

The Rise of Market Power and the Macroeconomic Implications*

Jan De Loecker[†]
KU Leuven
NBER and CEPR

Jan Eeckhout[‡]
UPF Barcelona (ICREA, GSE)
and UCL

Gabriel Unger[§]
Harvard University

November 22, 2018

Abstract

We document the evolution of markups based on firm-level data for the US economy since 1955. Initially, markups are stable, even slightly decreasing. In 1980, average markups start to rise from 21% above marginal cost to 61% now. The increase is driven mainly by the upper tail of the markup distribution: the upper percentiles have increased sharply, while the median is unchanged. In addition to fattening upper tail of the unweighted markup distribution, there is reallocation of market share from low to high markup firms. This rise occurs mostly within industry for *all* industries. We also find an increase in the average profit rate from 1% to 8%. While there is also an increase in overhead costs, the markup increase is in excess of overhead. We then discuss the macroeconomic implications of an increase in average market power, which can account for a number of secular trends in the last four decades, most notably the declining labor and capital shares as well as the decrease in labor market dynamism.

Keywords: *Markups; Market Power; Secular Trends; Labor Market; Declining Labor Share.*

JEL: E2, D2, D4, J3, K2, L1

*We would like to thank Mark Aguiar, Pol Antràs, John Asker, Eric Bartelsman, Steve Berry, Emmanuel Farhi, Bob Hall, John Haltiwanger, Xavier Gabaix, Eric Hurst, Loukas Karabarbounis, Patrick Kehoe, Pete Klenow, Thomas Philippon, Esteban Rossi-Hansberg, Chad Syverson, Jo Van Biesebroeck and Frank Verboven for insightful discussions and comments. Shubhdeep Deb provided invaluable research assistance. De Loecker gratefully acknowledges support from the FWO Odysseus Grant and Eeckhout from the ERC, Advanced grant 339186, and from ECO2015-67655-P. Gabriel joined the team in the second round revision in order to incorporate the Census Data analysis, as requested by the editors. This paper uses restricted data that were analyzed at the U.S. Census Bureau Research Data Center in Boston. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†] jan.deloecker@kuleuven.be

[‡] jan.eeckhout@upf.edu

[§] gunger@fas.harvard.edu

1 Introduction

Thriving competition between firms is a central tenet of a well functioning economy. The pressure of competitors and new entrants leads firms to set prices that reflect costs, which is to the benefit of the customer. In the absence of competition, firms gain market power and command high prices. This has implications for welfare and resource allocation. In addition to lowering consumer well-being, market power decreases the demand for labor and dampens investment in capital, it distorts the distribution of economic rents, and it discourages business dynamics and innovation. This in turn has ramifications for policy, from antitrust to monetary policy and income redistribution.

Despite the vital importance of market power in economics, surprisingly little is known about its systematic patterns for the aggregate economy and over time. In this paper, our main goal is to document the evolution of market power for the US economy since the 1950s. Based on firm-level data, we find that while market power was more or less stable between 1955 and 1980, there has been a steady rise in market power since 1980, from 21% above cost to 61% above cost in 2016. Over a 36 year period, that is an average increase in the price level relative to marginal cost of nearly 1% per year.

Once we have robustly established the fact, we discuss the macroeconomic implications of this rise in market power and the general equilibrium effects it has. We argue that the rise in market power is consistent with several secular trends in the last four decades, most notably the decline in the labor share and the decrease in the capital share. We also discuss how market power affects business dynamism and labor reallocation.

The evidence on market power we have to date comes from case studies of specific industries,¹ for which researchers have access to detailed data. In this approach championed by Bresnahan (1989) and Berry, Levinsohn, and Pakes (1995), the estimation of markups traditionally relies on assumptions on consumer behavior coupled with profit maximization, and an imposed model of how firms compete, e.g., Bertrand-Nash in prices or Cournot quantity competition. The fundamental challenge that this approach confronts is the notion that marginal costs of production are fundamentally not observed, requiring more structure to uncover it from the data. The firm's optimal pricing decision relates observed price data to estimates of substitution elasticities in order to uncover the marginal cost of production, and consequently the markup. This approach thus requires a combination of data on consumer demand (containing prices, quantities, characteristics, consumer attributes, etc.) as well as the need for specifying a model of conduct. All these requirements have limited the use of the so-called demand approach to particular markets.

In this paper we follow a radically different approach to estimate markups, the so-called production approach. Recent advances in the literature on markup estimation by De Loecker and Warzynski (2012), that in turn builds on Hall (1988), relies on individual firm output and input data, and in contrast to the demand approach described above, posits cost minimization

¹For example cars (Berry, Levinsohn, and Pakes (1995)), breakfast cereal (Nevo (2001)), or beer (Koujianou Goldberg and Hellerstein (2012))

by producers. A measure of the markup is obtained for each producer at a given point in time as the wedge between a variable input's expenditure share in revenue (directly observed in the data) and that input's output elasticity. The latter is obtained by estimating the associated production function. The advantage of this approach is twofold. First, the production approach does not require to model demand and/or specify conduct for many markets over a long period of time. Second, we can rely on publicly available production data. In particular, most of the information we need is available in the financial statements of firms. While there still exist many measurement issues and associated econometric challenges, to our knowledge there is no viable alternative to make progress. This is the only method that allows us to uncover the basic pattern of market power over a long period of time and across the entire economy.²

This paper starts by documenting the main patterns of markups in the US economy over the last six decades, and in doing so we provide new stylized facts on the cross-section and time-series of markups. This analysis is interesting and important in its own right as increasingly economic models allow for meaningful markup variation across producers and time; and allowing for this firm/time variation has substantially different implications for a variety of questions.³ The main analysis focuses on data from the financial statements of all publicly traded firms covering all sectors of the US economy over the period 1955-2016.⁴ And while publicly traded firms are relatively few compared to the total number of firms, they tend to be large. As of 2000, they account for 29% of total US private sector employees, excluding the self-employed and farm workers (Davis, Haltiwanger, Jarmin, and Miranda (2007)).

Our main finding is that while average markups were fairly constant between the 1950s and 1980 at around 1.2-1.3 (initially increasing and then slightly decreasing), there was a sharp increase starting in 1980 with average markups reaching 1.61 in 2016. In 2016, on average a firm charges prices 61% over marginal cost compared to only 21% in 1980.⁵ The increase was steep in the 1980s and 1990s, followed by a period of stable markups in the 2000s. At the end of the Great Recession of 2008, there has again been a sharp increase in market power. Second, there are marked changes in the distribution of markups over time. The increase occurs mainly in the top of the markup distribution. Markups in the top percentiles of the sales-weighted distribution go up most (from 1.5 in 1980 to 2.3 in 2016 for the 90th percentile), whereas markups at or below the median are flat or even decreasing. Third, there is no strong compositional pattern

²While the approaches – the demand approach and the production approach – differ, the obtained estimates should be similar. De Loecker and Scott (2016) apply both methods to estimate markups in the US beer industry, and find that they yield very similar markup estimates. In Appendix 5 we compare estimates from the literature using the demand approach to our estimates for the corresponding sector.

³For example, markup variability is found important in quantifying the gains from trade, Melitz and Ottaviano (2008) and Edmond, Midrigan, and Xu (2015), where trade reforms are best thought of as (exogenous) changes to competition and cost structures of firms.

⁴The data is from Compustat who extract the information from the Security and Exchange Commission (SEC) required public filing of financial statements. A handful of private firms are also included that have filing requirements.

⁵The only other attempts at measuring markups economy-wide that we have found in the literature are based on industry level aggregate data, and for the period up to the 1980s. Both Burnside (1996) and Basu and Fernald (1997) find little evidence of market power (nor of returns to scale or externalities), which is consistent with our finding that market power only picks up after 1980.

across industries and the increase occurs mainly within industry. The increase can therefore not be attributed to one particular industry such as technology. That of course does not exclude the possibility that technological change is driving the increase, only that the technology is affecting traditional sectors such as textiles for example as much as it is affecting more modern sectors. Finally, the rise in average weighted markups is due to an increase in the upper tail of the unweighted markups (roughly one third), combined with the reallocation of market share from low to high markup firms (about two thirds).

We then analyze market power. Markups measure how much higher prices are relative to marginal costs. However, markups larger than one do not necessarily imply firms are inefficiently exerting market power. If firms have high overhead costs, then a competitive firm will charge price over marginal cost exactly to offset the overhead. One major concern with the measured increase in markups is that this is due to a commensurate increase in overhead costs. Technological change may be one reason why overhead costs have grown as a share of total costs. Tech firms for example, invest a large amount in developing software that can then be reproduced at a very low marginal cost. We address this issue in two ways.

First, we introduce an alternative production technology where overhead is a factor of production, like capital. In this new technology we now have three factors of production: a bundle of variable inputs, and two fixed inputs, capital and overhead. While overhead as a factor of production still is not variable, in the long run it does adjust. In the traditional technology, overhead is a fixed cost that is incurred once at startup. In their financial statements, firms report overhead costs as Selling, General and Administrative Expenses (SG&A). These expenses are not directly related to production, and include sales, advertising, marketing, executive compensation,... and can in part be interpreted as expenses on intangible capital. We find indeed that overhead as a share of costs have increased, from 15% to 21% of total cost, and that increase comes mainly from large firms. When we estimate output elasticities for the technology with overhead as a factor of production, we find that there is a role for technological change. In this new technology, the elasticity of variable inputs decreases over time. As a result of this technological change, the estimated average markup rises from just under 1 to around 1.32 (the levels are lower because costs to compensate for factors of production now include overhead). Instead of an increase by 40 points, there is an increase by about 30 points. Technological change thus accounts for roughly one quarter of the increase in average markups. We also find that there are increasing returns to scale and that the returns to scale (measured as the sum of the Cobb-Douglas coefficients) increase by about 5-10 percentage points between 1980 and 2016.

The second way to address the role of overhead is to calculate the profit rate, which is total sales minus all cost (including overhead and the user cost of capital) as a share of sales. We find that the average profit rate has risen from close to 1% in 1980 to around 8% in 2016. While overhead costs have increased, markups have increased even more and firms charge an excess markup that more than compensates for overhead. In fact, it is the firms with the highest overhead costs that charge the highest excess markup and that therefore have the highest profits. The increase in the average profit rate is thus also driven by the right upper tail. We also find that the stock market valuation as a share of sales has risen over the same period. These facts

confirm that firms increasingly exert market power: they charge higher prices not merely to compensate for higher overhead costs, they also obtain higher profits. One important caveat here is that we assume that all expenses and overhead costs are properly accounted for. Firms may have incurred huge costs at startup that are thus not accounted for in present overhead costs and are now reaping the benefits from it.⁶

We perform a number of robustness exercises. The main one is to use data from the US Censuses to perform the same analysis for the full universe of firms. This addresses the major concern that our data on publicly traded firms is not representative of the whole economy. We focus on the Census of Manufacturing, the Census of Retail and the Census of Wholesale. The results we find mirror those we find in each of the sectors for the publicly traded firms. In Manufacturing and Retail, markups trend upwards since the 1980s, and down in Wholesale. And as with the publicly traded firms, the rise in average markups is driven by the increase in the upper percentiles. The important conclusion is that these results establish that the rise of market power is not restricted to the publicly traded firms and extends to universe of all firms in these sectors. We also find that this is mainly driven by some firms, for most firms markups do not change. This is even more pronounced in the Census data where the very upper tail of the distribution shows a sharp increase. It should be noted that while the Censuses are comprehensive for their coverage, the amount of data for each firm is substantially worse than that of the publicly traded firms. Most importantly, there is no information on overhead so we cannot analyze market power or calculate firm profits.

After we establish the main facts, we discuss the implications of the rise in market power for recent debates in the macro/labor literature. In particular, we analyze how the rise in markups naturally implies a decrease in the labor share. It follows immediately from the firm's optimization decision that high markups necessarily lead to lower expenditure on inputs such as labor. Hence the negative relation between markups and the labor share. We further discuss the role rising markups play in the decrease in the capital share, the decrease in low skilled wages, the decrease in labor market participation, and the decrease in labor reallocation and in interstate migration.

The analysis of markups and market power plays a central role in many literatures in economics, most notably in Industrial Organization, Macroeconomics and Labor Economics. As a result, it has always received due attention. Currently, there are several papers that touch on the aggregate dimension of market power that we stress here. Gutiérrez and Philippon (2017) analyze the Herfindahl-Hirschman Index (HHI) of concentration as a measure of market power (see also Grullon, Larkin, and Michaely (2016) and Brennan (2016)). They find that the increase in concentration is mainly driven by a decrease in domestic competition. This in turn leads to a decrease in firm-level investment, particularly in intangible assets by industry leaders. Our findings are consistent with theirs. Methodologically, our approach has the advantage that it

⁶What the increase in market power as we measure it tells us is that firms now would have incurred a much bigger cost up front than they did four decades ago, when it was close to zero. Also, the magnitude of this difference is around 7% of firm sales, extrapolating to the economy would imply roughly 14% of GDP (across firms in the economy, the ratio of sales to value added is around 2).

derives firm-level markups, which circumvents the limitations of the HHI measure.⁷ Hartman-Glaser, Lustig, and Zhang (2016) and Autor, Dorn, Katz, Patterson, and Van Reenen (2017) focus on the role of large firms. Hartman-Glaser, Lustig, and Zhang (2016) document that the firm-level capital share has decreased on average, even though the aggregate capital share for U.S. firms has increased. They explain the divergence with the fact that large firms now produce a larger output share even if the labor compensation has not increased proportionately. Autor, Dorn, Katz, Patterson, and Van Reenen (2017) document that the growing importance of large firms that dominate the market. They show that this leads to higher concentration and that it decreases the labor share. Like ours, their results are based on firm-level data, not macroeconomic aggregates. Our analysis explicitly derives markups as a measure of market power and we show that markups are directly related to the labor share at the firm level.

Since our first results came out, Traina (2018) and Karabarbounis and Neiman (2018) have criticized our findings arguing that the increase in markups is entirely offset by the increase in overhead as measured by SG&A. We disagree. In part, because their method does not use the correct notion of markups, as they bundle variable and fixed factors of production. But more importantly, we disagree because their conclusion is ill-founded. We show that there is indeed an increase in overhead, but that the increase in markups more than offsets this increase. We thus find that there are excess markups, which is consistent with the increase in profits that we document. Finally, also the macroeconomic implications of market power have recently received attention. In addition to the work already mentioned, Eggertsson, Robbins, and Wold (2018) start from our facts and document the implications for the macroeconomic Kaldor facts, in line with what we do. And Barkai (2017) analyzes the secular decline in capital ratios, while other research finds that there is a link between markups and aggregate outcomes such as IT pricing (Karabarbounis and Neiman (2014)) and housing (Rognlie (2016)).

The paper is organized as follows. In Section 2 we present the empirical framework to recover markups. In Section 3, we document the evolution of markups under different specifications of the technology. We show in Section 4 the extent to which the increase in markups also implies an increase in market power. We analyze the robustness of our findings in Section 5, and we discuss the macroeconomic implication in Section 6. We conclude in Section 7.

2 Empirical Framework

We start out by discussing the empirical framework that we build on to describe the patterns of markups in the US economy. In particular, we show that firm-level output and input data for firms across the US economy is sufficient to measure firm-level markups. To that end, we use minimal assumptions on producer behavior, yet without assumptions on product market competition and consumer demand. Once we have laid out the fundamentals of the production approach, we discuss how to implement it, with a particular focus on backing out output

⁷Most notably, concentration is not necessarily related to market power when products are differentiated (see Bresnahan (1989)), and an adequate concentration measure requires precise knowledge of what constitutes a market with information on all firms in that market.

elasticities either from production function estimation or from cost shares.

2.1 Obtaining Markups from Producer Behavior

Measuring markups is notoriously hard as marginal cost data is not readily available, let alone prices, for a large representative sample of firms. The standard approach in modern Industrial Organization is to specify a particular demand system that delivers price-elasticities of demand which, combined with assumptions on how firms compete, delivers measures of markups through the first order condition associated with optimal pricing. This approach, while powerful in other settings, is not useful here for two distinct reasons. First, we do not want to impose a specific model of how firms compete across a large dataset of firms active in very different industries, or commit to a particular demand system for *all* the products under consideration. Second, even if we wanted to make all these assumptions, there is simply no information on prices and quantities at the product level for a large set of sectors of the economy over a long period of time to successfully estimate price elasticities of demand, and specify particular models of price competition for all sectors.

We rely on a recently proposed framework by De Loecker and Warzynski (2012), based on the insight of Hall (1988) to estimate firm-level markups. The method uses information from the firm's financial statements, and does not require any assumptions on demand and how firms compete. Instead, markups are obtained by exploiting cost minimization of a variable input of production. This approach requires an explicit treatment of the production function.

Consider an economy with N firms, indexed by $i = 1, \dots, N$. Firms are heterogeneous in their productivity and otherwise have access to a production technology $Q_{it}(\cdot)$.⁸ In each period t , firm i minimizes the contemporaneous cost of production given the production function that transforms inputs into the quantity of output Q_{it} produced by the technology $Q_{it}(\cdot)$:

$$Q_{it} = Q_{it}(\Omega_{it}, \mathbf{V}_{it}, K_{it}), \quad (1)$$

where $\mathbf{V} = (V^1, \dots, V^J)$ is the vector of variable inputs of production (including labor, intermediate inputs, materials,...), K_{it} is the capital stock and Ω_{it} is the Hicks-neutral productivity term that is firm-specific. The key assumption is that within one period (a year in our data), variable inputs frictionlessly adjust, whereas capital is subject to adjustment costs and other frictions. Because in the implementation we will use information on a bundle of variable inputs, and not the individual inputs, in the exposition we treat the vector \mathbf{V} as a scalar V . We consider the Lagrangian objective function associated with the firm's cost minimization:

$$\mathcal{L}(V_{it}, K_{it}, \lambda_{it}) = P_{it}^V V_{it} + r_{it} K_{it} + F_{it} - \lambda_{it}(Q(\cdot) - \bar{Q}_{it}), \quad (2)$$

where P^V is the price of the variable input, r is the user cost of capital,⁹ F_{it} is the fixed cost, $Q(\cdot)$ is the technology specified in equation (1), \bar{Q} is a scalar and λ is the Lagrange multiplier.

⁸We derive the expression to compute markups in the most general case of firm-specific technologies, as long as the production function is twice differentiable. We subject our main empirical findings to various robustness checks precisely related to this production technology heterogeneity.

⁹Below, we will use lower case letters to denote logs, for example, $\log(P^V) = p^V$. For capital we use r because of general convention, but we assume $P^K = r$ so that we can denote $\log(r) = p^K$.

We assume that these input prices are given to the firm. It is important to note that this does not preclude that the input provider charges a markup over marginal cost, potentially leading to double-marginalization. The method for computing markups allows for arbitrary markups along the input-output table of the economy.¹⁰ We consider the first order condition with respect to the variable input V , and this is given by:

$$\frac{\partial \mathcal{L}_{it}}{\partial V_{it}} = P_{it}^V - \lambda_{it} \frac{\partial Q(\cdot)}{\partial V_{it}} = 0. \quad (3)$$

Multiplying all terms by V_{it}/Q_{it} , and rearranging terms yields an expression for the output elasticity of input V :

$$\theta_{it}^v \equiv \frac{\partial Q(\cdot)}{\partial V_{it}} \frac{V_{it}}{Q_{it}} = \frac{1}{\lambda_{it}} \frac{P_{it}^V V_{it}}{Q_{it}}. \quad (4)$$

The Lagrange multiplier λ is a direct measure of marginal cost – i.e. it is the value of the objective function as we relax the output constraints. We define the markup as $\mu = \frac{P}{\lambda}$, where P is the price for the output good, which depends on the extent of market power. Substituting marginal cost for the markup to price ratio, we obtain a simple expression for the markup:

$$\mu_{it} = \theta_{it}^v \frac{P_{it} Q_{it}}{P_{it}^V V_{it}}. \quad (5)$$

The expression of the markup is derived without specifying conduct and/or a particular demand system. Note that with this approach to markup estimation, there are in principle multiple first order conditions (of each variable input in production) that yield an expression for the markup. Regardless of which variable input of production is used, there are two key ingredients needed in order to measure the markup: the revenue share of the variable input, $\frac{P_{it}^V V_{it}}{P_{it} Q_{it}}$, and the output elasticity of the variable input, θ_{it}^v . While this approach does not restrict the output elasticity, when implementing this procedure it depends on a specific production function, and assumptions of underlying producer behavior in order to consistently estimate this elasticity in the data.

A crucial component to measure markups is to obtain an estimate of the output elasticity of a variable input of production (θ_{it}^v). We use two distinct methods to estimate the output elasticity of the production function. First, we estimate a parametric production function for each sector using recent techniques that take into account the well-known potential biases discussed in the literature. This method accounts for the fact that the variable factor of production V adjusts in response to a productivity shock, while the fixed factor K does not. This requires us to restrict the production function to a particular functional form to guarantee that the coefficients of interest – which determines the output elasticity – are identified. Second, we non-parametrically estimate the output elasticity using (constructed) cost shares. Both approaches have their advantages and disadvantages, which we discuss below. The focus of this paper is

¹⁰We do maintain the assumption that the input price is not a function of the input quantity demanded, either through bargaining, bulk discounting or long-term contracts. For a formal analysis allowing for such input price feedback see De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and for an application see Morlacco (2017).

to provide a robust description and analysis of markups across producers covering the entire economy, over more than six decades, rather than proposing one particular estimation routine to obtain output elasticities.

2.2 Output Elasticities: Production Function Estimation

We rely on what have become standard methods in production function estimation. However, we do depart from the literature by considering time-varying production function parameters. In particular, in the baseline model we estimate production functions with both time-varying and sector-specific coefficients, for each of the 22 sectors (i.e. 2 digit NAICS).¹¹ There are good reasons to believe that technology varies across sectors of the economy, from retail with giants like Walmart and Amazon, to highly specialized medical devices companies. Equally or more important for the evolution of markup is that the technology is time-varying. Over a period of seven decades, technology is likely to change. This is important for the estimation of markups since systematic technological change will imply a time-varying output elasticity θ_{it}^V . From inspection of equation (5), imposing a constant technology and hence a constant θ^V will therefore yield an overestimate of the markup if θ^V is decreasing and an underestimate if θ^V is increasing.¹²

We follow standard practice and rely on a panel of firms, for which we estimate production functions by industry. For the benchmark specification, we consider a sector-year-specific Cobb-Douglas production function, with a variable input bundle and capital as inputs.¹³ For a given industry s we consider the production function:

$$q_{it} = \theta_{st}^V v_{it} + \theta_{st}^K k_{it} + \omega_{it} + \varepsilon_{it}, \quad (6)$$

where lower cases denote logs and $\omega_{it} = \ln \Omega_{it}$, where q_{it} is a measure of realized firm's output, and ε_{it} is the unanticipated shock to output, or simply classical measurement error in output – i.e., $q_{it} = \ln(Q_{it} \exp(\varepsilon_{it}))$.

The challenges in estimating production functions, using any dataset, be it the Compustat data or plant-level manufacturing census data, can be grouped into two main categories: 1. dealing with *unobserved productivity* shocks (ω_{it}); and 2. extracting *units* of output and inputs from expenditure data. Both these issues are of course not independent, and we rely on methods that aim to deliver consistently estimated output elasticities, dealing with both adequately.

2.2.1 Unobserved productivity

We follow the literature and control for the simultaneity and selection bias, inherently present in the estimation of equation (6), and rely on a control function approach, paired with a law of

¹¹In principle we can consider industries at a lower level of aggregation, e.g. 3 digit NAICS, at the cost of pooling over longer periods of time, hereby keeping output elasticities constant over that same period.

¹²Because of data scarcity, we use a five year rolling window around the year where we estimate the technology. In Section 5 as well as in the Appendix, we discuss the estimation routine in further detail, and show the robustness of our findings with the baseline technology to different technological specifications.

¹³We consider various specifications of the production function, both at the level of the economy and the industry. In particular, below we also consider the more flexible translog production functions.

motion for productivity, to estimate the output elasticity of the variable input.¹⁴ In what follows, we work under the assumption that output and inputs are expressed in comparable units of quantity, either directly through deflating or indirectly, by imposing additional structure. We turn to the latter in the next subsection.

The attractive feature of the control function approach in our context is that it rests on an optimal input demand equation, which is immediate in the cost minimization framework used to recover an expression for the markup. In particular, the insight from Olley and Pakes (1996) is that the (unobserved) productivity term ω_{it} is given by a function of the firm's inputs and a control variable, d_{it} , such that $\omega_{it} = h_{st}(d_{it}, k_{it})$. The literature has suggested two types of control variables: static (e.g. intermediate inputs, see Levinsohn and Petrin (2003)) and dynamic (investment, see Olley and Pakes (1996)). Akerberg, Benkard, Berry, and Pakes (2007) provides an excellent treatment of the two approaches.

Regardless of which control variable is used, this method relies on a so-called two-stage approach. In the first stage, the measurement error and unanticipated shocks to output are purged using a non-parametric projection of output on the inputs and the control variable:¹⁵

$$q_{it} = \phi_t(v_{it}, k_{it}, d_{it}) + \varepsilon_{it}, \quad (7)$$

where $\phi = \theta_{st}^V v_{it} + \theta_{st}^K k_{it} + h_{st}(d_{it}, k_{it})$. The productivity process is given by $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$, and this gives rise to the following moment condition to obtain the industry-year-specific output elasticity:

$$\mathbb{E}(\xi_{it}(\theta_{st}^V) v_{it-1}) = 0. \quad (8)$$

In the second stage, $\xi_{it}(\theta_{st}^V)$ is obtained, given θ_{st}^V , by projecting productivity $\omega_{it}(\theta_{st}^V)$ on its lag $\omega_{it-1}(\theta_{st}^V)$, where productivity is in turn obtained from $\phi_{it} - \theta_{st}^V v_{it} - \theta_{st}^K k_{it}$, using the estimate ϕ_{it} from the first-stage regression. This approach identifies the output elasticity of a variable input under the assumption that the variable input use responds to productivity shocks, but that the lagged values do not, and more importantly, that lagged variable input use is correlated with current variable input use, and this is guaranteed through the persistence in productivity.

2.2.2 Units

A well known challenge when estimating production functions is that output and inputs are not measured in quantities.¹⁶ In almost any dataset, we fail to observe physical output and

¹⁴We estimate the production function, by industry, over an unbalanced panel to deal with the non-random exit of firms, as found important in Olley and Pakes (1996). However, the source of the attrition in the Compustat data is likely to be different than in traditional plant-level manufacturing datasets – i.e., firms drop out of the data due to both *exit*, and mergers and acquisitions, and as such the sign of the bias induced by the selection is ambiguous. We are, however, primarily interested in estimates of the variable output elasticity, while the selection bias is expected to impact the capital coefficient more directly.

¹⁵In the case of the static control, $d_{it} = v_{it}$, and the function ϕ collapses to $\phi_t(v_{it}, k_{it})$

¹⁶The presence of unobserved price heterogeneity, in both output and inputs, generates various biases. See De Loecker and Goldberg (2014) for a discussion and potential solutions. In addition, even when prices are readily observed in the data, there is still a potential bias due to unobserved quality differences among products or inputs.

input (q, v, k) .¹⁷ Instead we observe sales and expenditures on the various inputs. It is best practice to deflate these variables with the relevant industry-specific deflator. Denote the deflated input price (in logs) of input j , for $j = \{V, K\}$ by p_{it}^j , and the (deflated) output price (in logs) by p_{it} . This yields the following specification that we, in general, take to the data:

$$q_{it} + p_{it} = \theta_{st}^V \tilde{v}_{it} + \theta_{st}^K \tilde{k}_{it} + \omega_{it} + p_{it} - \sum_j \theta_{st}^j p_{it}^j + \varepsilon_{it}, \quad (9)$$

where \tilde{j} denotes (the log of) deflated expenditure on input j , e.g. $\tilde{v} = v + p^j$. This might seem an obvious problem, but the literature on production function estimation has only recently started to pay close attention to the implications of these unobserved price and quality components. Given our focus on markups, it is important to distinguish the bias of the output elasticities (θ) from the bias of the productivity residual ($\omega + p - \sum_h \theta_{st}^h p_{it}^h$). Given that we are only interested in estimating the output elasticity of the variable input θ^V for a given industry and time period, under certain modeling restrictions we can obtain consistent estimates of the production function without relying on separate price and quantity data.¹⁸ We now consider some specific cases of interest in estimating equation (9).

1. **TIME AGGREGATION.** If we consider the standard Cobb-Douglas specification adopted in the literature where the production function coefficients do not change over time, the bias induced by the price terms would only affect the level of the markup, but not the variation across producers in a given year, or over time.¹⁹ This immediately suggests an interesting observation: if we stick to the time-invariant output elasticity specification, then the time-series findings in this paper, which are at the core of our investigation, are not affected by unobserved output and input prices. Nor are the results about the within-sector changes affected by the inability to observe relevant physical output and input data. We do, however, potentially bias the findings across industries where the bias could depend on the level of price heterogeneity in a given industry.

2. **PASS-THROUGH.** In the most general case, given the production structure assumed throughout, we are left with (deflated) revenue and (deflated) expenditures, as specified in equation (9). Conditional on productivity shocks, the error terms cancels out if variation in input prices (scaled by their relevant output elasticity) are perfectly absorbed by the variation in output prices. This is the case if pass-through is complete, i.e., changes to input prices are fully passed on to prices, scaled by the relevant cost share of that input (i.e. the output elasticity). This can

¹⁷The exception being the US Census of Manufacturing, which for a handful of products consistently has comparable output quantities, see Foster, Haltiwanger, and Syverson (2008).

¹⁸As we will show below, perhaps somewhat counterintuitively, observing neither output or input prices, and therefore relating sales to expenditures can under certain conditions generate consistent estimates of the output elasticities. We subject our markup analysis to a variety of specifications and robustness checks to confront the inability to observe producer-level output and input prices.

¹⁹This is the approach adopted by Brandt, Van Biesebroeck, Wang, and Zhang (2017) to study the markup dynamics in the Chinese economy during the WTO accession.

be thought of as operating both within a producer over time, or across producers in any given year.²⁰

In the case of incomplete pass-through, the variable markup creates a wedge between the output price and the input price bundle. This implies that the output elasticity is in general biased. However, an alternative strategy presents itself by recognizing how the total error term in (9) relates to marginal cost. Under a constant returns to scale production function:

$$\ln \lambda_{it} = \sum_h \theta_j p_{it}^j - \omega_{it}. \quad (10)$$

Using the fact that $p_{it} = \ln \mu_{it} + \ln \lambda_{it}$, and plugging in the expression for the price in equation (9) yields

$$q_{it} + p_{it} = \theta_c \tilde{v}_{it} + \theta_k \tilde{k}_{it} + \ln \mu_{it} + \varepsilon_{it}. \quad (11)$$

This expression highlights that output elasticities can be consistently estimated using data on sales and expenditures as long as one can control for markups. If in fact the variation in markups is uniquely determined by cost-side heterogeneity, say productivity, then the approach discussed in the previous subsection applies.²¹ More generally, the variation in markups can be controlled for by a function, $\mathcal{M}(\cdot)$, that contains relevant determinants of markups.²² As expected, there is no one-size-fits-all control function specification, but rather we have to subject our results to a variety of robustness checks regarding which variables to include in $\mathcal{M}(\cdot)$. Empirically we consider equation (11) for each sector and year and we approximate $\mathcal{M}(\cdot)$ by a year-industry specific linear function, and include a firm's market share (in the reported 4 digit industry and productivity as determinants of markups). We treat the unobserved productivity shock ω_{it} like before. The output elasticities are identified and estimated using the same moment conditions as before, except that additional parameters are estimated – i.e., the parameters associated with the markup control function $\mathcal{M}(\cdot)$:

$$\mathbb{E}(\tilde{\xi}_{it}(\theta_{st}^v) \tilde{v}_{it-1}) = 0, \quad (12)$$

where $\tilde{\xi}_{it}(\cdot)$ is obtained from projecting $\omega_{it}(\cdot)$ on $\omega_{it-1}(\cdot)$, and where now $\omega_{it}(\cdot) = q_{it} + p_{it} - \theta_c \tilde{v}_{it} - \theta_k \tilde{k}_{it} - \mu(m_{it})$, and where m_{it} is the market share. The additional parameters to be estimated capture the relationship between a firm's market share and its markup.

²⁰De Loecker (2011) considers the case where the only source of price heterogeneity comes from the output side. This is of course an assumption, but it is motivated by the standard common input price model adopted in the productivity literature. This implies that we are left with *only* an output price bias. De Loecker (2011) relies on the insights of Klette and Griliches (1996) and introduces a demand system to separate the demand from the production patterns. The main insight is that observable demand shifters allow the identification of the elasticity of demand, and thus we can control for unobserved prices. This approach is less attractive in our case since we do not have plausible instruments shifting the demand for all relevant sectors and time periods.

²¹Under arbitrary returns to scale the relationship between sales and the input bundle expenditure depends on the *scale* of production, say γ . This yields $q_{it} + p_{it} = \theta_v \tilde{v}_{it} + \theta_k \tilde{k}_{it} + \ln \mu_{it} + \gamma q_{it} + \varepsilon_{it}$.

²²For example, in a standard Cournot homogeneous good model, the markup is only a function of the firm market share m_{it} . A related, and recently adopted approach by Atkeson and Burstein (2008), considers a nested-CES demand structure with Cournot competition, and this gives rise to a control function in a firm's market share, and the market share of the industry in the economy. Finally, in full-fledged models of oligopoly with consumer taste for characteristics (e.g. Berry, Levinsohn, and Pakes (1995)), this control function in addition contains the vector of product characteristics.

2.3 Cost Shares

An alternative to estimating production functions to obtain output elasticities is to rely on the cost shares, in particular that of the variable input V . This cost share is the share of expenditures on the variable input bundle in total cost. The cost-share-approach has the major advantage that it does not require to estimate the production and it consequently bypasses the thorny identification issues (unobserved productivity and prices).

However, this advantage does not come for free. First of all, this approach relies on each input of production to be variable and for production to occur under constant returns to scale. Only then will the cost share be equal to the output elasticity. This is the reason why prominent users of this approach, most notably Foster, Haltiwanger, and Syverson (2008), compute the average cost share across the sample period to try and eliminate the cross-sectional deviations due to adjustment costs. Of course, given that we require time-varying technology coefficients, that approach is less attractive. Second, in order to compute the total cost of production, a user cost of capital is required (r in our framework). This is not required for the estimation of production functions (where we use k , not rk) and it poses an empirical challenge to measure this user cost of capital accurately, due to the heterogeneity across firms and time. In addition, measurement error in the capital stock will directly affect the output elasticity, whereas in the production function (at least in the Cobb-Douglas production function we consider) the measurement error in capital is expected to mainly impact the capital coefficient.²³

While averaging the cost shares across producers and time has the potential benefit of eliminating the deviations due to adjustment costs of factors, it still requires time-invariant technology parameters and constant returns to scale in production. These are assumptions we do not wish to maintain in the context of our analysis. Time-varying technology parameters and returns to scale are fundamentally potential confounders of interpreting rising wedges of sales to variable input expenditures as rising markups, and consequently market power.²⁴ For these reasons, the cost shares approach offers a robustness check but should be interpreted with caution. At the same time we acknowledge that this approach yields a powerful, and transparent method to *measure* the output elasticities directly in the data, provided we impose a common user cost of capital.

We therefore resort to calculating year-specific cost shares, by industry or firm. We consider two levels of aggregation. First, we consider firm-time specific cost shares, which incidentally imply that our measure of the markup is simply the ratio of sales-to-total costs, a popular measure of markups in the literature. This highlights that this approach is subject to the caveats discussed above. In order to partly deal with the variation in adjustment cost across producers in a given period, in addition we consider the median cost share by industry-year.

²³Collard-Wexler and De Loecker (2016) perform Monte Carlo simulations to evaluate the impact of measurement in capital and they find a small spillover bias onto the variable input (labor).

²⁴Below, we follow Syverson (2004) and estimate the scale elasticity, subject to the other potential biases when estimating production functions, using $q_{it} = \gamma \ln Z_{it} + \omega_{it} + \varepsilon_{it}$, where $\ln Z_{it} = \alpha^V v_{it} + \alpha^K k_{it}$ and α denotes the relevant cost share, when we consider the traditional production function.

3 The Evolution of Markups

With the empirical framework laid out, we now document the evolution of markups for different estimation methods. We first describe the data in greater detail. Then we start by considering the traditional production function with fixed costs.

3.1 Data

The choice of data is driven entirely by the ability to cover the longest possible period of time, *and* to have a wide coverage of economic activity. This for example rules out exclusively using the US Census of manufacturing establishments, given the decreasing share of manufacturing in the overall employment of US economy, from about 25% in 1960 to 9% in 2014 (Baily and Bosworth (2014)). The Census also does not include information on overhead at the firm level. To our knowledge, Compustat²⁵ is the only data source that provides substantial coverage of firms in the private sector over a substantial period of time, spanning the period 1950 to 2015. While publicly traded firms are few relative to the total number of firms, because the public firms tend to be the largest firms in the economy, they account for 29% of private US employment (Davis, Haltiwanger, Jarmin, and Miranda (2007)).

There is a serious concern though that the sample of publicly traded firms is not representative of the distribution of the universe of firms. Listed firms are bigger, older, more capital intensive, and more skill intensive. They also involve a bigger role for multinationals. And the industry mix of Compustat firms differs from that of the private sector as a whole.²⁶ We deal with the selection from the publicly traded firms in two ways. In Section 3.6 we repeat our analysis on the US Censuses. For a number of sectors, there we have the universe of firms. Second, we use the population weights of each sector to adjust the weights in the Compustat sample (see Section 5). While we still only use publicly traded firms to calculate the markups, we account for any bias due to the sectoral composition.

The Compustat data contains information of firm-level financial statements, which allows us to rely on the so-called production approach for measuring markups. In particular, we observe measures of sales, input expenditure, capital stock information, as well as detailed industry activity classifications.²⁷ The item from the financial statement of the firm that we will use to measure the variable input is “Cost of Goods Sold” (COGS). It bundles all expenses

²⁵The data covers all firms that are traded on US stock exchanges and that have as a result the obligation to file with the Securities and Exchange Commission (SEC), as well as some private firms that have SEC filing requirements.

²⁶There are also pronounced trends in the number and character of listed firms in recent decades. These developments are well documented in the literature. To summarize briefly, there was a huge influx of riskier, younger firms in the 1980s and 1990s (See, e.g., Fama and French (2004), Davis, Haltiwanger, Jarmin, and Miranda (2007), and Brown and Kapadia (2007).) And, in something of a reversal, there has been a huge net decline in the number of U.S. listed firms since the early 2000s (see Gao, Ritter, and Zhu (2013) and Doidge, Karolyi, and Stulz (2017). And in the period since the mid 1990s, the average firm size has increased.

²⁷The Compustat data has been used extensively in the literature related to issues of corporate finance, such as CEO pay, e.g., Gabaix and Landier (2008), but also for questions of productivity and multinational ownership, e.g., Keller and Yeaple (2009).

directly attributable to the production of the goods sold by the firm and includes materials and intermediate inputs, labor cost, energy,...²⁸ In addition, we observe relevant, and direct accounting information of profitability and stock market performance. The latter information is useful to verify whether our measures of markups, as discussed below, also relate to the overall evaluation of the market. Table A.1 in the Appendix provides basic summary statistics of the firm-level panel data used throughout the empirical analysis.

From our data we construct a measure of the user cost of capital. We follow the standard procedure in the literature and use $r_t = (i_t - \pi_t) + \delta$, where i_t, π_t, δ are the nominal interest rate, the inflation rate and a depreciation rate. We use Gross Capital (PPEGT) that we adjust for the industry-level input price deflator (PIRIC from FRED), for the federal funds rate and for an exogenous depreciation rate and risk premium jointly that we set at 12%.²⁹

Our data also has a measure of Overhead, booked under “Selling, General and Administrative Expenses” (SG&A). This item includes selling expenses (salaries of sales personnel, advertising, rent,...), general operating expenses, and administration (executive salaries, general support related to the overall administration). We use SG&A to calculate total costs – not just the cost of factors of production – in order to measure the profits of the firms. In addition, below we will consider a production technology, different from the conventional technology, where we treat overhead as a factor of production.

CENSUSES. As a robustness exercise and to verify the extent of selection bias in our sample of publicly traded firms, we repeat exercise for the Economic Census. The Economic Census is administered every five years (in years ending in “2” and “7,” e.g. in 2012). It is composed of Censuses of different sectors: a Census of Manufactures, a Census of Retail Trade, a Census of Wholesale Trade, and so on. Within each sector, it covers the universe of employer establishments (establishments which hire workers, and are not just one-person sole proprietorships); compliance is legally required. “Establishment” denotes a discrete physical unit, but an establishment is also required to identify its firm if it is part of a multi-establishment firm. The microdata for the oldest Census (manufacturing) begins in the 1960s; coverage of most sectors begins in the 1970s or 1980s, and the most recently added Census, covering the financial service industries, begins in 1992. Together they now cover virtually the entire US private sector, with the exception of non-employer establishments, and of agriculture (the US Census of Agriculture is now administered separately by the USDA, instead of the US Census Bureau). Our data for all Economic Censuses ends in 2012.

²⁸The sample does not directly report a breakdown of the expenditure on variable inputs, such as labor, intermediate inputs, electricity, and others, and therefore we prefer to rely on the reported total variable cost of production. Alternatively, we could rely on imputed intermediate inputs as in Keller and Yeaple (2009). However, that requires additional assumptions by deriving a measure of intermediate input use from operating income before depreciations, and the total wage bill, where the latter is imputed from multiplying the reported total number of employees with industry-wide wage data.

²⁹Below, when we investigate the capital share (the user cost of capital divided by sales) we find, not surprisingly, that this measure is quite volatile. Gross Capital is a long term measure that adjusts at a lower frequency and that therefore is more subject to aggregate fluctuations. Also, in the 1970s there was a sudden drop in capital investment. Those were tumultuous financial times: inflation was high and financial frictions were considered higher.

The Census of Manufactures contains establishment-level data on sales, in addition to very comprehensive data on inputs (the total labor wage bill, capital, materials, and so on). However, most of the other sector censuses (retail, wholesale, etc.) only contain data on establishment-level sales and wage bill, and not other non-labor inputs. In section 3.6 we will analyze markups for Manufacturing, Retail and Wholesale. A more detailed description of the Census data is in the Appendix.

3.2 Traditional Production Function with Fixed Costs

The traditional production function (in equation (1), henceforth denoted by PF1) takes factors of production V (variable) and K (fixed) as inputs. The fixed cost F from overhead is not a factor of production and therefore does not matter for output, even though it matters for profits.

We use the production function estimation technique described above to derive the average markup. The estimation method for the production technology allows for flexibility in how disaggregated the data is. For the baseline measure of our markup, we estimate the technology with an elasticity that is time-varying and specific to each sector. We denote sectors by s . Then the output elasticity for all firms i in sector s is θ_{st}^V , and the markup (equation (5)) of a firm i in sector s is given by:

$$\mu_{it} = \theta_{st}^V \frac{P_{it} Q_{it}}{P_{it}^V V_{it}}. \quad (13)$$

This generates a markup estimate for each firm i . From the distribution of markups across all firms, we first derive the *average* markup, weighted by the share of sales of the firm i in the entire sample:

$$\mu_t = \sum_i m_{it} \mu_{it}, \quad (14)$$

where $m_{it} = \frac{S_{it}}{S_t}$ is the market share of sales in the sample with $S_t = \sum_i S_{it}$.

Figure 1 reports the evolution of our baseline measure of average markups across the economy over time. In the beginning of the sample period markups were relatively stable, initially slightly increasing to 1.34 in the 1960s and then decreasing to 1.21 in 1980. Since 1980 there has been a steady increase to 1.61. In 2016, the average markup charged is 61% over marginal cost, compared to 21% in 1980. In Online Appendix 4 we report a few examples of individual firms' markups.

To establish the robustness of the markup measure in the baseline, in Section 5 and the Appendix we report the average markup measure for different specifications of the production technology to estimate the output elasticity. We estimate the output elasticity that is: 1. time-invariant and sector-specific θ_s ; 2. time-varying and not sector-specific θ_t ; 3. time invariant and not sector-specific θ ; 4. We also estimate a technology that is translog and we report the input-weighted average markup from the baseline model.

We now repeat the analysis where we obtain the output elasticity from cost shares. For each firm we have an observation for the cost share $\alpha_{it}^V = \frac{p_t^V V_{it}}{p_t^V V_{it} + r_t K_{it}}$. Within an industry, we use

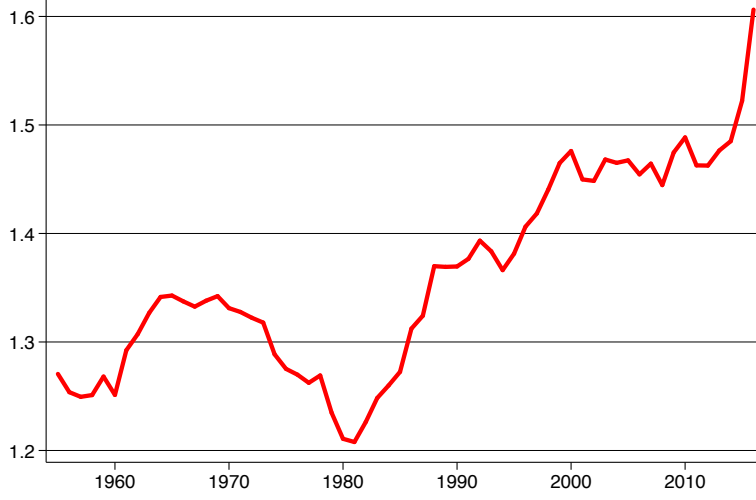


Figure 1: Average Markups for Conventional Production Function. Output elasticities θ_{st} from estimated PF1 are time-varying and sector-specific (2 digit). Average is sales weighted. Evolution 1955-2016.

the median of the distribution as the measure for the output elasticity:

$$\theta_{st} = \text{median}_{i \in s} \{\alpha_{it}^V\}. \quad (15)$$

Figure 2 reports the sales-weighted average of the markups with the output elasticity derived from the cost-share for the traditional production technology where overhead is a fixed cost and denoted by CS. The pattern is very similar to that in Figure 1. There is a moderate decrease from the 1960s and then an increase from 1980 up to 2016. The level is slightly higher and the increase by 50 points is somewhat more pronounced.

From inspection of the definition of markup in equation (5), the rise in the markup could potentially be attributed to two sources: 1. an increase in the ratio of sales to expenditure on variable inputs; or 2. technological change, an increase in the output elasticity θ^V over time. In Figure 3a we plot the average cost share of the factors of production V and K as well as the average output elasticity estimated from PF1. There is some volatility in the cost shares, and the estimated output elasticity is decreasing over the time period, but the magnitude of the changes are fairly small. This indicates that the steep increase in markups is driven by the increase of sales over expenditure on inputs. Firms are selling their goods at higher margins. This is also evident from inspection of Figure B.1 in the Appendix where we report the markup for a constant θ^V . The pattern there is virtually identical to the one in the baseline model in Figure 1. This tells us that the evolution of the average markup is mainly driven by the ratio of sales to expenditure on variable inputs and not by changes in the output elasticity.

Because there is a fixed cost, the increase in markups does not mean that there has been an increase in the profits. It may be that firms sell their goods a higher margin because they are incurring higher fixed costs. Figure 3b plots the cost share of variable factors in the *total cost*

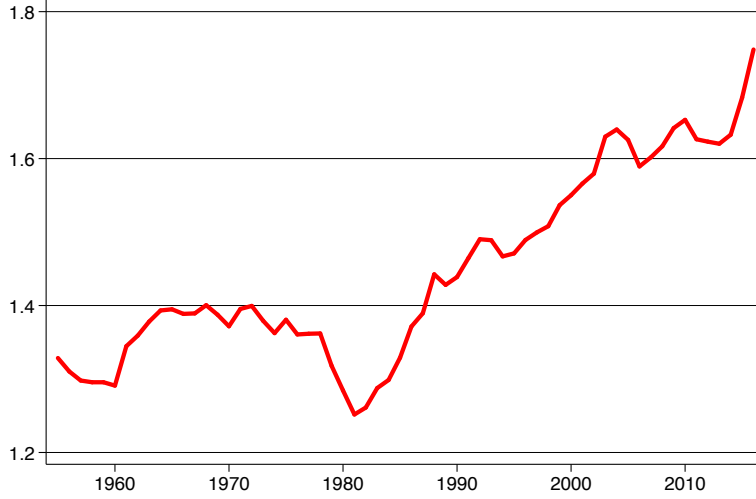
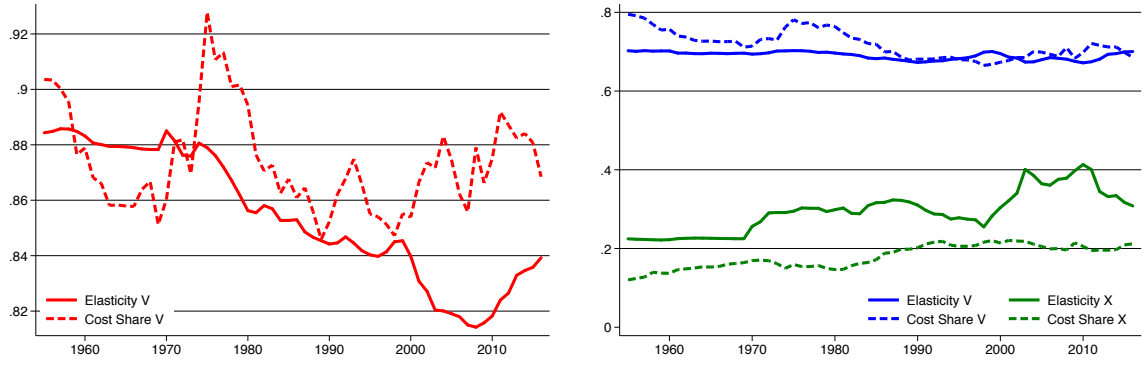


Figure 2: Average Markups for the Conventional Production Function using Cost Shares: CS1. Output elasticities θ_{st} are time-varying and sector-specific (2 digit, median). Average is sales weighted. Evolution 1955-2016.



(a) Output elasticity PF1 and Cost Shares: V

(b) Output elasticity PF2 and Cost Shares: V and X

Figure 3: Average Cost Shares and Average Output Elasticity (sales weighted average).

(consisting of variable factors, capital and overhead), as well as the cost share of overhead. We see that there is a slight decrease in the cost share of the variable factor of production from 80% in the beginning of the sample to 70% in 2016. The share of the fixed cost has increased from 18% at the beginning to 24% towards the end. This is indicative of the fact that the overhead cost, and thus the technology, has changed. We further investigate this technological change in the next section.

Following the initial circulation of our results, Traina (2018) and Karabarbounis and Neiman (2018) have argued that the increase in SG&A has fully offset the increase in the markup. However, they mix the notion of variable and non-variable factors of production. By summing the

variable factors of production (COGS) with the non-variable factors (SG&A), they do not measure the price to marginal cost ratio, or the markup, because they assume SG&A varies in the firm's optimization decision, which we show below is counterfactual. Instead, we treat SG&A either as a one-off fixed overhead cost, or a non-variable factor of production (in Section 3.3 below). In addition, by summing SG&A and COGS, these two factors of production are assumed to be perfect substitutes, which is not a realistic assumption. This affects the markups as well as the estimated output elasticities needed to calculate the markup.

3.3 Production Function with Overhead as a Factor of Production

The conventional production function uses as factors of production the variable input V and capital K . All other expenditures accounted for as not directly related to the production of the goods sold are overhead. They are considered fixed costs, a cost incurred that is independent of the output produced. This is the standard approach in the Industrial Organization literature on markup estimation.

In contrast to the conventional interpretation of the production technology, we propose an alternative interpretation where a portion of the overhead is a factor of production. Higher expenditure on getting more and better logistics managers will lead to an increase in the units produced. More sales people increases the units sold. Not all of this factor of production necessarily leads to more production, it may also mean just more sales. In order to be able to interpret overhead as a factor of production, we denote its expenditure by $p^X X$, where the quantity that enters the production technology is X and the unit price is p^X .

We now take this non-conventional interpretation of overhead as a factor of production seriously, and assume that all of it is a factor. The production function can then be written as $Q(V, K, X)$ and firm profits are $PQ(V, K, X) - P^V V - rK - P^X X$. We can apply the same cost-based method for the derivation of markups as laid out in section 2. We now treat X as a factor of production that enters the production function, but it is non-variable, just like capital K . The treatment of the variable input V remains as before. The difference relevant for the measurement of markups therefore stems from the production function estimation and the resulting estimate for θ^V . To differentiate we denote the estimates from this production function by PF2. When we calculate elasticities based on cost shares that take into account overhead as a factor of production, we refer to it as CS2. To further differentiate the graphical representation, PF1 and CS1 are in plotted red, and PF2 and CS2 are in blue.

In Figure 4a, we report the evolution of the average markup with this new production technology. Qualitatively, we see a similar pattern for the increase in the average markup starting in 1980. Initially slightly below one, it increases by around 30 points in 2016. The increase for this technology is some 10 percentage points lower than for the traditional production function (Figure 1). This difference is driven by the fact that the cost share of overhead (and the estimated output elasticity θ^X) is increasing over time (see Figure 3b). What matters for the markup estimate however is the elasticity θ^V . We know that it is roughly constant for the conventional production function (Figure 3a). For the production technology with overhead as a factor of production, θ^V is slightly decreasing (Figure 3b). Therefore the estimated markup

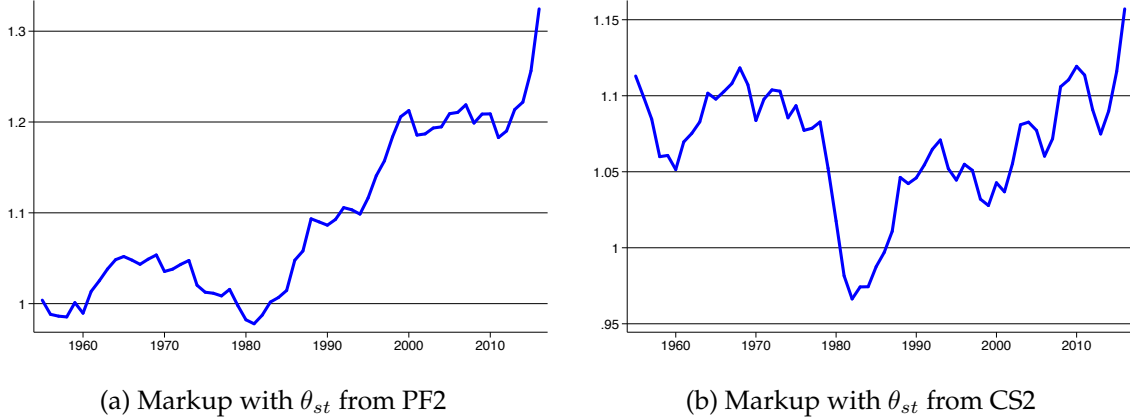


Figure 4: Average Markups for Production Function with Overhead as a factor. Output elasticities from estimated PF2 and from CS2: time-varying, sector-specific (2 digit) output elasticity θ_{st} (sales weighted average).

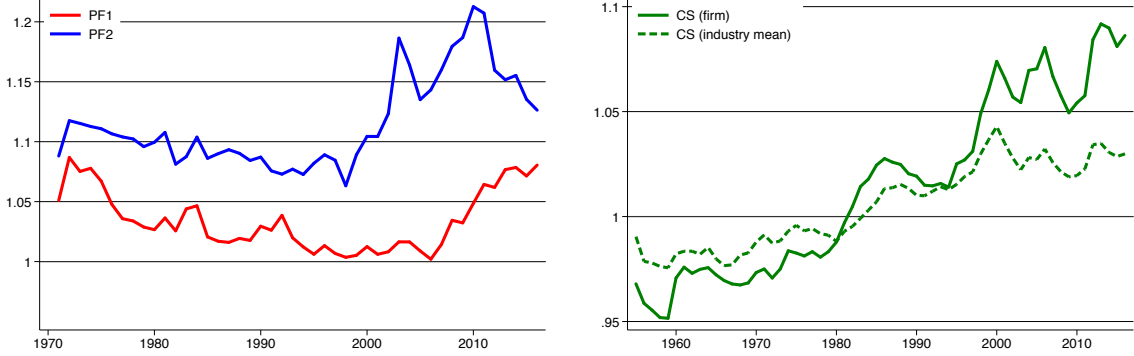
shows a more moderate increase (30 points) than under the conventional production technology (40 points).

In Figure 4b we plot the average markup where we use the cost shares as a measure for the output elasticity (CS2). The evolution of the average markup is markedly different. There is still a steady increase from 1980 onwards, but that increase is now by about 20 points only (instead of 40 points for PF1 and CS1, and 30 points for PF2). In addition, markups before 1980 are high relative to those after 1980. The reason for the discrepancy between the markup using PF2 and CS2 stems from the implicit assumptions underlying the technology that uses cost shares. While often hailed as non-parametric, using cost shares implicitly assumes a technology that imposes constant returns to scale. Only under constant returns (and Cobb-Douglas) is the cost share equal to the output elasticity. It turns out that ruling out returns to scale is an important factor in explaining why the average markup calculated with the cost shares increases less. We turn to returns to scale next.

3.4 Returns to Scale

With the estimated technologies (PF1 and PF2), we can evaluate any technological change that affects the returns to scale. Because the technology is Cobb-Douglas, the returns to scale are measured by the sum of the output elasticities: $\theta^V + \theta^K$ for PF1 and $\theta^V + \theta^K + \theta^X$ for PF2. We find that the estimated technology shows a rise in the degree of increasing returns over time. In Figure 5a we report the sum of the output elasticities for both technologies PF1 and PF2. For the conventional technology (PF1), from the start of the sample, the estimated returns to scale go from around 1.02 in 1980 to 1.08 in 2016. Instead, the returns to scale of the technology with overhead as a factor of production, there is an increase from 1.07 up to 1.13, reaching 1.22 in 2010. The fact that the production function with overhead as an input has higher returns to scale confirms that overhead X is in part a fixed cost that generates increasing returns.

Moreover, since those returns to scale are increasing more over time as overhead increases, establishes that the role of overhead as a source of returns to scale is growing.



(a) RTS (sum of output elasticities) of Estimated PF1 and PF2; sales weighted (b) Estimated RTS of Cost Shares: firm CS and sector average CS

Figure 5: Returns to Scale

The increase in the returns to scale also explains why the markup estimate based on cost shares only shows an increase of 20 percentage points (Figure 4b), whereas under the elasticity estimated from the production function the increase since 1980 is 30 percentage points (Figure 4a). By construction, cost shares add up to one and therefore the implied elasticities are derived under the assumption of constant returns. As a result, the increase in the elasticity θ^X due to an increase in the expenditure share of overhead must necessarily lead to a decrease in θ^V .³⁰ With θ^V decreasing, from equation (5), the increase in the markup must necessarily be dampened. This illustrates that directly using the cost shares can by construction not account for any change in the returns to scale in the technology.

However, we can rely on the method first used in Syverson (2004) to derive the returns to scale based on cost shares. He assumes the following functional form for the technology based on cost shares but without constant returns:

$$q = \gamma [\alpha_V v + \alpha_K k + \alpha_X x] + \omega, \quad (16)$$

with all variables in logs, where $\alpha_V = \frac{p^V V}{p^V V + r^K K + p^X X}$ is the cost share of the variable input, and likewise for α_K and α_X . While each cost share determines the output elasticity, the technology need not be constant returns and the curvature is captured by γ . In Figure 5b we plot two measures of the estimated γ , one for the average firm-level γ and one where we impose a common γ at the 2-digit industry level, both for the technology with overhead as a factor of production.³¹ Both reveal that also with this method, returns to scale have increased throughout the sample.

³⁰In principle, it could also lead to a decrease in θ^K , but θ^K is so small it cannot offset all of the increase in θ^X . We consistently find that the estimated θ^K is constant across all specifications.

³¹An important caveat here is that we estimate this technology by means of a simple regression, without accounting for endogeneity.

There were DRTS before 1980 and since 1980 returns to scale have been increasing, up to 1.05 at the end of the sample.³²

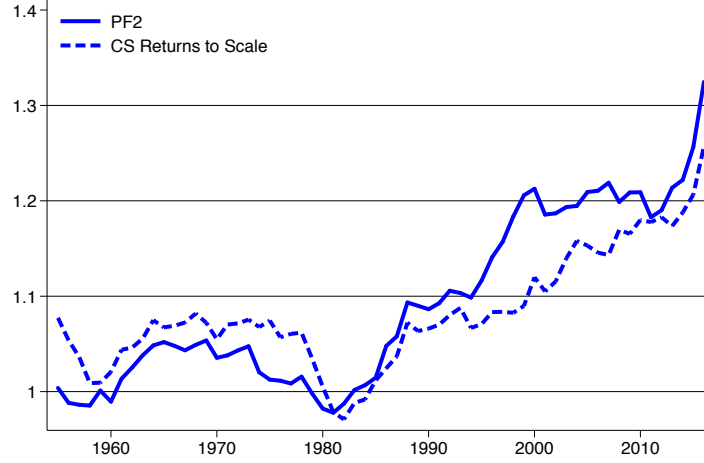


Figure 6: Average Markup: Benchmark (PF2) and Cost Shares with Returns to Scale (Syverson)

The evolution of returns to scale helps us understand the difference between Figures 4a and 4b. In the latter, we ignore the change in the returns to scale because Cost Shares are implicitly assuming Constant Returns to Scale. If instead we use the elasticities obtained for the Syverson (2004) technology in equation (16) which is equal to $\gamma\alpha_v$, we obtain an average markup (see Figure 6) that is very similar to the one using the elasticity estimated with the production function (PF2). The increase in γ in Figure 5b implies that the elasticity α_V used in Figure 4b is multiplied by γ .

3.5 The Distribution of Markups

So far, we have only provided evidence of the average markup. A key finding is that the increase in markups is driven by a few firms, without any increase for most. The advantage of our method is that we obtain a markup for each firm, so we have a distribution of markups. In this section, we investigate how the higher moments of this distribution have evolved. We do all exercises for both production functions, with overhead as fixed cost (PF1, panels a.) and overhead as a factor of production (PF2, panels b.).

THE DISPERSION OF MARKUPS. To get an idea of the evolution of the entire distribution of markups, we plot the kernel density of the unweighted markups for 1980 and 2016 (Figure 7). For both markup measures, we find that the variance has increased and that in particular, the upper tail has considerably fattened and become longer. It is the upper tail that drives the increase in the average markup.

³²In a recent paper (based on the original method proposed by Hall (1988) that works with aggregate data), Hall (2018) uses the expenditure shares (rather than the cost shares) to estimate markup.

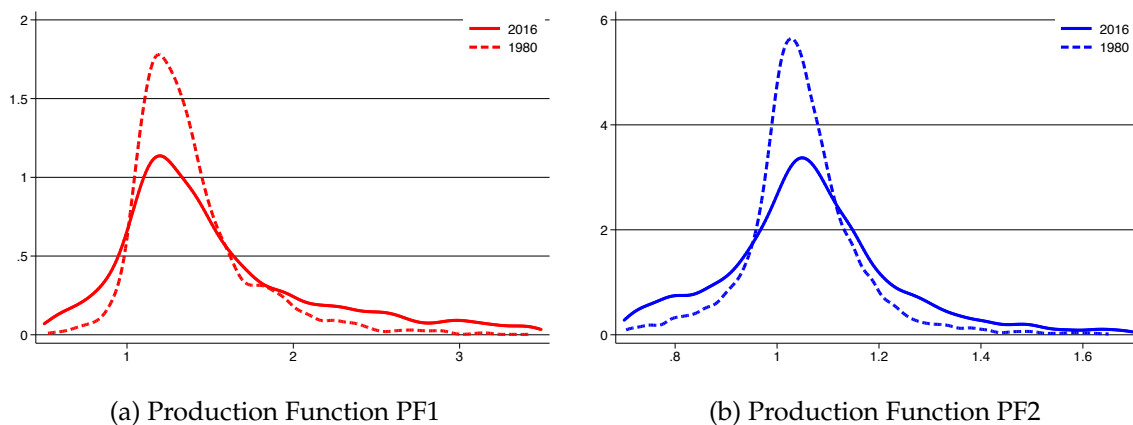


Figure 7: Distribution of Markups μ_{it} : Kernel Density Plots (unweighted)

Because the kernel density does not take into account the weights, we next plot the different moments of the distribution of *sales-weighted* markups over time (Figure 8). We rank the firms by markup, and to obtain the percentiles we weigh each firm by its market share in the entire sample. This makes the percentiles directly comparable to our share weighted average. The ranking is updated each year, so the firms at the top may be different each year (below, we investigate the persistence in the markup process).

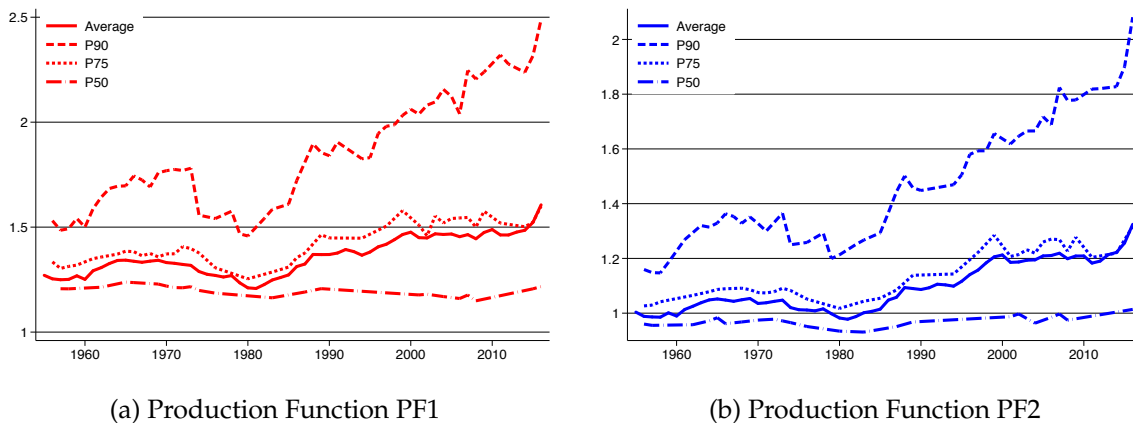


Figure 8: Percentiles of the Markup Distribution. (The percentiles of the Markup distribution are weighted by marketshare of sales in the sample.)

The increase in the average markup comes entirely from the firms with markups in the top half of the markup distribution. The percentiles below the median are largely invariant over time. For percentiles above the median, markups increase. For the 90th percentile in particular, the increase is sharpest. Between 1980 and 2016, it increases from 1.8 to 2.8 for the conventional production function (Figure 8a), and from 1.5 to 2.5 for the production function with overhead as an input (Figure 8b). This indicates that the change in average markup is largely driven by

a few firms that currently have much higher markups than decades ago.³³ We now turn to summarizing the evolution of markups in a simple stochastic process.

THE PROCESS OF MARKUPS. To capture some properties of the process that governs the evolution of markups, we assume the following autoregressive process for our measure of markups (with the conventional production function) as well as for the data on sales and employment:

$$z_{it} = \rho x_{it-1} + \varepsilon_{it}, \quad z \in \{\log \mu, \log S, \log L\}. \quad (17)$$

Figure 9 shows the evolution of the cross-sectional standard deviation of the shocks in the markup, sales and employment processes. Starting in 1980, there is clearly a sharp rise in the standard deviation of the markup μ and a more moderate increase in that of sales S . Interestingly, there is much less of an increase in the standard deviation of employment L . If anything, there is a decline from 2000 onwards. This is because the increase in the standard deviation of markups is precisely driven by the increase in the wedge between the volatility of sales (increasing) and inputs, in this case labor (fairly constant and then decreasing).³⁴

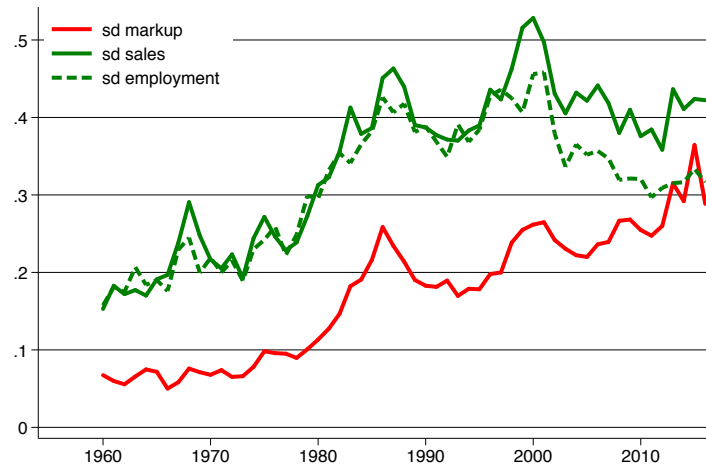


Figure 9: The Evolution of the Standard Deviation of Markups, Sales and Employment (1960 - 2016). (AR(1) in logs on their lag with year and industry fixed effects; The estimated persistence is 0.84).

This increasing wedge is consistent with the evidence in Decker, Haltiwanger, Jarmin, and Miranda (2014) that the shock process itself has not changed much but that the transmission of the shocks to inputs (labor) has.

DECOMPOSITION WITHIN VERSUS BETWEEN SECTORS. We now decompose the *change over time* in the markup by firm size. Is the increase in markup over time due to a change of markup

³³This is consistent with the evidence in Kehrig (2011). He studies the cyclicalty of productivity and finds that the dispersion in TFPR is increasing, especially in the upper tail.

³⁴As robustness checks, we have also included higher order terms of the persistence and find that those are not important.

at the industry level (Δ within), due to a change in the composition of the firms – there are more firms with a high markup – (Δ between), or due to the joint change in markup and the firm composition (Δ cross term). This can be expressed in the following formula:

$$\Delta\mu_t = \underbrace{\sum_s m_{s,t-1} \Delta\mu_{st}}_{\Delta \text{ within}} + \underbrace{\sum_s \mu_{s,t-1} \Delta m_{s,t}}_{\Delta \text{ between}} + \underbrace{\sum_s \Delta\mu_{s,t} \Delta m_{s,t}}_{\Delta \text{ cross term}} \quad (18)$$

We consider the change over 10 year periods starting in 1956.³⁵ As we already observed in Figure 1, our baseline measure of markup has slightly decreased in the run up to the 1980s, and has since then increased at an increasing rate. The decomposition shows that the change in markup is mainly driven by the change within industry. Most of the Δ Markup is driven by Δ Within. There is some change in the composition between industries, but that is relatively minor compared to the within industry change. The change due to reallocation, the joint effect, is mostly small.

Production Function PF1					
	Markup	Δ Markup	Δ Within	Δ Between	Δ Cross
1966	1.337	0.083	0.057	-0.017	0.041
1976	1.270	-0.067	-0.055	0.002	-0.014
1986	1.312	0.042	0.035	0.010	-0.003
1996	1.406	0.094	0.098	0.004	-0.008
2006	1.455	0.049	0.046	0.007	-0.005
2016	1.610	0.154	0.133	0.014	0.007

Production Function PF2					
	Markup	Δ Markup	Δ Within	Δ Between	Δ Cross
1966	1.051	0.061	0.044	-0.012	0.028
1976	1.014	-0.037	-0.030	-0.004	-0.003
1986	1.048	0.035	0.028	0.008	-0.001
1996	1.141	0.093	0.096	0.002	-0.006
2006	1.211	0.070	0.068	0.004	-0.002
2016	1.329	0.118	0.106	0.004	0.008

Table 1: Decomposition of 10 year change in Markup.

The fact that most of the increase in markups is driven by the increase *within* sectors and not for particular sectors is an important unexpected discovery. Intuitively, we would expect that certain sectors such as Tech would see a much bigger increase in the markup. But as the sector specific markups in Figure 10.1 in the Online Appendix illustrate, there are no sectors that systematically have higher market power. This confirms that the increase in market power occurs in all sectors and industries.

³⁵The decomposition for the 3 and 4-digit industry classification is reported in Table 1 in the Appendix.

DECOMPOSITION AT THE FIRM LEVEL. We now address to what extent the rise in aggregate markups is driven by an increase in the firm level markups, or by an increase in the sales share of firms with high markups. Though not in all, in most theories of market power, firms that have higher market power also increase their market share (in the Cournot model in particular, the market share is a sufficient statistic of market power).

Our objective here is to decompose the change in the aggregate markup. We know from inspection of Figure 7a that the distribution of markups has changed, with a slight increase in the mean and a substantial increase in the variance. Now even if the markup distribution remained unchanged, we could observe an increase in the average weighted markup if high markup firms saw an increase in their market shares. Because reality is most likely a combination of both, we decompose the average markup at the firm level as follows:

$$\Delta\mu_t = \underbrace{\sum_i m_{i,t-1} \Delta\mu_{it}}_{\Delta \text{ within}} + \underbrace{\sum_i \tilde{\mu}_{i,t-1} \Delta m_{i,t}}_{\Delta \text{ market share}} + \underbrace{\sum_i \Delta\mu_{i,t} \Delta m_{i,t}}_{\Delta \text{ cross term}} + \underbrace{\sum_{i \in \text{Entry}} \tilde{\mu}_{i,t} m_{i,t} - \sum_{i \in \text{Exit}} \tilde{\mu}_{i,t-1} m_{i,t-1}}_{\text{net entry}}, \quad (19)$$

$\Delta \text{ reallocation}$

where $\tilde{\mu}_{it-1} = \mu_{it} - \mu_{t-1}$.³⁶

We apply the insights from the productivity-decomposition literature, and while this decomposition appears very similar to that in (18), it is different, first because it has one additional term, and second because its interpretation is very different. There is an additional term here because there is entry and exit of firms, whereas in the sectoral decomposition the number of sectors is fixed. But also the interpretation differs. Following Haltiwanger (1997), we consider a theoretical counterfactual where the ‘ Δ within’ term measures the average change that is due merely to a change in the markup, while keeping the market shares unchanged from last period. Instead, the ‘ Δ market share’ term measures the change due to an increase in market share while keeping the markup fixed. If this term is increasing, it captures the fact that firms with higher markups now have a higher market share and hence there is increase in the weight of the high markup firms. This in turn raises the average markup without raising the markup itself. The ‘ Δ cross term’ measures the joint change in markups and market share. We denote by ‘ Δ reallocation’ the joint effect of ‘ Δ market share + Δ cross term’.³⁷ Finally, the new last term measures the effect of entry and exit on markups. This captures the change in the composition of firms in the market. If the entering firms have higher markups than the exiting firms for example, then this term will be positive.

We first perform this decomposition across firms in the entire economy. This implies that the reallocation component captures movements of firms across all sectors. This is in contrast to the sectoral analysis presented in Table 1, where sector-level aggregate markups were

³⁶We demean the (lagged) markups by the appropriate aggregate (share-weighted) level, in order to correctly identify the role of the reallocation term. With a slight abuse of notation the markup for entrants is given by $\tilde{\mu}_{it} - \mu_{t-1}$. See Haltiwanger (1997) for more discussion.

³⁷The ‘ Δ cross term’ is virtually zero in the experiments we perform.

calculated before exploring the within and between dimensions. We then perform the same decomposition for each of the broad sectors of the economy.

In Figure 10 we plot the average markup (in red), as well as three counterfactual experiments based on the decomposition starting in 1980. We set the initial level to 1980 and then cumulatively add the changes of each of the component terms in equation (19).³⁸

The first experiment (plot in blue) shows the evolution of the average markup as if there was only component ‘ Δ within’ and all other components were zero. This shows that the rise in average markups in the 1980s and 1990s from 1.21 to 1.3 in 2000 is about one third of the total increase from 1.21 to 1.47. From 2000 onwards, this term decreases and picks up again after the great recession. The change in the average markup is also evident from Figure 7, where we see an increase in the upper right tail.

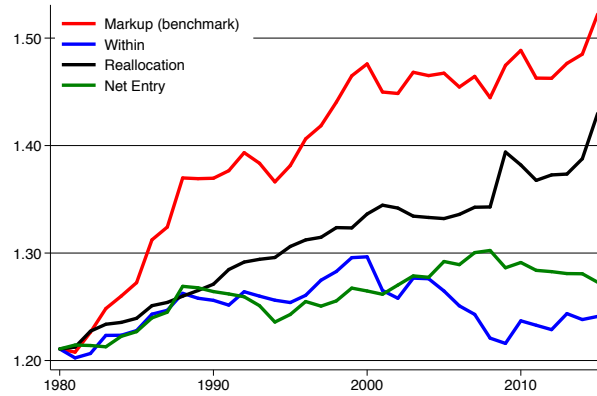


Figure 10: Decomposition of markup growth at the firm level.

The second experiment (plot in black) show the path of the markup if the only change had been due to ‘ Δ reallocation’. All markups remain unchanged from last period, and we apply only the change in the market shares. The plot shows that accumulated over the whole time period, reallocation accounts for about two thirds of the change in the weighted markup. The main takeaway here is that there are two forces at work. On the one hand, the markup (the within term) increases, which is an indication of the change in pricing power of firms. In De Loecker, Eeckhout, and Mongey (2018) we show that this can be due to a change in the market structure (less competition) or due to technological change (bigger spread in firm productivity). On the other hand, there is also a reallocation of sales activity away from low markup firms towards high markup firms (the reallocation term). This is entirely consistent with a model of imperfect competition where firms with higher markups also attract a higher market share. This reallocation effect is in accordance with the findings in Autor, Dorn, Katz, Patterson, and Van Reenen (2017) and Hartman-Glaser, Lustig, and Zhang (2016) who establish that large firms have grown in size relative to small firms, and those firms tend to operate in

³⁸In the Online Appendix section 3 we tabulate the measured yearly changes of each of the four components for all years between 1955 and 2016. The cumulative representation in Figure 10 shows decomposition of the change in markups in a more concise way.

more concentrated markets. While we find that reallocation term is important, it is not the only force at work. Unweighted markups have gone up (measured by the ' Δ within' term and visualized by the density of markups in Figure 7, especially in the upper tail), which is an important force behind the rise in market power. In a general equilibrium model with input-output linkages, Baqaee and Farhi (2017b) find a similar decomposition of the within and the reallocation component.

The third experiment (plot in green) shows the evolution of markups if the only change was net entry of firms. The net entry component rises early on and is more or less constant afterwards, indicating that the rise in markup is not exclusively driven by the changing composition of firms in the sample. The net entry component can simply be driven by the fact that the panel of firms is not balanced and more firms enter than exit. In part, it can also be driven by mergers and acquisitions. Consider two firms that merge. If their joint market share is unchanged but they now charge higher markups, then the net entry term will be positive. Or it could be driven by the fact that the net entry accounts for a higher market share than the sum of the individual pre-merger shares.

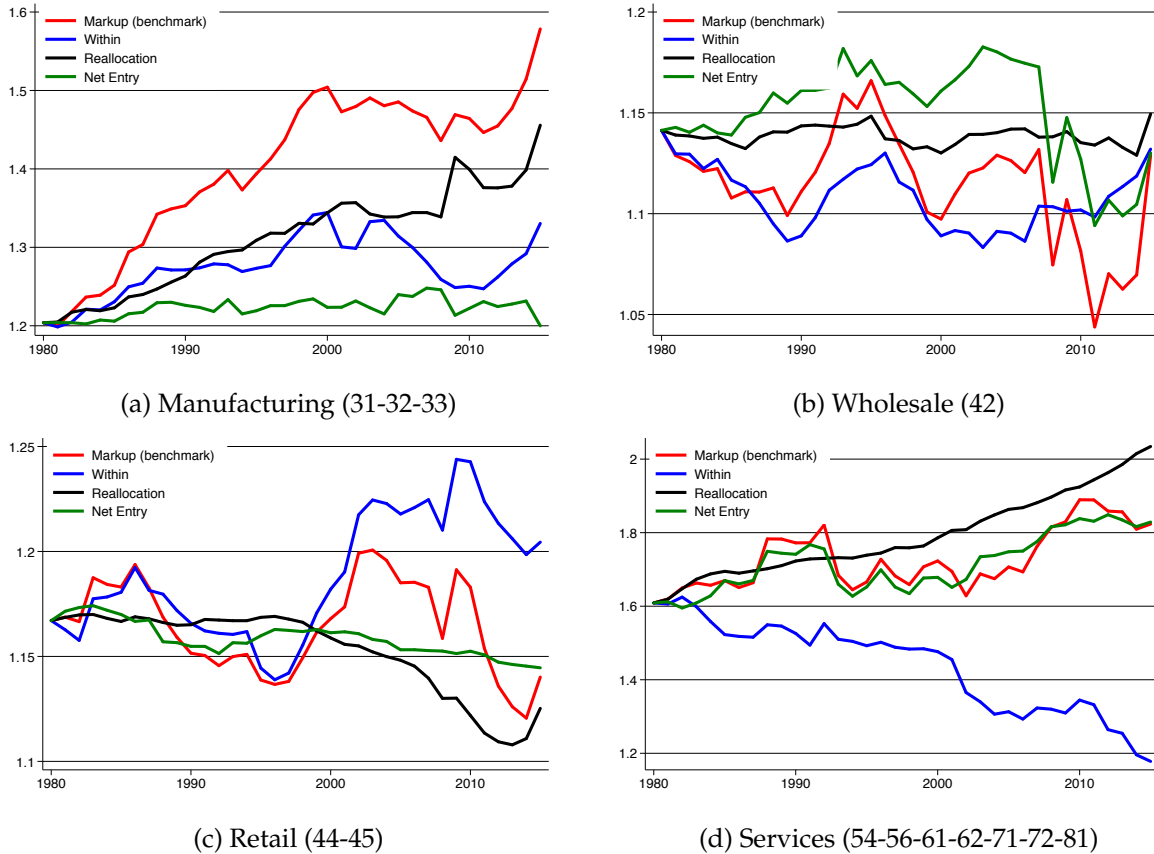


Figure 11: Firm Level Decomposition by Sectors

In summary, the rise in aggregate markups is driven in part by a change in the markup

distribution itself, by a reallocation from low markup firms to high markup firms, and by some net entry. The first decade of the 1980s, all three forces are equally at work. But the end of the period, reallocation dominates.

In Figure 11 we repeat the decomposition exercise with the counterfactuals for a few leading (broad) sectors separately – i.e., Manufacturing, Wholesale, Retail and Services. In the manufacturing sector both the within and the reallocation component contributed positively to the sharp rise in the sector’s aggregate markup. Had market shares been kept at their initial level (here at their lagged levels, in each period we consider), the aggregate markup would only have increased by about 12 percentage points, leaving a role for the reallocation of market share towards high markup firms.

The Wholesale sector looks markedly different, where the aggregate markup is almost uniquely driven by the pattern of within markup growth.³⁹ The other sectors, FIRE, Agriculture, Mining and Utilities all broadly follow the same trend – i.e. the markup growth component tracks the observed aggregate markup closely.⁴⁰

The Retail sector’s aggregate markup, on the other hand, is uniquely driven by overall markup growth leading up to the late nineties. From there onward, the negative reallocation term offsets a strong growth in markups. The latter suggests that high markup firms saw their market share decline in this last period of the sample period. The Services sector, arguably the most heterogeneous group, stands out in that markups consistently decline, at initial market shares. Had it not been for the reallocation of market share towards high markup firms, the sector’s aggregate markup would have declined from about 1.6 in 1980 to about 1.2 in 2016.

The sectoral analysis indicated that the increase in the aggregate markup was largely due to the rise of markups within sectors, and not the reallocation across sectors. Once we consider the micro-data, however, we find that there is, in addition to markup growth, a substantial role for reallocation of economic activity. In almost all sectors and periods this takes the form of a reallocation of market share towards the high markup firms; with the exceptions of the Retail sector in the period 2000-2016, where market shares moved towards relatively lower markup firms. The latter is consistent with the increasing share of online sales.

3.6 Results from the US Censuses

The data on publicly traded firms suffers from selection. So far, what we have analyzed cannot be generalized to the entire US economy. The publicly traded firms tend to be large, and the number of firms (less than ten thousand) is small relative to the approximately six million firms in the economy. Moreover, entry and exit in the sample of publicly traded firms is non-random. Even though the shares of GDP and of employment are large (because the firms are large), we want to find out whether our results are representative for the entire economy.

To that end, we repeat the exercise we have done above for the different Censuses in different industries. The advantage of the Censuses is that they represent the universe of firms

³⁹The net-entry term contribution is harder to interpret given the inherent selection in and out of Compustat.

⁴⁰We present the decompositions for the remaining sectors (Finance, insurance and real estate; Agriculture, Mining and Utilities) in Figure D.1.

within a sector and are therefore representative of the whole economy in that sector. We focus on three Censuses: Manufacturing (NAICS codes 31-32-33), Wholesale (NAICS code 42) and Retail (NAICS codes 44-45). We provide more detail on the sample construction and measurement of the key variables in Appendix A.

The measurement of markups in the census data relies on the framework outlined in section 2. The implementation, however, differs because we do not observe the same detailed information as in Compustat regarding a firm's balance sheet and income and loss statement, with the exception of the census of manufacturing for which we do observe most of the traditional production and cost variables. The analysis of the manufacturing sector will therefore closely track the analysis applied to the universe of Compustat firm. There remains one big difference: there is to our knowledge no analogue to the reporting of SG&A (or overhead cost) in the census data.⁴¹ We therefore restrict our attention to the analysis of the conventional production function (PF1) without overhead as a factor of production, and the associated interpretation of the markups.

With the exception of the census of manufacturing data, we only observe the wage bill and sales consistently across plants and time. This implies that output elasticities cannot be measured or estimated due to the limited information on costs. For Manufacturing, where there is more detailed reporting of costs, we use the industry-time specific cost shares as measures for output elasticities. For Retail and Wholesale, we cannot impute the cost shares. Instead, we use the sector and time-specific output elasticities that we estimated from the publicly traded firms.

In the Census of Manufacturing we use the cost shares to construct the output elasticity of any variable input (labor and materials) at the 4 digit NAICS industry level (denoted by n , by census year.⁴² This leads to the standard recovery of the output elasticity for the variable input:

$$\theta_{nt}^V = N_{nt}^{-1} \sum_{j \in n} \frac{P_{jt}^V V_{jt}}{P_{jt}^V V_{jt} + r_{nt} K_{jt}}, \quad (20)$$

where j denotes a plant active in industry n ; in this case a unique 4 digit NAICS code.⁴³ For Manufacturing we can use information on materials as well as on the wage bill for the variable input V . This allows us to check the robustness of our findings. For the other Censuses, we only observe the wage bill. In the absence on information on cost shares, we infer the output elasticities of labor using the cost-share approach in Compustat. In particular for each 2-digit NAICS sector (s), we compute the median labor cost share, by year, for the sample of active firms, as in equation (15).

⁴¹Some components such as marketing and advertising costs are in principle recorded, but not items such as brand value, research and development, executive compensation packages, etc.

⁴²These are made available by Foster, Haltiwanger, and Syverson (2008) and are accessed through the census file. Alternatively the output elasticities can be obtained by estimating the 86 distinct production functions using an approach as outlined in section 2. We opted to rely on the cost-share approach to minimize the impact of measurement error, and imputed data in obtaining reliable estimates of the output elasticity. See Syverson (2004) and Nishida, Petrin, Rotemberg, and White (2017) for a discussion of these issues.

⁴³We remind the reader that there are four distinct levels of aggregation in our analysis: plants (j), firms (i),

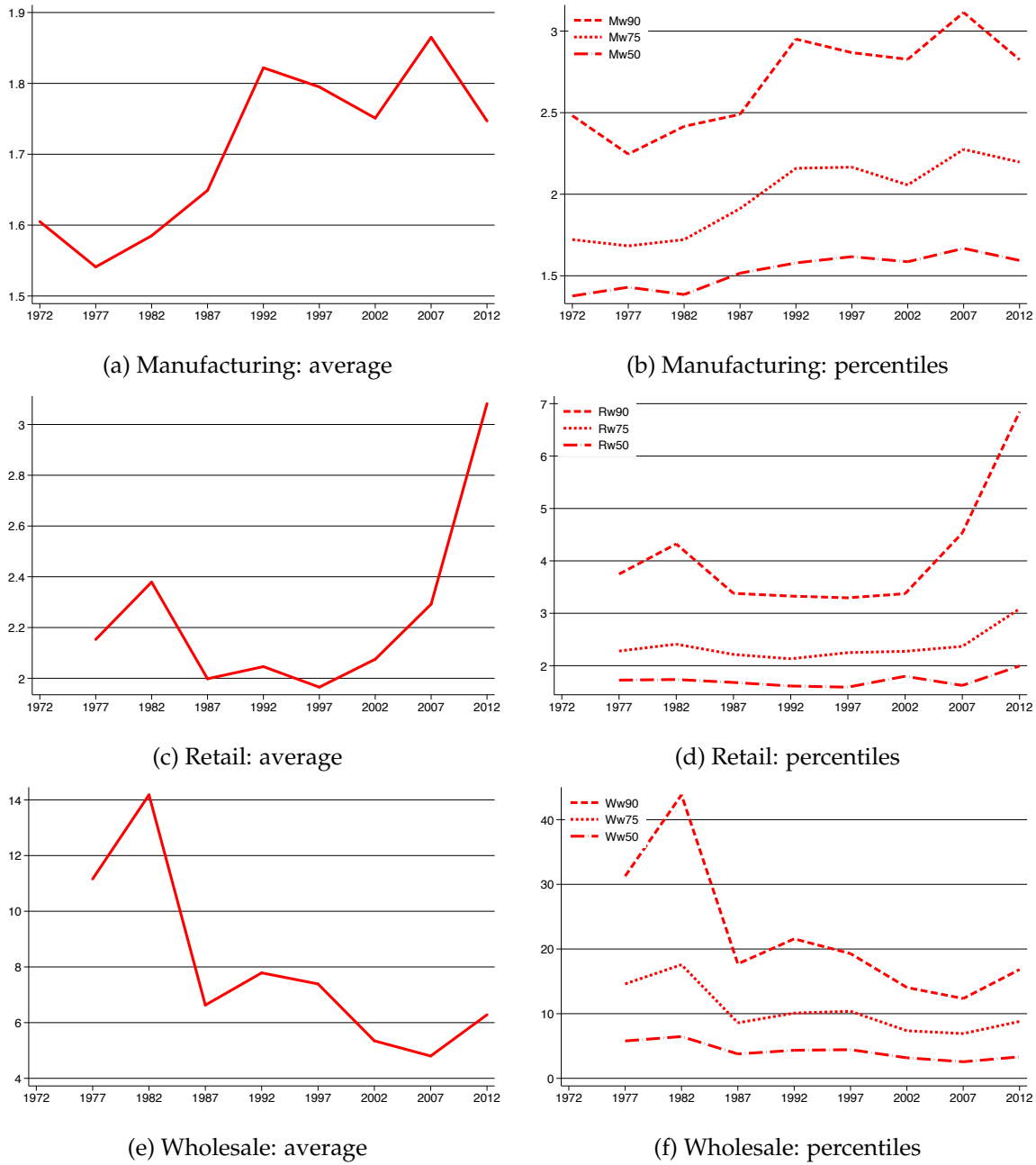


Figure 12: Markups in the US Censuses: Manufacturing, Retail and Wholesale. The variable input is employment. Averages and percentiles are sales weighted. Manufacturing firm-level markups rely on the industry-specific cost shares. Retail trade relies on the output elasticities computed in the Compustat sample. Wholesale relies on a calibrated output elasticity.

Finally, we aggregate the plant-level markups to obtain firm-level markups, the ultimate industries (n – i.e., 4 digit NAICS) and sectors (s – i.e., 2 digit NAICS).

object of interest in this analysis. This also makes our results consistent with the analysis performed for the Compustat sample.⁴⁴ More specifically, we compute markups at the plant-level and aggregate to the firm level using plant-level revenue shares. The sector-specific aggregate markup is computed as before, using a firm's share in total sectoral sales.

Figure 12 reports the weighted average (left panels) for each of the three Censuses, as well as the percentiles of the markup distribution (right panels), weighted by sales (the equivalent of Figure 8a). With data only in five-year intervals, the patterns are obviously less detailed.

Starting with Manufacturing (Figures 12a and 12b), we see average markups that start to increase from 1977 onwards, from around 1.55 up to around 1.8. This pattern mirrors what we find in the whole sample of publicly traded firms as well as in the publicly traded firms in manufacturing in Figure 11(a) (the red line).

We also calculate the markup using manufacturing as the variable input, instead of employment, and we find a very similar pattern. In the Compustat sample we cannot separate the labor and material expenditures, instead we have to rely on the bundle COGS. The results indicate that all three series (Compustat COGS-based, Census labor-based and Census materials-based) indicate the same pattern of rising aggregate markups.⁴⁵

Like for the publicly traded firms, the pattern in Retail (Figures 12c and 12d) until 2002 is flat or only slightly increasing. This is the case also for the percentiles. There is instead a sharp increase of the weighted average in 2012 that we do not observe in the publicly traded firms.

The figures for the wholesale are again in line with the series obtained from our analysis in the Compustat sample. We observe a continuous decline in the aggregate markup until the year 2002, after which we see an increase of about 15 percentage point in the markup over the course of ten years. The percentiles highlight again that the rise is concentrated at the top of the (weighted) markup distribution. In contrast to the results for the manufacturing and retail census, we could not rely on reliable labor cost shares to approximate the time-specific output elasticity. We describe the procedure, and compare the results to reported (aggregate) profit margins in the Online Appendix section 12, but the same message as before still holds: the time-series markup pattern is dominated by the dynamics in the sales-to-expenditure (here the wage bill) ratio, and the output elasticity mostly affects the level.

RETURNS TO SCALE. In all three sectors, we rely on the cost share of labor (also materials in the case of the manufacturing sector) to measure the output elasticity of labor (or materials). This implies that the aggregate markups for the various Censuses (manufacturing, wholesale and retail trade) are obtained assuming constant returns to scale in production. From our analysis on the Compustat sample (see section 3.4), we know that this assumption fails to hold, at least in that sample, and more importantly we documented increasing returns during the period of an increase in aggregate markup in the Compustat sample.

Outside of manufacturing, we do not observe reliable measures of capital and other vari-

⁴⁴In the case of the Wholesale and Retail sectors, the output elasticity is measured at the level of the 2 digit NAICS code – i.e., NAICS 42 and 44-45 combined, respectively.

⁴⁵For more details on the use of multiple variable inputs in the manufacturing sector, see the Online Appendix section 11.

able inputs across a wide range of producers and time to reliably estimate the returns to scale parameters. In order to facilitate comparison across the three Census sectors, and ultimately aggregate across all three of them, we apply the following correction: We impose the time-varying returns to scale parameter estimated on the Compustat sample (see equation (16)). This parameter γ_t was estimated applying the approach of Syverson (2004), which directly relies on the cost shares for each input of production. This approach is a natural aggregate version of the approach taken to compute plant, and subsequently firm-level markups.

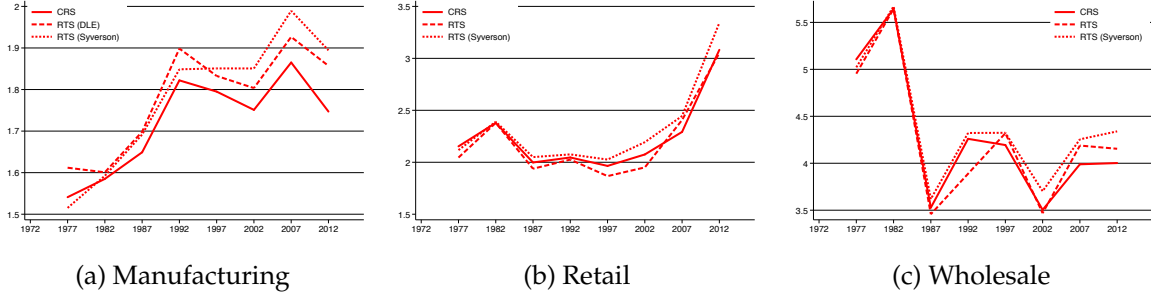


Figure 13: Markups in the US Censuses adjusted for Returns To Scale.

While this correction is applied to the observed labor (and materials) cost shares, as reported in the census data, for the manufacturing sector, for the two other sectors the correction is performed using uniquely Compustat information. First of all, labor cost shares are not reported, or cannot be computed due to the inability to observe the cost of the other inputs of production (capital, and intermediate inputs). Therefore we have to rely on (share-weighted) labor cost shares of the respective firms (in wholesale or retail trade) in the Compustat sample. The returns to scale correction is applied as in the case of manufacturing, by multiplying the aggregate markup by the term γ_t .

The results are reported in Figure 13. Applying the returns to scale correction yields a pattern of markups that is similar to that with constant returns, though with a somewhat more pronounced increase, especially in the later years. This is the case for all three Censuses.

3.7 Collecting evidence

We find agreement between the analysis from Compustat and the micro-census data. This is perhaps not surprising given the fact that in the census data the patterns are driven by the top firms capturing a large share of sectoral total sales (i.e. the upper percentiles), and these are more often than not captured by the Compustat data. However, we also find that, while covering in principle the population of plants and firms in the US, the census data outside of manufacturing substantially limits the ability to estimate markups in a flexible way (relative to the Compustat sample).

The biggest restriction is that one has to rely on labor as a variable input in production. While we were able to formally check materials as a variable input for the manufacturing sector, and found very similar markup patterns, this is not expected to hold outside manufacturing;

especially not for the wholesale sector. The latter is well-known to be a high capital intensive sector where employment is mostly overhead, serving headquarter tasks.

In the absence of any other cost data (on either materials or capital) we cannot estimate the rich production structures with overhead costs as performed here on the Compustat firms, and importantly we cannot entertain the possibility of changing returns to scale (again as we found to matter in the Compustat data).

With those caveats in mind, we do, however find a robust set of facts: aggregate markups increase, and this increase is mostly due to the rise of margins at the top of the weighted markup distribution. There are, however, sectoral differences, and the rise of markups happens at different points in time across the sectors of the economy.

4 Market Power

The rise in markups that we have documented does not necessarily imply that firms have higher profits. Whatever the reason for the increase in markups (a decrease in marginal costs, an increase in demand or in its elasticity, a change in the market structure,...), that does not necessarily imply an inefficiency or a decline in welfare. If at the same time technology has changed to increase fixed overhead costs, then the higher prices are needed to avoid making losses. Consider for example high tech firms that produce software products that need one big up front investment and can be scaled nearly without any additional cost. Such technological change will lead to higher markups (due to lower marginal costs), but prices will not drop because firms need to generate revenue to cover fixed costs. As a result, profits will continue to be low and higher markups do not imply higher market power. From Figure 3b we have learned that indeed, the share of overhead in total expenditure has increased. So the question is whether higher markups simultaneously lead to higher profits.⁴⁶

In part, the objective of analyzing the technology with overhead as a factor of production above (PF2) was to incorporate the change in the fixed cost. In this section, we analyze the implied net economic profits under the two technologies, with overhead as a fixed cost or with overhead as a factor of production. To calculate profits, we use the markup measure and properly account for all costs, including the fixed costs. We then interpret this profit rate as a measure of market power.

Let $\Pi_i = S_{it} - P_t^V V_{it} - r_t K_{it} - P_t^X X_{it}$ denote net profits, then the net profit rate $\pi_{it} = \frac{\Pi_{it}}{S_{it}}$ can be written as:

$$\pi_{it} = 1 - \frac{\theta_{st}}{\mu_{it}} - \frac{r_t K_{it}}{S_{it}} - \frac{P_t^X X_{it}}{S_{it}}, \quad (21)$$

where we have substituted the expenditure on variable inputs as a share of sales with the output elasticity over the markup, from equation (5). This measure of the profit share is differ-

⁴⁶Profits do not necessarily derive exclusively from market power. There could be capital market imperfections that constrain investment and lead to higher profits. However, in a model with both market power and financial frictions, Cooper and Ejarque (2003) find that profitability is explained entirely by market power and none by financial frictions.

ent from the accounting profits because it uses a measure of capital that is obtained from the balance sheet, not the income statement. Given adjustment frictions, the accounting measure does not adequately reflect the expenditure on capital. Note also that our measure of profits incorporates the output elasticity of the production technology, which takes into account that the variable factor of production V adjust while the fixed factors K (and X in the production specification where overhead is a factor) do not.

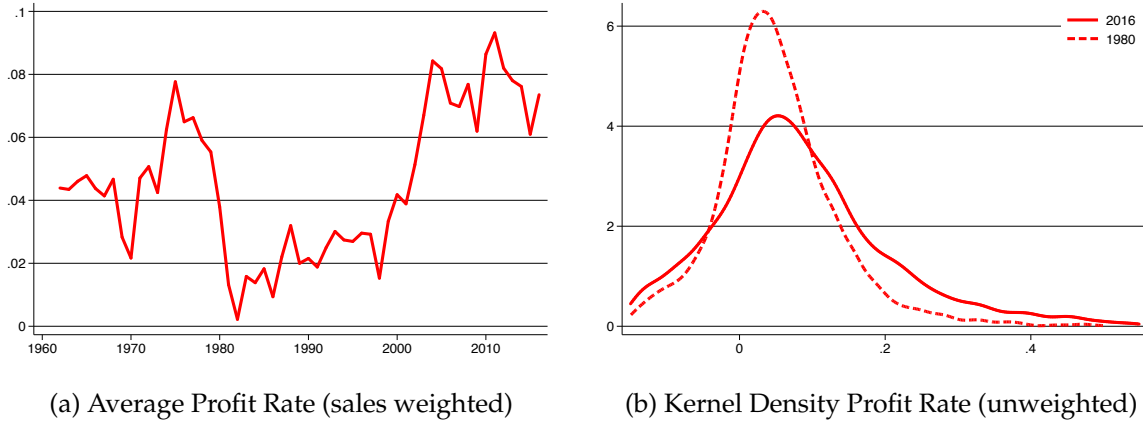


Figure 14: Average Profit Rate and Profit Rate Distribution.

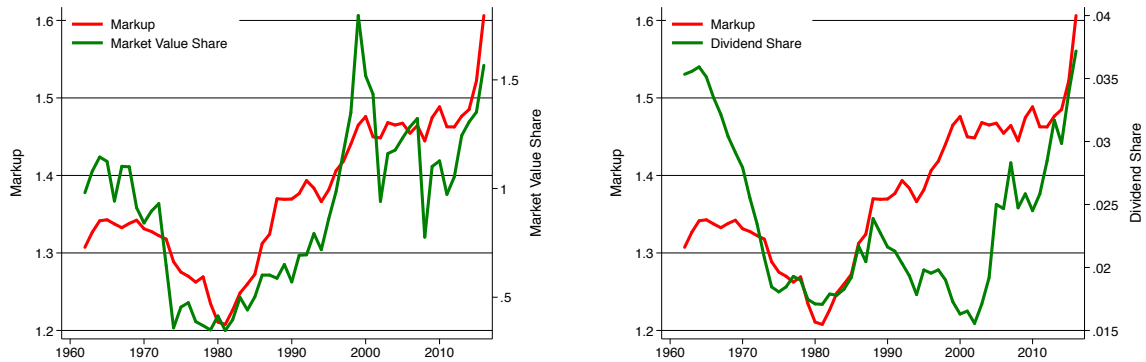
Figure 14a plots the average sales weighted profit rate for the data in our sample. We find that profits have gone up by about 7 percentage points between 1980 and 2016.⁴⁷ Underlying the rise in profits is the increase in the upper tail of the profit distribution. In Figure 14b we plot the kernel density of the unweighted profit rate distribution in 1980 and 2016. The rise in average profit rate is nearly exclusively driven by the increase in the upper percentiles of the profit distribution. More firms have extremely high profit rates of 15% and higher.

Our measure of the profit rate is the firm profits as a share of sales, which effectively scales those profits by the firm size as measured by its revenue. From an investment viewpoint, we may want to measure the return on assets. The return on assets is calculated as the firm profits divided by its assets. We define profits by sales minus all costs, COGS, SG&A and the user cost of capital. Because the user cost of capital is included, our measure of return on assets is the return over and above r , which includes the inflation adjusted risk free rate, as well as an adjustment for depreciation and risk. Therefore, it is the excess return on assets. We plot this in Figure 7.1a in the Online Appendix together with our baseline profit rate. The return on assets is remarkably similar to the profit rate, with an increase starting in 1980 and rising from around 1% to around 8% in 2016. This average return on assets is weighted by the capital of each firm. When we weigh it by the sales of each firm (Figure 7.1b), then the average return on assets is higher and also rising faster. Firms with high sales have higher returns on assets, and the large firms have seen bigger rise in their returns.

⁴⁷This increase may not appear remarkable given the high profits in the 1970s. However, that dramatic spike in profits in the mid 1970s is entirely driven by the drop in capital expenditure during a period of high inflation. The role of the drop in the user cost of capital due to inflation can also be seen in Figure 6.1 in the Online Appendix.

All this seems to suggest that at least based on the flows reported in the accounting data, starting in 1980 there is an increase in the profitability of firms, and therefore an increase in market power. Note that the profit rate we have reported accounts for the increase in contemporaneous overhead costs as measured by SG&A. Of course, some costs may have been incurred earlier. Still, it is not clear what those startup costs may be as they are not booked in the firms' accounts, and firms have incentives to book as many costs as possible to reduce corporate taxes on profits. The only possibility is that those startup costs were incurred before the firms were observed in our data. As a result, profits based on contemporaneous costs may therefore be overstated. What the data is indicating however, is that if such costs are incurred earlier, then there must be an increase in those startup costs as a share of the sales of a firm since 1980. With free entry and hence zero ex ante expected profits, what we expect is that over the last four decades, the unmeasured startup cost as a share of future sales has gone up from 1% of sales to 8% of sales (roughly from 2% of value added to 16%). Some of those costs could be R&D cost that were incurred before the firms were observed in our data. We turn to the impact of recorded R&D costs below.

The flow of profits may not be the best measure of profitability of the firm, because it mixes up the firm's result with investment decisions. To that effect, we consider as a measure of profitability based on what firms generate as a return to their shareholders. For that we have two measures: 1. the market value (or market capitalization); and 2. dividends. Our second measure, dividends, is the return an investor receives on holding equity in the firm. Of course, dividends may vary for reasons that have nothing to do with the actual flow of profits. In particular, they will be closely related to the investment opportunities that the firm has. Still, over a long enough horizon and averaging out over a large number of firms, we would expect that dividends are a good indicator of profits. Our first measure, market value, is essentially the discounted sum of dividends, since a shareholder who sells shares in a firm gives up the opportunity value of receiving the indefinite stream of dividend payments. In contrast to the actual dividends, the market price is more a measure of future expected profits, not just contemporaneous profits, since it takes into account the flow of all expected future dividends.



(a) Average Market Value (share of Sales), Markup (b) Average Dividends (share of Sales), Markup

Figure 15: Market Value and Dividends.

Figure 15a shows the evolution of the market value as a share of sales, averaged by the sales share in the entire economy: $\sum_i \frac{S_i}{\sum_i S_i} \frac{MktVal_i}{S_i} = \frac{\sum_i MktVal_i}{\sum_i S_i}$. Unlike standard composite indices of stock market values like the S&P 500, this measure is a “rate” that can be interpreted in conjunction with the profit rate π (profits as a share of sales) from our model. As such, first, it is not affected by inflation⁴⁸ and second, this measure is independent of the size of firms or the composition of firms since it is normalized by sales. For example, even if there are 500 firms in the index, the index will artificially grow when firms become larger, e.g., due to mergers.⁴⁹

If the flow profits and dividends as a share of sales were constant, then the market value that reflects the discounted stream of dividends would be constant as a share of sales. This is clearly not the case. Market value as a share of sales rises from less than 50% in 1980 to over 150% in 2016 (Figure 15a, right scale). A similar pattern arises for dividends, where dividends as a share of sales increases from 1.7% in 1980 to over 3.5% in 2016 (Figure 15b).⁵⁰

This is not just an artifact of the aggregate data. At the individual firm level, firms with higher markups also have higher market values and dividends. In Table 2 we report the regression results.⁵¹ Not surprisingly, contemporaneous firm-level markups are correlated with both market value and dividends. For all specification, the coefficient is a highly significant (even in the presence of firm fixed effects, see columns (4) and (8)). At the firm level, this is consistent with the fact that higher markups reflect higher profits and therefore higher dividends and market values.

Based on the evidence from the firm’s fixed overhead as measured by SG&A and the resulting profits as well as by market value and dividends, we find evidence that the rise in markups is associated with the rise in market power.

In addition to technological change that gives rise to an increase in the fixed overhead cost, technological change can also lead to higher returns to scale. As we have established above (Figure 5), we find evidence of an increase in the returns to scale of the technology. This does of course not necessarily imply that the technological change is socially optimal.

To complete this Section, we investigate the relation between profits, markups and overhead costs (SG&A). In Figure 16a we plot the relation between the share of sales of SG&A and the markup for different percentiles in the (unweighted) markup distribution. This shows that the firms with a higher SG&A share of sales have higher markups. For a given year, the higher percentiles in the distribution of markups have higher overhead shares. This is as expected in a competitive economy: higher prices relative to marginal cost are required in order to offset

⁴⁸The increase of the Dow Jones in the 1970s for example is misleading because during that period of high inflation, once adjusted for inflation the real index is actually decreasing.

⁴⁹Interpreting the market valuation of a firm as the discounted stream of profits obviously imposes a set of assumptions. The most important one is the fact that the discount rate has remained constant over the period. We know that the risk free rate has decreased, especially since the 1990s. While the interest rate is not the discount rate, a preference parameter, it is quite feasible that the risk free rate affects the valuation of stocks. And of course, changes in legislation affect tax incentives and therefore firm valuation (see Smith, Yagan, Zidar, and Zwick (2017)).

⁵⁰Note that the reason for the decline in dividends the 1990s and sudden increase until the early 2000s is due to tax incentives for firms to issue dividends. Until the 2003 tax reform, dividends were taxed at the individual’s income tax rate, and at 15% thereafter.

⁵¹In Table 13.1 in the Appendix we report the same regressions for our markups estimated with the technology that includes overhead as factor of production, PF2. The coefficients are very similar.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\ln\left(\frac{\text{Market Value}}{\text{Sales}}\right)$				$\ln(\text{Market Value})$			
$\ln(\text{Markup PF1})$	0.71 (0.03)	0.64 (0.02)	0.56 (0.02)	0.17 (0.03)	0.71 (0.02)	0.65 (0.02)	0.58 (0.02)	0.27 (0.02)
$\ln(\text{Sales})$					0.81 (0.00)	0.81 (0.00)	0.83 (0.00)	0.68 (0.01)
Year Fixed Effects		Y	Y	Y		Y	Y	Y
Sector Fixed Effects			Y				Y	
Firm Fixed Effects				Y				Y
R^2	0.05	0.13	0.21	0.68	0.68	0.71	0.73	0.89

	$\ln\left(\frac{\text{Dividends}}{\text{Sales}}\right)$				$\ln(\text{Dividends})$			
$\ln(\text{Markup PF1})$	1.05 (0.04)	0.97 (0.03)	0.80 (0.04)	0.26 (0.05)	1.03 (0.04)	0.93 (0.04)	0.78 (0.04)	0.26 (0.05)
$\ln(\text{Sales})$					0.94 (0.01)	0.92 (0.01)	0.93 (0.01)	0.76 (0.02)
Year Fixed Effects		Y	Y	Y		Y	Y	Y
Sector Fixed Effects			Y				Y	
Firm Fixed Effects				Y				Y
R^2	0.06	0.11	0.17	0.70	0.66	0.68	0.70	0.89

Table 2: firm-level Regressions: market values and dividends on markups (clustered standard errors by firm in brackets).

the overhead and avoid making losses. In addition, over time, the overhead share is increasing which automatically implies that the markup increases, even in a competitive economy. Note that if we plot the markup against the share of COGS in Sales, then by construction this relation is downward sloping, indicating that unlike SG&A, COGS is a variable input.

Now we want to evaluate whether the increase in markups that we observe is merely to offset the rise in overhead. To that effect, we calculate a fictitious markup, denote it by μ^* , that corresponds to zero profits. We obtain that markup from setting profits π_{it} to zero in equation (21) and solving for μ :

$$\mu_{it}^* = \frac{\theta_{st}}{1 - \frac{r_t K_{it}}{S_{it}} - \frac{P_t^X X_{it}}{S_{it}}}. \quad (22)$$

This zero profit markup is a weak upper bound, however, and the true zero profit markup is weakly lower (provided there are no costs in addition to COGS, SG&A and capital). This is because we do not know what sales S_{it} would be under competition. To predict sales under

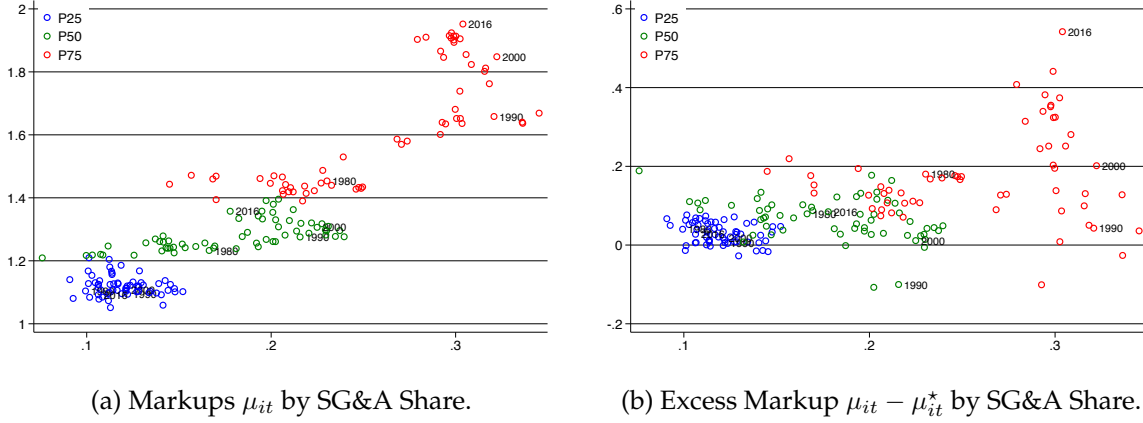


Figure 16: Markup, Excess Markup and SG&A Share (Markup PF2).

perfect competition, we need to know the properties of demand. Only in the case of unit elasticity demand will sales be invariant for different markups. In all other cases, however, sales under perfect competition will be lower than when there is market power. This is due to the fact that firms are charging higher prices only if the marginal revenue is positive, which by definition necessarily implies higher sales for higher markups. Therefore, sales under perfect competition (S_{it}^*) will be weakly lower than than under market power. Since under our assumption that in the short run, K_{it} and X_{it} are not variable, the expression in equation (22) where we use S_{it} instead of S_{it}^* is weakly higher than the true zero profit markup.

In Figure 16b we also plot $\mu_{it} - \mu_{it}^*$ for different percentiles in the markup distribution. Because μ_{it}^* is the upper bound of the zero profit markup, the gap between the actual markup and μ_{it}^* indicates the extent of the excess markup, over and above the markup that arises under perfect competition. We see that the excess markup is highest for the highest percentiles of the markup distribution, where incidentally the SG&A share is the highest as well. High overhead firms have high markups, but also high excess markups; and this became stronger over time (excess markup from about 0.2 in 1980 to about 0.6 in 2016).

When we analyze the relation between markups (and profits) and overhead at the individual firm level, we find a strong positive relation, as expected. As we have pointed out all along, one of the reasons for raising prices and markups is that overhead has increased. The elasticity is 0.56 (see Table 3): only over half of the SG&A increases are passed on to markups. In a competitive economy this should be 1. Interestingly, firms with higher SG&A also have higher profits. In a competitive market, this coefficient should be zero. We can decompose the change in SG&A into R&D expenditure and Advertising expenditure. These are often signaled as the components of SG&A that are important for intangible capital. Indeed, R&D expenditure has risen from 5% in 1980 to 20% of SG&A, and advertising from 4% to 10%. Even in 2016, these remain relatively minor shares of SG&A. The majority is still sales related and administrative expenditure. We find that the elasticity of R&D Expenditure on markups is 16% and 5% for Advertising Expenditure. Interestingly, most of that effect remains when the dependent variable

	Markup (log)			Profit Rate (log)	
	(1)	(2)	(3)	(4)	(5)
SG&A (log)	0.56 (0.01)			0.15 (0.03)	
R&D Exp. (log)		0.16 (0.01)			0.10 (0.01)
Advertising Exp. (log)		0.05 (0.00)			0.03 (0.01)
R&D dummy			0.06 (0.01)		
Advertising dummy			-0.00 (0.03)		
R ²	0.61	0.07	0.43	0.04	0.05
N	26,743		247,615	26,743	

Table 3: Regressions: effect of SG&A, R&D Expenditure and Advertising Expenditure on markups and Profit Rate; Extensive margin effect of R&D and Advertising.

is the profit rate. This elasticity should be zero under competition. Most of R&D and Advertising Expenditures translate in profits as much as they do in higher markups. These are all at the intensive margin. When we evaluate the extensive margin – whether a firm does or does not have expenditures on R&D or Advertising – we find an elasticity of 6% from R&D and no significant effect from Advertising (since nearly all firms have Advertising expenditure, there is not enough variation; only about 10% of the firms report R&D expenditure).

5 Discussion and Robustness

We now report a number of robustness exercises. One of the problems with the Compustat sample is that it is not representative of the US population of firms. We already established that relying on the micro-Census data confirms the results. Still, the US Censuses we analyze do not encompass all sectors in the economy and there is no data on overhead or capital allowing the estimation of production functions. We therefore also perform robustness exercises using the data for the publicly traded firms. In particular we consider the following checks: 1) Reweigh the Compustat sample according to the actual economy-wide weights; 2) Exploit the Compustat segment data is used to distinguish consolidated and unconsolidated accounts; 3) Consider different production technologies (translog); 4) Apply different aggregation weights; 5) Compute markups using the sub-sample of firms reporting the wage bill and 6) Contrast variable and fixed factors of production and the implications for markup estimation.

USING ECONOMY-WIDE WEIGHTS TO ADJUST THE SAMPLE. Because the Compustat sample is not representative of the US population of firms, we perform the analysis on the Compustat sample while correcting for the aggregate economy firm distribution across sectors. We

use data from the Bureau of Economic Analysis (BEA) on the economy-wide firm population weights.⁵² While the information on individual markups in a sector is still only based on the publicly traded firms, the weight we give to each sector is the exact population weight of that sector in the economy-wide distribution. This has two advantages. First, the average we calculate accurately represents the sectors in the economy, so we do not overweigh those sectors that are overrepresented in the Compustat sample. Second, and possibly most importantly, the evolution of the average markup is now not marred by an evolution in our sample that is different from the evolution economy-wide. For example, suppose markups in each sector remain unchanged and sectors in the economy are stable, but a high markup sector becomes more prevalent over time in the Compustat sample. Then measured average markups go up simply because of this bias in the composition. Using the economy-wide sector weights adjusts for this.

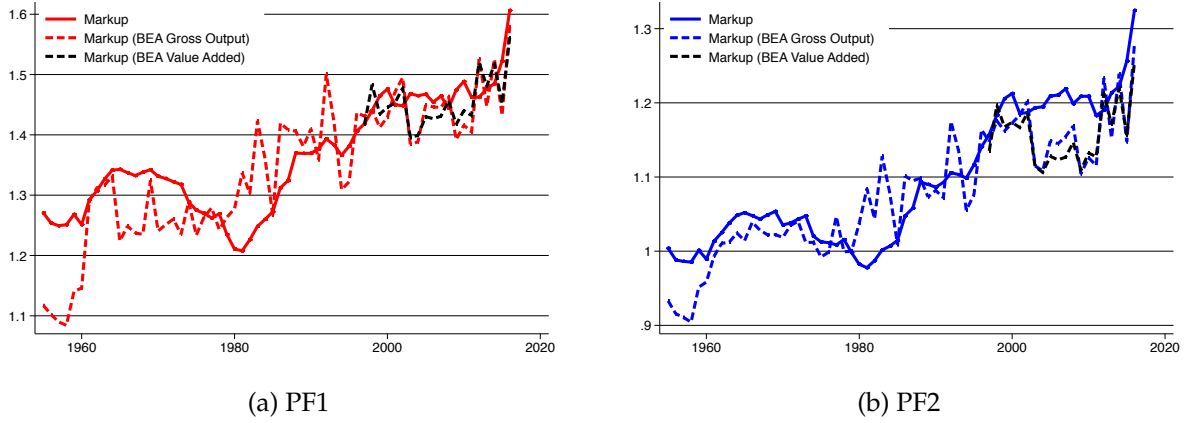


Figure 17: Markup of Compustat firms with economy-wide sample weights for 2-digit sector (from BEA).

We use the Bureau of Economic Analysis 2-digit economy-wide sales distribution to construct the sample weights (we use both the BEA value added and gross output weights).⁵³ Once we adjust for the economy-wide sample weights, we obtain a measure for markups (both for PF1 in Figure 17a and PF2 in Figure 17b) that is more noisy but close to the one obtained with the weights from our data of publicly traded firms. This indicates that the sectoral composition selection of the publicly traded firms does not engender a systematic and substantial bias in the estimated markups. There is no level difference that would indicate that the Compustat composition tends to favor sectors with higher/lower markups. Nor is there a systematic difference in the evolution of the adjusted market, which would indicate that the composition of sectors in Compustat is changing relative to the economy (say from low markup sectors to high

⁵²We are grateful to Eric Hurst and two referees who suggested using the BEA weights.

⁵³In Figure 1.1 in the Online Appendix, we plot the scatter plot of the Compustat weights against the BEA sample weights. The samples are not identical: 'Manufacturing' in particular is overrepresented in the publicly traded firms, and so are 'Information' and 'Mining, Quarrying, and Oil and Gas Extraction'. All the others are underrepresented.

markup sectors, which would imply that the BEA-weights adjusted measure would be flatter). This is consistent with the findings above that most of the increase in markups is driven by an increase within sector, not between sectors. Of course, the markups within sector are still calculated for publicly traded firms only, even if we adjust the sectoral weights, but this was addressed using the US Censuses data.

CONSOLIDATED ACCOUNTS: GEOGRAPHICAL DECOMPOSITION. One of the concerns in using the publicly traded firms is that the data is for consolidated accounts. Especially for large firms active in many different markets, this has implications for the aggregation. In the consolidated accounts, a firm is classified to belong to the sector where it has its main activity, even if this firm is active in other sectors. Likewise, the activities of firms that have foreign subsidiaries are all bundled together with the domestic activities (sales and cost of goods sold). Since 2009, the Compustat data provides segment data that reports both the geographical as well as the sectoral decomposition of the firm's activities. Unfortunately, the segments data is sparsely populated, so we can only consider the unconsolidated accounts for a small subset of firms.

To that effect, we analyze the markup of those firms that report foreign activity and focus only on those sales and costs that are reported to be domestic. To calculate the markup we use the output elasticity estimated in the benchmark model. We thus obtain a measure for the firm's individual markup based on its domestic activity only, denoted by μ_{it}^D .

Figure 2.1a in the Online Appendix reports the scatter plot of this measure of the domestic markup and our benchmark measure of the markup, with one observation for each firm in this subsample and for each year. The scatter points are weighted by the firm's sales share in the full sample for that year. All observations line up along the 45 degree line, indicating that there is little systematic difference between the markup based on the consolidated accounts and the markup based on the domestic accounts only.

To get an idea how the average markup compares, we calculate the average markup for the subset of firms for which we have geographically disaggregated information. Denote that subset by \mathcal{G} . We measure the ratio between the two measures of markup, one weighed by the sales shares of the entire sample of firms, the other weighted only by the share of firms in the subsample \mathcal{G} .⁵⁴ Both measures give very similar ratios, that are very close to one (Figure 2.1b). Again, this confirms that at least in the subset of firms for which we have geographically disaggregated data, there is little reason to suspect that the consolidated accounts lead to a bias in the markup estimates. The average markup based on the domestic activities is within half a percentage point of the average markup based on the consolidated accounts, except for the first year where the data is available (2009).

Finally, we restrict the sample to those firms that are incorporated in the United States. Figure 2.2 in the Online Appendix. We find that the pattern of the average markup is very

⁵⁴Formally, these ratio are obtained as:

$$\frac{\sum_{i \in \mathcal{G}} \frac{S_{it}}{S_t} \mu_{it}}{\sum_{i \in \mathcal{G}} \frac{S_{it}}{S_t} \mu_{it}^D} \text{ and } \frac{\sum_{i \in \mathcal{G}} \frac{S_{it}}{S_t(\mathcal{G})} \mu_{it}}{\sum_{i \in \mathcal{G}} \frac{S_{it}}{S_t(\mathcal{G})} \mu_{it}^D}. \quad (23)$$

similar to that for the full sample. The country of incorporation of course does not preclude a firm from having foreign sales, but the notion is that firms incorporated abroad may have a higher share of their sales outside the United States. Note that this sample excludes firms incorporated in tax havens (especially the Cayman Islands) even if they could be considered American firms. Since the tax incentives to incorporate in a tax haven are largest for those firms with the highest profits, this sample restriction tends to bias the sample towards low profit firms.

TRANSLOG PRODUCTION FUNCTION. The translog production function is a common technology specification that includes higher order terms. While the estimated parameters of the technology are constant, the implied elasticities are time varying because the output elasticity is a function of those constant parameters as well as the level of the input (see Appendix B.2). The technology allows for returns to scale and those can be varying over time. This is a parsimonious way to estimate time-varying technologies with returns to scale while using the full sample. There are of course strong structural assumptions, so it is not necessarily the case that the estimated translog technology will generate the same patterns for average markups as our baseline model.

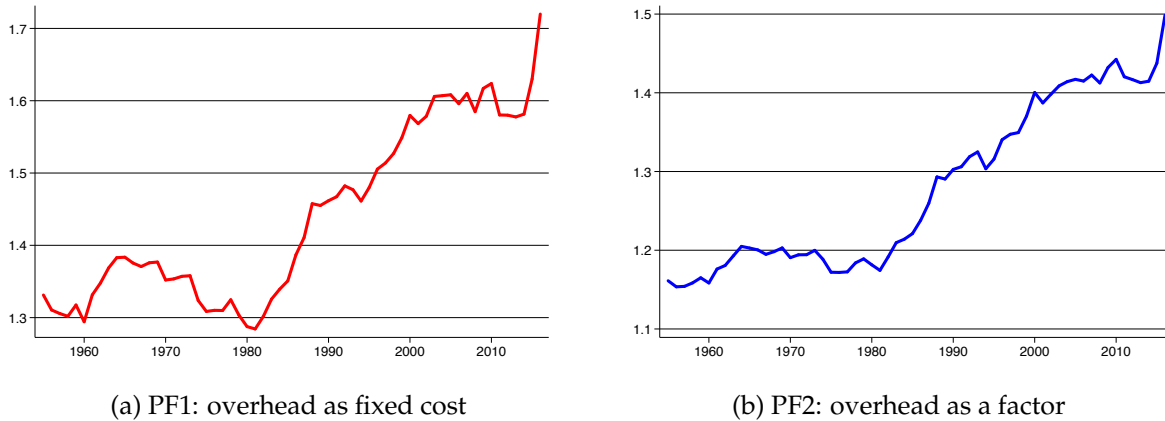


Figure 18: Average Markup with the Translog Production Technology.

In Figure 18a we report the average markup for the conventional translog production function with overhead as a fixed cost and in Figure 18b for the translog production function with overhead as a factor. In both figures, the level of the markup is 10 to 15 percentage points higher than in the baseline (Figures 1 and 4a), which is due to the estimated output elasticity being systematically higher under translog. But what is remarkable is that the evolution for both technologies (with overhead as a fixed cost and overhead as a factor) is very similar to that in the baseline model. There is a hump-shaped pattern with minor variation before, and a steady increase in the average markup after 1980. The increase for the traditional translog production function with overhead as fixed cost is 40 points and for the translog production with overhead as a factor it is 30 points. In both cases, these increases are very similar to those estimated in the baseline model. Recall that the baseline specification is considerably more flexible

because we estimate an output elasticity for each year and for each 2-digit sector. Nonetheless, the translog production function generates remarkably similar elasticity estimates and therefore similar average markup estimates as our baseline specification.

INPUT WEIGHTS AND JOINT DISTRIBUTIONS (CONTOUR PLOTS). The pattern of aggregate markups depends on the aggregation weight used. Each firm is given a weight that reflects the importance of that firm. So far, we have used the share of Sales. This is the most common, it is used for example to construct the HHI. So far we made no assumption on demand and hence there is no welfare measure that guides us which aggregation weight to use. In Appendix C we report the joint distribution (by means of a contour plot) of the individual level markups together with sales. We also include the contour plot of markups with employment and the variable input (COGS). In addition, we report the the average markup using the input or employment weights. From the contour plots and the weighted average markups calculated with the employment and the input (COGS) weight. There are differences in the magnitudes, which is not surprising since we have found that markups increase both due to the increase of markups themselves (especially in the upper tail) and the reallocation of sales towards high markup firms. Overall however, we find that the pattern of the increase in markups starting in 1980 is robust across the weighting measures used: COGS, Capital, Total Cost, Employment and the Harmonic Sales weighted Mean.⁵⁵

An important conclusion to take away from these alternative measures for average markups is that they are different moments of a much richer distribution of markups. We have documented that the distribution has a fairly constant median, that the upper tail has become a lot fatter, and that within a market, larger firms tend to have higher markups. For example, input weighted markups are lower than sales weighted markups because as market power rises, firms raise prices and sales, but they produce less, and as a result they reduce inputs (employment and materials).

MARKUPS BASED ON WAGE BILL. We have chosen to do our analysis on a bundle of variable inputs, Cost of Goods Sold. This is imperfect because it assumes that everything in the bundle is variable, they are assumed to be perfectly substitutable, and there is one output elasticity. Instead, if we had a vector of variable inputs (say materials M and labor L), we could allow for imperfect substitutability between these two inputs.

Unfortunately, Compustat data is highly sketchy to decompose the bundle of variable inputs in materials and labor. While most firms report the number of employees, less than 10% of the firms report the wage bill.⁵⁶ In addition, a portion of employment such as salaries of management, sales personnel,... is booked under overhead (SG&A) and not under Cost of Goods Sold. Nonetheless, as a robustness exercise, we attempt to construct a sales weighted average markup based on employment rather than COGS. In Figure 19 we report the markup estimate for different specifications.

⁵⁵It can be shown analytically that for the Cobb-Douglas production function, the input (COGS) weighted mean is identical to the Harmonic Sales weighted mean.

⁵⁶Reporting the wage bill is not an SEC requirement.

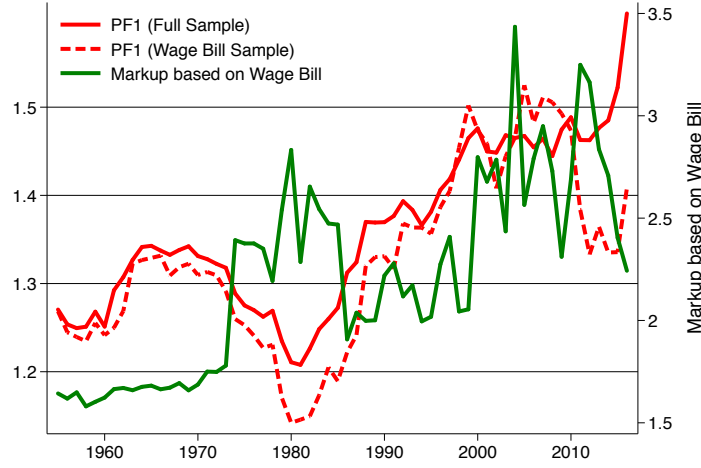


Figure 19: Average Markup derived from Wage Bill: PF1 (full sample); PF1 (report XLR);

VARIABLE VERSUS FIXED INPUTS: IMPLICATIONS FOR MARKUP MEASUREMENT. As discussed above, our methodology indicates that *in principle* we can rely on multiple first order conditions and thus multiple variable inputs to compute markups. If we were to observe multiple such variable inputs, and they satisfy the conditions (no adjustment costs or no frictions), then the implied aggregate markups would be identical.

Throughout, we have assumed that our variable input V , cost of goods sold (in Compustat) or labor and intermediate inputs (in Census), adjusts instantaneously, while Capital and Overhead do not.⁵⁷ We have not specified a model for how Capital and Overhead adjust, whether there are frictions, adjustment costs, delays,... Instead, let us investigate the premise whether Capital and Overhead can be treated as variable inputs. If they can, then applying the first order condition (5) to K and X instead of to V would give us an expression for the markup μ_{it}^K and μ_{it}^X :

$$\mu_{it}^K = \theta_{it}^K \frac{P_{it}Q_{it}}{P_{it}^K K_{it}} \quad \text{and} \quad \mu_{it}^X = \theta_{it}^X \frac{P_{it}Q_{it}}{P_{it}^X X_{it}}. \quad (24)$$

If K and X are variable inputs that adjust within one period, then the markups in (24) are identical and equal to the markup μ_{it} that we obtained for V (provided V is variable and fully adjusts within one period). This offers an over-identifying restriction of the model, and a simple way to evaluate whether K and X are variable. In Figure 20a we report the markups μ_{it}^K and μ_{it}^X , if these inputs were indeed variable and statically optimized.⁵⁸

⁵⁷It is only due to the data restriction in Compustat that we do not further break up cost of goods sold into materials, wages and energy expenditures. In the Online Appendix section 11 we consider the census of manufacturing plants and consider material inputs separately.

⁵⁸The output elasticities are obtained from estimating PF2 for each sector as discussed before. We thus abstract from the modeling decision whether the fixed factors even belong in the production to begin with. We allow for the most extreme scenario, and of course if SG&A is not included in the production function, the entire discussion on what the correct markup pattern is becomes mute. The cost-share approach is of course by construction not useful, as each markup measures collapses to the sales-total cost ratio, due to the imposition of constant returns to scale

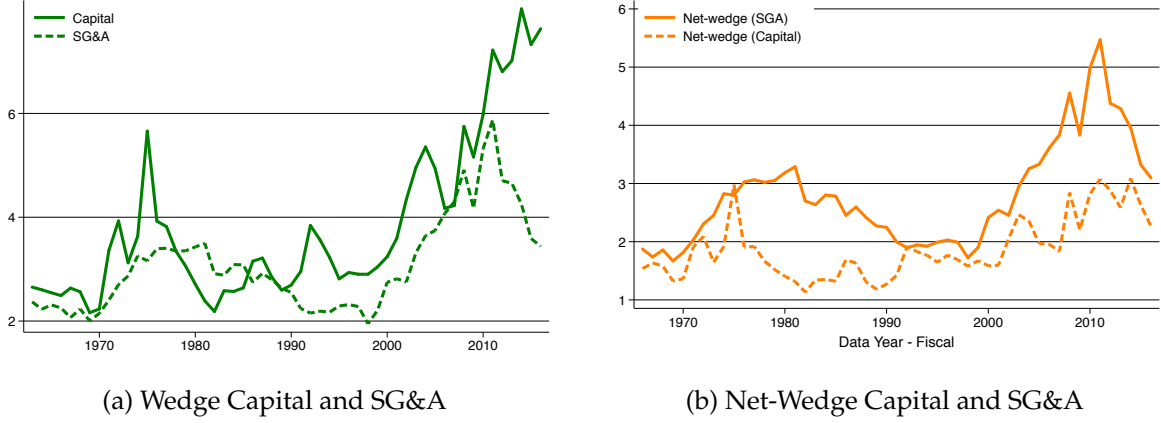


Figure 20: Fixed Factors and Markup Measurement

We draw three conclusions. First, the level of these markups is a lot higher than that of our baseline measure. This tells us that K and X do not adjust as flexibly as V . Second, the increase over the period is a lot steeper, from 2 to up to 8 for K and reaching 6 in 2010 for X . While we are agnostic about the model, this indicates that in the long run these measures of markups are increasing, and if anything, more steeply than the baseline measure. Third, there is a delay. The increase for μ_{it}^K and μ_{it}^X does not really take off until the end of the 1990s, nearly two decades after the benchmark measure started to increase. Incidentally, the 2000s is the decade when the increase in the benchmark measure μ_{it}^V slows down. The delay appears to indicate that there are indeed adjustment frictions of Capital and Overhead.

The different levels and time profile is easily understood if we go back to the underlying cost minimization framework. For any factor of production that is fixed in nature, or faces any friction or adjustment cost, the first order conditions is invalid. In particular, even in a perfectly competitive environment, there would be a wedge between this input's revenue share and the output elasticity. This seems highly plausible for the items listed under SG&A; such as managerial compensation, advertising, marketing, R&D, sales expenses, etc.

In our framework we can identify the reduced-form net-wedge, net of the markup wedge. To see this consider the reduced-form first-order condition for the fixed factors (K and X) which generates an expression for the net-wedge (ψ) in terms of data and the (estimated) markup:

$$\psi_{it} = \frac{\hat{\mu}_{it}}{\mu_{it}} = \frac{\theta^K P_{it}^K K_{it}}{\theta^V P_{it}^V V_{it}}, \quad (25)$$

where ψ_{it} denotes the firm-level net-wedge, which can be interpreted as adjustment costs and any other friction preventing the static optimization of the factor of production. In Figure 20b we plot the share-weighted average over time, for both SG&A and capital. A wedge equal to one implies that the factor K or X is as variable as the variable factor V . Both wedges are substantially larger than one and have been increasing since 1990. The wedge for SG&A (X) is

and *all* factors of production to be variable.

substantially higher than that for K , reaching levels higher than 5 around 2010.

We believe this conclusion is important in the light of the approach that Karabarbounis and Neiman (2018) and Traina (2018) take. They derive markups under the assumption that K and X are variable, in addition to the assumptions that they are perfect substitutes and that they satisfy constant returns to scale. We reject all three assumptions, and this is important to correctly measure the pattern of markups, across firms and time, and how it relates to salient secular trends in the aggregate economy. While summing a variable and a fixed factor of production in any given year, is invalid, it does not invalidate inspecting the long term trend of the fixed factor's wedge between the revenue share and the output elasticity.

6 The Macroeconomic Implications

The focus of our analysis so far has been on documenting in detail the time series and cross section evolution of market power. We now turn to discuss the macroeconomic implications of the rise in market power in the last decades.

THE SECULAR DECLINE IN THE LABOR SHARE. In the national accounts, the labor share of income measures the expenditure on labor (the wage bill) divided by the total income generated (value added). While there are business cycle fluctuations, the labor share has been remarkably constant since the second world war up to the 1980s, at around 62%. Since 1980, there has been a secular decline all the way down to 56% (Bureau of Labor Statistics Headline measure).⁵⁹ The decline since the 1980s occurs in the large majority of industries and across countries (see Karabarbounis and Neiman (2014) and Gollin (2002)).

Economists have not found conclusive evidence for the mechanism behind the decline in the labor share. There are several candidate explanations. Karabarbounis and Neiman (2014) hypothesize that the decrease in the relative price of investment goods, due to information technology, can explain half of the decline.⁶⁰ An alternative explanation is the composition of manufacturing and services. Manufacturing tends to use a higher labor share than services, so it seems natural that with a change in the composition of industry shifting from manufacturing to services, the labor share will decrease. However, this transition does not coincide with the timing of the decrease in the labor share. In fact, most of the transition of manufacturing to services happened before the 1980s (between 1950 and 1987, the share of manufacturing in output dropped from 26.8% to 18.1%, while the share of services doubled, from 21 to 40%, and that transition has slowed down since 1987, see Armenter (2015)).

Koh, Santaaulia-Llopis, and Zheng (2017) offer yet another (and appealing) explanation, which is based on the increasing importance of intangible capital and its incomplete measurement as part of capital in aggregate data. Firms now invest substantially more in intellectual

⁵⁹There are issues of measurement. See Elsby, Hobijn, and Şahin (2013) on the role of how labor income of the self-employed is imputed. Even after adjusting for measurement issues, the labor share still exhibits a secular decline.

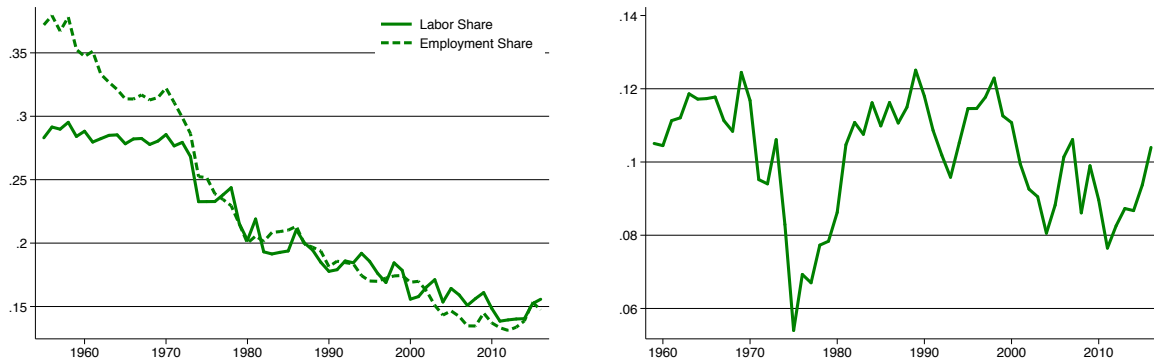
⁶⁰Antras (2004) finds evidence that the technology is not Cobb-Douglas, which when estimated under competition, is consistent with what we observe when there is a change in the markup.

property products and this leads to a lower expenditure on labor.⁶¹ However, in their world with perfect competition, this measurement issue should not lead to an increase in the total profit share. As we have documented above, there is a substantial increase in the profit rate. If intangibles play a role, it must allow firms to exert more market power, which is the central thesis of our paper. We do find evidence that expenditure on overhead has increased (see Figure 3b), which could certainly include intangibles, but we also find that economic profits increase even if we interpret overhead (and hence intangibles) as a factor of production (see Figure 14a). Finally, Elsby, Hobijn, and Şahin (2013) find little support for capital-labor substitution, nor for the role of a decline in unionization. They do find some support for off-shoring labor-intensive work as a potential explanation.

In the context of our setup, the change in the markup has an immediate implication for the labor share. While we have calculated the markup from all variable inputs, we could do so as well for labor alone. Then rewriting the First Order Condition (5) where $V = L$, $P^V = w$ and $\theta^V = \theta^L$, the output elasticity of labor, we obtain that at the firm level the labor share satisfies

$$\frac{w_t L_{it}}{P_t Q_{it}} = \frac{\theta_{it}}{\mu_{it}}. \quad (26)$$

Observe that if there are multiple inputs that are fully variable, the estimated markup should be the same. So even if the markup is calculated for the bundle V , it should hold for L as well as long as both V and L are variable. Profit maximization by individual firms thus implies the labor share is inversely proportional to the markup. As markup increases, we expect to see an decrease in the labor share.



(a) Average Labor and Employment Shares (= Employees/Sales; normalized to 0.2 in 1980). (b) The Evolution of the Average Capital Share in our sample of firms.

Figure 21: Labor and Capital Shares.

Unfortunately Compustat does not have good data for the wage bill. Because reporting compensation to the SEC is not compulsory, the variable XLR for total compensation is heavily underreported.⁶² Because of selection in the sample of those firms that do report total compen-

⁶¹Intangible assets are non-physical assets including patents, trademarks, copyrights, franchises,... that grant rights and privileges, and have value for the owner.

⁶²Less than 10% of the year-firm observations include XLR.

sation, we need to be to be cautious interpreting the aggregate labor share outcomes. In Figure 21a we report the sales weighted average labor share for those firms that report it in our sample. Observe that the level is a lot lower than the labor share reported in the macro literature, because conventionally it is calculated as the share of labor expenditure in value added, not gross output or sales as we do here.

The magnitude would roughly be double, though ratio of value added to gross output in the aggregate has grown from 0.5 in 1980 to 0.58 in 2016. Moreover, the variable XLR for staff costs is less inclusive than the BEA measure for labor cost (in the aggregate, the BEA includes self employed or proprietor's labor compensation (corporations, partnerships and sole proprietorships), and some firms in our sample do not include commissions, bonuses or incentive compensation). We observe a decline over the duration of the whole sample. Again, we do not know whether this is driven by the composition of firms. Many more firms report employment. In the same Figure we also report employment as a share of sales, where we normalize its value to 0.2 in 1980. Effectively this assumes a constant real wage. Observe that the pattern between 1980 and 2016 is remarkably similar. The decrease in the labor share between 1980 and 2016 is about 4 percentage points (from 20% to 16%).

	Labor Share (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Markup (log)	-0.24 (0.03)	-0.23 (0.03)	-0.20 (0.03)	-0.24 (0.03)	-0.68 (0.02)	- 0.73 (0.02)
Cost Share (log)					0.91 (0.01)	0.96 (0.01)
Year F.E.		X	X	X	X	X
Industry F. E.			X		X	
Firm F.E.				X		X
R ²	0.02	0.08	0.21	0.88	0.93	0.99
N	24,838					

Table 4: Regressions: log (Labor Share) on log(Markup). F.E. = Fixed Effects; 4-digit Industries; Standard Errors (in parentheses) clustered at the firm level.

Despite the shortcomings of our data, we can nonetheless verify the firm's optimization condition (26) at the firm level. In Table 4 we report the regression coefficients of the log of the labor share on the log of the firm's markup. The first four specifications only differ in the fixed effects that are included. We consistently find a negative coefficient of around -0.20–0.24. As a firm's markup increases by say 10%, its labor share decreases by 2-2.4%.

To extrapolate these firm-level results to the aggregate economy, we need to use caution because there is no such a thing as a representative firm in this context. The rise of average markups is distributed unequally, and increasingly so. If we were to envisage an increase of the markup between 1980 and 2016 of 33% (from 1.21 to 1.61) for a representative firm, our regression would predict a decrease in the aggregate labor share of -7–8%. Compare this to the decrease of the economy's aggregate labor share by around -9% (from 62% to 56%), and the firm

level relation appears to be in accordance with the decline in the labor share in the aggregate. This calculation has no validity however because it is based on a representative firm hypothesis. Most importantly, this calculation abstracts from the reallocation of market share away from low markup firms towards high markup firms. As a result, the decline in the aggregate labor share is a combination of the rise in markups within the firm and the reallocation of market share towards firms with high markups. Our findings for the firm level markups are thus consistent with those in Autor, Dorn, Katz, Patterson, and Van Reenen (2017) and Kehrig and Vincent (2017) for the Census of manufacturing. In sum, we find firm level evidence of the direct inverse relation between markups and the labor share that we obtain from the first order condition (26).

In the table we also analyze whether we can reject any evidence that there is perfect competition. Rows (5) and (6) report the same regression where we now include the log of the cost share (labor over total cost) as a covariate. Under perfect competition, the coefficient is one. Here we find a coefficient significantly smaller than one, indicating there is a wedge between sales and costs. Equally important, any other covariate (in this case markup) should be insignificant. We find instead that the coefficient on the markup is highly significant and negative. This indicates that there is evidence of non-competitive price setting.

THE SECULAR DECLINE IN THE CAPITAL SHARE. The same logic for the decline in the labor share also applies to materials M , i.e. variable inputs that are used in production. Those are included in our variable cost measure COGS. Now if we consider the evolution of capital investment, which is not included in our measure of variable cost and which adjusts at a lower and more long run frequency, then the increase in markup has implications for the capital share.⁶³ In the long run and once the adjustment frictions are taken into account, higher output prices and lower output quantities eventually will lead to a decrease in the capital share. While the decline in the labor share is widely discussed, the decline in the capital share has received much less attention.⁶⁴

Assuming a static environment, the following equality has to hold:

$$\frac{P^V V}{PQ} + \frac{rK}{PQ} = 1 - \frac{P^X X}{PQ} - \frac{\Pi}{PQ}, \quad (27)$$

The labor share and the capital share sum up to the one minus the profit share minus the overhead share. We have established above that both the profit share and the overhead share increase, so the right hand side decreases. With Capital and Variable inputs complementary, and over a long enough time horizon for capital to adjust, the user cost of capital rK as a share of output will be decreasing over time. In fact, if capital were fully flexible, it would adjust

⁶³This is independent of the frequency at which capital adjusts. Implicit in our assumptions is the fact that variable inputs, which consist of labor L and material inputs M , fully adjust within a year, our unit of time. This assumption allows us to calculate the markup. Capital may or may not adjust. From Figure 20a we have inferred that capital is not equally flexible as the variable inputs.

⁶⁴A notable exception is Barkai (2017). He uses aggregate data: value added and compensation from the National Income and Productivity Accounts (NIPA), and capital from the Bureau of Economic Analysis Fixed Asset Table. Instead, we use firm-level data.

according to the equivalent of first order condition (5) $\frac{rK}{PQ} = \frac{\theta^K}{\mu}$ which relates the capital share to the inverse of the markup.

In Figure 21b we document the evolution of the capital share for the firms in our data. Not surprisingly this measure is quite volatile because it is a long term measure that adjusts at a lower frequency and that therefore is more subject to aggregate fluctuations. Also, before the 1980s, capital investment was particularly low because of tumultuous financial times: inflation was high and financial frictions were considered higher. What we learn from the figure is a decrease in the capital share from around 12% in 1980 to 8-10% towards the end of the sample. In the aggregate, the capital share is correlated with the inverse of our markup measure. With a long enough horizon, capital investment adjusts and hence there will be a reduction in capital investment as markups increase.⁶⁵

	Capital Share (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Markup (log)	0.03 (0.02)	0.03 (0.02)	-0.02 (0.01)	-0.14 (0.02)	-0.90 (0.00)	-0.86 (0.00)
Cost Share (log)					1.13 (0.00)	1.11 (0.00)
Year F.E.		X	X	X	X	X
Industry F. E.			X		X	
Firm F.E.				X		X
R ²	0.00	0.02	0.31	0.83	0.98	1.00
N			242,692			

Table 5: Regressions: log (Capital Share) on log(Markup). F.E. = Fixed Effects; 4-digit Industries; Standard Errors (in parentheses) clustered at the firm level.

As with the labor share, we can also investigate the firm-level relation between the capital share and markups. In Table 5 we report the regression coefficients for different specifications. We find that without firm fixed effects, there is no significant relation between markups and the capital share. This may be indicative of the adjustment costs that firms face when investing in capital. Instead, with firm fixed effects, there is a significant negative effect, with an elasticity of -0.14. When we include the cost share, the coefficient on the cost share is larger than one. Under variable adjustment of capital, perfect competition would impose this to be equal to one, and less than one with market power (see for example Table 4 for the labor share). The fact that the coefficients for on the cost share here are larger than one indicates that capital does not adjust frictionlessly.

THE SECULAR DECLINE IN LOW SKILL WAGES AND LABOR FORCE PARTICIPATION. An increase in market power and the corresponding increase in prices of goods sold implies a de-

⁶⁵A more detailed analysis of the impact of market concentration on business investment is in Gutiérrez and Philippon (2017). In particular, they show within manufacturing that there is a positive investment response to competition from China.

crease in aggregate output produced, whenever demand is not perfectly inelastic. Lower output produced then implies lower demand for labor. This results in both lower labor force participation and lower wages. Even if supply is perfectly elastic, real wages decrease with market power because the price of the output goods has increased.

There is ample evidence of the stagnation of wages in the lower half of the distribution. The median weekly wage in constant prices has changed barely since 1980, from \$330 to \$345 (1982 prices, source Current Population Survey (CPS)). But there has been technological progress, and the share of median wages out of GDP has nearly halved, because over three and a half decades, GDP has nearly doubled. In the last few decades, also labor force participation has been decreasing from 67% in the 1990s to 63% now. Most strikingly, while the gender gap has continued to close, in the last two decades also female labor force participation is decreasing.

The quantitative investigation of the effect of market power on low skill wages and labor force participation is beyond the scope of the current paper. In De Loecker, Eeckhout, and Mongey (2018) we construct an oligopolistic framework for firm dynamics that quantitatively accounts for these general equilibrium implications of the rise in market power.

THE SECULAR DECLINE IN LABOR REALLOCATION AND MIGRATION RATES. It is well-known that in an environment with market power, shocks to productivity and costs are not translated one for one into prices. In a competitive market, firms face a perfectly elastic demand and any decrease in costs is passed on to the consumer, where prices decrease by the same amount as the decrease in costs. With market power however, the passthrough of cost shocks to prices is generally incomplete.⁶⁶ Crucial for our finding is that the higher the degree of market power by firms, the lower the passthrough.

Now consider an environment where firms have market power and face shocks to their productivity. With positive shocks, firms face lower costs and adjust their inputs (say labor) upwards. With negative shocks, they adjust inputs downwards. Because passthrough is lower in the presence of higher market power, the rise in market power will give rise to lower degree of adjustment of the variable inputs, including labor, for the same shock process.

This is precisely what Decker, Haltiwanger, Jarmin, and Miranda (2014) find for the US economy over the last three decades. The volatility of shocks has not decreased, but rather the *responsiveness* of firm's output and labor force decisions to the existing shocks.⁶⁷ The rise in market power thus can rationalize the decrease in labor reallocation across firms, even if the observed shocks to firm productivity has remained constant.

The decrease in labor market dynamism is evident in the decrease of labor reallocation as well as in the decrease of job to job transitions, non-employment to employment transitions and

⁶⁶Most of the evidence comes from studies that measure the impact of changes in the exchange rate or reductions in tariffs, see for example Campa and Goldberg (2005). More recently, incomplete pass-through has been documented in a domestic setting. E.g. Ganapati, Shapiro, and Walker (2018) reports incomplete pass-through of energy input price changes across industries of the US manufacturing sector.

⁶⁷Independent evidence at business cycle frequency by Berger and Vavra (2017) establishes that the volatility of prices is due to firms' time-varying responsiveness to shocks rather than to the time-varying nature of the shocks themselves. Their identification strategy is derived from the exchange rate passthrough of volatility on prices.

employment to non-employment transition.⁶⁸ The decrease in market power and the resulting decrease in labor reallocation can also rationalize the fact that migration rates across US States and metropolitan areas has decreased to nearly half from around 3% in 1980 to 1.5% in 2016.⁶⁹ If firms are based in different local labor markets and a fraction of all job relocation decisions are between local labor markets, then lower job flow rates will automatically give rise to lower migration rates. We assess the quantitative significance of the impact of the rise of market power on labor reallocation and migration in companion work (De Loecker, Eeckhout, and Mongey (2018)).⁷⁰

7 Concluding Remarks

Using firm level data on the accounts of all publicly traded firms and of the census of private firms (in manufacturing, retail and wholesale trade) in the US, we estimate markups for each firm and we document the properties of the distribution of markups. We find that average markups have been relatively constant between 1955 and 1980 at around 20% to 30% above marginal cost. From 1980 onwards, there has been marked change in this pattern with markups steadily rising from 21% to nearly 61% in 2014, an increase of 40 points. There was a sharp increase in the 1980s and 1990s, a period of stagnation in the 2000s, and again a steep increase following the Great Recession.

We attribute this rise in the average markup nearly exclusively to the increase of the markup of the firms with the highest markups already. The distribution of markups has become more skewed with a fat upper tail while the median of the distribution remains unchanged. Because of this increasingly more skewed distribution, we must be cautious not to use the average markup as that of a representative firm to draw any conclusion about the aggregate economy. When markets are non-competitive, aggregation is generally non-linear. In particular, the rise in sales weighted markups is due in part to the rise of the markups themselves, and in part to the reallocation of sales shares from low to high markup firms. Across industries, we do not find a strong compositional pattern and the increase occurs mainly within industry.

We further establish that the rise in markups is not merely to offset a rise in overhead costs. While overhead costs have risen, the rise in markups exceeds that of overhead. We thus find that there are excess markups, and the excess markups are highest for those firms with high overhead costs. This is consistent with the increase in our measure of profits. The profit rate

⁶⁸There are several potential alternative explanations for the decline in job flows: demographic change (aging workforce, Fallick, Fleischman, and Pingle (2010) and Engbom (2017)), a more skilled workforce, lower population growth, decreased labor supply (Karahana, Pugsley, and Şahin (2016)), technological change (Eeckhout and Weng (2017)), changed volatility of production, and government policy (such as employment protection legislation, licensing,...; see Davis and Haltiwanger (2014)). Hyatt and Spletzer (2013) show that demographic changes can explain at most one third of the decline in job flows.

⁶⁹See also Kaplan and Schulhofer-Wohl (2012), amongst others.

⁷⁰Recent work by Baqaee and Farhi (2017a) also draws attention to the fact that firm-level productivity shocks can give rise to a nonlinear impact on macroeconomic outcomes. For example, models with network linkages such as Gabaix (2011) give rise to such non-linearities. The framework in De Loecker, Eeckhout, and Mongey (2018) establishes that also market power in the presence of incomplete passthrough gives rise to non-linearities.

has risen from 1% of sales in 1980 to 7-8% recently. We also find substantial increases in the market value as a share of sales. All this indicates that the rise in markups is evidence of a rise in market power.

We use our evidence to investigate the macroeconomic implications of the rise of markups. We focus our attention on the decrease in the labor share. From the first order condition of the firm optimization problem, there is a negative relation between the labor share and the markup. We establish that this negative relation exists at the firm level. This provides a compelling justification for the secular decline in the labor share that the aggregate US economy has experienced. We further discuss the impact of the rise in market power on the decrease in the capital share, on the decrease in low skill wages and labor force participation, and on the decrease in labor market dynamism and migration rates.

Markups of some firms are reaching heights multiple times higher than ever seen, at least since the second world war when our data start. It is open to speculation whether this trend will continue, but for now there are no signs that markups will decrease substantially any time soon.

Appendix A Data: Summary Statistics

A.1 Compustat

We obtain firm-level financial variables of all US-incorporated publicly listed companies active at any point during the 1950-2016. We access the Compustat North America – Fundamentals Annual (through WRDS), and download the annual accounts for all companies. The results in this paper are obtained with a download on March 25 2018. We keep unique records for each firm, and assign a firm to a unique 2 digit industry, as reported. When a firm reports both an Industry Format and a Financial Format, we keep the Industry Format; and we exclude firms that do not report an industry code. All financial variables are deflated with the appropriate deflators. The main results, unless reported otherwise, rely on the sample of firms over the period 1950-2016, where we eliminate firms with reported cost-of-goods to sales, and SG&A to sales ratio's in the top and bottom 1 percent, where the percentiles are computed for each year separately. Our results are invariant to trimming up to the 5 percent (bottom and top).⁷¹ As such a firm-year observation requires information on both sales and cost-of-goods sold, two essential ingredients to measure markups. The Table below presents a few basic summary statistics for a few leading variables used in our analysis (sales, cost-of-good sold, capital, wage bill, employment and SG&A), for two samples. Sample A, observations with information on sales, cost-of-goods sold and SG&A; and Sample B, observations with information the wage bill.

Table A.1: Summary Statistics (1955-2016)

	Acronym, var.	Sample A			Sample B		
		Mean	Median	Nr Obs	Mean	Median	Nr Obs
Sales	<i>SALE, PQ</i>	1,922,074	147,806	247,644	5,894,779	578,912	28,116
Cost of Good Sold	<i>COGS, V</i>	1,016,550	55,384	247,644	2,970,693	195,087	28,116
Capital Stock	<i>PPEGT, K</i>	1,454,210	57,532	247,644	5,193,319	345,592	28,116
SG&A	<i>XSG&A, X</i>	342,805	29,682	247,644	926,542	78,487	28,116
Wage Bill	<i>XLR, WL</i>	1,093,406	130,486	28,116	1,093,406	130,486	28,116
Employment	<i>EMP, L</i>	8,363	863	221,121	24,861	4,522	25,527

Notes: Million USD deflated using the GDP Deflator with base year 2010. For each variable we list: the Compustat acronym, the associated notation (in levels) used throughout the manuscript.

A.2 Economic Censuses

The focus of our analysis of the Census data is on Manufacturing (NAICS codes 31, 32, 33), Retail (NAICS codes 44, 45) and Wholesale (NAICS code 42). In 2012, Manufacturing consists of about 297,000 establishments, Retail of about 1,060,000 establishments, and Wholesale of

⁷¹Rather than winsorizing the tails of the distribution, we find similar results when doing structural error correction to purge the measurement error from sales using the specification of the control function in equation (7).

about 420,000 establishments. These establishments aggregate into about 650,000 retail firms, about 314,000 wholesale firms, and about 250,000 manufacturing firms. Together these three sectors make up a little over 20% of US GDP. In principle, each Economic Census spans the universe of every single employer establishment in its sector, across the size distribution; only non-employer establishments (sole proprietorships with no employees) are omitted.

The other Censuses that we do not use are Census of Services, the Census of Construction Services, the Census of Mining, the Census of Transportation, Communications, and Utilities, the Census of Finance, Insurance, and Real Estate, and the Census of Auxiliary Establishments.⁷²

The data is organized around the most discrete unit of production in the microdata, an “establishment”, which is a single physical plant. Establishments can be aggregated to the EIN-level (“Employer Identification Number”: the most discrete legal unit of production – an EIN is a unique tax ID associated with a distinct legal entity), and higher up, to the firm level (major corporations are usually collections of EINs, which in turn are each collections of multiple establishments). The microdata associates each establishment both with an EIN and a “firm ID”: the EIN is considered part of the firm if the firm has complete or majority ownership of the EIN.

Perhaps the most common way of defining “firm” in the recent firm heterogeneity literature is to say that all of a firm’s establishments in a given 4-digit SIC industry (roughly equivalent to a 6-digit NAICS industry) are a distinct firm (this is the approach taken by Hsieh and Klenow (2009), Autor, Dorn, Katz, Patterson, and Van Reenen (2017), and others). Under this definition, Walmart’s establishments listed as e.g. SIC 5411 (Retail - Grocery Stores) are one firm, and Walmart’s establishments listed in SIC 5412 (Retail - Convenience Stores) are a separate firm. Our preferred default approach is to define “firm” as all of the firm’s establishments in a single sector Census (e.g. all of Walmart’s firms in all of retail, NAICS codes 44-45, are a single firm).

Appendix B Production Technology

We present alternative specifications for the underlying technology – i.e., the production function, which directly impacts the measurement of markups through the choice of the variable input, and the associated output elasticity. We first show the aggregate markup using a calibrated scalar for the output elasticity (0.85), to highlight the time-series pattern of the market-share-weighted average sales-to-cost-of-good-sold ratio. Second, we consider a sector-specific translog production function, yielding firm- and time-specific output elasticities. Finally, we consider a Leontief production function, and rely on labor as the variable input of production (for the sub-sample reporting employment).

⁷²Note that the Census of Auxiliary Establishments includes many corporate support units that do not directly face customers or take in revenue.

B.1 Markup with Time-invariant and Sector-invariant Output Elasticity

Figure B.1 plots the average markup with a constant elasticity. By construction (equation (5)), all of the change is due to the evolution of the expenditure on variable inputs to sales.

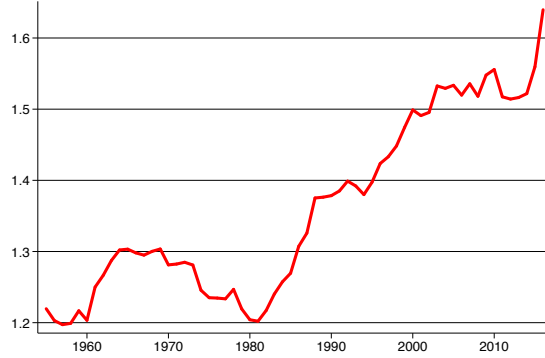


Figure B.1: Average Markup with a time-invariant and sector-invariant elasticity (calibrated to $\theta_{ts}^V = 0.85$ for each year t and every sector s).

B.2 Translog production function

We explore the sensitivity of our results and consider an industry-specific translog production function. As discussed in De Loecker and Warzynski (2012), this functional form permits time-varying output elasticities while preserving the identification results, and it constitutes a parsimonious way to generate firm/time-specific output elasticity differences, which ultimately impact the estimate of markups. The moment conditions are as before and exploit the static optimization of the variable inputs:

$$\mathbb{E} \left(\xi_{it}(\beta) \begin{bmatrix} v_{it-1} \\ v_{it-1}^2 \end{bmatrix} \right) = 0, \quad (\text{B.1})$$

where now the vector β contains two additional parameters.

In particular, variation in output elasticity over time (or firms), will no longer be attributed to markup variation. We consider the following translog production function for each industry s :

$$q_{it} = \beta_{v1}v_{it} + \beta_{k1}k_{it} + \beta_{v2}v_{it}^2 + \beta_{k2}k_{it}^2 + \omega_{it} + \varepsilon_{it} \quad (\text{B.2})$$

This yields an estimate for the output elasticity of the composite variable input: $\theta_{it}^V = \beta_{v1} + 2\beta_{v2}v_{it}$, and we calculate the markup as before. We do not consider the interaction term between v and k to minimize the potential impact of measurement error in capital to contaminate the parameter of most interest – i.e. the output elasticity. See Collard-Wexler and De Loecker (2016) for a discussion and implications of the presence of measurement error in capital. The results are, however, not sensitive to this omission. Figure B.2 plots the evolution of the output elasticities estimated from the Translog production technology. It shows that the evolution in both cases is very similar to the time-varying Cobb-Douglas specification.

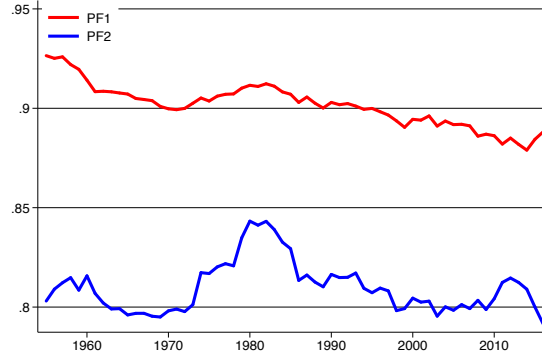


Figure B.2: Evolution of the Average Output Elasticities over Time (Translog Production Technology: PF1 and PF2).

In the main text, we plot the share-weighted aggregate markup obtained with the translog model alongside the one obtained with time invariant production functions (by industry). The time-series pattern is very similar, and especially the increase of the overall markup starting in 1980 is identical. The only difference is in the actual level of the markup, which is not direct interest, while the change over time is again very similar.

B.3 An Alternative Production Function: Unbundling COGS

We have used a bundle of inputs (COGS) as well as the Wage Bill only, which is included in COGS, to estimate markups. We now unbundle COGS into the Wage Bill and Materials, which we calculate as the residual of COGS minus the Wage Bill, denoted by \tilde{M} . In doing so, we face the well-known limitation in the Compustat sample that the wage bill is not reported only by a small number of firms.

In order to compute the markup using the FOC on labor we require an output elasticity of labor. This elasticity is obtained as before by estimating the production function. However, now we have to distinguish between labor and intermediate inputs, taken together the variable input bundle V , and this requires a modeling choice as to how intermediate inputs enter the production function. There are two distinct technologies to consider in our setting: a fixed-proportion (Leontief) technology in the intermediate variable, and a Cobb-Douglas gross-output specification.

LEONTIEF TECHNOLOGY. This is the case considered in De Loecker and Scott (2016), Akerberg, Caves, and Frazer (2015) and Gandhi, Navarro, and Rivers (2011), and avoids the potential identification issues surrounding intermediate inputs in the classical setting. Consider:

$$Q_{it} = \min\{\theta^M M_{it}, L_{it}^{\theta_L} K_{it}^K \Omega_{it}\} \quad (\text{B.3})$$

We estimate this production function by sector, for each year with sector fixed effects. There are not enough observations to reliably estimate the production function by sector/year. The main

insight is that we do not require to observe intermediate inputs to estimate the production function, but instead we project gross output on labor and capital. To compute the markup, however, we do have to include intermediates as the marginal cost of production requires the appropriate increase in intermediate inputs when increasing labor. We derive the markup from the first order condition accounting for the fact that the non-differentiable technology (B.3) requires input choices in fixed proportions:

$$\mu_{it}(L) = \frac{1}{\mu_{it}^{-1} + \frac{M_{it}}{Q_{it}}} \quad (\text{B.4})$$

where μ_{it} is obtained with the standard formula, and where the output elasticity is the estimated labor coefficient – i.e., $\mu_{it} = \theta^L \frac{S_{it}}{wL_{it}}$.

We report the Average Markup: green Leontief technology (s: sector specific; t: time varying); C: Cobb-Douglas (time-varying); Red: baseline markup (PF1) for the selected sample with data on wage bill.

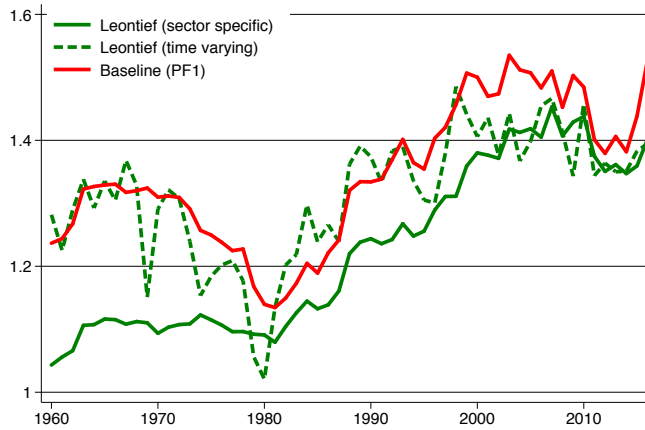


Figure B.3: Industry Specific Average Markups

Appendix C Alternative Aggregation Weights

As is standard in the literature, we have calculated the average markup using the sales share of the firm in the entire sample. Different aggregation weights will affect the average. Here we consider alternative variables to sales to calculate the weights.

First in Figure C.1 we plot the contour plot of the kernel density of the joint distribution of the firm's markup and the aggregation weight for the three weights and for two years 1980 and 2016. This allows us to inspect the raw data. We do this for Sales, COGS and Employment. The three variables confirm the same pattern: markups have increased in the top half of the distribution and are more or less invariant in the bottom half; the increase in the upper half of the distribution leads to the large dispersion of markups; across all sectors of the economy,

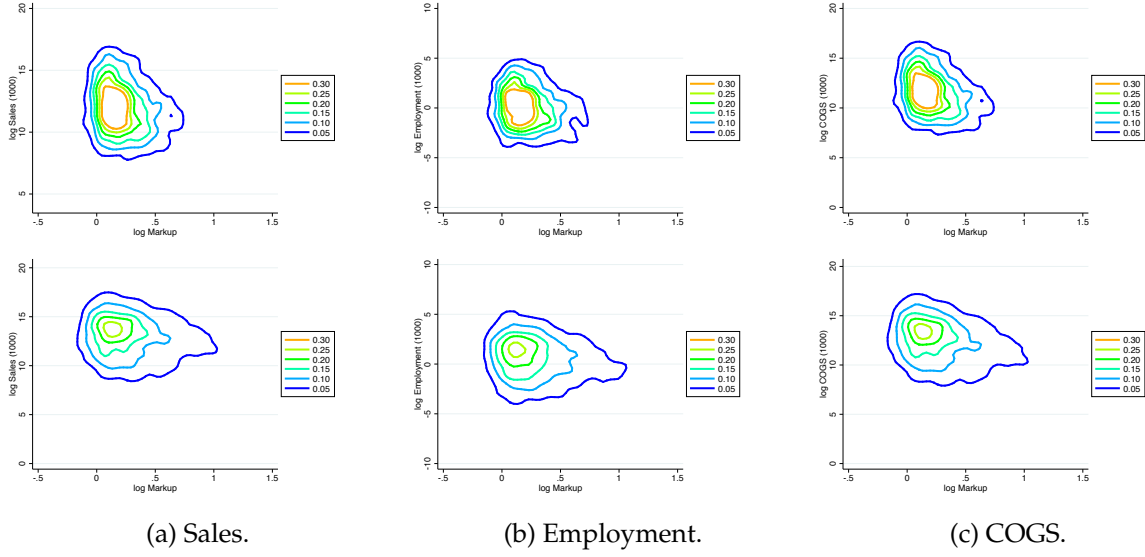


Figure C.1: Joint Distribution of the firm-level Markup and the Share of Sales PQ , Employment L and the value of the variable input $P^V V$. The year 1980 is on the first row, and the year 2014 on the second row.

markups are negatively correlated with size of sales, employment and COGS, which is evident from the fact that the ridge of the contour is negatively downward sloping.

Based on these alternative weights, in Figure C.2 we calculate the aggregate markup for different input weights and for the two production function specifications: COGS, Capital, Employment, Total Cost and the Harmonic Mean of Sales. Which one is the most relevant depends on the question at hand, and for the purpose of welfare analysis, the appropriate measure depends on the model specification of demand and technology. For that reason we include the harmonic mean of sales because in the Atkeson and Burstein (2008) model of imperfect competition (see also Edmond, Midrigan, and Xu (2015) and Grassi (2017)), the appropriate aggregation is to use the harmonic mean of sales. We include the Harmonic Mean of as an alternative for our benchmark measure in Figure 1. The evolution of the average markup under these different weights follows our benchmark closely with an initial hump shape between 1950 and 1980 and a steady rise from 1980 to 2014. The level of the employment weighted markup is very similar, while the level of the value-of-input weighted markup is lower. This is as expected since the gap between sales and input value widens as the markup increases. After all, this is how markup is defined. Therefore less weight is put on high markups with input value than in with sales. Because we have seen a shift towards extreme markups, this explains the lower average with input weights.

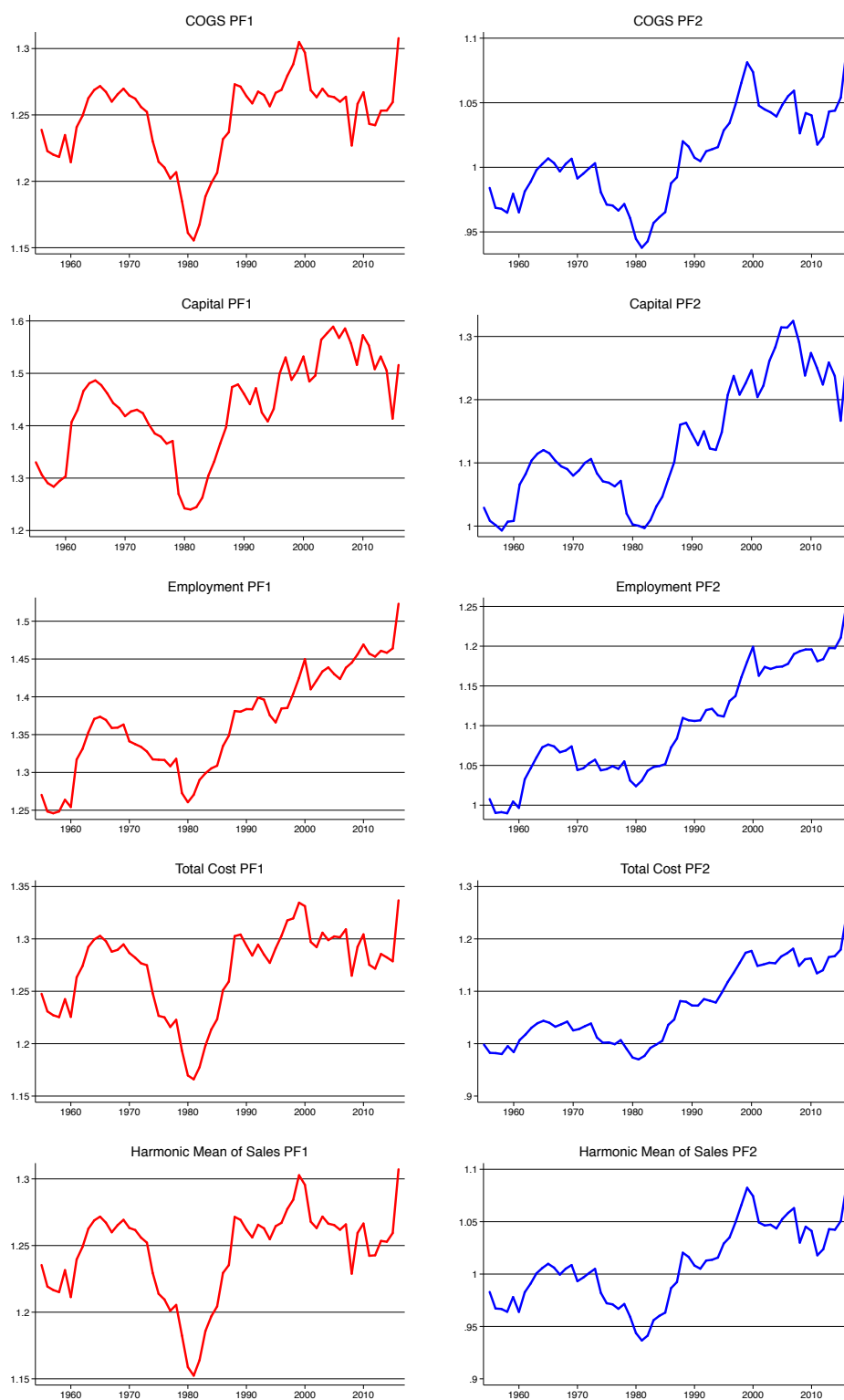


Figure C.2: Arithmetic Mean Markup (PF1 and PF2), weighted by expenditure share on variable inputs (COGS), Capital, Employment, Total Cost as well as the Harmonic Mean of Sales. It can be shown analytically that the Harmonic Mean of Sales and the COGS weighted measures are identical.

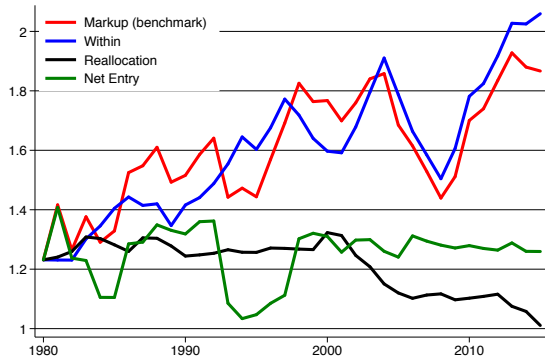
Appendix D Results by sector

We present additional sector-level results on the time-series of markups, and the underlying components using the decomposition introduced in the main text. We present the aggregate markup by sector (2 digit) for the last year in our data (2016), alongside the change between 1980 and 2016, and the sector's share in total sales.

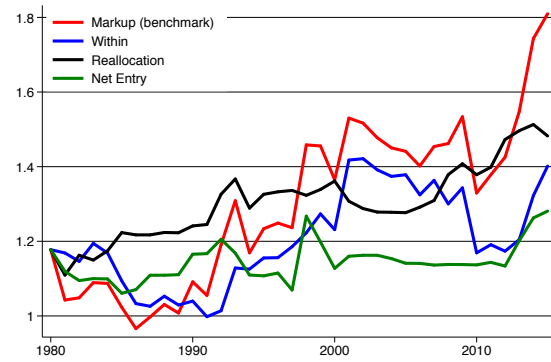
Table D.1: Markups by Sector, 1980-2016: PF1.

NAICS	Sector	Markup Change		Sales Share 2016
		2016	1980-2016	
11	Agriculture, Forestry, Fishing and Hunting	1.81	0.00	0.00
21	Mining, Quarrying, and Oil and Gas Extraction	1.49	0.34	0.03
22	Utilities	1.41	0.07	0.00
23	Construction	1.16	0.03	0.01
31	Manufacturing (1)	1.81	0.05	0.06
32	Manufacturing (2)	1.94	0.10	0.23
33	Manufacturing (3)	1.38	0.04	0.24
42	Wholesale Trade	1.12	-0.01	0.04
44	Retail Trade (1)	1.01	-0.05	0.07
45	Retail Trade (2)	1.27	0.01	0.03
48	Transportation and Warehousing (1)	1.28	0.03	0.04
49	Transportation and Warehousing (2)	0.93	-0.20	0.00
51	Information	2.14	0.06	0.11
52	Finance and Insurance	1.77	-0.06	0.05
53	Real Estate and Rental and Leasing	2.64	0.21	0.01
54	Professional, Scientific, and Technical Services	1.66	0.11	0.02
56	Admin. & Support & Waste Mgt & Remediation Serv.	1.76	0.06	0.01
61	Educational Services	1.67	-0.04	0.00
62	Health Care and Social Assistance	1.11	0.01	0.02
71	Arts, Entertainment, and Recreation	1.27	0.01	0.00
72	Accommodation and Food Services	1.12	-0.01	0.01
81	Other Services (except Public Administration)	1.25	0.02	0.00

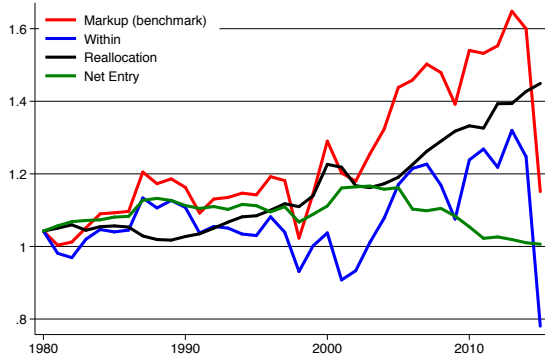
We also report the decomposition of sector-specific aggregate markup, into the various components, for the four remaining sectors not presented in the main text (Finance, real estate and insurance; Agriculture; Mining; and Utilities).



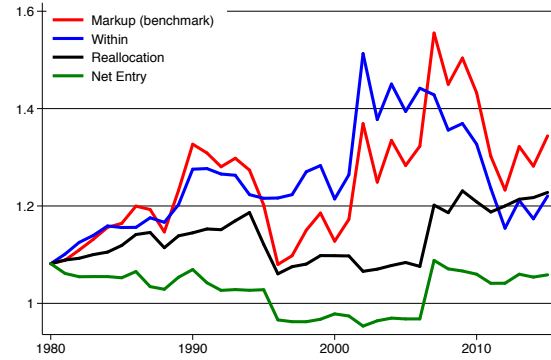
(a) FIRE: Finance, Insurance, Real Estate (52-53)



(b) Agriculture (11)



(c) Mining (21)



(d) Utilities (22)

Figure D.1: Firm Level Decomposition by Sectors

References

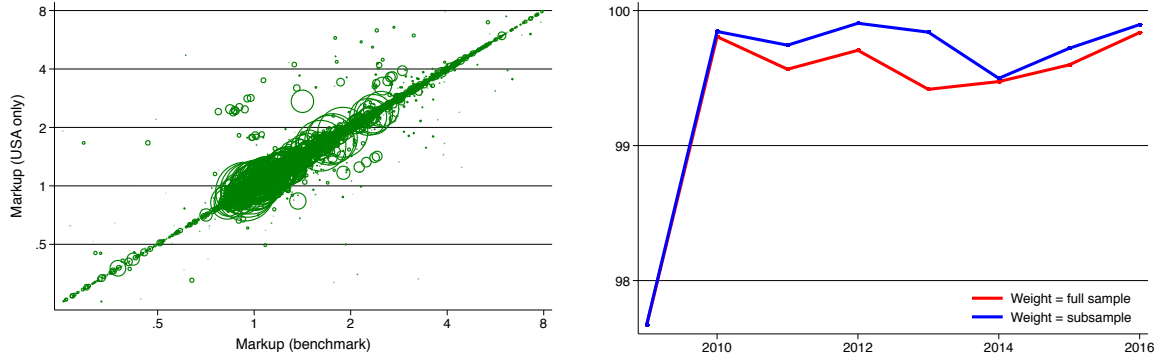
- ACKERBERG, D., C. L. BENKARD, S. BERRY, AND A. PAKES (2007): "Econometric tools for analyzing market outcomes," *Handbook of econometrics*, 6, 4171–4276.
- ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification properties of recent production function estimators," *Econometrica*, 83(6), 2411–2451.
- ANTRAS, P. (2004): "Is the US aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution," *Contributions in Macroeconomics*, 4(1).
- ARMENTER, R. (2015): "A bit of a miracle no more: the decline of the labor share," *Business Review*, Q3, 1–9.
- ATKESON, A., AND A. BURSTEIN (2008): "Pricing-to-market, trade costs, and international relative prices," *American Economic Review*, 98(5), 1998–2031.
- AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. VAN REENEN (2017): "The Fall of the Labor Share and the Rise of Superstar Firms," Discussion paper, Centre for Economic Performance, LSE.
- BAILY, M. N., AND B. P. BOSWORTH (2014): "US manufacturing: Understanding its past and its potential future," *The Journal of Economic Perspectives*, 28(1), 3–25.
- BAQAEE, D. R., AND E. FARHI (2017a): "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," Discussion paper, National Bureau of Economic Research.
- (2017b): "Productivity and Misallocation in General Equilibrium.," Discussion paper, National Bureau of Economic Research.
- BARKAI, S. (2017): "Declining Labor and Capital Shares," Chicago Booth mimeo.
- BASU, S., AND J. G. FERNALD (1997): "Returns to scale in US production: Estimates and implications," *Journal of political economy*, 105(2), 249–283.
- BERGER, D., AND J. VAVRA (2017): "Shocks vs. Responsiveness: What Drives Time-Varying Dispersion?," Discussion paper, National Bureau of Economic Research.
- BERRY, S., S. KORTUM, AND A. PAKES (1996): "Environmental change and hedonic cost functions for automobiles," *Proceedings of the National Academy of Sciences*, 93(23), 12731–12738.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 841–890.
- BRANDT, L., J. VAN BIESEBROECK, L. WANG, AND Y. ZHANG (2017): "WTO accession and performance of Chinese manufacturing firms," *American Economic Review*, 107(9), 2784–2820.
- BRENNAN, J. (2016): "Rising Corporate Concentration, Declining Trade Union Power, and the Growing Income Gap: American Prosperity in Historical Perspective," *Levys Economics Institute*.
- BRESNAHAN, T. F. (1989): "Empirical studies of industries with market power," *Handbook of industrial organization*, 2, 1011–1057.
- BROWN, G., AND N. KAPADIA (2007): "Firm-specific risk and equity market development," *Journal of Financial Economics*, 84(2), 358–388.
- BURNSIDE, C. (1996): "Production function regressions, returns to scale, and externalities," *Journal of monetary Economics*, 37(2), 177–201.

- CAMPA, J. M., AND L. S. GOLDBERG (2005): "Exchange rate pass-through into import prices," *Review of Economics and Statistics*, 87(4), 679–690.
- COLLARD-WEXLER, A., AND J. DE LOECKER (2016): "Production function estimation with measurement error in inputs," Discussion paper, National Bureau of Economic Research.
- COOPER, R., AND J. EJARQUE (2003): "Financial frictions and investment: requiem in q ," *Review of Economic Dynamics*, 6(4), 710–728.
- DAVIS, S. J., AND J. HALTIWANGER (2014): "Labor market fluidity and economic performance," Discussion paper, National Bureau of Economic Research.
- DAVIS, S. J., J. HALTIWANGER, R. JARMIN, AND J. MIRANDA (2007): "Volatility and dispersion in business growth rates: Publicly traded versus privately held firms," in *NBER Macroeconomics Annual 2006, Volume 21*, pp. 107–180. MIT Press.
- DE LOECKER, J. (2011): "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 79(5), 1407–1451.
- DE LOECKER, J., J. EECKHOUT, AND S. MONGEY (2018): "Quantifying the Effect of Market Power: An Oligopolistic Framework for Firm Dynamics," Mimeo.
- DE LOECKER, J., AND P. K. GOLDBERG (2014): "Firm performance in a global market," *Annu. Rev. Econ.*, 6(1), 201–227.
- DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNİK (2016): "Prices, Markups and Trade Reform," *Econometrica*, 84(2).
- DE LOECKER, J., AND P. T. SCOTT (2016): "Estimating market power Evidence from the US Brewing Industry," Discussion paper, National Bureau of Economic Research.
- DE LOECKER, J., AND F. M. P. WARZYŃSKI (2012): "Markups and Firm-level Export Status," *American Economic Review*, 102(6), 2437–2471.
- DECKER, R., J. HALTIWANGER, R. JARMIN, AND J. MIRANDA (2014): "The secular decline in business dynamism in the US," University of Maryland mimeo.
- DOIDGE, C., G. A. KAROLYI, AND R. M. STULZ (2017): "The US listing gap," *Journal of Financial Economics*, 123(3), 464–487.
- EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2015): "Competition, markups, and the gains from international trade," *American Economic Review*, 105(10), 3183–3221.
- EECKHOUT, J., AND X. WENG (2017): "The Technological Origins of the Decline in Labor Market Dynamism," mimeo.
- EGGERTSSON, G. B., J. A. ROBBINS, AND E. G. WOLD (2018): "Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States," Discussion paper, Brown University.
- ELSBY, M. W., B. HOBIJN, AND A. ŞAHİN (2013): "The decline of the US labor share," *Brookings Papers on Economic Activity*, 2013(2), 1–63.
- ENGBOM, N. (2017): "Firm and Worker Dynamics in an Aging Labor Market," Discussion paper, Working Paper.
- FALLICK, B., C. FLEISCHMAN, AND J. PINGLE (2010): "The effect of population aging on the aggregate labor market," in *Labor in the New Economy*, pp. 377–417. University of Chicago Press.

- FAMA, E. F., AND K. R. FRENCH (2004): "New lists: Fundamentals and survival rates," *Journal of financial Economics*, 73(2), 229–269.
- FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 98(1), 394–425.
- GABAIX, X. (2011): "The granular origins of aggregate fluctuations," *Econometrica*, 79(3), 733–772.
- GABAIX, X., AND A. LANDIER (2008): "Why has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.
- GANAPATI, S., J. S. SHAPIRO, AND R. WALKER (2018): "The incidence of carbon taxes in U.S. Manufacturing: Lessons from Energy Cost Pass-Through," Discussion paper, Yale University.
- GANDHI, A., S. NAVARRO, AND D. A. RIVERS (2011): "On the identification of production functions: How heterogeneous is productivity?," mimeo.
- GAO, X., J. R. RITTER, AND Z. ZHU (2013): "Where have all the IPOs gone?," *Journal of Financial and Quantitative Analysis*, 48(6), 1663–1692.
- GOLLIN, D. (2002): "Getting income shares right," *Journal of political Economy*, 110(2), 458–474.
- GRASSI, B. (2017): "IO in IO: Competition and volatility in input-output networks," *Unpublished Manuscript, Bocconi University*.
- GRULLON, G., Y. LARKIN, AND R. MICHAELY (2016): "Are US Industries Becoming More Concentrated?," *Unpublished Working Paper*.
- GUTIÉRREZ, G., AND T. PHILIPPON (2017): "Declining Competition and Investment in the US," Discussion paper, National Bureau of Economic Research.
- HALL, R. (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96(5), 921–947.
- HALL, R. E. (2018): "New Evidence on the Markup of Prices over Marginal Costs and the Role of Mega-Firms in the US Economy," Discussion paper, National Bureau of Economic Research.
- HALTIWANGER, J. C. (1997): "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence," *Federal Reserve Bank St. Louis Review*, 79(3), 55–77.
- HARTMAN-GLASER, B., H. LUSTIG, AND M. X. ZHANG (2016): "Capital Share Dynamics When Firms Insure Workers," Discussion paper, National Bureau of Economic Research.
- HSIEH, C.-T., AND P. J. KLENOW (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124(4), 1403–1448.
- HYATT, H. R., AND J. R. SPLETZER (2013): "The recent decline in employment dynamics," *IZA Journal of Labor Economics*, 2(1), 5.
- KAPLAN, G., AND S. SCHULHOFER-WOHL (2012): "Understanding the long-run decline in interstate migration," Discussion paper, National Bureau of Economic Research.
- KARABARBOUNIS, L., AND B. NEIMAN (2014): "The Global Decline of the Labor Share*," *Quarterly Journal of Economics*, 129(1).

- (2018): “Accounting for Factorless Income,” Discussion paper, National Bureau of Economic Research.
- KARAHAN, F., B. PUGSLEY, AND A. ŞAHİN (2016): “Demographic Origins of the Startup Deficit,” NY Fed mimeo.
- KEHRIG, M. (2011): “The cyclicalities of productivity dispersion,” *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*.
- KEHRIG, M., AND N. VINCENT (2017): “Growing productivity without growing wages: The micro-level anatomy of the aggregate labor share decline,” Duke mimeo.
- KELLER, W., AND S. YEAPLE (2009): “Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States,” *Review of Economics and Statistics*, 91(4), 821–831.
- KLETTE, T. J., AND Z. GRILICHES (1996): “The inconsistency of common scale estimators when output prices are unobserved and endogenous,” *Journal of Applied Econometrics*, pp. 343–361.
- KOH, D., R. SANTAELLA-LLOPIS, AND Y. ZHENG (2017): “Labor Share Decline and Intellectual Property Products Capital,” Discussion paper, Washington University mimeo.
- KOUJIANOU GOLDBERG, P., AND R. HELLERSTEIN (2012): “A structural approach to identifying the sources of local currency price stability,” *Review of Economic Studies*, 80(1), 175–210.
- LEVINSOHN, J., AND A. PETRIN (2003): “Estimating production functions using inputs to control for unobservables,” *The Review of Economic Studies*, 70(2), 317–341.
- MELITZ, M. J., AND G. I. OTTAVIANO (2008): “Market size, trade, and productivity,” *The review of economic studies*, 75(1), 295–316.
- MORLACCO, M. (2017): “Market Power in Input Markets: Theory and Evidence from French Manufacturing,” Discussion paper, Yale.
- NEVO, A. (2001): “Measuring market power in the ready-to-eat cereal industry,” *Econometrica*, 69(2), 307–342.
- NISHIDA, M., A. PETRIN, M. ROTEMBERG, AND T. WHITE (2017): “Are We Undercounting Reallocation’s Contribution to Growth?,” Discussion paper, US Census.
- OLLEY, G. S., AND A. PAKES (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica: Journal of the Econometric Society*, pp. 1263–1297.
- ROGNLIE, M. (2016): “Deciphering the fall and rise in the net capital share: accumulation or scarcity?,” *Brookings papers on economic activity*, 2015(1), 1–69.
- SMITH, M., D. YAGAN, O. ZIDAR, AND E. ZWICK (2017): “Capitalists in the Twenty-first Century,” *UC Berkeley and University of Chicago Working Paper*.
- SYVERSON, C. (2004): “Market structure and productivity: A concrete example,” *Journal of Political Economy*, 112(6), 1181–1222.
- TRAINA, J. (2018): “Is Aggregate Market Power Increasing? Production Trends using Financial Statements,” Chicago Booth mimeo.

Appendix 2 Consolidated Accounts: Geographical Decomposition



(a) Scatter plot of μ_{it} and μ_{it}^D (size of scatter point is (b) Ratio of average consolidated to domestic (USA only) markup weight of the yearly sales share)

Figure 2.1: Markups for domestic versus consolidated activities. The coefficient on μ_{it} in a log-linear regression of μ_{it}^D on μ_{it} is 0.98 (0.01) and the R-squared is 0.94.

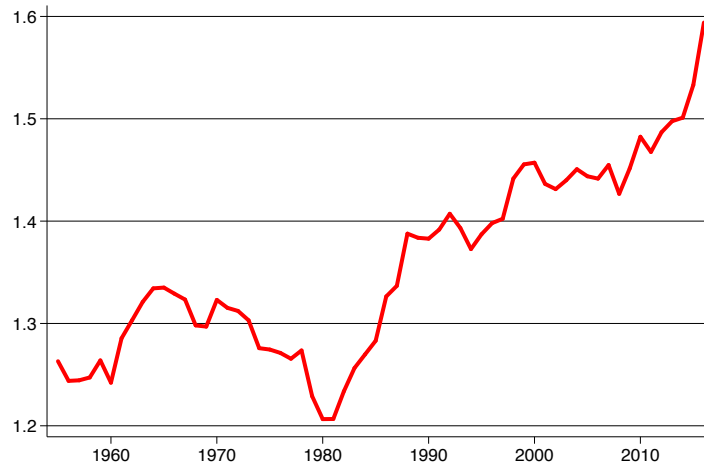


Figure 2.2: Average Markup (PF1) using only those firms in the sample that are incorporated in the United States (84.3% of all firm-year observations).

Appendix 3 Decomposition: details (equation (19))

year	Markup	Δ Markup	Δ Within	Δ ms	Δ cross	Net entry
1981	1.21	0.00	-0.01	0.00	0.00	0.01
1982	1.23	0.02	0.00	-0.05	0.00	0.07
1983	1.25	0.02	0.02	-0.01	0.00	0.02
1984	1.26	0.01	0.00	-0.03	0.00	0.05
1985	1.27	0.01	0.00	-0.01	0.00	0.02
1986	1.31	0.04	0.02	0.03	0.00	0.00
1987	1.32	0.01	0.00	0.00	0.00	0.01
1988	1.37	0.05	0.02	0.03	0.00	0.00
1989	1.37	0.00	0.00	-0.02	0.00	0.02
1990	1.37	0.00	0.00	0.01	0.00	-0.01
1991	1.38	0.01	0.00	0.05	0.00	-0.04
1992	1.39	0.02	0.01	-0.02	0.00	0.03
1993	1.38	-0.01	0.00	0.02	0.00	-0.02
1994	1.37	-0.02	0.00	-0.06	0.00	0.05
1995	1.38	0.02	0.00	-0.01	0.00	0.03
1996	1.41	0.02	0.01	-0.02	0.00	0.04
1997	1.42	0.01	0.01	0.02	0.00	-0.01
1998	1.44	0.02	0.01	0.04	0.00	-0.03
1999	1.46	0.02	0.01	0.03	0.00	-0.02
2000	1.48	0.01	0.00	0.07	0.00	-0.06
2001	1.45	-0.03	-0.03	0.02	0.00	-0.02
2002	1.45	0.00	-0.01	-0.02	0.00	0.03
2003	1.47	0.02	0.02	-0.03	-0.02	0.05
2004	1.47	0.00	0.00	-0.01	0.00	0.01
2005	1.47	0.00	-0.01	0.07	0.00	-0.06
2006	1.45	-0.01	-0.01	0.03	0.00	-0.03
2007	1.46	0.01	-0.01	0.04	0.00	-0.02
2008	1.44	-0.02	-0.02	-0.01	0.00	0.01
2009	1.47	0.03	0.00	0.05	0.00	-0.02
2010	1.49	0.01	0.02	-0.01	0.00	0.00
2011	1.46	-0.03	0.00	-0.02	0.00	0.00
2012	1.46	0.00	0.00	0.02	0.00	-0.01
2013	1.48	0.01	0.01	0.01	0.00	-0.01
2014	1.49	0.01	-0.01	0.01	0.00	0.00
2015	1.52	0.04	0.00	0.06	0.01	-0.03

Table 3.1: Decomposition of aggregate markups at the firm level (equation (19)).

Appendix 4 A Selection of Firms' Individual Markups

	Markup μ_i					Empl. L_i thousands	Sales S_i millions (2010 \$) 2016
	1980	1990	2000	2010	2016		
1-800-FLOWERS.COM	.	.	1.41	1.43	1.61	4	1,053
ALKERMES PLC	.	.	.	3.37	6.66	2	669
ALPHABET INC	.	.	.	2.50	2.23	72	81,025
AMAZON.COM INC	.	.	1.19	1.13	1.50	341	122,058
AMERICAN AIRLINES GROUP INC	.	.	.	0.98	1.31	122	36,064
ANHEUSER-BUSCH COS INC	1.21	1.46	1.60
ANHEUSER-BUSCH INBEV	.	.	.	2.43	2.83	207	40,855
APPLE INC	1.51	1.90	1.18	1.39	1.50	116	193,059
AT&T INC	.	2.31	.	1.83	1.53	269	146,989
BRIGHT SCHOLAR EDU-ADR	1.36	6	140
CAMPBELL SOUP CO	1.21	1.37	2.01	1.63	1.45	16	7,146
CAVIUM INC	.	.	.	2.65	3.39	2	542
COCA-COLA CO	1.58	2.32	3.17	2.86	2.54	100	37,575
COSTCO WHOLESALE CORP	.	.	0.99	0.99	1.05	218	106,559
DISNEY (WALT) CO	1.33	195	49,934
DR PEPPER SNAPPLE GROUP INC	.	.	.	2.35	2.32	20	5,780
GENERAL ELECTRIC CO	1.27	1.51	.	2.39	1.35	295	107,428
GOODYEAR TIRE & RUBBER CO	1.15	1.17	1.12	1.00	1.31	66	13,605
HARLEY-DAVIDSON INC	.	1.15	1.40	1.61	1.58	6	5,382
HEWLETT PACKARD ENTERPRISE	1.31	195	44,989
INTEL CORP	2.08	1.97	3.04	3.53	3.29	106	53,304
JOHNSON & JOHNSON	1.85	2.76	3.38	3.05	3.55	126	64,526
KELLOGG CO	1.45	1.88	2.03	1.68	1.54	37	11,653
KRAFT HEINZ CO	1.41	1.55	1.60	1.50	1.60	41	23,774
LEVI STRAUSS & CO	1.44	1.50	1.60	1.92	1.92	13	4,086
MADISON SQUARE GARDEN CO	1.21	9	1,001
MICROSOFT CORP	.	4.66	7.67	4.87	2.27	114	76,581
MYRIAD GENETICS INC	.	.	.	7.66	4.65	2	677
NYX GAMING GROUP LTD	5.96	1	147
OFFICE DEPOT INC	.	1.12	1.22	1.24	1.22	38	9,892
PEPSICO INC	1.90	2.02	2.55	2.18	2.17	264	56,366
PFIZER INC	1.92	2.70	6.21	4.28	4.59	96	47,413
SEADRILL PARTNERS LLC	.	.	.	2.75	3.55	1	1,436
SEAWORLD ENTERTAINMENT INC	1.22	13	1,207
STEEL PARTNERS HOLDINGS LP	1.40	5	1,044
TESLA INC	.	.	.	1.28	1.33	30	6,283
TEXAS INSTRUMENTS INC	1.26	1.11	2.08	2.07	2.52	30	12,001
TRIPADVISOR INC	14.93	3	1,328
UNITED AIRLINES INC	1.13	.	.	.	1.27	.	32,812
VERIZON COMMUNICATIONS INC	.	.	.	1.96	1.75	161	113,076
WHOLE FOODS MARKET INC	.	.	1.27	1.36	1.23	87	14,113
WORLD WRESTLING ENTMT INC	.	.	1.37	1.34	1.21	1	655
YELP INC	.	.	.	11.71	8.46	4	640

Table 4.1: Individual Firms' Markup (baseline measure PF1)

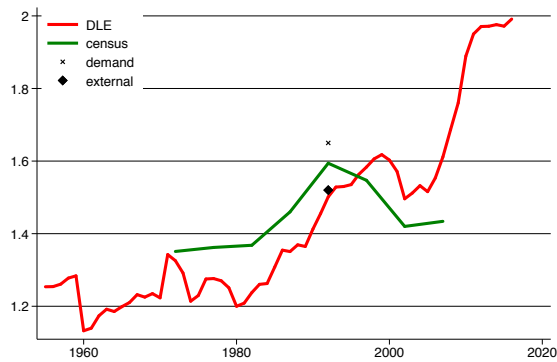
Appendix 5 Comparing markup measures: 7 case studies

We compare our markup estimates with alternative measures. Those either use the so-called *demand approach* (briefly discussed in the introduction, but we refer to De Loecker and Scott (2016) for an extensive discussion and how it relates to the *production approach* used in this paper), the production approach applied to census data, or are based on external evidence on profit margins (either drawn from the Consumer reports or industry sources).

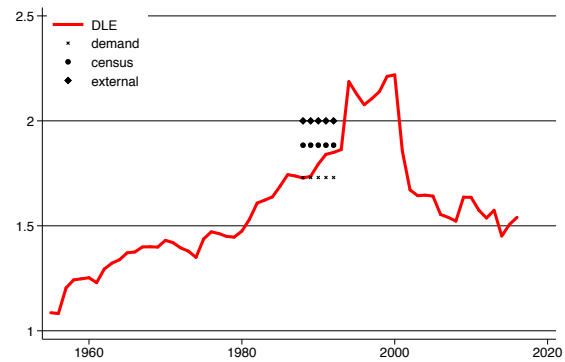
An important observation made in De Loecker and Scott (2016) is that a strict comparison of various approaches is difficult for at least two separate reasons. First, except for a handful of sectors, there is no overlap (in either time or geography) between the data required for applying the production approach and the demand approach. Second, even if there is overlap (as is the case for the Brewing industry analyzed in De Loecker and Scott (2016), both approaches rely on non-nested assumptions (on either market structure, consumer demand and technology).

Given these restrictions, we compare markups for a variety of industries spanning manufacturing, services and retail. In particular we report markups using (if available) the various approaches and data sources for the Brewing industry (NAICS 3121), the Automobile industry (NAICS 3361), the Steel industry (NAICS 3311), the Airline industry (NAICS 4811), the RTE Cereal industry (NAICS 311230), the Department Store industry (NAICS 4521) and the Electronic Shopping and Mail-order Houses industry (NAICS 4541). We plot the share-weighted aggregate markup for the various 4 digit NAICS industries using in the Compustat sample, and this series is reported in red, the production-based approach using the full census of relevant producers is indicated in green (whenever a time-series is available), the values based on the demand-approach are indicated with a full circle, while external markups are indicated with diamonds.

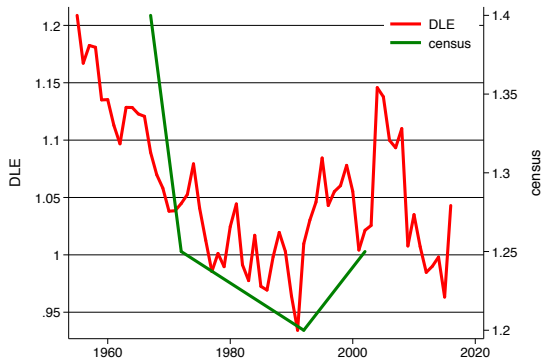
The underlying sources are: *Brewing*: Compustat: share-weighted aggregate markup, demand and census based on analysis in De Loecker and Scott (2016), and external evidence from the 1996 Consumer Reports; *RTE Breakfast Cereal*: Compustat: share-weighted aggregate markup for Kellogs and Quaker), for the years 1988-1992 using three different sources (all reported in Nevo (2001)): production-based approach using the full census of cereal producers, the demand-approach using market-level data, and reported markups using industry reports (discussed in Nevo (2001)); *Steel*: Compustat: share-weighted aggregate markup, census based on analysis in Collard-Wexler and De Loecker (2016); *Automobile*: Compustat: share-weighted aggregate markup, demand approach based on analysis in Berry, Levinsohn, and Pakes (1995), and census approach is obtained by applying De Loecker and Warzynski (2012) to the results reported in Berry, Kortum, and Pakes (1996); *Airlines*: Compustat: share-weighted aggregate markup, the external evidence comes from a 2005 Hearing before the Subcommittee on Aviation Committee on Transportation and Infrastructure United States House of Representatives, by Morrison and Winston; *Department stores*: share-weighted aggregate markup, and census: calculations of the authors based on census reports; *Electronic shopping and mail-order houses*: Compustat: share-weighted aggregate markup, and census: calculations of the authors based on census reports.



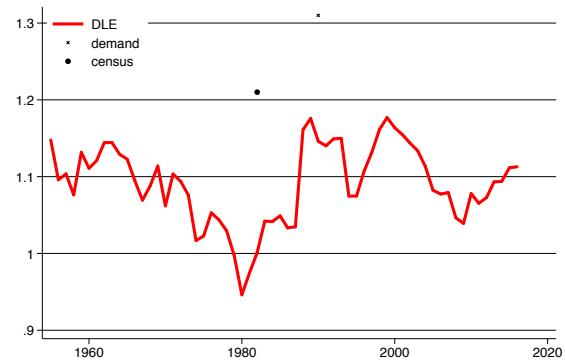
(a) Markups: US Beer Industry



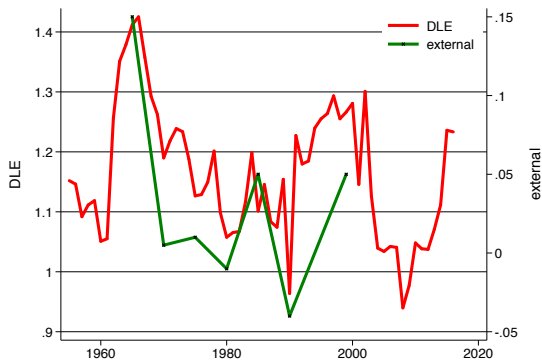
(b) Markups: US RTE Breakfast Cereal Industry



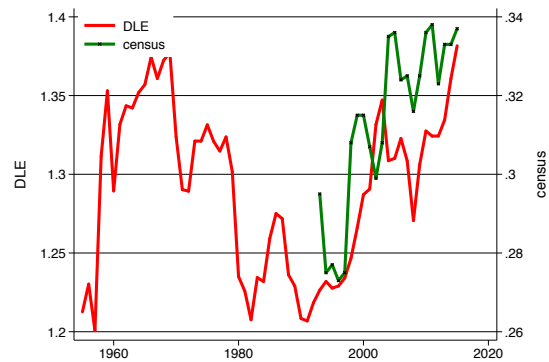
(c) Markups: US Steel Industry



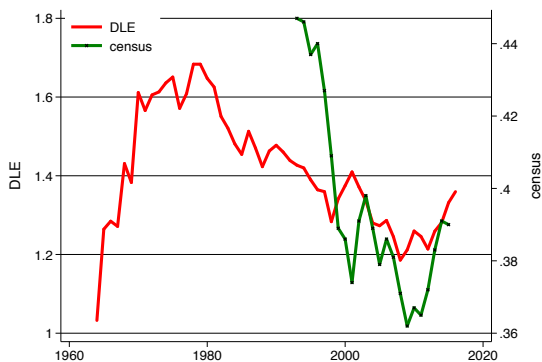
(d) Markups: US Automobile Industry



(e) Markups: US Airlines Industry



(f) Markups: US Department Store Industry



(g) Markups: US Electronic Shopping and Mail⁶
Order Houses Industry

Figure 5.1: Markups: DLE and alternative sources

Appendix 6 Profit Rate without Capital

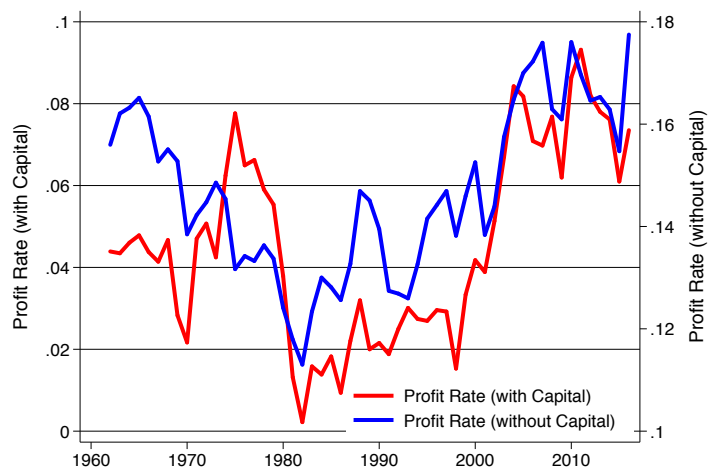
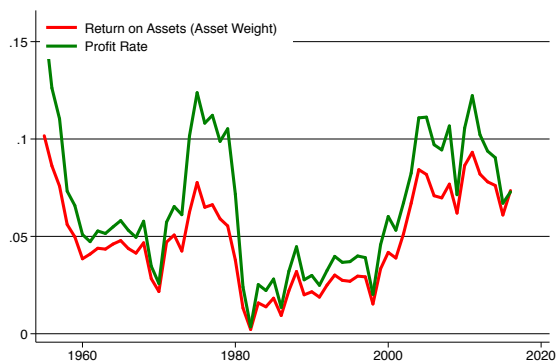
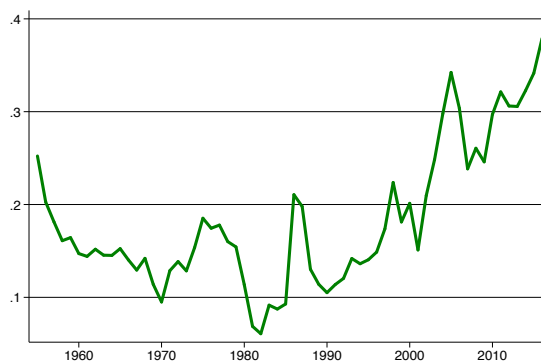


Figure 6.1: Profit Rate: with and without Capital.

Appendix 7 Return on Assets



(a) Return on Assets (asset weight) and Profit Rate.



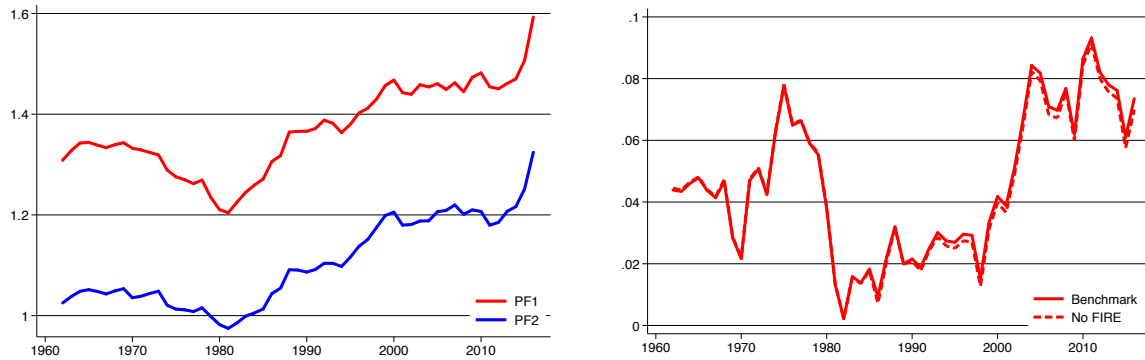
(b) Return on Assets (sales weight).

Figure 7.1: Average Return on Assets.

Appendix 8 Excluding FIRE

We repeat the calculation of markups and the profit rate after excluding 2-digit sectors 52 Finance and Insurance and 53 Real Estate. The argument is that these sectors are very different from the rest of the economy. We also report the markups and the share of sales below in the sector-specific plots.

The markup and the profit rate for the sample without FIRE is reported in Figure 8.1. We find little difference in the pattern of the average markups for both specifications PF1 and PF2 (Figure 8.1a) compared to the sample with FIRE. Likewise for the profit rate (Figure 8.1b).



(a) Average Markup (PF1 and PF2) without FIRE

(b) Profit Rate without FIRE

Figure 8.1: Markup and profit rate without Finance, Insurance and Real Estate (FIRE).

Appendix 9 Decomposition of change in markup at different sectoral level of aggregation

	Markup	Δ Markup	Δ Within	Δ Between	Δ Realloc.
2-digit sector					
1966	1.337	0.083	0.057	-0.017	0.041
1976	1.270	-0.067	-0.055	0.002	-0.014
1986	1.312	0.042	0.035	0.010	-0.003
1996	1.406	0.094	0.098	0.004	-0.008
2006	1.455	0.049	0.046	0.007	-0.005
2016	1.610	0.154	0.133	0.014	0.007
3-digit sector					
1966	1.337	0.041	0.029	-0.021	0.033
1976	1.270	-0.070	-0.060	-0.000	-0.010
1986	1.312	0.043	0.013	0.032	-0.002
1996	1.406	0.094	0.075	0.020	-0.000
2006	1.455	0.065	0.089	0.011	-0.035
2016	1.610	0.156	0.103	0.021	0.031
4-digit sector					
1966	1.337	0.002	0.016	-0.048	0.034
1976	1.270	-0.074	-0.061	-0.001	-0.011
1986	1.312	0.043	0.001	0.050	-0.009
1996	1.406	0.096	0.060	0.047	-0.011
2006	1.455	0.081	0.056	0.045	-0.021
2016	1.610	0.171	0.086	0.061	0.024

Table 9.1: Decomposition of 10 year change in Markup at different levels of sectoral aggregation. Estimated Markup using PF1 (where elasticities are estimated at the 2-digit sectoral level). Note: for the 3 and 4-digit Δ Markup does not coincide with the ten year difference in the markup because some firms do not report 3 or 4-digit sector codes.

Table 9.1 reports the decomposition from Table 1 in the text, and in addition the decomposition at the 3 and 4-digit sectoral level. The average markup for a given year at each sector level is the same, and so is the change in the markup over the preceding 10 years. The variation Within, Between and the Reallocation are different depending on the disaggregation because equation (18) sums over different size sectors. Not surprisingly, the within sector variation that accounts for the change in markups is less sizable as we move from 2 to 3 to 4 digit sectors.

Appendix 10 Industry-specific trends

10.1 Markups by 2-digit Sector

We repeat the same exercise as in our benchmark model for 2-digit NAICS industries. We have found in Table 1 that most of the pattern of increasing markups stems from within industry increases. As a result, we expect the increase in markups to hold for different individual industries. The following figure documents this for all 2-digit industries. Below, we also list a table with the markups and their change between 1980 and 2016 for each of the 2-digit sectors.

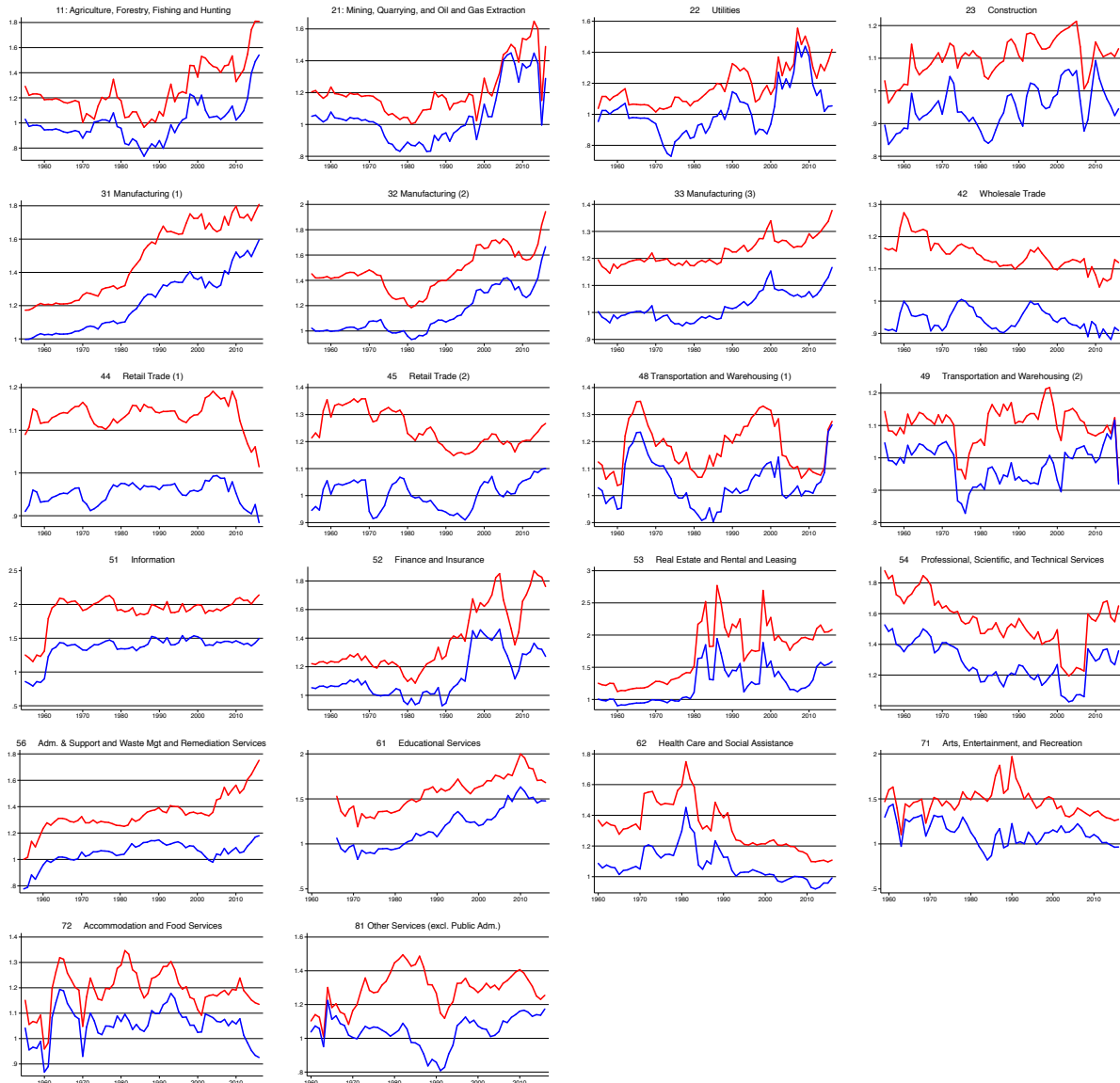


Figure 10.1: Industry Specific Markups.

10.2 Estimated Output Elasticities by 2-digit Sector



Figure 10.2: Industry Specific Output Elasticities.

10.3 Sales Shares by 2-digit Industry

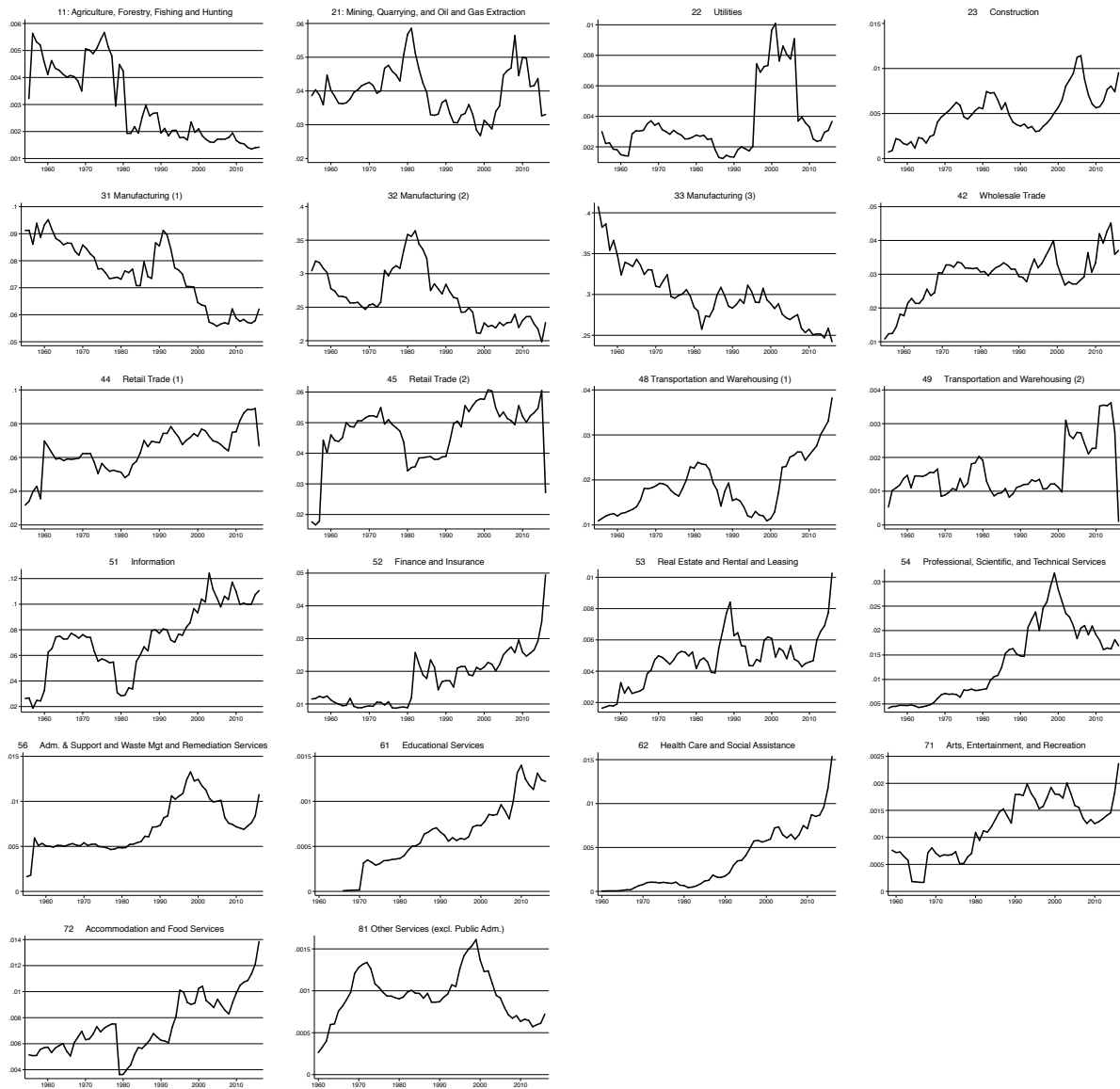


Figure 10.3: Industry Specific Market Shares.

10.4 Decomposition Within Firm by 2-digit Industry

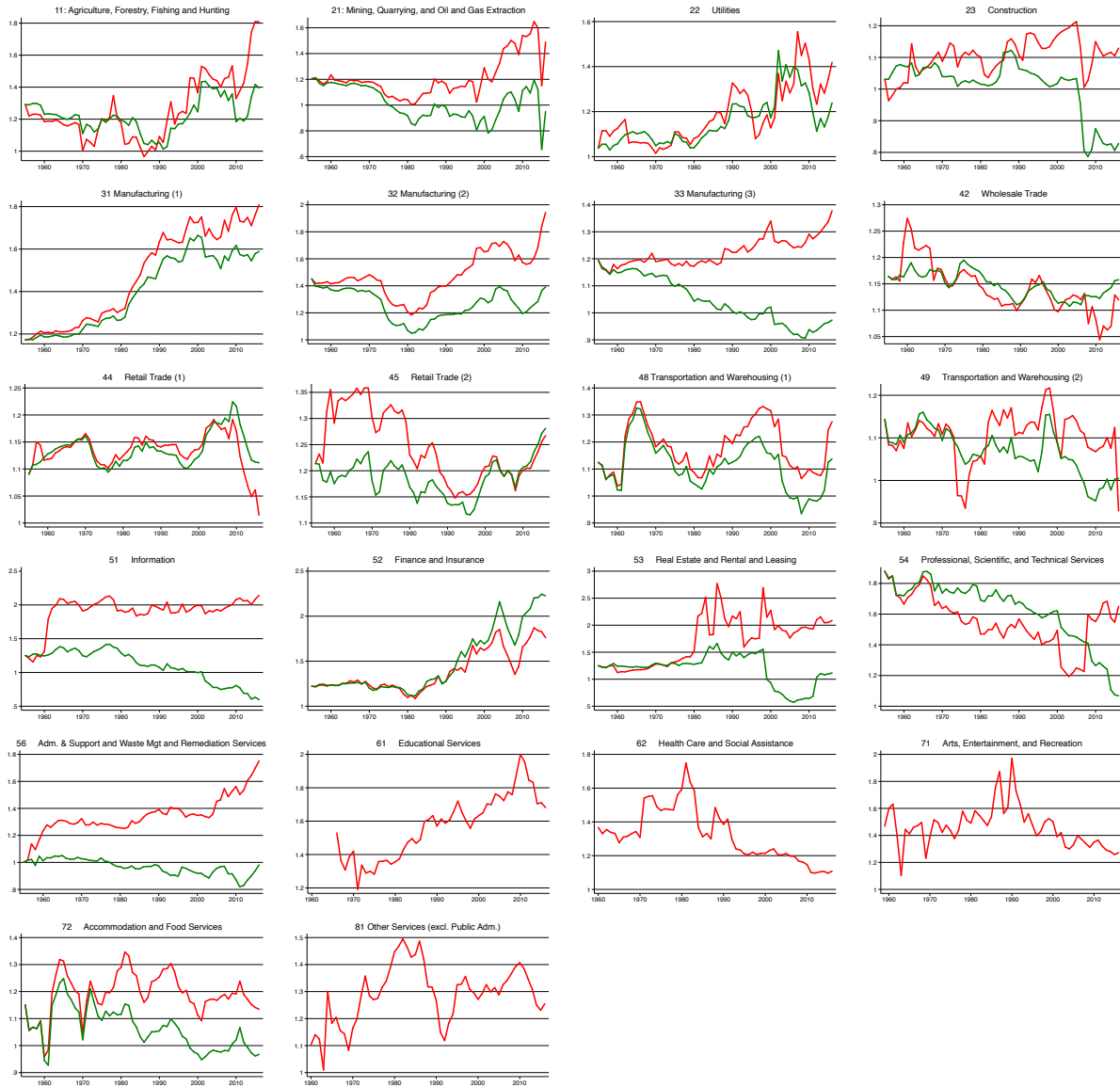


Figure 10.4: Decomposition Within Firm by 2-digit Industry.

For sectors 61, 62, 71, and 81 there is not enough data for at least one year to calculate the time difference. Recall that to calculate the elasticities, for that reason we use a five year rolling windows.

Appendix 11 A closer look at the Manufacturing sector

We repeat the analysis for the Census of Manufacturing where we treat Materials as the variable input instead of employment. The pattern in Figure 11.1 both of the average markup as well as the percentiles is similar to the markups that we calculate using employment as the variable input, and those from the publicly traded firms. Markups start increasing in the early 1980s. There is however a difference in the level of the markup, with the markups based on manufacturing being higher.

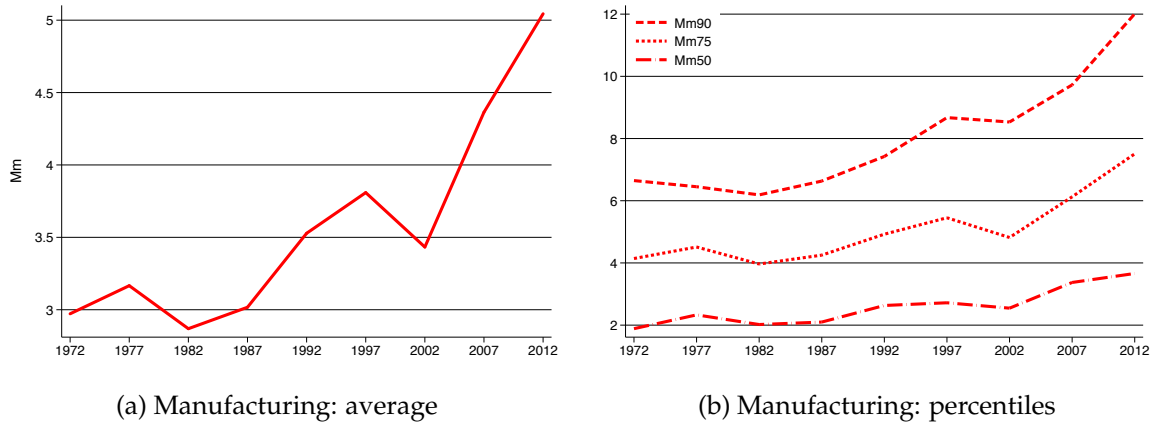


Figure 11.1: Markups in the US Census of Manufacturers. The variable input is materials. Averages and percentiles are sales weighted.

For the manufacturing sector we can compare the results from the Compustat and census datasets in greater detail. This is because in the census of manufacturing data we observe both expenditures on materials and employment. As discussed in the main text, one of the restrictions we face using the Compustat data is that we do not separately observe these; but rather rely on the bundle cost of goods sold (which of course captures additional expenses).

In particular, we compare the aggregate markup using labor and materials as candidate variable inputs, at census, with the aggregate markup obtained using COGS in Compustat. In Figure 11.2 the aggregate manufacturing markup (in red) using PF1 is repeated, and we compare it to two census-based series. First, the aggregate markup using materials as the variable input (the dashed green line), and second, the aggregate markup using labor as the variable input (the solid green line).

Given that all three specifications rely on distinct production technologies and therefore different output elasticities, we re-scale the census-based series to the value of the 1978 Compustat aggregate markup (around 1.26). The main focus is on the time series comparison; and the three series line up very closely, providing a very robust pattern of increasing margins during the eighties and nineties. Interestingly enough, the two census-based series strongly diverge during the last two census years (2007 and 2012). At this level of aggregation we cannot distinguish whether this due to compositional changes rather than firm-level changes in the input mix.

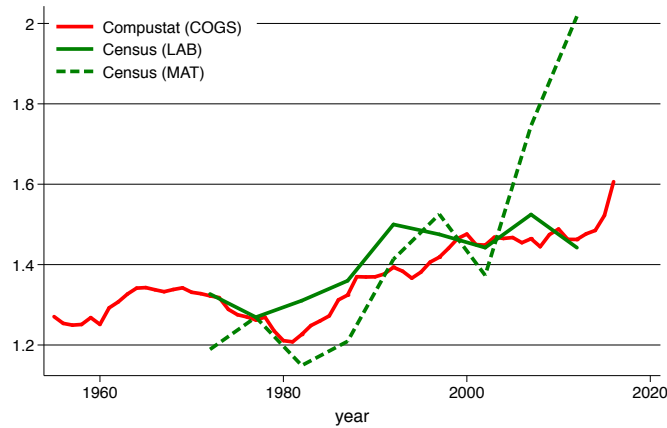


Figure 11.2: Aggregate Markup in Manufacturing: Variable inputs

Appendix 12 Wholesale trade in Census

The strategy to rely on the set of Compustat wholesale firms, and read off their labor cost shares, is not useful in the case of the wholesale sector, and this for two reasons. First, there are only a handful of firms reporting total labor cost, and judging from their identities they more often than not contain global firms, potentially combining wholesale with other activities, such as manufacturing. This implies that the level of the cost shares is expected to be a very large upper bound to the actual cost share, and as such of the output elasticity. This is inconsistent with reporting of high capital-intensive wholesale plants and firms, with very few employees, of which a large share are plausible fixed in nature and therefore constitute an overhead. The latter further invalidates the use of the first order condition on employment to compute the markup.

Taking these set of limitations as given, we perform the following comparison. In Figure 12.1 we compare the aggregate markup series for wholesale as obtained by applying our approach to the set of Compustat firms, relying on the variable input cost of goods sold (red line), to the publicly reported gross margins of the entire wholesale sector (black line).⁷³ Both series paint a very similar picture of the aggregate variable profit margins in the total sector.

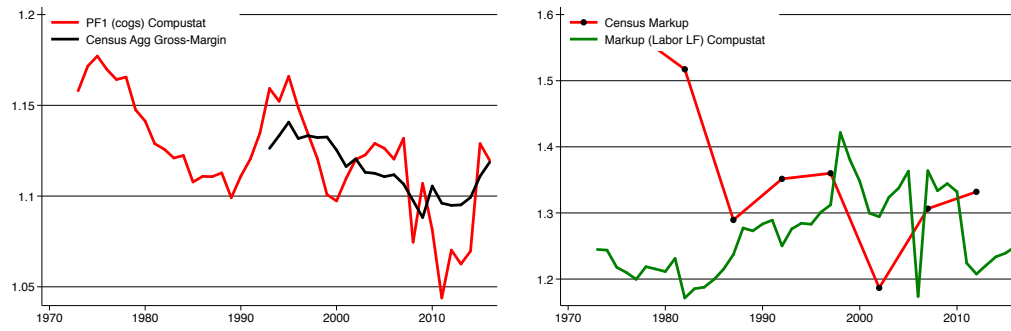
Finally, we consider the sector-aggregate markup using the first order condition of employment. The green line in the second panel in the figure below denotes the series obtained with the set of Compustat firms.⁷⁴ We overlay the aggregate markup obtained using the micro-census data (using a calibrated output elasticity of 0.085). Again the series are in broad agreement with each other. In fact in five out the seven census years, the values and patterns are remarkably similar. The census year 2002 has a sharper dip in the micro-census data, and a

⁷³The latter is obtained from accessing the public files <https://www.census.gov/data/tables/2016/econ/awts/annual-reports.html>, and converting the reported gross margins to our definition of the markup ($\mu = P/c$).

⁷⁴The difference between this green line in the right panel, and the red line in the left panel, is largely due to the different sample of firms reporting the cost of employment.

stronger drop from the 1982 to the 1987 census year, but the trajectories are in fact the same.

Figure 12.1: Wholesale trade: margins in census and Compustat



Appendix 13 Regressions of Market Value and Dividends on Markup Measure PF2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\ln\left(\frac{\text{Market Value}}{\text{Sales}}\right)$				$\ln(\text{Market Value})$			
$\ln(\text{Markup PF2})$	0.71 (0.03)	0.64 (0.02)	0.56 (0.02)	0.18 (0.03)	0.73 (0.03)	0.65 (0.03)	0.59 (0.02)	0.29 (0.03)
$\ln(\text{Sales})$					0.81 (0.00)	0.81 (0.00)	0.83 (0.00)	0.68 (0.01)
Year Fixed Effects		Y	Y	Y		Y	Y	Y
Sector Fixed Effects			Y				Y	
Firm Fixed Effects				Y				Y
R^2	0.05	0.12	0.21	0.68	0.68	0.71	0.73	0.89

	$\ln\left(\frac{\text{Dividends}}{\text{Sales}}\right)$				$\ln(\text{Dividends})$			
$\ln(\text{Markup PF2})$	1.03 (0.04)	0.93 (0.04)	0.79 (0.04)	0.23 (0.05)	1.01 (0.04)	0.90 (0.04)	0.77 (0.04)	0.25 (0.05)
$\ln(\text{Sales})$					0.94 (0.01)	0.92 (0.01)	0.93 (0.01)	0.76 (0.02)
Year Fixed Effects		Y	Y	Y		Y	Y	Y
Sector Fixed Effects			Y				Y	
Firm Fixed Effects				Y				Y
R^2	0.06	0.10	0.17	0.70	0.66	0.68	0.70	0.89

Table 13.1: firm-level Regressions: market values and dividends on markups (PF2) (clustered standard errors by firm in brackets).