondents to provide comments.

[33] Here, the price elasticity of demand is defined as the negative change in consumption given a change in price, in line with the definition used in the previous sections.

Table B.1: Sample characteristics

	Sample (%)	Population (%)
Female	51.93	51.64
18-25 y.o.	10.90	10.96
25-35 y.o.	17.92	17.91
35-45 y.o.	16.60	16.61
45-55 y.o.	16.29	16.26
55-65 y.o.	17.02	17.01
White	74.37	74.22
Black	12.24	12.28
Asian	5.91	5.99
No college degree	67.98	68.21
Bachelor's degree	20.06	19.95
Married	51.20	51.07
Single	29.56	29.62
Employed	63.46	63.47
Unemployed	2.84	2.85
Owner with mortgage	43.96	43.95
Owner without mortgage	25.93	25.92
Household income < 40k	22.90	22.88
Household income $\in$ [40k, 80k)	27.68	27.66
Household income $\in$ [80k, 120k)	20.10	20.09

Note: The table displays summary statistics for the sample and the U.S. population.
Source: Statistics for the U.S. population are taken from the 2019 Current Population Survey.

increase, nor change their consumption of food at home if the price of groceries increases. In contrast, the demand for furniture, appliances, audio and visual equipment and services, alcohol is elastic for most households.

Estimation results. The main empirical exercise is based on the following linear probability model

$$\xi_{ij} = \alpha + \beta e_i + \gamma z_i + \epsilon_{ij}, \tag{B.1}$$

where ξ_{ij} is a dummy capturing respondent i's perception of their price elasticity of demand for product j, e_i is the household's income, and z_i is a set of demographic characteristics. Two cases are considered for the dummy of the price elasticity of demand. The first case is where ξ_{ij} equals 1 if the individual's demand is elastic, i.e., the respondent answered she would reduce her consumption if prices increased, and 0 otherwise. The second case is where ξ_{ij} equals 1 if the individual's demand is inelastic, i.e., the respondent answered they would consume the same amount despite the price increase, and 0 otherwise. The set of demographic characteristics includes the age of the respondent, their employment status, gender, household size, industry of employment, occupation, race, and relationship status. Observations are scaled by their survey weight. For robustness, a probit model is also estimated.

Table B.3 presents the estimated coefficient capturing the income effect on the price elasticity of demand (β in equation (B.1) above) for each category of goods and services (each column) for both the linear probability and probit models (only statistically significant estimates are shown). The upshot is that demand is more likely to be elastic for lower-income households and thus those households are more likely to reduce their consumption when prices increase, in particular for food at home (column 1), food away (2), apparel (3), public transportation (4), vehicle insurance (5), medical and

Table B.2: Distribution of price elasticities of demand, ξ, by categories

Category	Avg. spend. (yearly, US$)	% with $\xi = 0$	% with $\xi \in (0,1)$	% with $\xi = 1$	% with $\xi > 1$
Mortgage payments and rent	13,747	50.54	4.79	20.53	13.46
Health insurance	5,065	42.30	15.11	20.09	11.61
Personal insurance (e.g. life insurance, accident and disability)	5,032	34.24	13.47	24.69	21.57
Food at home (e.g. cereals, meats, dairy products, fruits, vegetables)	4,704	46.00	13.15	17.64	10.22
Utilities (electricity, natural gas, water and trash collection, telephone/mobile services)	4,694	41.22	16.63	21.64	7.72
Vehicle insurance	3,918	41.21	16.77	16.31	17.63
Medical and dental services, drugs and medical supplies	3,629	34.53	16.42	23.16	16.93
Gasoline	3,490	37.72	16.76	20.77	13.31
Home insurance	2,717	43.46	13.13	17.48	11.95
Food away (e.g. fast food, take-out, delivery, full-service restaurants, excl. alcoholic bev.)	2,036	16.04	7.33	34.02	34.92
Housekeeping expenses (e.g. cleaning, post., stationary, garden., pest control, storage)	1,893	19.40	15.70	30.68	27.22
Tuition and related expenses (e.g. elementary, high-school, college, books, supplies)	1,832	24.93	11.96	38.44	15.92
Apparel (e.g. men, women, boys, girls, footware, watches, jewelry)	1,616	11.49	7.88	31.21	43.92
Vehicle purchases, maintenance and repairs, leases and rental charges	1,585	20.62	17.76	28.04	26.81
Other lodging expenses out of town (e.g. hotels, vacation homes)	1,493	14.74	7.74	26.34	44.31
House maintenance and repairs	1,295	22.97	7.98	35.89	26.29
Child care, preschool tuition, or care of the elderly	1,110	29.03	7.65	18.50	27.86
Entertainment fees and admissions, hobbies, pets, and toys	879	11.43	13.33	28.92	42.01
Audio and visual equipment and services (e.g. tvs, smartphones, cable, musical instr.)	613	8.33	12.58	23.83	49.96
Alcohol (at home or away from home)	602	12.72	9.68	22.52	47.54
Tobacco and other smoking products	484	4.74	19.32	21.35	32.21
Furniture (indoor, outdoor, floor coverings)	436	9.25	8.38	28.43	50.54
Appliances (e.g. refrigerators, dishwashers, ovens, vacuum cleaners, air-conditioners)	382	12.24	8.43	24.57	50.43
Public transportation (e.g. mass transit, buses, trains, airlines, taxis, school buses)	313	23.06	18.84	10.70	36.09

Note: The price elasticities of demand, ξ, are for individuals who reported positive expenditures on that category.

dental services, drugs and medical supplies (6), health insurance (7), child care, preschool tuition, or care of elderly (8), school and college tuition and related expenses (9), personal insurance (10), audio and visual equipment and services (14), and vehicle purchases, maintenance and repairs, leases and rental charges (15).

Figure B.1 displays the predicted probability of adjusting demand for (a) child care, preschool tuition, or care of elderly, (b) food away, and (c) vehicle purchases, maintenance and repairs, leases and rental charges, along the household income distribution based on the probit estimation for households with expenses in that category and the set of controls used in equation (B.1). In response to a price increase for child care, 87% of households whose income is between $50,000 and $60,000 would reduce their demand as opposed to less than 38% for households earning between $150,000 and $200,000. In response to a price increase of fast food, take-out, delivery, or full-service restaurants, 85% of households whose income is between $50,000 and $60,000 would reduce their demand as opposed to less than 59% for households earning between $150,000 and $200,000. In response to a price increase of vehicles, maintenance and repairs, or leases and rental charges, 81% of households whose income is between $50,000 and $60,000 would reduce their demand as opposed to less than 52% for households earning between $150,000 and $200,000.

Similarly, households who are more likely not to adjust their consumption level when prices rise are wealthier, as Table B.3 shows. For them, the demand for food at home (column 1), food away (2), apparel (3), personal insurance (10), appliances (13), and vehicle purchases, maintenance and repairs, leases and rental charges (15) is perfectly inelastic. The only exception is the demand for tobacco and other smoking products (11). In that case, less well-off households are more likely not to adjust their consumption level if prices were to increase.

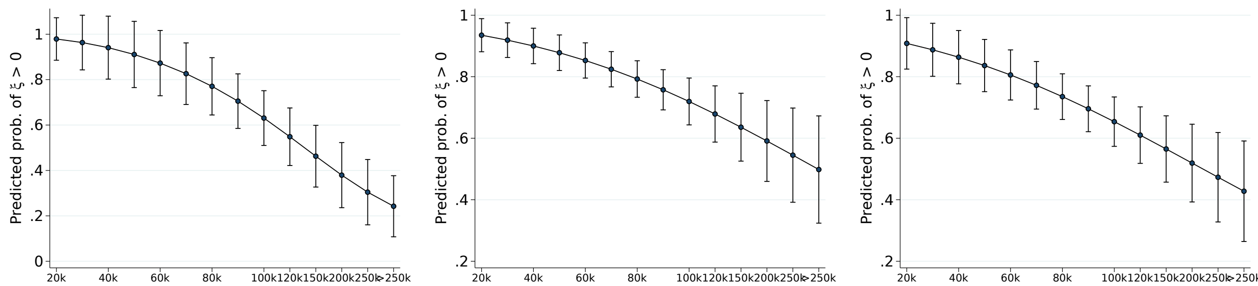
Figure B.2 presents the predicted probability of *not* adjusting demand for (a) tobacco and other smoking products, (b) food away, and (c) vehicle purchases, maintenance and repairs, leases and rental charges, along the household income distribution based on the probit estimation for households with expenses in that category and the set of controls used in equation (B.1). For instance, in response to a price increase of tobacco, only 1% of households earning between $150,000 and $200,000 would not change their demand for tobacco, while 31% of households whose income is between $50,000 and $60,000 would consume the same amount of tobacco in spite of the price hike.

Table B.3: Effect of income on price elasticity of demand, $\widehat{\beta}$

Specification	Price elasticity of demand							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Elastic (LPM)	-0.020*	-0.034***	-0.014*	-0.037**	-0.024**	-0.027**	-0.027*	-0.079**
	(0.011)	(0.010)	(0.008)	(0.014)	(0.012)	(0.012)	(0.014)	(0.034)
Observations	581	555	476	86	513	444	428	52
Adjusted R^2	0.059	0.126	0.190	0.478	0.144	0.115	0.120	0.465
Elastic (Probit)	-0.053*	-0.125***	-0.083**	-0.177***	-0.073**	-0.079**	-0.077**	-0.386***
	(0.029)	(0.033)	(0.036)	(0.050)	(0.032)	(0.031)	(0.038)	(0.121)
Observations	581	555	476	86	513	444	428	52
Adjusted R^2	0.044	0.125	0.207	0.446	0.121	0.094	0.093	0.458
Inelastic (LPM)	0.019*	0.021***	0.014**	0.014	0.015	0.017	0.015	0.029
	(0.010)	(0.008)	(0.007)	(0.010)	(0.011)	(0.012)	(0.013)	(0.027)
Observations	581	555	476	86	513	444	428	52
Adjusted R^2	0.040	0.111	0.108	0.521	0.129	0.074	0.082	0.218
Inelastic (Probit)	0.051*	0.105***	0.111**	0.098*	0.049	0.056*	0.043	-0.042
	(0.028)	(0.033)	(0.044)	(0.053)	(0.031)	(0.033)	(0.036)	(0.105)
Observations	581	555	476	86	513	444	428	46
Adjusted R^2	0.030	0.154	0.198	0.521	0.114	0.067	0.066	0.188
	(9)	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	
Elastic (LPM)	-0.032**	-0.054***	-0.012	-0.019	-0.016	-0.028***	-0.037***	
	(0.015)	(0.013)	(0.016)	(0.012)	(0.011)	(0.009)	(0.011)	
Observations	105	270	76	370	279	383	464	
Adjusted R^2	0.300	0.162	0.607	0.168	0.153	0.135	0.143	
Elastic (Probit)	-0.120**	-0.160***	-0.045	-0.067*	-0.069*	-0.128***	-0.125***	
	(0.059)	(0.042)	(0.066)	(0.040)	(0.042)	(0.038)	(0.035)	
Observations	105	270	76	370	279	383	464	
Adjusted R^2	0.260	0.130	0.620	0.159	0.171	0.161	0.134	
Inelastic (LPM)	0.009	0.034***	-0.016**	0.004	0.012*	0.004	0.024**	
	(0.013)	(0.012)	(0.008)	(0.007)	(0.007)	(0.005)	(0.009)	
Observations	105	270	76	370	279	383	464	
Adjusted R^2	0.342	0.144	0.163	0.118	0.151	0.085	0.068	
Inelastic (Probit)	0.051	0.117***	-0.294***	0.019	0.104**	0.034	0.094***	
	(0.060)	(0.044)	(0.088)	(0.039)	(0.050)	(0.044)	(0.032)	
Observations	105	270	54	370	279	383	464	
Adjusted R^2	0.321	0.132	0.385	0.206	0.325	0.171	0.075	

Note: The estimated coefficients are for the effect of household income on the price elasticity of demand for each specification, $\widehat{\beta}$. *Elastic* stands for the case in which the price elasticity of demand is positive, while *Inelastic* stands for the case in which the price elasticity of demand is equal to 0. The regressions are estimated for each category of goods and services separately and include the following set of controls: age, employment status, gender, household size, industry, occupation, race, relationship status. Each column is for a category: (1) food at home; (2) food away; (3) apparel; (4) public transportation; (5) vehicle insurance; (6) medical and dental services, drugs and medical supplies; (7) health insurance; (8) child care, preschool tuition, or care of elderly; (9) school and college tuition and related expenses; (10) personal insurance; (11) tobacco and other smoking products; (12) other lodging expenses out of town; (13) appliances; (14) audio and visual equipment and services; (15) vehicle purchases, maintenance and repairs, leases and rental charges. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

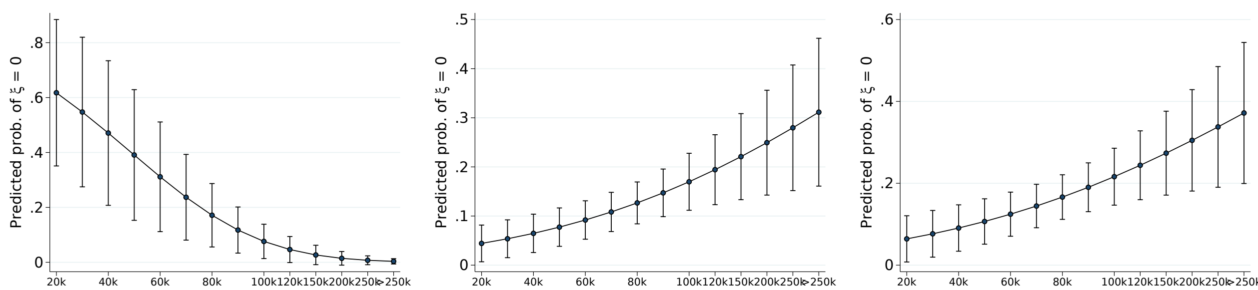
Figure B.1: Who is more likely to adjust demand in response to a price increase?



(a) Child care (b) Food away (c) Vehicle purchases

Note: Panel (a) shows the probability of adjusting child care demand in response to a price increase along different household income levels, based on the probit estimation for households with expenses in that category and the set of controls. Panel (b) shows the same response for food away. Panel (c) shows the same response for vehicle purchases, maintenance and repairs, leases and rental charges.

Figure B.2: Who is more likely *not* to adjust demand in response to a price increase?



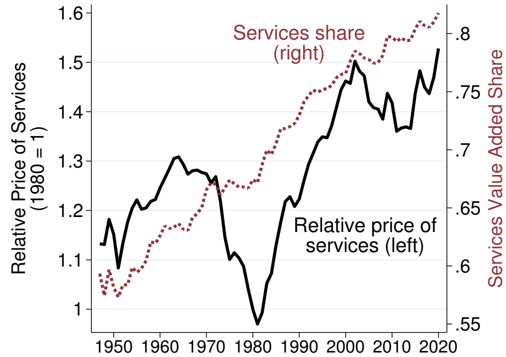
(a) Tobacco (b) Food away (c) Vehicle purchases

Note: Panel (a) shows the probability of *not* adjusting tobacco and other smoking products demand in response to a price increase along different household income levels, based on the probit estimation for households with expenses in that category and the set of controls. Panel (b) shows the same response for food away. Panel (c) shows the same response for vehicle purchases, maintenance and repairs, leases and rental charges.

C Empirics

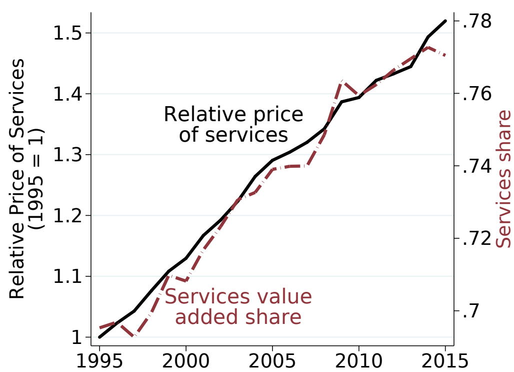
C.1 Additional figures

Figure C.1.1: Structural change in the U.S. starting in 1947

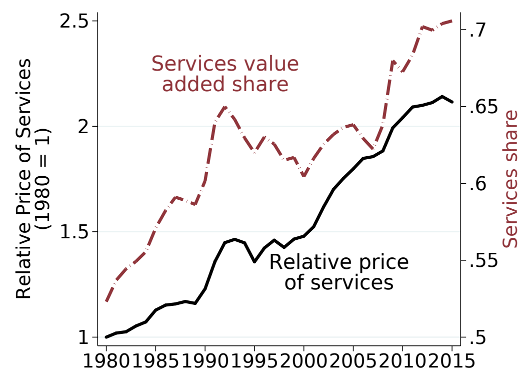


Note: The figure shows the services value added share (red) and the relative value added price of services (black) for the United States starting in 1947.

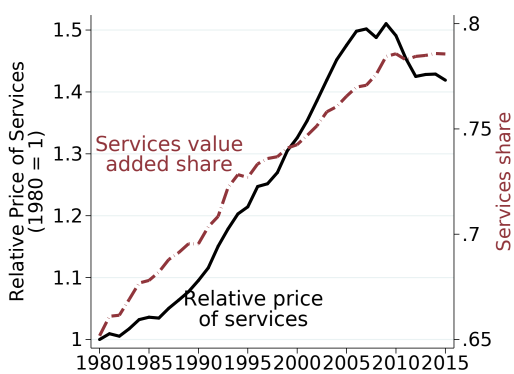
Figure C.1.2: Structural change across European countries



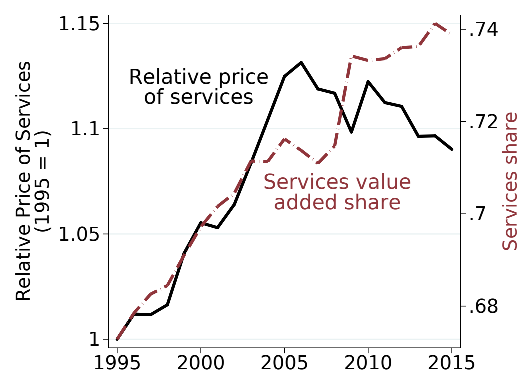
(a) Belgium



(b) Finland



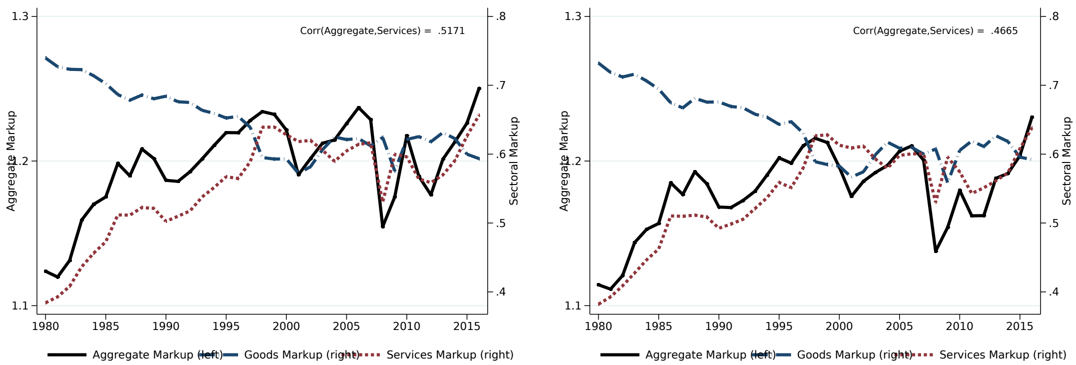
(c) France



(d) Italy

Note: The figure shows the services value added share (red) and the relative value added price of services (black) for Belgium (a), Finland (b), France (c), and Italy (d).

Figure C.1.3: Markups and superstars

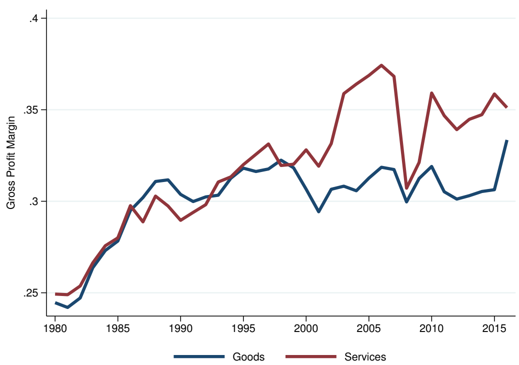


(a) Dropping top 5th percentile

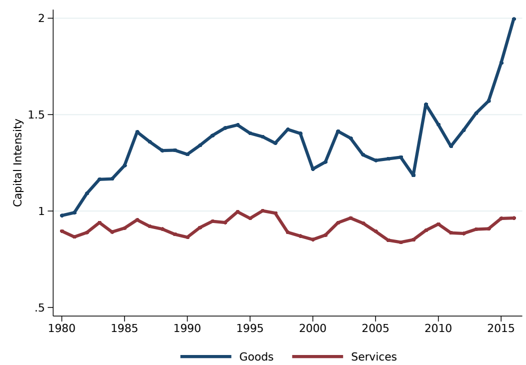
(b) Dropping top 10th percentile

Note: Panel (a) shows the aggregate markup (black) and its decomposition in terms of services (red) vs. non-services (blue) when firms in the top 5th percentile are removed from the sample. Panel (b) shows the same trends when firms in the top 10th percentile are removed from the sample.

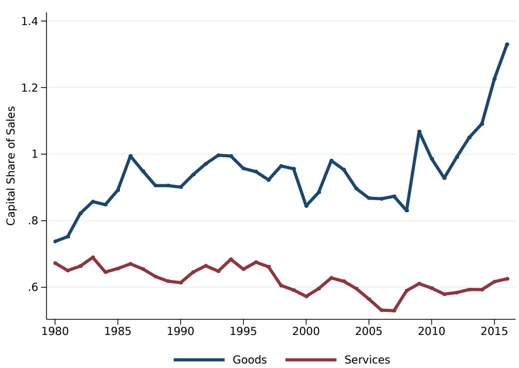
Figure C.1.4: Firm characteristics across sectors



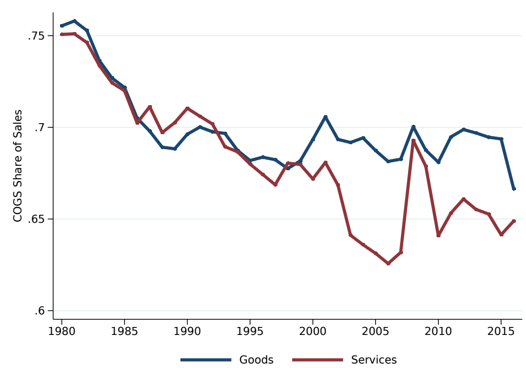
(a) Gross profit margin



(b) Capital intensity



(c) Capital share of sales



(d) Cost of goods sold share of sales

Note: The figure shows the gross profit margin, capital intensity, capital share of sales, and cost of goods sold as a share of sales for the services sector (red) and the non-services sector (blue).

C.2 Additional tables

Table C.2.1: Firms' markups and their characteristics

	Firm-level markups (in logs)	
	Non-services	**Services**
	(1)	(2)
Capital share	0.028***	0.017***
	(0.007)	(0.005)
Intangible capital share	0.146***	0.009
	(0.023)	(0.010)
Cogs share	-1.072***	-1.228***
	(0.093)	(0.055)
Fixed cost share	0.126*	0.156***
	(0.074)	(0.028)
Time FE	Yes	Yes
Observations	97,351	74,197
Adjusted R^2	0.732	0.847

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Firms are weighted by their cost shares. Data is taken from Compustat for the 1980-2015 period.

D Quantitative Analysis

D.1 Assumptions on the indirect subutility

Assumption D.1. (SECTORAL INDIRECT SUBUTILITY) *The sector-specific indirect subutility $v_j(e, p_j, q_j)$ satisfies the standard properties of indirect utilities, namely: $v_j(e, p_j, q_j)$ is continuous on $\mathbb{R}^3$; decreasing in prices, $\frac{\partial v_j}{\partial p_j(\omega)} \leqslant 0$; strictly increasing in income, $v'_{j_e} > 0$; homogeneous of degree 0 in $(e, p_j(\omega))$; convex, and hence quasiconvex, in $(e, p_j(\omega))$ up to a choke price, which is the the maximum willingness to pay for each variety of commodity j (common to all households and possibly infinite). For any price above that choke price, the indirect subutility is such that $v_j = v'_{j_p} = v'_{j_e} = 0$ (and it is thus assumed that $v_j > 0$ for any price below). It is further assumed that v_j is at least thrice differentiable, with $v''_{j_{p,p}} > 0$, $v''_{j_{p,w}} < 0$, $v'''_{j_{p,p,p}} < -\frac{v''_{j_{p,p}}}{p_j(\omega)} < 0$, and $v'''_{j_{p,p,w}} < 0$, which ensures that the price elasticity of demand is positive and that commodities consumed conform to the law of demand.*

D.2 Proof of Proposition 3.1 (From the indirect to the direct utility)

Start from the household's consumption demand for some variety ω of commodity $j \in \{G, S\}$ using Roy's identity

$$c_{j_t}(\omega) \quad = \quad -\frac{\left(\partial \hat{v}_j\left(e_t, p_{j_t}(\omega), q_{j_t}(\omega)\right)/\partial p_{j_t}(\omega)\right)\ e_{i_t}}{\left(v_j(e_t, \mathbf{p}_{j_t}, \mathbf{q}_{j_t})/\lambda_j\right)\ \Phi_t(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})},$$

where $\partial\widehat{v}_j\left(e_t, p_{j_t}(\omega), q_{j_t}(\omega)\right)/\partial p_{j_t}(\omega) = -\frac{1}{e_t}\left(\frac{\phi_j e_t - p_{j_t}(\omega)}{e_t}\right)^{\gamma_j} q_{j_t}(\omega)^{\delta_j(1+\gamma_j)}$. Rearrange this expression to write

$$\left(\frac{\phi_j e_t - p_{j_t}(\omega)}{e_t}\right)^{1+\gamma_j} = -\left[\frac{v_j(e_t, \mathbf{p}_{j_t}, \mathbf{q}_{j_t})\,\Phi_t(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})\, c_{j_t}(\omega)}{\lambda_j}\right]^{\frac{1+\gamma_j}{\gamma_j}} q_{j_t}(\omega)^{-\frac{\delta_j(1+\gamma_j)^2}{\gamma_j}}.$$

Use this in the sectoral indirect utility (equation (3.3)) to write it as a function of consumption and quality. This results in the following sectoral *direct* utility u_j given by

$$u_j(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})^{-\frac{1}{\gamma_j}} = -\frac{1}{1+\gamma_j}\left[\frac{\Phi_t(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})\,\widetilde{C}_{j_t}}{\lambda_j}\right]^{\frac{1+\gamma_j}{\gamma_j}},$$

where $\widetilde{C}_{j_t} = \left(\int_0^{N_{j_t}}\left[\frac{c_{j_t}(\omega)}{q_{j_t}(\omega)^{\delta_j}}\right]^{\frac{1+\gamma_j}{\gamma_j}} d\omega\right)^{\frac{\gamma_j}{1+\gamma_j}}$. To derive an expression of the income elasticity of the indirect utility, Φ, as a function of consumption, use the equations above to write the consumption spending share on commodity j as

$$\frac{\int_0^{N_{j_t}} p_{j_t}(\omega)c_{j_t}(\omega)d\omega}{e_t} = \phi_j\int_0^{N_{j_t}} c_{j_t}(\omega)d\omega + \left[\frac{u_j(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})\,\Phi_t(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})}{\lambda_j}\right]^{\frac{1}{\gamma_j}} \widetilde{C}_{j_t}^{\frac{1+\gamma_j}{\gamma_j}}.$$

Note that from the definition of the direct utility above, we have

$$\frac{\int_0^{N_{j_t}} p_{j_t}(\omega)c_{j_t}(\omega)d\omega}{e_t} = \phi_j\int_0^{N_{j_t}} c_{j_t}(\omega)d\omega - (1+\gamma_j)\frac{\lambda_j}{\Phi_t(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})}.$$

Next, use that expression in the budget constraint to get

$$\widehat{C}_t - \frac{1}{\Phi_t(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t})}\left[(1+\gamma_G)\lambda_G + (1+\gamma_S)\lambda_S\right] = 1,$$

where $\widehat{C}_t = \phi_G\int_{\Omega_{G_t}} c_{G_t}(\omega)d\omega + \phi_S\int_{\Omega_{S_t}} c_{S_t}(\omega)d\omega$. Now that an expression for Φ was obtained as a function of consumption and parameters, we can replace it in the definition of the sectoral direct utility according to

$$u_j(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t}) = -\frac{1}{1+\gamma_j}\left[\frac{\lambda_j(1+\gamma_j)}{\lambda_G(1+\gamma_G)+\lambda_S(1+\gamma_S)}\right]^{1+\gamma_j}\left(\frac{\widehat{C}_t - 1}{\widetilde{C}_{j_t}}\right)^{1+\gamma_j}.$$

To get the direct utility, aggregate the two sectoral direct utilities using the Cobb-Douglas weights λ_G and λ_S according to

$$u(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t}) = \psi\left(\frac{\widehat{C}_t - 1}{\widetilde{C}_{G_t}}\right)^{\lambda_G(1+\gamma_G)}\left(\frac{\widehat{C}_t - 1}{\widetilde{C}_{S_t}}\right)^{\lambda_S(1+\gamma_S)},$$

where $\psi = (1+\gamma_G)^{-\lambda_G}(1+\gamma_S)^{-\lambda_S}\left[\frac{\lambda_G(1+\gamma_G)}{\lambda_G(1+\gamma_G)+\lambda_S(1+\gamma_S)}\right]^{\lambda_G(1+\gamma_G)}\left[\frac{\lambda_S(1+\gamma_S)}{\lambda_G(1+\gamma_G)+\lambda_S(1+\gamma_S)}\right]^{\lambda_S(1+\gamma_S)}$.

D.3 Proof of Proposition 3.2 (Two-sector CES)

Assume $\phi_j = 0$, $\gamma_j < -1$, and $\delta_j < 0$ for $j = \{G, S\}$. Denote the elasticity of substitution across varieties by $\vartheta_j = -\gamma_j$ and let P_{j_t} denote the sectoral ideal price index given by

$$P_{j_t} = \left[\int_0^{N_{j_t}} p_{j_t}(\omega)^{1-\vartheta_j} q_{j_t}(\omega)^{\delta(1-\vartheta_j)} d\omega\right]^{\frac{1}{1-\vartheta_j}}.$$

The direct utility is then given by

$$u(\mathbf{c}_{G_t}, \mathbf{c}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t}) = \psi \widetilde{C}_{G_t}^{\lambda_G(\vartheta_G - 1)} \widetilde{C}_{S_t}^{\lambda_S(\vartheta_S - 1)}$$

and the indirect utility by

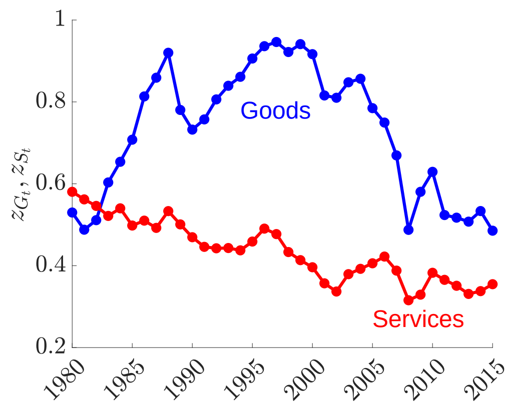
$$\begin{aligned} v(e_t, \mathbf{p}_{G_t}, \mathbf{p}_{S_t}, \mathbf{q}_{G_t}, \mathbf{q}_{S_t}) = & \left[\int_0^{N_{G_t}} \frac{1}{\vartheta_G - 1} \left[\frac{e_t}{(\phi_G e_t - p_{G_t}(\omega))\, q_{G_t}(\omega)^{\delta_G}}\right]^{\vartheta_G - 1} d\omega\right]^{\lambda_G} \\ & \left[\int_0^{N_{S_t}} \frac{1}{\vartheta_S - 1} \left[\frac{e_t}{(\phi_S e_t - p_{S_t}(\omega))\, q_{S_t}(\omega)^{\delta_S}}\right]^{\vartheta_S - 1} d\omega\right]^{\lambda_S}. \end{aligned}$$

The consumption demand for variety ω of commodity j is

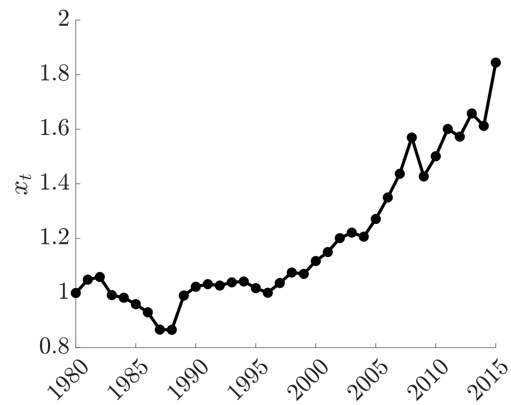
$$c_{j_t}(\omega) = \left[\frac{p_{j_t}(\omega)}{P_{j_t}}\right]^{-\vartheta_j} q_{j_t}(\omega)^{\delta_j(1-\vartheta_j)}\, \widetilde{C}_{j_t}.$$

D.4 Additional figures

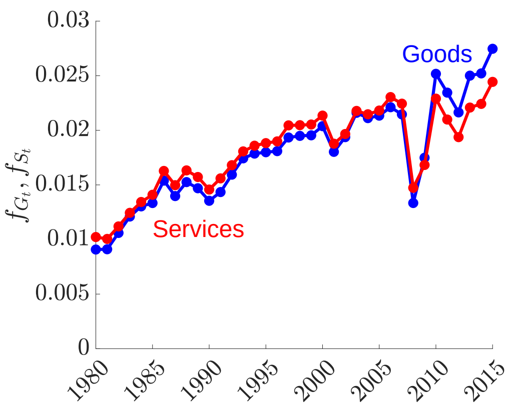
Figure D.4.1: Model-implied trends in productivity, entry costs, and high-skilled households



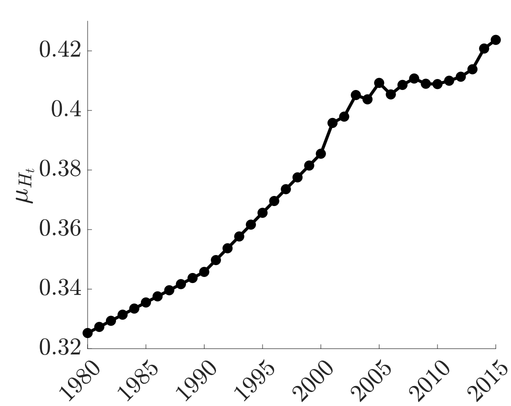
(a) Neutral productivity



(b) Skill-biased productivity



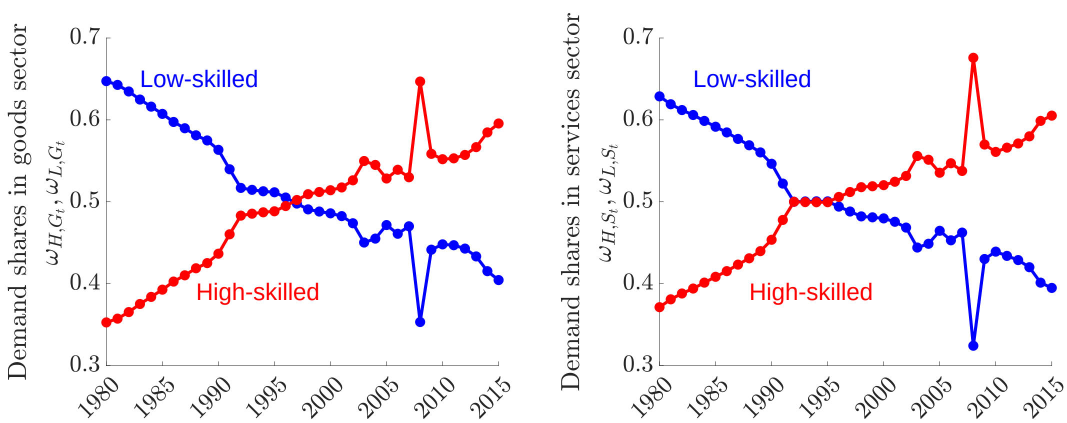
(c) Entry costs



(d) High-skilled share

Note: Panel (a) shows the evolution of neutral productivity in the goods (blue; z_{G_t}) and services (red; z_{S_t}) sectors. Panel (b) shows the evolution of skill-biased productivity in both sectors. Panel (c) shows the evolution of entry costs in the goods (blue; f_{G_t}) and services (red; f_{S_t}) sectors. Panel (d) shows the evolution of the share of high-skilled households in the economy.

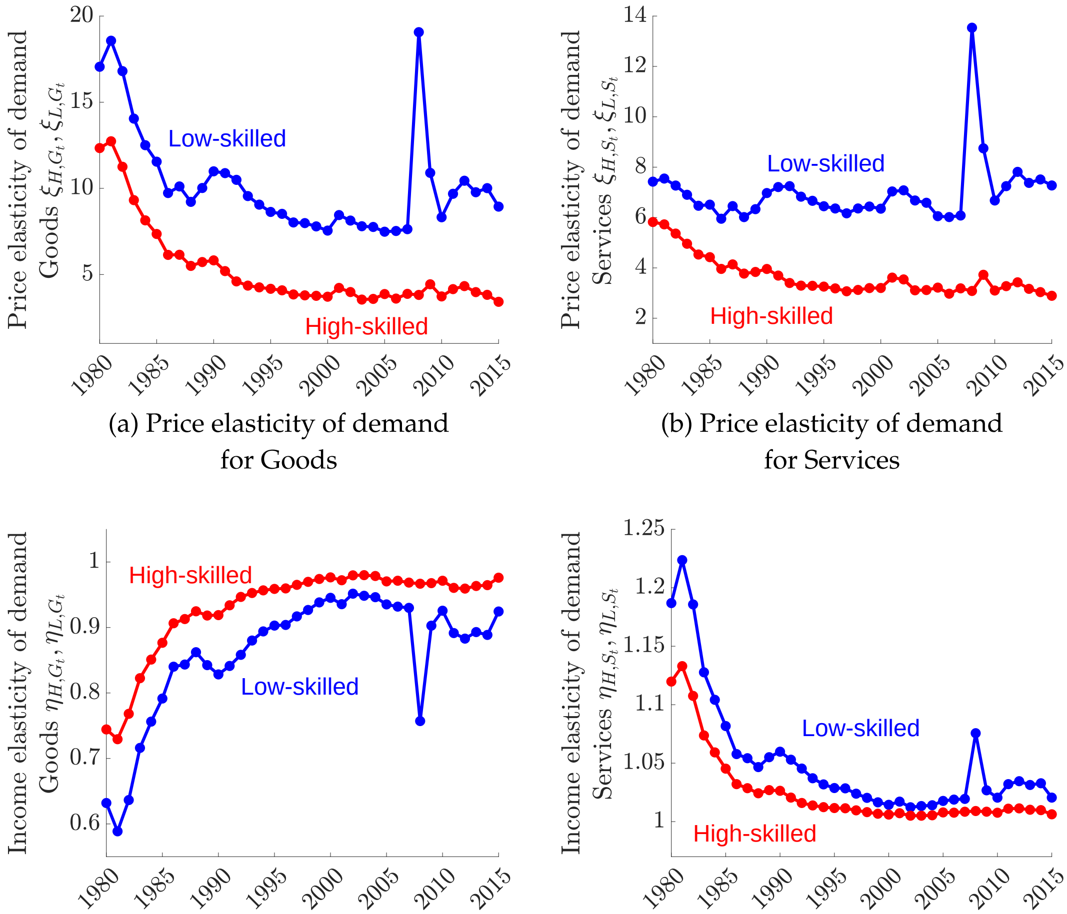
Figure D.4.2: Firms' demand shares



(a) Non-services (b) Services

Note: Panel (a) shows the evolution of goods firms' demand shares from high (red) and low-skilled (blue) consumers (ω_{H,G_t} and ω_{L,G_t}). Panel (b) shows the evolution of services firms' demand shares from high (red) and low-skilled (blue) consumers (ω_{H,S_t} and ω_{L,S_t}).

Figure D.4.3: Price and income elasticities of demand



(a) Price elasticity of demand for Goods

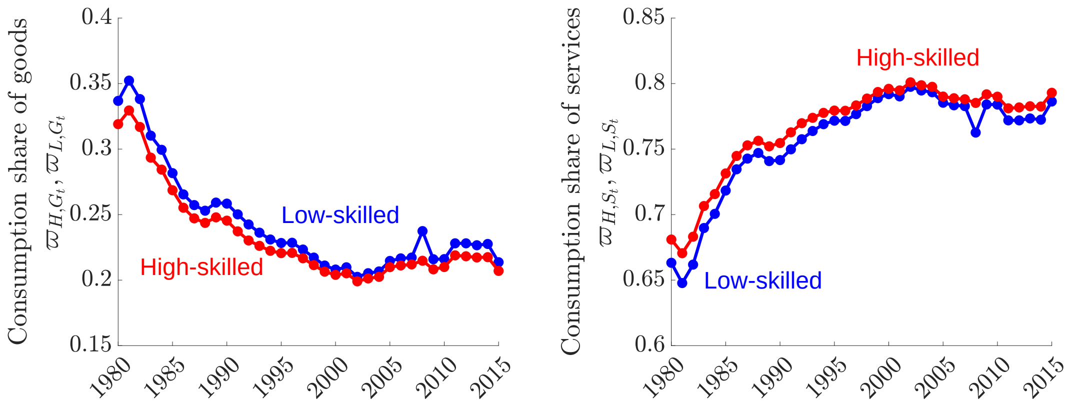
(b) Price elasticity of demand for Services

(c) Income elasticity of demand for Goods

(d) Income elasticity of demand for Services

Note: Panel (a) shows the evolution of the price elasticities of demand for goods for high (red) and low-skilled (blue) households in the model (ξ_{H,G_t} and ξ_{L,G_t}). Panel (b) shows the evolution of the price elasticities of demand for services (ξ_{H,S_t} and ξ_{L,S_t}). Panel (c) shows the evolution of the income elasticities of demand for goods for high (red) and low-skilled (blue) households in the model (η_{H,G_t} and η_{L,G_t}). Panel (d) shows the evolution of the income elasticities of demand for services (η_{H,S_t} and η_{L,S_t}).

Figure D.4.4: Consumption shares of income

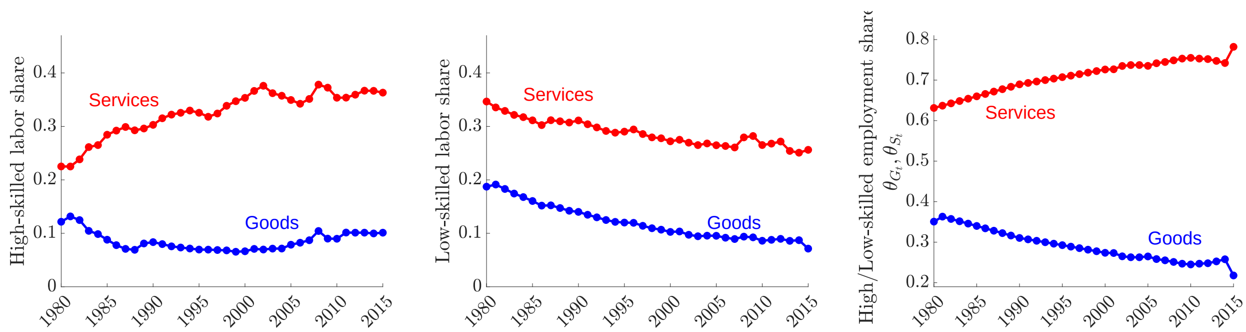


(a) Non-services

(b) Services

Note: Panel (a) shows the evolution of the goods consumption share of income (ϖ_{H,G_t} and ϖ_{L,G_t}) for high (red) and low-skilled (blue) consumers. Panel (c) shows the evolution of the services consumption share of income (ϖ_{H,S_t} and ϖ_{L,S_t}) for high (red) and low-skilled (blue) consumers.

Figure D.4.5: Labor shares and employment



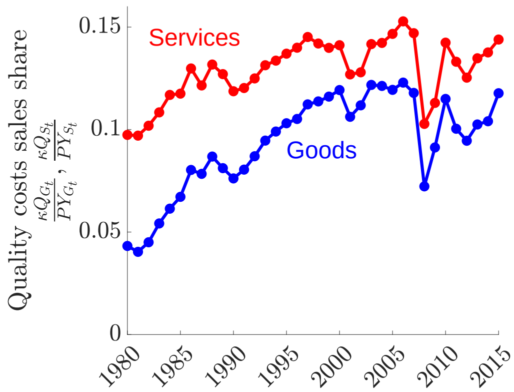
(a) High-skilled labor share

(b) Low-skilled labor share

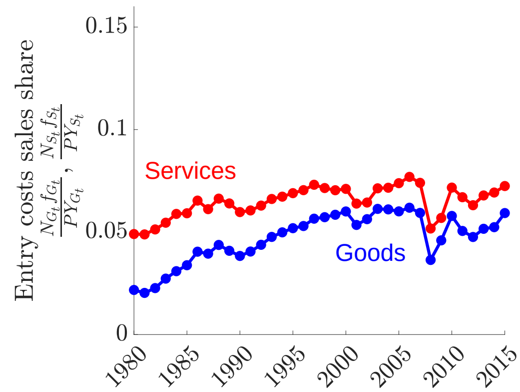
(c) Employment shares

Note: Panel (a) shows the evolution of the high-skilled labor share of income in the non-services sector $\left(\frac{VC_{H,G_t}}{PY_t}\text{, in blue}\right)$ and in the services sector $\left(\frac{VC_{H,S_t}}{PY_t}\text{, in red}\right)$. Panel (b) shows the evolution of the low-skilled labor share of income in the non-services sector $\left(\frac{VC_{L,G_t}}{PY_t}\text{, in blue}\right)$ and in the services sector $\left(\frac{VC_{L,S_t}}{PY_t}\text{, in red}\right)$. Panel (c) shows the evolution of the employment shares of high and low-skilled workers in the non-services sector (θ_{H,G_t} and θ_{L,G_t}, in blue) and in the services sector (θ_{H,S_t} and θ_{L,S_t}, in red).

Figure D.4.6: Fixed/entry costs as share of sales



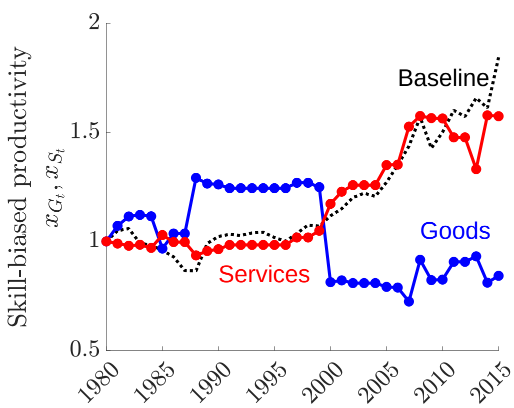
(a) Quality costs



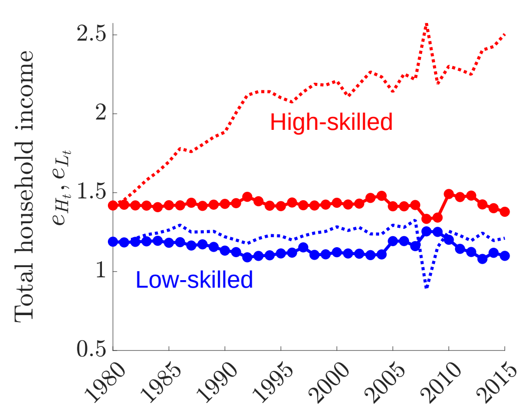
(b) Entry costs

Note: Panel (a) shows the evolution of fixed costs related with quality as a share of sales in the non-services sector ($\frac{FC_{G_t}}{PY_{G_t}}$, in blue) and in the services sector ($\frac{FC_{S_t}}{PY_{S_t}}$, in red). Panel (b) shows the evolution of entry costs as a share of sales in the non-services sector ($\frac{N_{G_t} f_{G_t}}{PY_{G_t}}$, in blue) and in the services sector ($\frac{N_{S_t} f_{S_t}}{PY_{S_t}}$, in red).

Figure D.4.7: Constant incomes experiment



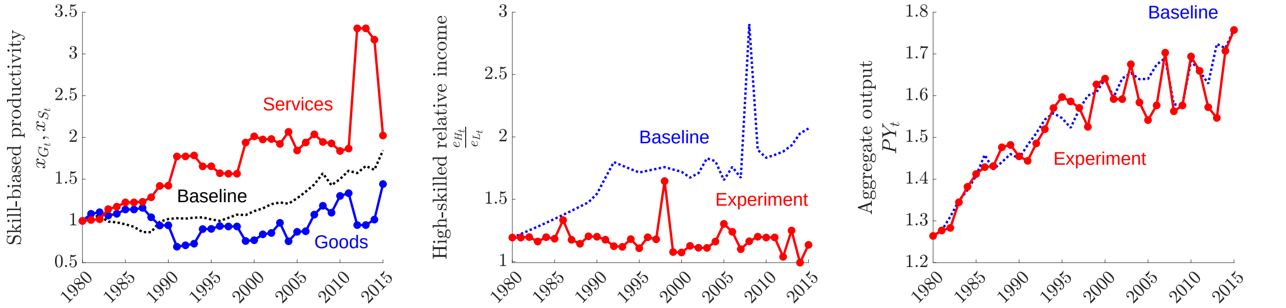
(a) Skill-biased productivity



(b) Households' income

Note: Panel (a) shows the evolution of skill-biased productivity in the baseline economy (in black), in the non-services sector (in blue) and in the services sector (in red). Panel (b) shows the evolution of the income of high-skilled households in the baseline (red dashed line) and in the experiment (red dotted line) as well as the income of low-skilled households in the baseline economy (blue dashed line) and in the counterfactual economy (blue dotted line).

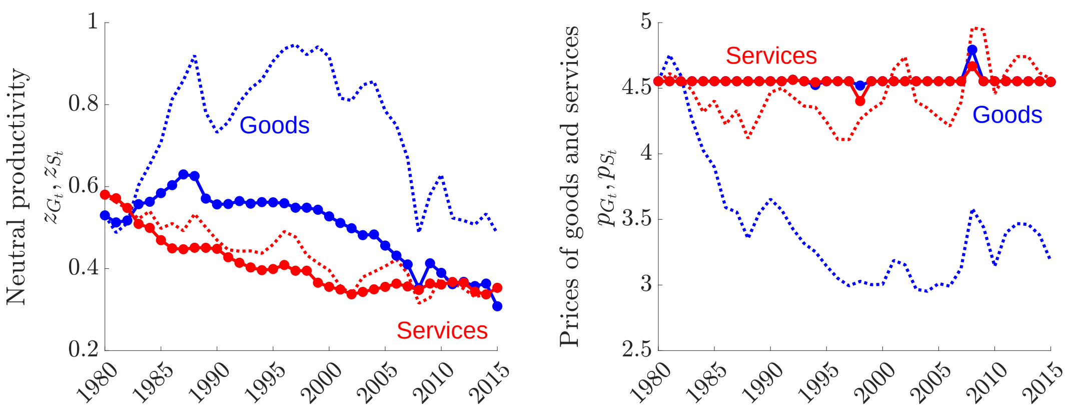
Figure D.4.8: Constant inequality experiment



(a) Skill-biased productivity (b) Relative households' income (c) Aggregate output

Note: Panel (a) shows the evolution of skill-biased productivity in the baseline economy (in black), in the non-services sector (in blue) and in the services sector (in red). Panel (b) shows the evolution of the income of high-skilled households relative to low-skilled households in the baseline economy (in blue) and in the experiment (red). Panel (c) shows the evolution of aggregate output in the baseline economy (in blue) and in the experiment (red).

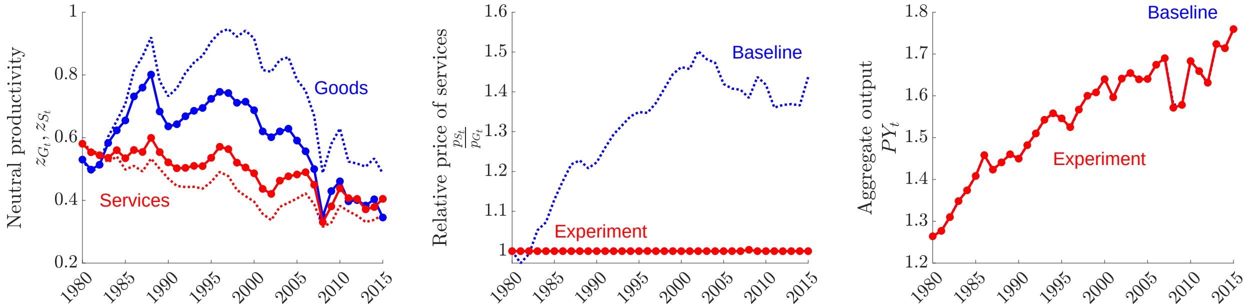
Figure D.4.9: Constant price experiment



(a) Total factor productivity (b) Prices

Note: Panel (a) shows the evolution of total factor productivity in the goods sector in the baseline economy (blue dashed line) and the counterfactual economy (blue dotted line) as well as in the services sector in the baseline economy (red dashed line) and the counterfactual economy (red dotted line). Panel (b) shows the evolution of the price of goods in the baseline economy (blue dashed line) and the counterfactual economy (blue dotted line) as well as the price of services in the baseline economy (red dashed line) and the counterfactual economy (red dotted line).

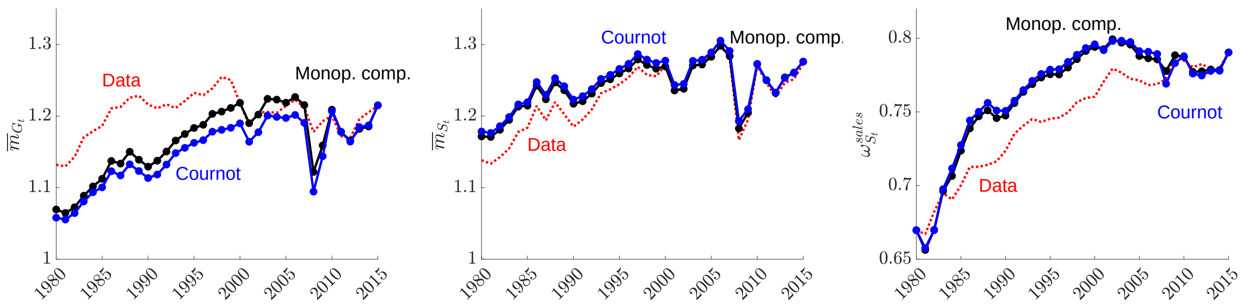
Figure D.4.10: Constant relative price experiment



(a) Total factor productivity (b) Relative price of services (c) Aggregate output

Note: Panel (a) shows the evolution of total factor productivity in the goods sector in the baseline economy (blue dashed line) and the counterfactual economy (blue dotted line) as well as in the services sector in the baseline economy (red dashed line) and the counterfactual economy (red dotted line). Panel (b) shows the evolution of the relative price of services in the baseline economy (in blue) and in the experiment (red). Panel (c) shows the evolution of aggregate output in the baseline economy (in blue) and in the experiment (red).

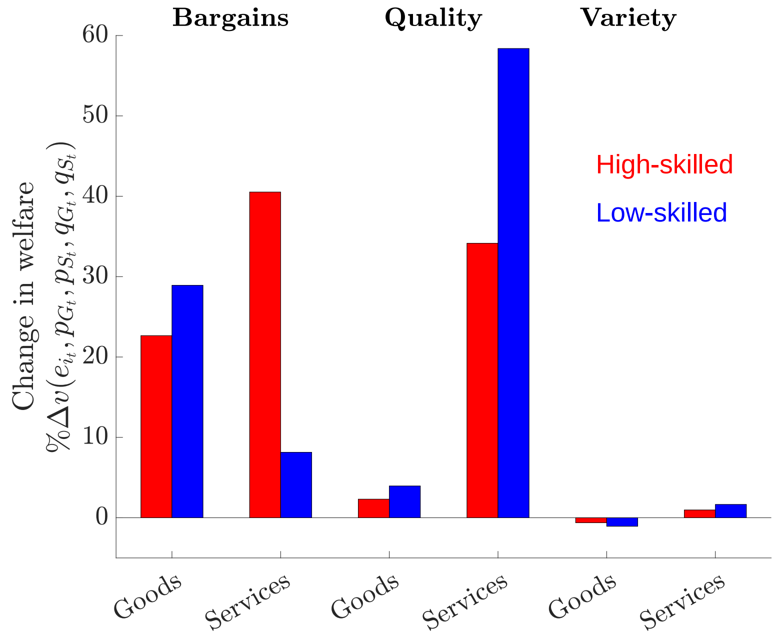
Figure D.4.11: Average markups and services sales share (Cournot)



(a) Average markup of goods (b) Average markup of services (c) Services sales share

Note: Panel (a) shows the average markup within the goods sector ($\overline{m}_{G_t}$) in the model with monopolistic competition (black), with Cournot (blue), and in the data (red). Panel (b) shows the average markup within the services sector ($\overline{m}_{S_t}$) in the model with monopolistic competition (black), with Cournot (blue), and in the data (red). Panel (c) shows the services sales share ($\omega^{\text{sales}}_{S_t}$) in the model with monopolistic competition (black), with Cournot (blue), and in the data (red).

Figure D.4.12: Decomposing welfare gains (Cournot)



Note: The figure shows the contribution of each term in equation (5.1) to the change in indirect utilities for high-skilled (red) and low-skilled (blue) consumers.

D.5 Additional tables

Table D.5.1: Experiments (Cournot): Income and prices

Variable	Description	(1) Baseline (Cournot) 1980, 2015	(2) e_{H_t}, e_{L_t} 2015	(3) $\Delta PY_t, \frac{e_{H_t}}{e_{L_t}}$ 2015	(4) p_{G_t}, p_{S_t} 2015	(5) $\Delta PY_t, \frac{p_{S_t}}{p_{G_t}}$ 2015
Markups						
M_t	Aggregate markups	1.136, 1.263	1.203	1.296	1.239	1.263
$\overline{m}_{G_t}$	Average goods markups	1.058, 1.215	1.123	1.098	1.125	1.143
$\overline{m}_{S_t}$	Average services markups	1.178, 1.276	1.237	1.500	1.296	1.320
Relative prices and income						
$\overline{p}_{S_t}/\overline{p}_{G_t}$	Relative price of services	1.000, 1.437	1.260	0.585	1.000	1.000
w_{H_t}/w_{L_t}	Skill premium	1.347, 1.928	1.172	1.828	1.928	1.928
ϵ_{H_t}	High-skilled income share	0.365, 0.603	0.423	0.634	0.600	0.604
$\omega_{S_t}^{\text{sales}}$	Services share	0.670, 0.790	0.719	0.569	0.697	0.708
Employment and labor shares						
θ_{H,S_t}	High-skilled empl. share in services	0.645, 0.782	0.729	0.627	0.667	0.678
θ_{L,S_t}	Low-skilled empl. share in services	0.645, 0.782	0.673	0.308	0.667	0.678
$VC_{H,G_t}/PY_t$	High-skilled labor share in goods	0.123, 0.101	0.104	0.165	0.158	0.150
$VC_{H,S_t}/PY_t$	High-skilled labor share in services	0.224, 0.363	0.281	0.278	0.316	0.315
$VC_{L,G_t}/PY_t$	Low-skilled labor share in goods	0.189, 0.071	0.146	0.228	0.111	0.106
$VC_{L,S_t}/PY_t$	Low-skilled labor share in services	0.345, 0.256	0.301	0.101	0.223	0.222
Entry						
$\kappa Q_{G_t}/PY_{G_t}$	Sales share of quality costs in goods	0.037, 0.118	0.073	0.060	0.074	0.083
$\kappa Q_{S_t}/PY_{S_t}$	Sales share of quality costs in services	0.101, 0.144	0.127	0.222	0.152	0.161
N_{S_t}/N_{G_t}	Relative number of services firms	4.059, 5.180	5.027	5.511	5.325	5.269

Table D.5.2: Equivalent variations (Cournot)

ε_i^{ev}, %	High-skilled	Low-skilled
Baseline economy, 1980 vs. 2015		
From 1980 to 2015	397.5	58.1
Baseline economy vs. Counterfactual economy, 2015		
Incomes constant at 1980 values	-24.8	-10.2
Prices constant at 1980 values	205.5	24.7
Entry costs constant at 1980 values	21.2	46.6
High-skilled share constant at 1980 values	16.1	40.5