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 15, 2019

Abstract

We document the evolution of market power based on firm-level data for the US economy since 1955. We measure both markups and profitability. 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. Quite strikingly, the median is unchanged. In addition to the fattening upper tail of the markup distribution, there is *reallocation* of market share from low to high markup firms. This rise occurs mostly within industry. 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: *Market Power; Markups; Profits; 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, Susanto Basu, Steve Berry, Tim Bresnahan, Emmanuel Farhi, Xavier Gabaix, Bob Hall, John Haltiwanger, Eric Hurst, Loukas Karabarbounis, Patrick Kehoe, Pete Klenow, Christian Michel, Ariel Pakes, Thomas Philippon, Esteban Rossi-Hansberg, Chad Syverson, James Traina, 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 the ERC, Consolidator Grant 816638, and Eeckhout from the ERC, Advanced grant 339186, and from ECO2015-67655-P. 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. First, we analyze markups, the most common measure of whether firms are able to price their goods above marginal cost. Traditionally in the Industrial Organization literature, this measure is of importance because it is informative about the technology that firms use and whether or not there is the efficiency in production. Based on firm-level data, we find that while aggregate markups was more or less stable between 1955 and 1980, there has been a steady rise since 1980, from 21% above cost to 61% above cost in 2016. More important than the increase in the aggregate markup, the main insight is that the *distribution of markups* has changed: the median is constant, and the upper percentiles have gone up substantially. This rise in markups by a few firms has gone together with *reallocation* of economic activity. Few firm have high markups and are large, the majority firms see no increase in markups and lose market share.

Markups alone do not tell the full story about market power. For example, markups may be high because overhead costs or fixed costs are high. In that case, the firm charges prices well above marginal costs in order to cover fixed costs. We therefore also analyze measures of *profitability*, that take into account not only the marginal cost, but total costs, including the expenditure on capital and fixed or overhead costs. Because measuring profitability is challenging, we document a rise in different measures, ranging from accounting profits to stock market performance. We show that both measures, markups and profitability, are related. While we do find that there is an increase in overhead costs, the rise of markups cannot exclusively be attributed to overhead. Markups have gone up more, and as a result, so has profitability. The increase in both markups and profitability provides evidence that *market power has increased.*

Once we have robustly established the facts, 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 and capital shares, as well as the decrease in business dynamism and labor reallocation.

Measuring market power is notoriously hard. The most widely used measures of market power such as concentration ratios, e.g., the Herfindahl-Hirschman Index (HHI), have serious pitfalls because they are sensitive to the definition of a market. This is especially problematic when analyzing market power in the aggregate across different industries and over long time periods, where market definitions change. While HHI is a good measure under certain circumstances – especially when the market definition is stable and when firms compete à la Cournot for example – and it is widely used, it is not an adequate measure of market power for the macroeconomy across time and space.[1]

The evidence on market power we have to date comes from case studies of specific industries,[2] 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. This approach 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 and prohibit its applicability for macroeconomic questions.

In this paper we follow a radically different approach to estimate markups, the so-called production approach. Building on Hall (1988), recent advances in the literature on markup estimation by De Loecker and Warzynski (2012) relies on individual firm output and input data, and posits cost minimization 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 heterogeneous markets over a long period of time. Second, we can rely on publicly available accounting 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 on backing out economy wide measures of market power.[3]

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. 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)). We also perform on Census data where for selected industries we have the universe of firms.[5] We find that the distribution of markups changes dramatically since 1980: most firms see no rise

[1] See Bresnahan (1989) and Syverson (2019)

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

[3] While the approaches – the demand approach and the production approach – differ, the obtained estimates should be similar. In Appendix 7 we compare estimates from the literature using the demand approach to our estimates for the corresponding sector.

[4] The 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.

[5] The 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.

in markups, while those in the upper tail experience a sharp rise. At the same time, there is a reallocation of economic activity towards high markup, large firms, consistent with the superstar firm effect that Autor, Dorn, Katz, Patterson, and Van Reenen (2017) find.

We then analyze firm profitability. The objective is to analyze whether markups have not increased exclusively due to a rise in overhead costs.[6] To address this issue, we calculate the profit rate, which is total sales minus all cost (including overhead and the expenditure on 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 from 15% to 21% of total cost, markups have increased even more and firms charge an excess markup that more than compensates for overhead. In fact, we find that the firms with the highest overhead costs charge the highest excess markup and therefore have the highest profits. Like markups, the increase in the average profit rate is driven by a change in the distribution, especially the 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.

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 find that due to reallocation of economic activity towards high markup firms, the decline in the economy-wide labor share is predominantly driven by large, high markup firms that have individually low labor shares. This is consistent with the findings in Autor, Dorn, Katz, Patterson, and Van Reenen (2017) and Kehrig and Vincent (2017) that large firms drive the decline in the aggregate labor share. Our finding is a slightly nuanced version: market power as a common cause determines both the increase in firm size and the decline in 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 derives firm-level markups, which circumvents the limitations of the HHI measure.[7]

[6]While we find that the rise in markups has been accompanied by a rise in market power, even if the rise in aggregate markups we document here was purely a function of rising overhead costs and came with no change in market power, this finding would still be deeply significant. Markups are a fundamental variable throughout macroeconomics – from the benchmark New Keynesian model, to any standard endogenous growth model – as they are central to understanding technology, the efficient allocation of resources between firms, and how we think about trade-offs between static and dynamic efficiency.

[7]Most notably, concentration is not necessarily related to market power when products are differentiated (see

Hartman-Glaser, Lustig, and Zhang (2016), Autor, Dorn, Katz, Patterson, and Van Reenen (2017) and Kehrig and Vincent (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) show 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, as does Kehrig and Vincent (2017). Like ours, their results are based on firm-level data, not macroeconomic aggregates.

We share share with these three papers that the reallocation of economic activity towards large firms has substantial implications that resolve a number of puzzles in macroeconomics, most notably the decline in the labor share. We argue that market power and the rise of markups is the *common cause* of both the reallocation towards large firms and the decline in the labor share. The decline in the labor share holds at the firm level, from firm optimization: as markups increase, firms spend less on labor. With an economy-wide increase in market power, enough firms reduce their expenditure on labor which translates in an aggregate decline in the labor share, as observed in the macro aggregates.

In this paper we focus on robustly establishing the facts regarding the evolution of market power and are agnostic about the origins of the rise in market power and the corresponding reallocation of economic activity towards high markup firms. The most prominent explanations are technological change and the change in the market structure (for example due to the decline in antitrust enforcement, as argued by Gutiérrez and Philippon (2018)). In a companion paper, De Loecker, Eeckhout, and Mongey (2018) derive quantitatively that *both* technological innovation and a change in market structure are at the root cause of the rise in market power. Ex ante, the effect on welfare is ambiguous: large, high markup firms are more productive, but they also extract more rents from the customer and affect the labor market adversely through lower wages. In our quantitative exercise, we find that the net effect is negative.

Finally, while we focus in this paper exclusively on the United States, there is evidence of a rise in market power around the world. Using data on publicly traded firms around the world, in De Loecker and Eeckhout (2018a) we find remarkably similar patterns of the rise in market power since 1980. As for the US publicly traded firms, there is a sharp rise between 1980 and 2000, a period of stagnating markups in the 2000s followed by another sharp rise starting around 2010. The markup for the publicly traded firms increases from 1.1 in 1980 to 1.6 in 2016.

2 Empirical Framework and Data

We present the empirical framework that allows us to derive a markup measure for each firm covering the entire economy, over more than six decades. Our empirical framework uses the cost minimization approach where firms choose the optimal bundle of variable inputs of production. This reasonable assumption on firm behavior only relies on firm-level revenue and input expenditure data for firms across the US economy. As such we do not impose restrictions on product market competition and consumer demand.

Bresnahan (1989)), and an adequate concentration measure requires precise knowledge of what constitutes a market with information on all firms in that market.

In this section, we first present the model, and then discuss the particular implementation in the datasets we use. Our focus is to provide a robust description and analysis of markups across producers using different methods and approaches.

2.1 Obtaining Markups from Producer Behavior

The markup is commonly defined as the output price divided by the marginal cost. Measuring markups is notoriously hard as marginal cost data is not readily available, let alone prices. There exist three distinct approaches to measure markups. First, the accounting approach relies on directly observable gross (or net) margins of profits. While this approach is straightforward to implement, it suffers from well-known problems, chief among them is the inability to directly measure marginal cost of production. A straightforward way to circumvent this problem is to equate average to marginal costs, but this imposes strong and unrealistic restrictions on firm-level cost structures.[8]

The second approach was developed in the modern Industrial Organization literature (see Berry, Levinsohn, and Pakes (1995) and Bresnahan (1989)), and relies on the specification of a demand system that delivers price-elasticities of demand. Combined with assumptions on how firms compete, the demand approach delivers measures of markups through the first order condition associated with optimal pricing. This approach, while powerful in other settings, is not applicable in our setting 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. This to successfully estimate price elasticities of demand, and specify particular models of price competition for all sectors.

Instead, we rely on a third approach, the production approach. This approach is based on the insight of Hall (1988) to estimate markups from the firm's cost minimization decision. Hall (1988) used industry aggregates, De Loecker and Warzynski (2012) recently proposes 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, however, an explicit treatment of the production function to obtain the output elasticity of at least one variable input of production.

Before we discuss the production approach, on which we rely to measure markups, it is instructive to go back to the underlying assumptions of the accounting and so-called demand approach. Throughout we define markups as the price-to-marginal cost ratio:

$$\mu \equiv \frac{P}{c} \tag{1}$$

In essence the simplicity of the accounting approach is to simply multiply through by total output (Q) and obtain:

$$\frac{P}{c} = \frac{PQ}{cQ}. \tag{2}$$

[8] See Karabarbounis and Neiman, 2018 for a recent implementation of this approach. The discussion around the merits of the use of accounting markups (and profits) dates back to Bresnahan (1989).

The entire approach rests on the assumption that the object cQ is directly observable in the data. There are three main assumptions, and therefore complications. First, this approach relies crucially on marginal and average cost of production to be equal. This requires constant returns to scale (CRS) in production, and the absence of economies of scale, i.e., there are no fixed costs. Second, this approach implicitly relies on all relevant factors of production to be perfect substitutes in production. Third, and related, the measure of cost (cQ) is not equal to marginal cost if it includes cost items that do not vary with output. Note that in the accounting markup equals the profit rate when all cost items (including fixed factors like capital, and investment activities such as R&D and advertising) are included in the measure cQ.

The demand approach relies on an estimated demand curve (having data separately on prices and quantities for all products in a pre-specified market) and a particular model of competition to back-out c from a first-order condition resulting from profit maximization. The production approach frees up all these restrictions on conduct and demand by computing the marginal cost of production directly from the cost minimization condition for a single variable input of production.

2.2 The Production Approach

Consider an economy with N firms, indexed by $i = 1, ..., N$. Firms are heterogeneous in terms of their productivity Ω_{it} and production technology $Q_{it}(.)$.[9] In each period t, firm i minimizes the contemporaneous cost of production given the production function:

$$Q_{it} = Q_{it}(\Omega_{it}, \boldsymbol{V}_{it}, K_{it}), \tag{3}$$

where $\boldsymbol{V} = (V^1, ..., V^J)$ is the vector of variable inputs of production (including labor, intermediate inputs, materials,...), K_{it} is the capital stock and Ω_{it} is productivity . 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 use information on a bundle of variable inputs, and not the individual inputs, in the exposition we treat the vector $\boldsymbol{V}$ as a scalar V.[10] We consider the Lagrangian objective function associated with the firm's (conditional) cost minimization:[11]

$$\mathcal{L}(V_{it}, K_{it}, \lambda_{it}) = P_{it}^V V_{it} + r_{it} K_{it} + F_{it} - \lambda_{it}(Q(\cdot) - \overline{Q}_{it}), \tag{4}$$

where P^V is the price of the variable input, r is the user cost of capital,[12] F_{it} is the fixed cost, $Q(\cdot)$ is the technology specified in equation (3), $\overline{Q}$ is a scalar and λ is the Lagrange multiplier. We

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

[10] We can of course equally consider multiple inputs facing adjustment frictions – see De Loecker and Warzynski (2012) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a discussion. We do exactly that when we use labor as the variable input in our robustness exercise.

[11] The conditional statement refers to the fact that we condition on the factors of production that are chosen dynamically. E.g. if capital faces adjustment costs or simply time to build, the firm chooses variable inputs to minimize cost, given the level of capital that was set in the previous period.

[12] Below, we will use lower case letters to denote logs, for example, $\log(P^V) = p^V$.

assume that variable input prices are given to the firm.[13] We consider the first order condition with respect to the variable input V, and this is given by:

$$\frac{\partial \mathcal{L}_{it}}{\partial V_{it}} = P_{it}^V - \lambda_{it}\frac{\partial Q(\cdot)}{\partial V_{it}} = 0. \tag{5}$$

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

$$\theta_{it}^v \equiv \frac{\partial Q(\cdot)}{\partial V_{it}}\frac{V_{it}}{Q_{it}} = \frac{1}{\lambda_{it}}\frac{P_{it}^V V_{it}}{Q_{it}}. \tag{6}$$

The Lagrange multiplier λ is a direct measure of marginal cost (tracing out the value of the objective function as we relax the output constraint), and we define the markup as price-marginal cost ratio $\mu = \frac{P}{\lambda}$, where P is the output price. Substituting marginal cost for the markup to price ratio, we obtain a simple expression for the markup:

$$\mu_{it} = \theta_{it}^v \frac{P_{it}Q_{it}}{P_{it}^V V_{it}}. \tag{7}$$

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.

The markup formula (7) derived under the production approach highlights that the marginal cost of production is derived from a single variable input in production, without imposing any particular substitution elasticity with respect to other inputs (variable or fixed) in production, or returns to scale. It is instructive to contrast it to the accounting approach introduced above: only in the case of a CRS single variable input (V) production function without fixed costs will the correct markup be measured by the sales to the variable input expenditure.

An important component of the markup formula under the production approach is therefore to obtain the output elasticity θ_{it}^v. In Appendix A we discuss in detail the different approaches we take and we appraise the merits and shortcomings of each approach. We distinguish between obtaining output elasticities from estimating the production function, and from cost shares.

2.3 Data

In order to cover the longest possible period of time, *and* to have a wide coverage of economic activity we use data on publicly traded firms. To our knowledge, Compustat is the only data

[13]This approach does not preclude input providers charging 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. We 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). In fact the production approach can be used to recover the input-price elasticity by exploiting multiple first order conditions across a set of variable inputs – for applications of this approach see Morlacco (2017), Mertens (2019) and Rubens (2019).

source that provides substantial coverage of firms in the private sector over a substantial period of time, spanning the period 1950 to 2016. 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.[14] We deal with the selection from the publicly traded firms in two ways. In Section 3.4 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 1.2). 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.[15] 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 directly attributable to the production of the goods sold by the firm and includes materials and intermediate inputs, labor cost, energy,...[16] 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 B.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%.[17]

Our data also has a measure of Overhead, booked under "Selling, General and Admin-

[14] There 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.

[15] The 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).

[16] The 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.

[17] Below, when we investigate the capital share (the expenditure on capital divided by sales) and 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.

istrative 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. 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.

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. The Census does not include information on overhead directly.[18] In section 3.4 we will analyze markups for Manufacturing, Retail and Wholesale. A detailed description of the Census data is in Appendix B.

3 The Evolution of Markups in the US economy

The bulk of our analysis is for the Compustat data where we observe firms across a wide range of sectors and time. Because we have firm-level markups, the main focus of attention is on the evolution of the *distribution* of markups. We first report the average markup, then detailed properties of the distribution, and finally we decompose the average markup in order to single out the reallocation of economic activity towards high markup firms.

3.1 Aggregate Markups

The measure of markups in equation (7) is the product of the output elasticity θ and the inverse of the variable input's revenue share $\frac{PQ}{P^V V}$. The latter is directly measured in the firm's income statement, and we estimate the former. Our estimated output elasticities are sector and time specific and thus capture technological differences across sectors and time.

We calculate the average markup as follows:

$$\mu_t = \sum_i m_{it}\mu_{it}, \tag{8}$$

where m_{it} is the weight of each firm. In our main specification, we use the share of sales in the sample as the weight. 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

[18]However, one can obtain multiple sources of information about overhead costs in the Census data. We leave this for future work.

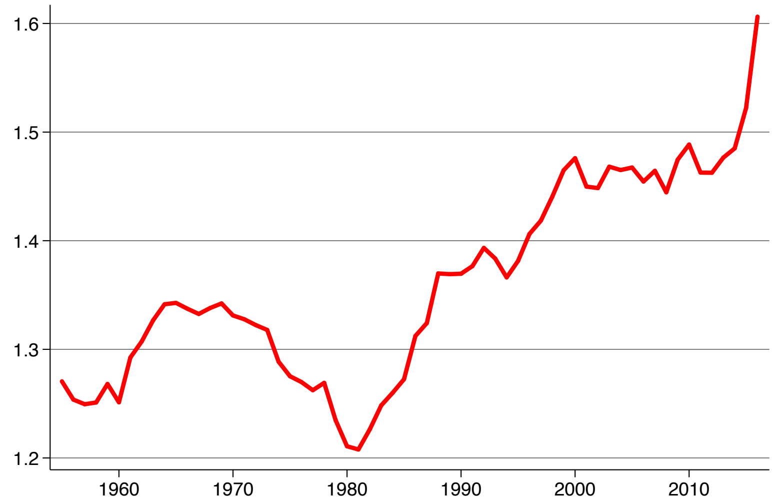


Figure 1: Average Markups. Output elasticities θ_{st} from estimated production function are time-varying and sector-specific (2 digit). Average is revenue weighted. Evolution 1955-2016.

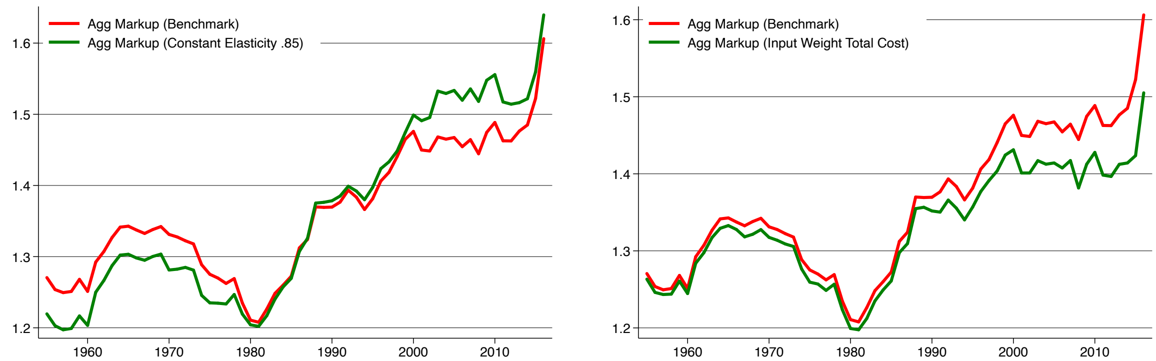
1980. Since 1980 there has been a steady increase to 1.61. In 2016, the average markup charged is 61% over marginal cost, compared to 21% in 1980. In Online Appendix 5 we report a few examples of individual firms' markups.

In broad terms, there are three sources that can account for this rise in aggregate markups: 1. the inverse ratio of the cost share of sales; 2. the output elasticity; 3. the weight. To show the sensitivity of the average markups to each of these determinants, in Figure 2 we plot the average markup with input weights and the average markup with a fixed, time-invariant output elasticity.

When we fix the output elasticity to be time-invariant (calibrated to 0.85, the average cost share), we find that the pattern of markups (Figure 2a) is similar to that in the benchmark with estimated output elasticities. This tells us that the rise in markups is not due to the change in the estimated output elasticity, which captures technological change under our production function specification. Consistent with this evidence, we find that the output elasticities vary very little over time (see also Figure 12b below).

Next, we investigate the role of the input weight, the importance of which has first been flagged by Grassi (2017) and Edmond, Midrigan, and Xu (2019). When firms have market power, they charge higher prices, and as a result, dampen demand. With lower demand, the quantity sold and the inputs used to produce are lower. Nonetheless, revenue (price times quantity) is higher. As a result, firms with higher markups tend to have higher revenue weights relative to their input weights.

This is exactly what we see in Figure 2b. The level of markups is lower throughout, and the rise is less pronounced, which indicates that the gap between inputs and sales has grown. The widening gap indicates that there is a change in the equilibrium outcome and the market structure. Moreover, as we will see in the next two sections, the widening of the markup distribution and the reallocation of sales towards high markup firms can explain why the gap has widened.



(a) Constant elasticity

(b) Input weighted (Total Cost)

Figure 2: Aggregate Markup

Here we use the total cost (computed as the sum of COGS, SG&A and rK) as the input weight and for different weighting measures, the gap between the sales and the input weighted aggregate markup is larger.[19] Since we are interested in the properties of the *entire* distribution of markups, we believe it is instructive to show as many different moments as possible. In particular, the gap between the input and the revenue weighted aggregate markup informs us about the underlying mechanism, about the underlying distribution, and about the reallocation.

We use as our benchmark the revenue weighed markup for following reasons. First, a substantial portion of what is going on in the output market is reallocation (see below) of revenues towards high markup firms. We cannot capture this crucial phenomenon with input weighted markups. The revenue weighted markup therefore informs us about the economic mechanism and we show in a companion paper (De Loecker, Eeckhout, and Mongey (2018)) that this is important determinant in explaining the rise of market power. Second, in order to study market power, we link markups with profit rates (Section 4 below). Profit rates are traditionally aggregated with revenue weights, and consistency then calls for revenue weighting of the markup as well. Finally, revenue weighting is a common benchmark that is commonly used, most notably for widely used economic indicators such as GDP and in the context of market power, HHI.

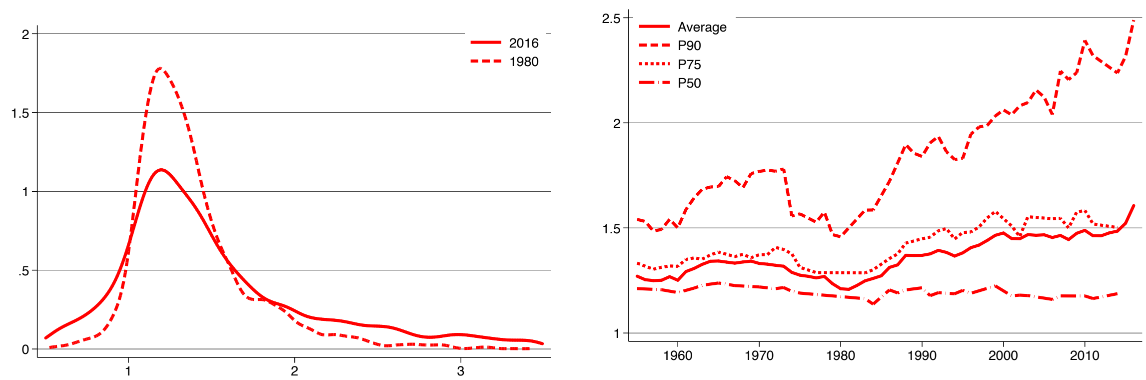
The bottom line is that our finding for the benchmark measure of aggregate markups is robust. This implies that the bulk of the action comes from the increase in the wedge of sales to COGS. The rise is not driven by technological change (changing output elasticities) and the weighting scheme informs us about the underlying mechanism where the increasing gap between the revenue and input weighted aggregate markup tells us that firms spend less on variable inputs.

[19] We revisit the weighting extensively in the Robustness Section 6 after we have introduced the markup distribution, reallocation and profit measures.

3.2 The Distribution of Markups

While average markups make for a good headline, they do not fully capture the underlying distributional change in markups. The advantage of our method to calculate markups is that we obtain one for each firm, so we have a distribution of markups. A key finding is that the increase in markups is driven by a few firms, without any increase for most.

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 3a). 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.



(a) Kernel Density (unweighted)

(b) Percentiles Markup Distribution (revenue weight)

Figure 3: The Distribution of Markups μ_{it}

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 3b). 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).

The increase in the average markup comes entirely from the firms with markups in the top half of the markup distribution. The median (P50) and the percentiles below the median are invariant over time. Most firms see no increase in markups.[20] For the higher percentiles, markups increase. For the 90th percentile in particular, the increase is sharpest. Between 1980 and 2016, it increases from 1.5 to 2.5. This indicates that the change in average markup is largely driven by a few firms that currently have much higher markups than decades ago.[21]

[20]Because the distribution is revenue weighted and the larger firms tend to have higher markups, this implies that the vast majority of firms see no rise in markups.

[21]This is consistent with the evidence in Kehrig (2011). He studies the cyclicality of productivity and finds that the dispersion in TFPR is increasing, especially in the upper tail, where TFPR captures both markups and cost-side heterogeneity. In Appendix E we further explore the distributional change by modeling the markup (as well as sales and employment) as an autoregressive process and confirm the rise in the standard deviation.

3.3 Reallocation of Economic Activity

The rise in the average markup is driven by a few firms in the top of the distribution. Most firms see no increase in markups, few firms see a large increase. We can further decompose the increase in the weighted average markup into the component that is due to the increase in the markup itself, and the component that is due to the reallocation of economic activity towards high markup firms.

Inspection of Figure 3a already shows that there is a change in the distribution of unweighted markups. The fatter tail is evidence that more firms have higher markups. But even if the distribution of unweighed markups had remained unchanged, the weighted aggregate markup could have gone up if the firms with higher markups now obtain a higher share of the market. This reallocation of economic activity towards higher markup firms is important to understand the implication that market power has on the concentration of economic activity in the hands of a few, dominant firms. 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).

Because the change in aggregate markups is a combination of both the rise in unweighed markups and a reallocation of economic activity, we decompose the average markup at the firm level as follows:

$$\Delta \mu_t = \underbrace{\sum_i m_{i,t-1} \Delta \mu_{it}}_{\Delta \text{ within}} + \underbrace{\underbrace{\sum_i \tilde{\mu}_{i,t-1} \Delta m_{i,t}}_{\Delta \text{ market share}} + \underbrace{\sum_i \Delta \mu_{i,t} \Delta m_{i,t}}_{\Delta \text{ cross term}}}_{\Delta \text{ reallocation}} + \underbrace{\sum_{i \in \text{Entry}} \tilde{\mu}_{i,t} m_{i,t} - \sum_{i \in \text{Exit}} \tilde{\mu}_{i,t-1} m_{i,t-1}}_{\text{net entry}}, \tag{9}$$

where $\tilde{\mu}_{it} = \mu_{it} - \mu_{t-1}$, and $\tilde{\mu}_{it-1} = \mu_{it-1} - \mu_{t-1}$.[22]

We apply the insights from the productivity-decomposition literature, and while this decomposition appears very similar to that in (10), 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.[23] But also the interpretation differs. Following Haltiwanger (1997), we consider a theoretical counterfactual where the 'Δ within' term measures the average change that is merely due 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'.[24] 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

[22]We demean the (lagged) markups by the appropriate aggregate (share-weighted) level, in order to correctly identify the role of the reallocation term – see Haltiwanger (1997) for more discussion.

[23]Entry and exit in the data of publicly traded firms comprises entry and exit in the database. This consists of firms listing and delisting, as well as merger and acquisition activity.

[24]The 'Δ cross term' is virtually zero in the experiments we perform.

for example, then this term will be positive.

We perform this decomposition across firms in the entire economy. To best present this decomposition, in Figure 4 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 (9).[25]

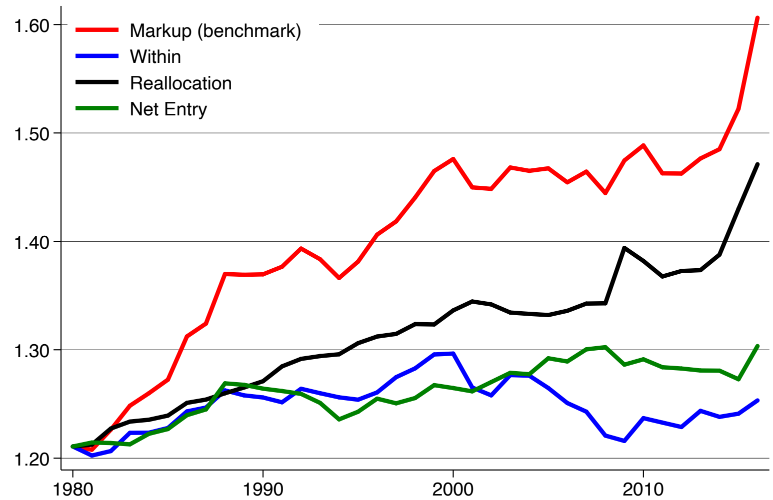


Figure 4: Decomposition of markup growth at the firm level.

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 3a, where we see an increase in the upper right tail.

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

[25] In the Online Appendix section 4 we tabulate the measured yearly changes of each of the four components for all years between 1955 and 2016. The cumulative representation in Figure 4 shows decomposition of the change in markups in a more concise way.

that large firms have grown in size relative to small firms, and those firms tend to operate in 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 3a, 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 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.

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. Cumulatively over the whole period, reallocation accounts for two thirds of the rise in markups.

The decomposition exercise above implies that the reallocation component captures movements of firms across all sectors. In Appendix 4 we perform the same decomposition for each of the broad sectors of the economy, where reallocation of economic activity is measured within sector.

In contrast to the firm level decomposition (economy-wide and within sector), we also analyze the decomposition of the rise of markups by firm size at the *sectoral level*, i.e., within and between sectors. Is the increase in markup over time due to a change of markup 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:

$$\Delta\mu_t = \underbrace{\sum_s m_{s,t-1}\Delta\mu_{st}}_{\Delta \text{ within}} + \underbrace{\sum_s \mu_{s,t-1}\Delta m_{s,t}}_{\Delta \text{ between}} + \underbrace{\sum_s \Delta\mu_{s,t}\Delta m_{s,t}}_{\Delta \text{ cross term}}. \tag{10}$$

We consider the change over 10 year periods starting in 1956 in Table 1.[26] 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.

In sharp contrast with the *firm level* decomposition where most of the increase is due to reallocation between firms, the *sectoral* decomposition shows that most of the increase in markups

[26]The decomposition for the 3 and 4-digit industry classification is reported in the Online Appendix.

	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

Table 1: Sectoral decomposition of 10 year change in Markup.

occurs *within all sectors*, not between sectors. This 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 12.1 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.

Further evidence that most of the rise in market power occurs *within* industry comes from comparison of our results with those based on aggregate data (industry level or economy-wide). Using national accounts data by sector, Hall (2018), extending his original work Hall (1988), finds a rise in market power but only by about 20 points, half of the increase we find with firm level data.

To investigate where the discrepancy from using of aggregate data comes from, we employ our firm level data and aggregate them at the industry level. In Figure 5, we plot our benchmark aggregate markup together with three series of industry averages, based on our firm-level data, summed up to industry averages: one were we treat the entire economy as one industry (blue), one where we aggregate at the industry level with constant elasticities (dashed black), and one where we aggregate at the industry level and use the estimated, time-varying and industry-specific elasticities.

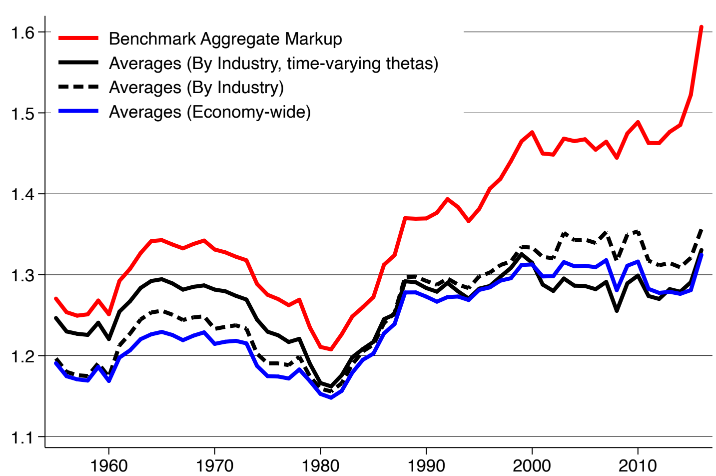


Figure 5: Using Industry and Economy-wide Averages versus Aggregating Micro Data

The three series with averages look similar. The average markups is below our benchmark, and it grows at half the rate. The increase between 1980 and 2016 is from 1.15 to 1.35 approximately, by about 20 points as in Hall (2018). This clearly establishes that a substantial part of the increase occurs within industry and that some of that change is lost when taking averages. To see how that can occur, consider the following comparison. To make the comparison as transparent as possible, let us abstract from any technological change or sectoral-heterogeneity in output elasticities and simply keep θ constant throughout. We compare our aggregate markup with the one obtained using aggregate data:

$$\sum_i m_{it} \frac{S_{it}}{P^V_{it} V_{it}} \neq \frac{\sum_i S_{it}}{\sum_i P^V_{it} V_{it}} \tag{11}$$

The reason that the two objects are not equal to each other is due to the heterogeneity in markups across firms.[27] Aggregation of a non-linear function (Jensen's inequality) leads to different outcomes. This is the case for any cross-section, but importantly with the reported increasing skewness in the underlying markup distribution, this difference becomes larger over time. The widening gap between the micro and the macro ratios is simple economics: if market share is reallocating towards the higher markup firms, this reinforces the process of increased skewness, due to the increased correlation of markups and market share (in a given industry or in the entire economy depending on the focus).

It is clear from the figure that the aggregate-based series trend up but to a much lesser extent and this is to be expected given the increased dispersion. This tells us that the dispersion and skewness of the distribution has increased over time. Much of the rise that we observe in the average markup disappears once we use industry or economy-wide averages. This tells us that most of the heterogeneity in markups is within industry and that the reallocation of market shares (see also below) occurs mainly within industries.

3.4 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 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 B.

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,

[27] With identical firms the market share is $m_{ij} = N^{-1}$ and both ratio's are identical.

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.[28]

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.[29] This leads to the standard recovery of the output elasticity for the variable input:

$$\theta^V_{nt} = N_{nt}^{-1} \sum_{j \in n} \frac{P^V_{jt} V_{jt}}{P^V_{jt} V_{jt} + r_{nt} K_{jt}}, \tag{12}$$

where j denotes a plant active in industry n; in this case a unique 4 digit NAICS code.[30] 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 (6.1).

Finally, we aggregate the plant-level markups to obtain firm-level markups, the ultimate object of interest in this analysis. This also makes our results consistent with the analysis performed for the Compustat sample.[31] 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 6 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 3b). With data only in five-year intervals, the patterns are obviously less detailed.

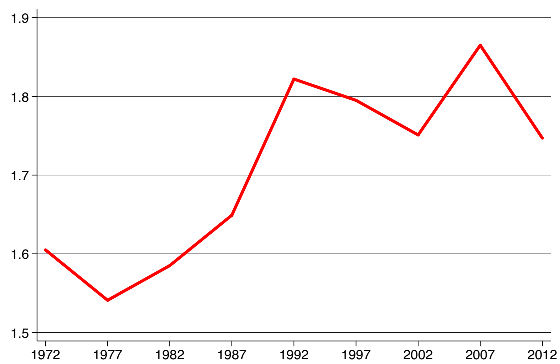
Starting with Manufacturing (Figures 6a and 6b), we see average markups that start to increase from 1977 onwards, from around 1.55 up to around 1.8. This pattern mirrors what

[28]Some 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.

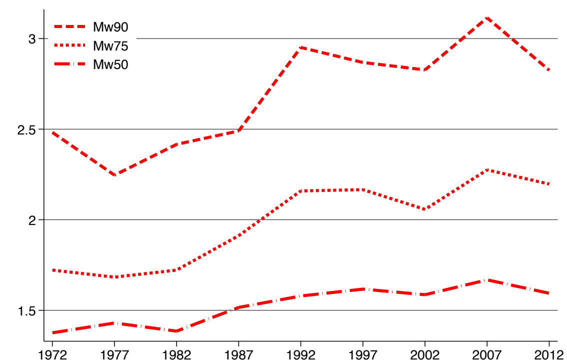
[29]These are made available by Foster, Haltiwanger, and Syverson (2008) and are accessed through the census file. Alternatively the output elasticities can be obtained by estimating the 86 distinct production functions using an approach as outlined in section 2. We opted to rely on the cost-share approach 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.

[30]We remind the reader that there are four distinct levels of aggregation in our analysis: plants (j), firms (i), industries (n – i.e., 4 digit NAICS) and sectors (s – i.e., 2 digit NAICS).

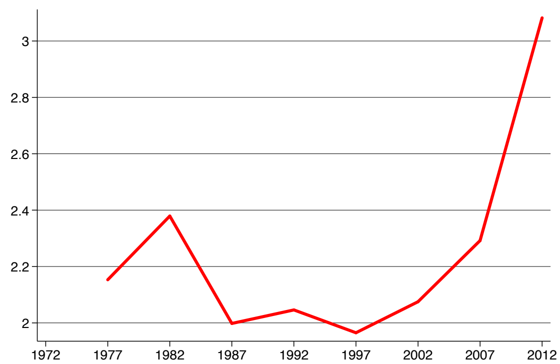
[31]In the case of the Wholesale and Retail sectors, the output elasticity is measured at the level of the 2 digit NAICS code – i.e., NAICS 42 and 44-45 combined, respectively.



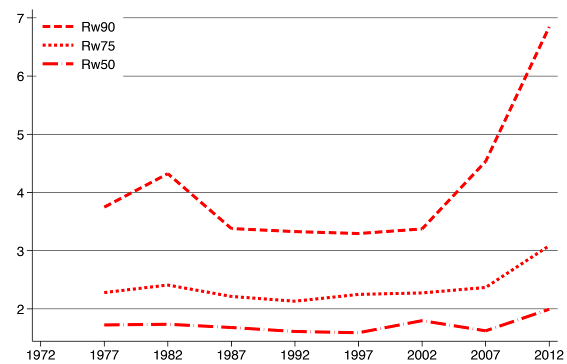
(a) Manufacturing: average



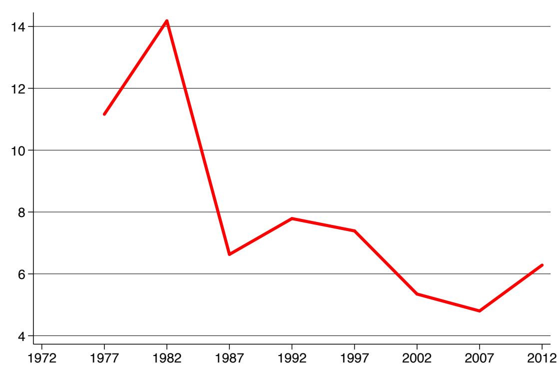
(b) Manufacturing: percentiles



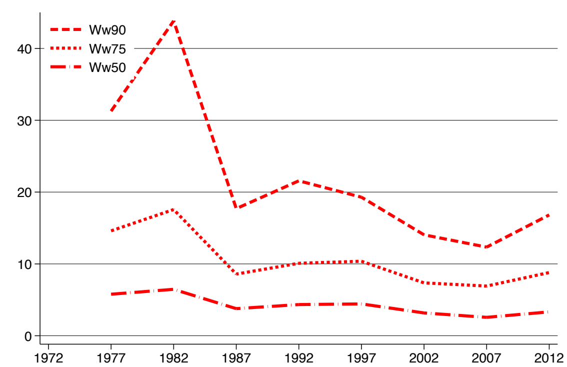
(c) Retail: average



(d) Retail: percentiles



(e) Wholesale: average



(f) Wholesale: percentiles

Figure 6: Markups in the US Censuses: Manufacturing, Retail and Wholesale. The variable input is employment. Averages and percentiles are revenue 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.

we find in the whole sample of publicly traded firms as well as in the publicly traded firms in manufacturing.

We also calculate the markup using materials 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.[32]

Like for the publicly traded firms, the pattern in Retail (Figures 6c and 6d) 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 18, 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.

4 Market Power and Profitability

The documented rise in markups does not necessarily imply that firms have more market power, and therefore higher economic profits. In fact, increasing markups can come from a variety of reasons that are not associated with a decline in aggregate welfare.[33] For example, a decrease in marginal costs, an increase in fixed costs or innovation, an increase in demand or in its elasticity, a change in the market structure, new product varieties, all lead to increasing markups without necessarily implying higher profits.

While the textbook definition of market power is the case whereby a firm can command a price above the marginal cost of production (markup), any conclusions regarding whether market power increased will greatly depend on the pattern of overhead costs, or any other factor affecting the cost structure of firms (like innovation activities such as R&D). Therefore, before we can conclude whether the higher markups are associated with market power, we need to analyze profits. In the absence of detailed data, the mapping from markups to market power (and therefore welfare) can only be done through a particular model of the economy.

With the accounting data available, we assume that we can observe profits as the wedge between sales and all variable and fixed costs (including innovation, advertising, and others). In what follows we consider higher market power a situation whereby a firm can generate higher profits.[34]

Key here is the evolution of overhead and capital as a share of expenditure. If those have

[32]For more details on the use of multiple variable inputs in the manufacturing sector, see the Online Appendix section 17.

[33]It does, however, potentially generate distributional implications.

[34]Profits 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. In this paper, we do, however, abstract away from such frictions.

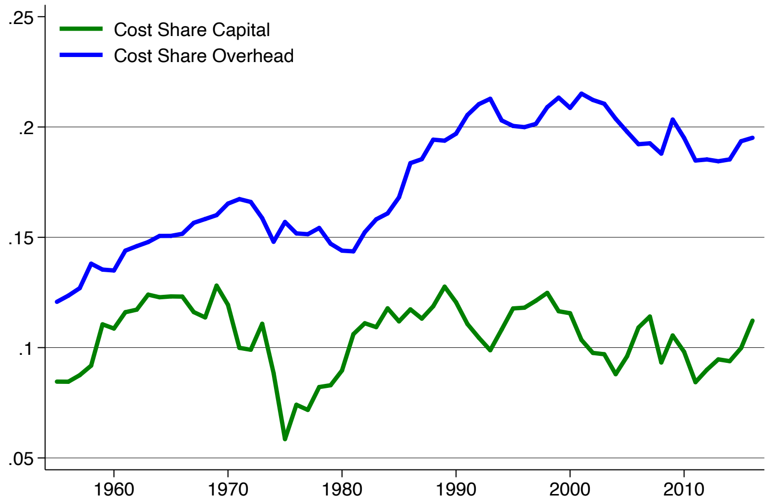


Figure 7: Aggregate Overhead and Capital Cost Shares of Total Cost

increased and markups have increased at the same rate, then the higher markups are charged only to recover the higher overhead costs and capital investment. In Figure 7 we plot the evolution of overhead and of capital as a share of total costs.[35] In our data, the capital share has been fairly constant, in line with the findings by Barkai (2017). Instead, overhead as a share of total expenditure has seen an increase. The rise in overhead costs thus requires us to analyze profits in order to conclude whether the rise in markups is associated with a rise in market power.

We proceed in two steps. First, we relate markups to recorded profits at the firm level; and contrast the observed markups to counterfactual markups generated by a zero-profit condition. Second, we consider aggregate profits and ask whether these are consistent with our estimates of firm-specific markups and recorded fixed costs.

4.1 Markups and profits at the firm level

To calculate profits, we use the markup measure and properly account for all costs, including the overhead (or fixed) costs and the expenditure on capital. We then interpret this profit rate as a measure of market power.

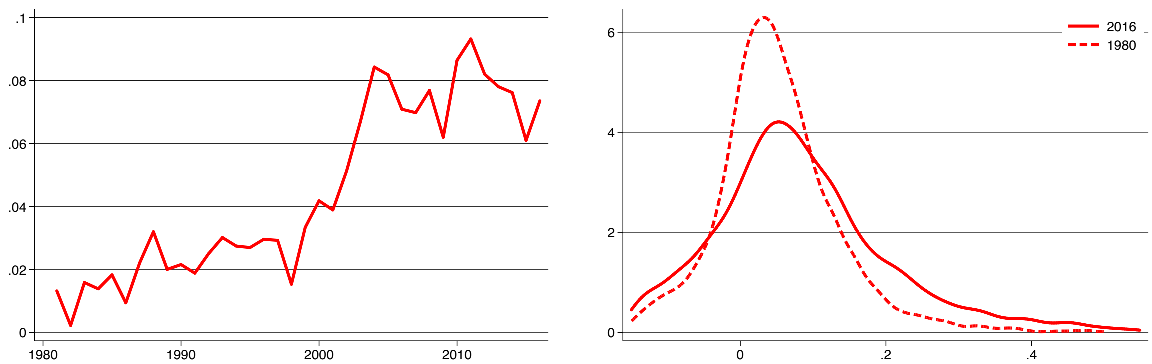
Let $\Pi_i = S_{it} - P_t^V V_{it} - r_t K_{it} - P_t^X X_{it}$ denote net profits, then the net profit rate $\pi_{it} = \frac{\Pi_{it}}{S_{it}}$ can be written as:

$$\pi_{it} = 1 - \frac{\theta_{st}}{\mu_{it}} - \frac{r_t K_{it}}{S_{it}} - \frac{P_t^X X_{it}}{S_{it}}, \tag{13}$$

where we have substituted the expenditure on variable inputs as a share of sales with the output elasticity over the markup, from equation (7). This measure of the profit share is different

[35]The series are obtained by taking the ratio between total overhead (capital) and aggregate total cost, alternatively this can be interpreted as weighing each firm with its share of total cost in aggregate total cost

from the accounting profits because it uses a measure of capital that is obtained from the balance sheet, not the income statement. With 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 do not.



(a) Average Profit Rate (revenue weighted) (b) Kernel Density Profit Rate (unweighted)

Figure 8: Average Profit Rate and Profit Rate Distribution.

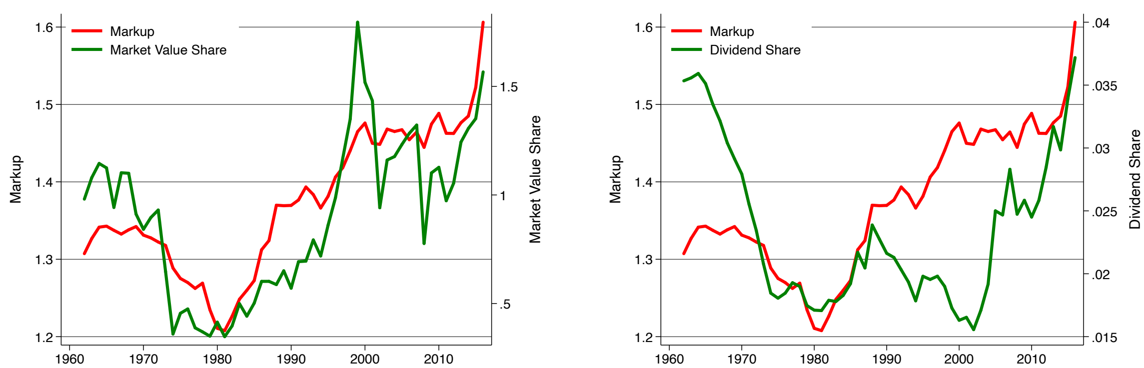
Figure 8a plots the average revenue weighted profit rate for the data in our sample. We find that profits have gone up by about 7 percentage points between 1980 and 2016.[36] Underlying the rise in profits is the increase in the upper tail of the profit distribution. In Figure 8b we plot the kernel density 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. Consistent with the results on markups, the average profit rate increase is driven in part by the reallocation of economic activity towards high profit, dominant firms.

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 expenditure on capital. Because the expenditure on 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 9.1a 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 9.1b), 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.

[36]Our measure of profits was also high in the mid 1970s (see Figure 8.1 in the Online Appendix), but that is entirely driven by the drop in capital expenditure during a period of high inflation. Once we consider gross profits, without subtracting the expenditure on capital, there is no such spike in the profit rate in the 1970s.

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 9: Market Value and Dividends.

Figure 9a 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 inflation[37] 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.[38]

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 9a, 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 9b).[39]

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.[40] 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.

To complete this Section, we investigate the relation between profits, markups and overhead costs (SG&A). In Figure 10a 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 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 $\mu^\star$, that corresponds to zero profits. We obtain that markup from setting profits π_{it} to zero in equation

[37]The 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.

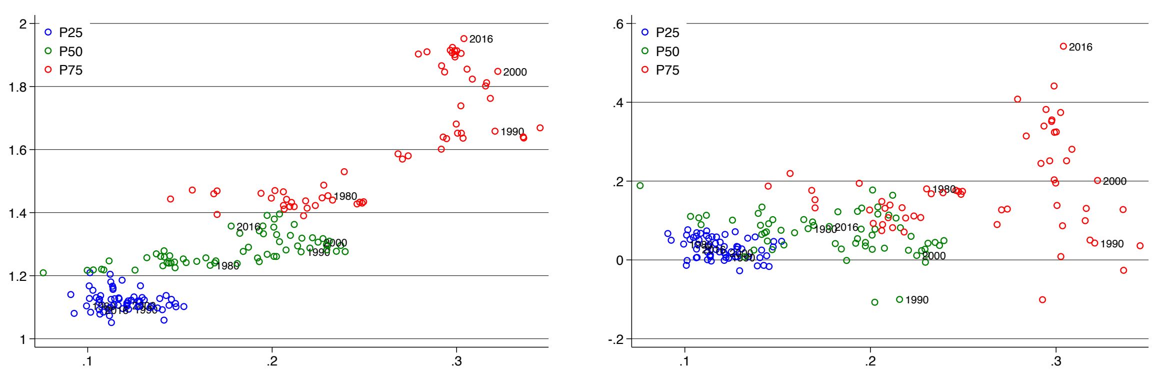
[38]Interpreting the market valuation of a firm as the discounted stream of profits obviously imposes a set of assumptions. The most important one is the fact that the discount rate has remained constant over the period. We know that the risk free rate has decreased, especially since the 1990s. While 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)).

[39]Note 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.

[40]In Table 19.1 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(Market Value)		
ln(Markup)	0.71	0.64	0.56	0.17	0.71	0.65	0.58	0.27
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
ln(Sales)					0.81	0.81	0.83	0.68
					(0.00)	(0.00)	(0.00)	(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(Dividends)		
ln(Markup)	1.05	0.97	0.80	0.26	1.03	0.93	0.78	0.26
	(0.04)	(0.03)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
ln(Sales)					0.94	0.92	0.93	0.76
					(0.01)	(0.01)	(0.01)	(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).



(a) Markups μ_{it} by SG&A Share.

(b) Excess Markup $\mu_{it} - \mu^{\star}_{it}$ by SG&A Share.

Figure 10: Markup, Excess Markup and SG&A Share (Markup PF2).

(15) and solving for μ:

$$\mu^{\star}_{it} = \frac{\theta_{st}}{1 - \frac{r_t K_{it}}{S_{it}} - \frac{P^X_t X_{it}}{S_{it}}}. \tag{14}$$

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 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 ($S^{\star}_{it}$) 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 (14) where we use S_{it} instead of $S^{\star}_{it}$ is weakly higher than the true zero profit markup.

In Figure 10b we also plot $\mu_{it} - \mu^{\star}_{it}$ for different percentiles in the markup distribution. Because $\mu^{\star}_{it}$ is the upper bound of the zero profit markup, the gap between the actual markup and $\mu^{\star}_{it}$ 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).

	Markup (log)			Profit Rate (log)	
	(1)	(2)	(3)	(4)	(5)
SG&A (log)	0.56			0.15	
	(0.01)			(0.03)	
R&D Exp. (log)		0.16			0.10
		(0.01)			(0.01)
Advertising Exp. (log)		0.05			0.03
		(0.00)			(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.

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 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).

In sum, at the firm level, we find consistent evidence that profits and the market valuation of firms have gone up together with markups. Markups are not higher only to compensate for higher fixed costs, they are also higher because *firms exert market power.*

4.2 Aggregate Profits and Markups

Even though markups and profit rates are different concepts – most notably because of the inclusion in profits of total costs, including overhead costs – they are related. In particular, there is an identity that links profit rates and markups and that holds for any technology $C(Q)$, as has been pointed out by Syverson (2019) and De Loecker and Eeckhout (2018b):

$$\pi_{it} = \frac{P_{it}Q_{it} - C(Q_{it})}{P_{it}Q_{it}} = 1 - \frac{AC_{it}}{\mu_{it}MC_{it}}, \tag{15}$$

where $\frac{AC_{it}}{MC_{it}}$ is the ratio of average cost to marginal cost and since $AC_{it} = \frac{C(Q_{it})}{Q_{it}}$ and $\mu_{it} = \frac{P_{it}}{MC_{it}}$.

Now there is a puzzle. The aggregate markup of 1.61 that we calculate in 2016 cannot be reconciled with the profit rate of 8%. In particular, Basu (2019) has pointed out that something must be wrong with our markup measure, because the implied profit is too high. If we plug in the aggregate markup in 2016 and assume that the ratio of average cost to marginal cost is equal to 1, then the implied profit rate is 38%.[41]

There are two problems with this argument. The first is that in this thought experiment, we have assumed that the average to marginal cost ratio is constant and equal to one. We know from Figure 7 that the fixed cost is sizable and has gone up. Therefore the average to marginal cost ratio is neither constant nor equal to one.

The second problem with this argument is that it erroneously relies on a representative firm framework. Equation (15) strictly holds at the firm level. In the aggregate, this translates into:

$$\pi_t = \sum_i m_{it}\pi_{it} = 1 - \sum_i m_{it}\frac{AC_{it}}{\mu_{it}MC_{it}} \neq 1 - \frac{AC_t}{\mu_t MC_t} \tag{16}$$

[41]Basu (2019) performs a slightly different exercise. He rewrites equation (15) as $\mu_{it} = \frac{1}{1-\pi_{it}}\frac{AC_{it}}{MC_{it}}$ and takes the ratio between of this expression evaluated at any two years, say 2016 and 1980: $\frac{\mu_{2016}}{\mu_1 980} = \frac{1-\pi_{1980}}{1-\pi_{2016}}$. He lets $\pi_{1980} = 0$ (which is close to the profit rate of 1% we find), then $\frac{1.61}{1.21} = \frac{1}{1-\pi_{2016}} \Rightarrow \pi_{2016} = 25\%$. This leads to a profit rate of 25%. This is completely unrealistic, especially since in our sample we find a profit rate of around 8% in 2016.

where $\mu_t = \sum_i m_{it}\mu_{it}, AC_t = \sum_i m_{it}AC_{it}, MC_t = \sum_i m_{it}MC_{it}$. Therefore, the premise of a representative firm framework is counterfactual.

Once we correct for these counterfactual assumptions – that the average cost to marginal cost ratio has increased and that we properly aggregate without assuming a representative agent framework – the implied average profit rate of 8% in 2016 and the markup of 1.61 are indeed consistent.

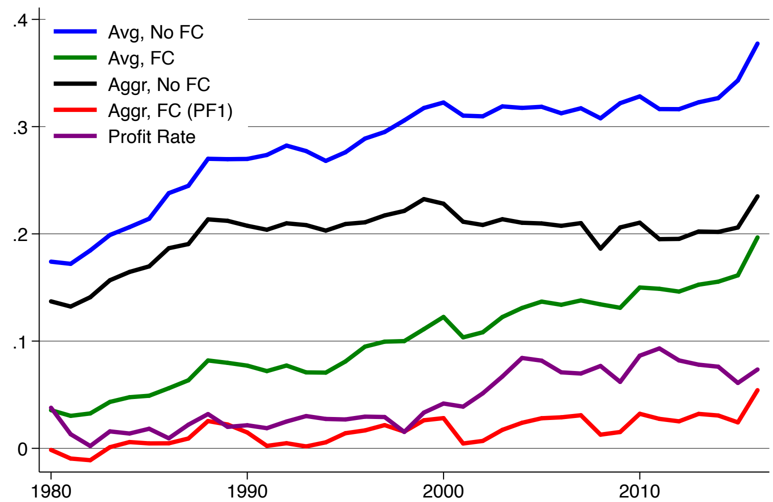


Figure 11: Decomposition of Equation (15) due to Overhead Costs and Aggregation.

In Figure 11, we decompose equation (15). We also report the actual values in Table 4.[42] When we assume *both* a representative firm and a constant average to marginal cost ratio equal to one (no Fixed Cost), we see profits rise from 18% in 1980 to 38% in 2016 (blue). When we adjust for the observed average to marginal cost ratio but keep the representative firm assumption (green), the profits drop by more than half over the entire period. When we adjust for proper aggregation (drop the representative firm) and keep a constant average to marginal cost ratio (black), profits drop by about one third. Note that the gap is larger towards 2016 than in 1980, which is consistent with the fact that the distribution of firm sizes and markups has become more more dispersed, resulting in a bigger gap between the aggregate and the average (due to Jensen's inequality). Finally, when we adjust for *both* proper aggregation and the observed average to marginal cost ratio (purple), profits are close to the observed profits in the data (red).

Overall, the relation that predicts average profit rates as a function of markups fits the data once we properly account for returns to scale (fixed costs) and once we we properly aggregate. This indicates that our measure of markups does not predict an outlandish profit rate. What it does confirm is that markups and profit rates are different objects and that we should be careful comparing them. Too often, they are used interchangeably.

Finally, Traina (2018) proposes a different measure of market power that includes both COGS and SG&A. His measure is therefore closely related to the profit rate. Denote by τ:

[42] We use the technological specification of our benchmark model with Cobb-Douglas production and a fixed cost.

		Profit Rate	
		1980	2016
Use Average	No Fixed Cost	0.17	0.38
Use Average	Fixed Cost	0.04	0.20
Aggregation	No Fixed Cost	0.14	0.24
Aggregation	Fixed Cost	0.00	0.05

Table 4: Decomposition of the average profit rate from equation (15).

$$\tau_{it} = \theta^{V+X} \frac{S_{it}}{P_{it}^V V_{it} + P_{it}^X X_{it}}, \tag{17}$$

where $p^V V$ is the expenditure on the variable input, $p^X X$ is the expenditure on overhead as measured by SG&A and where $\theta^{V+X} \approx 0.95$ (though he estimates a separate elasticity for each sector).

This ratio is directly related to the operation profit rate, the definition of which is

$$\pi_{it}^{OPX} = \frac{S_{it} - P_{it}^V V_{it} - P_{it}^X X_{it}}{S_{it}} = 1 - \frac{P_{it}^V V_{it} + P_{it}^X X_{it}}{S_{it}}. \tag{18}$$

We can therefore write the measure τ_{it} as

$$\tau_{it} = \theta^{V+X} \frac{1}{1 - \pi_{it}^{OPX}}. \tag{19}$$

Given this identity, this measure is closely related to the aggregate operating profit rate $\pi_t^{OPX} = \sum_i m_{it} \pi_{it}^{OPX}$ (see Figure F.1 in the Appendix F). We find an increase in the operating profit rate between 1980 and 2016 of about 7-8 percentage points, and for the measure τ, which we interpret as an alternative measure of the profit rate, we see an increase of about 10 points, from 1.08 to 1.18.

In sum, aggregate markups and profitability are both increasing. Therefore the rise in markups is not exclusively due to the rise in overhead costs. This is evidence of the rise in market power.

5 The Macroeconomic Implications

The focus of our analysis so far has been on documenting in detail the time series and cross sectional evolution of markups and profitability. 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).[43] 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 struggled to understand the mechanism behind the decline in the labor share. One obvious hypothesis, ex ante, would be a within-firm substitution of labor for capital. This hypothesis is explored most prominently in Karabarbounis and Neiman (2014), which argues that a secular decrease in the relative price of investments goods led firms to substitute away from labor towards capital, and can explain half of the decline in labor's share of income. The basic problem with this mechanism is that it rests, crucially, on a high elasticity of substitution between capital and labor (higher than 1). While Karabarbounis and Neiman (2014) claim that this elasticity is 1.25, the overwhelming majority of several decades of empirical studies (Antras (2004), among many others find that this elasticity is much lower than 1. The combination of a low elasticity of substitution between capital and labor, with the fact of a declining labor share of income, has been especially puzzling.

Koh, Santaeulalia-Llopis, and Zheng (2017) offer yet another 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 property products and this leads to a lower expenditure on labor.[44] 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 13b), 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 8a). 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 (7) 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

$$\frac{w_t L_{it}}{P_t Q_{it}} = \frac{\theta^L_{it}}{\mu_{it}}. \tag{20}$$

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.

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

[43]There 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.

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

underreported.[45] Because of selection in the sample of those firms that do report total compensation, we need to be cautious interpreting the aggregate labor share outcomes.

	Labor Share (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Markup (log)	-0.24	-0.23	-0.20	-0.24	-0.68	- 0.73
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Cost Share (log)					0.91	0.96
					(0.01)	(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 5: 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 (20) at the firm level. In Table 5 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 keep in mind that there is no such a thing as a representative firm in this context. The rise of average markups is distributed unequally, and increasingly so. Most importantly, since two thirds of the rise in market power is due to reallocation of economic activity towards towards high markup firms, the effect of markups on the labor share in the aggregate is predominantly driven by a few large firms with high markups and a low labor share. 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 (20).

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 ex-

[45] Less than 10% of the year-firm observations include XLR.

penses, 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.[46] 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.[47]

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

$$\frac{P^V V}{PQ} + \frac{rK}{PQ} = 1 - \frac{P^X X}{PQ} - \frac{\Pi}{PQ}, \tag{21}$$

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 expenditure on capital rK as a share of output will be decreasing over time. In fact, if capital were fully flexible, it would adjust according to the equivalent of first order condition (7) $\frac{rK}{PQ} = \frac{\theta^K}{\mu}$ which relates the capital share to the inverse of the markup.

In Figure 14.1b 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.[48]

As with the labor share, we can also investigate the firm-level relation between the capital share and markups. In Table 6 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 5 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.

[46]This 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 1.3a we have inferred that capital is not equally flexible as the variable inputs.

[47]A 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.

[48]A 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.

	Capital Share (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Markup (log)	0.03	0.03	-0.02	-0.14	-0.90	-0.86
	(0.02)	(0.02)	(0.01)	(0.02)	(0.00)	(0.00)
Cost Share (log)					1.13	1.11
					(0.00)	(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 6: Regressions: log (Capital Share) on log(Markup). F.E. = Fixed Effects; 4-digit Industries; Standard Errors (in parentheses) clustered at the firm level.

THE SECULAR DECLINE IN LOW SKILL WAGES AND LABOR FORCE PARTICIPATION. An increase in markups implies a decrease 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. We find that market power indeed has an effect on equilibrium wages, and that quantitatively, that effect is large. Our quantitative model predicts that real wages as a share of GDP drop by over 26%, consistent with what we see in the data.

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.[49] Crucial for our finding is that the higher the degree of market power

[49] Most 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.

by firms, the lower the passthrough.

Now consider an environment were 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.[50] 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 employment to non-employment transition.[51] 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.[52] 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)).[53]

6 Discussion and Robustness

We now further discuss the features of our model and report a number of robustness exercises.

6.1 Cost shares

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 $\alpha_{it}^V = \frac{p_t^V V_t}{p_I^V V_{it} + r_t K_{it}}$. Within an industry, we use the median of the distribution as the measure for the output elasticity: $\theta_{st} = \text{median}_{i \in s}\{\alpha_{it}^V\}$.

[50]Independent 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.

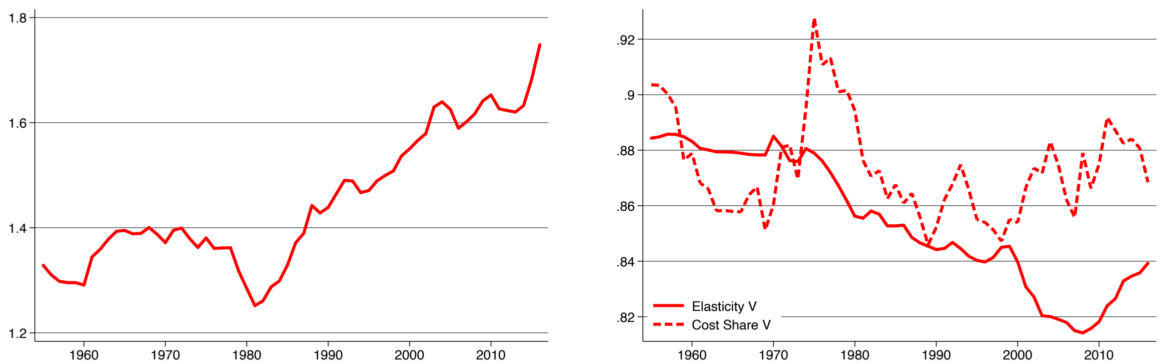
[51]There are several potential alternative explanations for the decline in job flows: demographic change (aging workforce, Fallick, Fleischman, and Pingle (2010) and Engbom (2017)), a more skilled workforce, lower population growth, decreased labor supply (Karahan, Pugsley, and Şahin (2016)), technological 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.

[52]See also Kaplan and Schulhofer-Wohl (2012), amongst others.

[53]Recent 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.

Figure **??** 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 (7), 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 12b we plot the average cost share of the factors of production V and K as well as the average output elasticity estimated from the benchmark technology. There is some volatility in the cost shares, but they are in line with the estimated output elasticity. 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 2b, confirming again that the evolution of the share-weighted average markup is mainly driven by the ratio of sales to expenditure on variable inputs and not by changes in the output elasticity.



(a) Aggregate Markup (CS)

(b) Output elasticity and Cost Shares: V

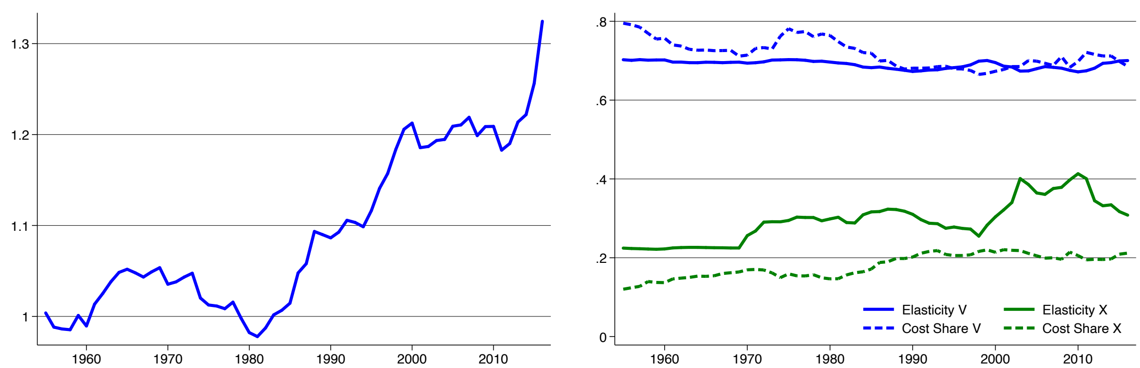
Figure 12: Cost-Share Based Aggregate Markups and Technology. Panel (a) reports the aggregate Markups using Cost Shares (Median cost shares for each year and sector). Panel (b) compares the sector-weighted cost share to the estimated output elasticities; weights are sectoral total sales.

6.2 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. 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.



(a) Markup with θ_{st} from PF2

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

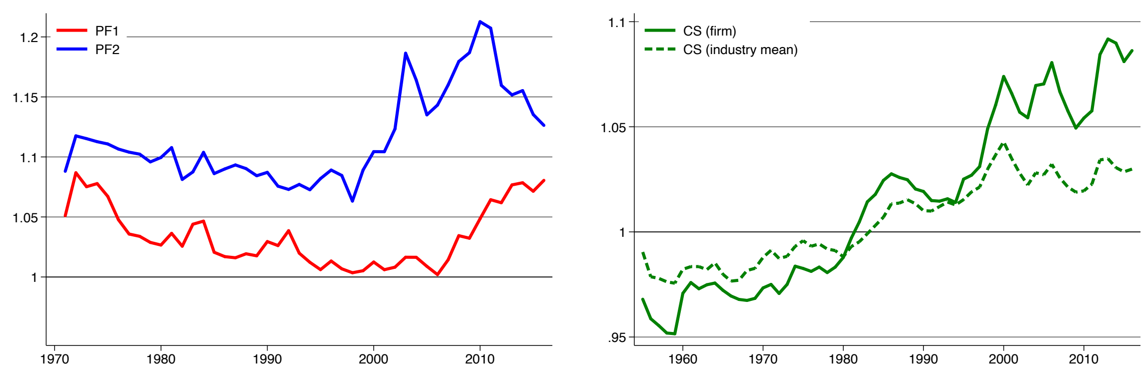
Figure 13: Average Markups, Elasticities and Cost Shares for Production Function with Overhead as a factor. Output elasticities from estimated PF2 and from CS2: time-varying, sector-specific (2 digit) output elasticity θ_{st} (revenue weighted average).

Figure 13b plots the cost share of variable factors in the *total cost* (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. The estimated output elasticities confirm this pattern, although importantly they do not necessarily have to sum to one (including rK of course).

In Figure 13a, 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 around one, the average markup increases by about 30 points in 2016. The increase for this technology is 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 13b). 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 12b). For the production technology with overhead as a factor of production, θ^V is slightly decreasing (Figure 13b). Therefore the estimated markup shows a more moderate increase (30 points) than under the conventional production technology (40 points).

6.3 Returns to Scale

With the estimated technologies, 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 14a 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; revenue weighted

(b) Estimated RTS of Cost Shares: firm CS and sector average CS

Figure 14: Returns to Scale

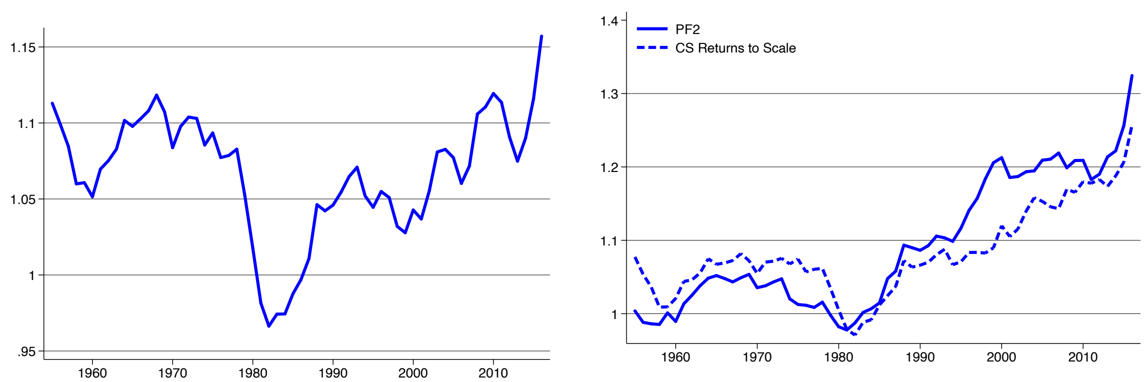
An alternative way to measure returns to scale is with a method first used in Syverson (2004). While using cost shares implicitly assumes that the technology is constant returns, Syverson (2004) adjusts the technology based on cost shares and derives the returns to scale. He assumes the following functional form for the technology based on cost shares but without constant returns:

$$q = \gamma \left[\alpha_V v + \alpha_K k + \alpha_X x\right] + \omega, \tag{22}$$

with all variables in logs, where $\alpha_V = \frac{p^V V}{P^V V + rK + P^X X}$ 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 14b 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.[54] Both graphs reveal that also with this method, returns to scale have increased throughout the sample. There were DRTS before 1980 and since 1980 returns to scale have been increasing, up to 1.05 at the end of the sample.[55]

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 15a), whereas under the elasticity estimated from the production function the increase since 1980 is 30 percentage points (Figure 13a). 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.[56] With θ^V decreasing, from equation (7), 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.



(a) Markup with θ_{st} from CS2

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

Figure 15: 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} (revenue weighted average).

The evolution of returns to scale helps us understand the difference between Figures 13a and 15a. 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 (22) which is equal to $\gamma\alpha_v$, we obtain an average markup (see Figure 15b) that is very similar to the one using the elasticity estimated with the

[54] An important caveat here is that we estimate this technology by means of a simple regression, without accounting for endogeneity, because the cost share approach implicitly assumes all inputs adjust within the time period.

[55] In 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.

[56] In 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.

production function (PF2). The increase in γ in Figure 14b implies that the elasticity α_V used in Figure 15a is multiplied by γ.

Finally, in Appendix 13 we also analyze the returns to scale using the data from the Censuses.

6.4 Input Weights and Joint Distributions

We have shown in Section 3 that in order to calculate aggregate markups, the choice of the weighing measure matters. Because we are interested in the entire joint distribution of markups and firm characteristics (revenue, costs, inputs,...), having information on as many moments as possible provides more detailed insights in the evolution of markups. [57]

Output Elasticity		Revenue Weight	Input Weight		
			1 Input		All Inputs
		$m_{it}=\frac{R_{it}}{R_t}$	$m_{it}=\frac{COGS_{it}}{COGS_t}$	$m_{it}=\frac{L_{it}}{L_t}$	$m_{it}=\frac{TC_{it}}{TC_t}$
Economy-wide	θ	$\theta\sum_i m_{it}\frac{R_{it}}{COGS_{it}}$	$\theta\frac{R_t}{COGS_t}$	$\theta\sum_i m_{it}\frac{R_{it}}{COGS_{it}}$	$\theta\sum_i m_{it}\frac{R_{it}}{COGS_{it}}$
Sector-specific	θ_{st}	$\sum_i \theta_{st} m_{it}\frac{R_{it}}{COGS_{it}}$	$\sum_s \theta_{st}\frac{R_{st}}{COGS_{st}}$	$\sum_i \theta_{st} m_{it}\frac{R_{it}}{COGS_{it}}$	$\sum_i \theta_{st} m_{it}\frac{R_{it}}{COGS_{it}}$

Note: $TC_{it}=COGS_{it}+r_tK_{it}+SGA_{it}$.

Table 7: Aggregate Markups: Variation by Technology (θ) and Weighing (m_{it})

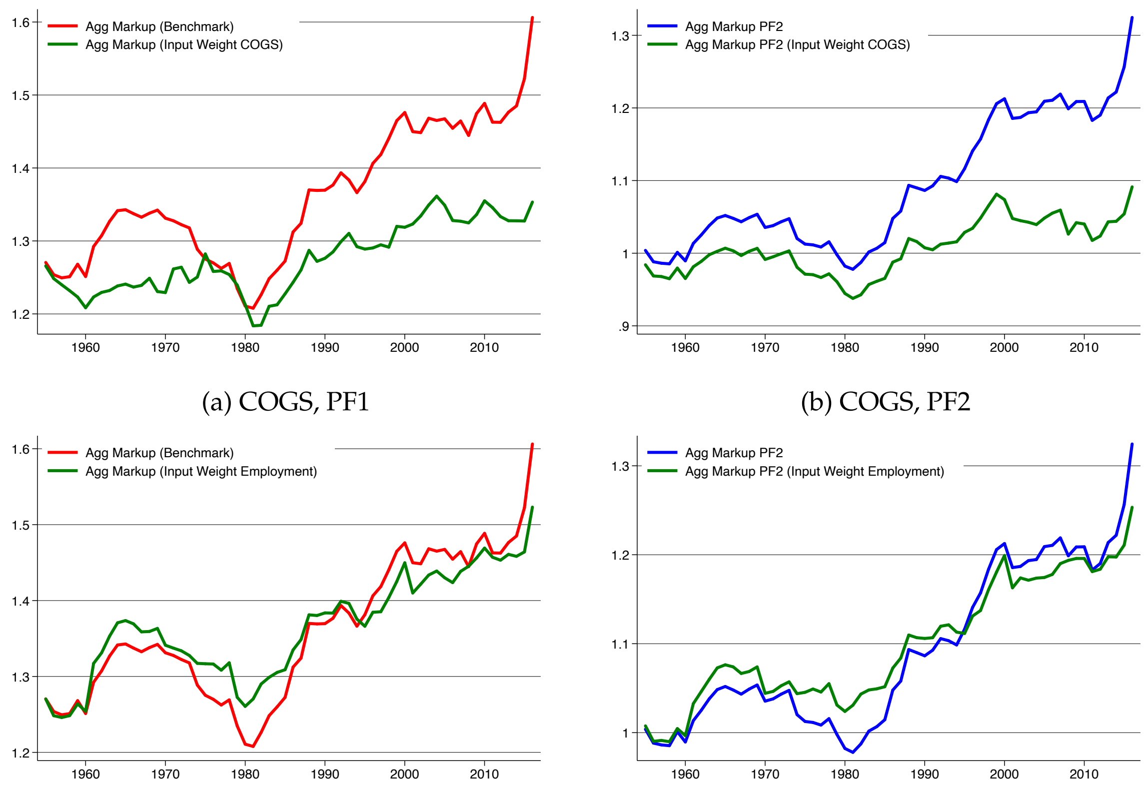
Here we report the the average markup using different weights. To get a better idea of the definition of each of the weights, we report them in Table 7.

For our data, above in Section 3 we have already reported the Revenue weighted aggregate markups in Figures 1 and 2a, as well as the input weighted markup with total cost as the input weight (Figure 2a).

Here we plot the aggregate markups with 1 input weight and for a sector-specific output elasticity (the results are very similar with economy-wide, constant output elasticities). We present two versions of the production technology (PF1 and PF2). In row 1 (Figures 16a and 16b) we plot the aggregate markup measure proposed by Edmond, Midrigan, and Xu (2015) and Grassi (2017) on the grounds of its representativeness of welfare measures in a setting with CES preferences in models such as monopolistic competition and Atkeson and Burstein (2008). Here we do not take a stance on welfare, but we do in De Loecker, Eeckhout, and Mongey (2018).

We find a rise in the COGS weighted aggregate markups that is only about half of the rise in revenue weighted markups and substantially lower than the total cost weighted aggregate markup (Figures 16a and 16b). We can see from Table 7 why. With one input equal to the variable input used for calculating the markup, the aggregate markup is simply the ratio of

[57] In the Online Appendix 6 we report the contour plots of the joint distributions of markups and revenue and inputs. This gives us a view of the 'raw' data.



(a) COGS, PF1

(b) COGS, PF2

(c) Employment, PF1

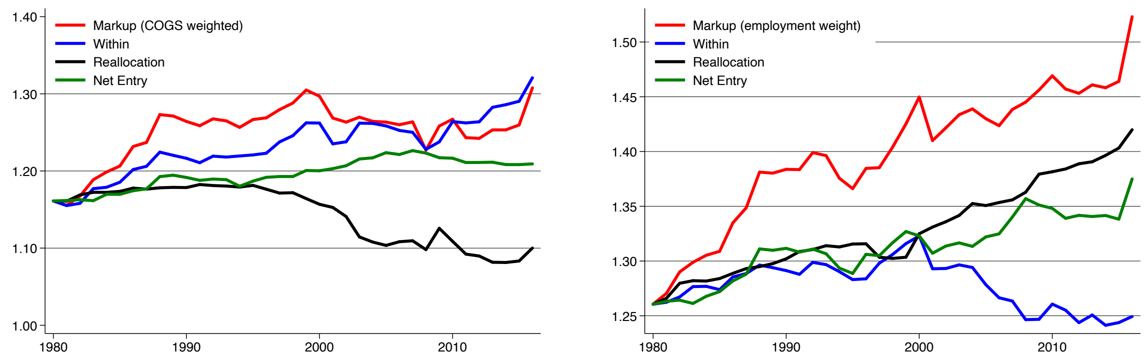
(d) Employment, PF2

Figure 16: Markups with input weights: COGS and employment for the benchmark technology and PF2.

revenue over COGS (multiplied by the output elasticity). As a result, the markup is a function of aggregates only. This implies that the aggregate is not sensitive to within sector (or within economy) variation. In fact, the aggregate markup measure is identical to that obtained by Hall (1988) and that we report in Figure 5.

Rather using expenditure, we can also use quantities. While we do not have quantities of COGS (because we do not know the unit prices), we have quantities of labor, as measured by the number of employees. In Figures 16c and 16d for the two technologies we use employment weights. Unlike aggregate markups with weights from expenditure shares of inputs, those with employment weights track the benchmark aggregates. This seems to indicate that at least for employment, the quantities do not adjust as much as the input prices (in this case wages) do. When we quantify the economy in De Loecker, Eeckhout, and Mongey (2018) we find indeed large general equilibrium effects on wages and smaller effects on labor supply.

Finally, we also do the same decomposition exercise on our input weighted measures in Figure 17. Not surprisingly, for the COGS weighted aggregate, there is no role for reallocation (the reallocation term is even negative). Because by construction, the COGS weighted measure is one based on averages only, there is no impact of within industry reallocation and we have shown earlier that the reallocation occurs predominantly within industry. What the decomposition shows is that virtually all of the change in the COGS weighted markups is driven by the within term, the rise in markups themselves.



(a) COGS

(b) Employment

Figure 17: Decomposition of input weighted average markups

Instead, for the employment weighted measure, the picture looks much more similar to that of the benchmark revenue weighted aggregate markup. About two thirds of the rise in the employment weighted average markup can be attributed to reallocation.

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. The different weights give us further insights into the joint distribution of markups, revenue and all inputs.

6.5 Comparison of our Estimates with those in the Literature

In Section 7 of the Online Appendix, we compare our estimates with those obtained in the literature using the Demand Approach (Berry, Levinsohn, and Pakes (1995)) for seven industries for which there is data: Beer, breakfast cereal, steel, Autos, airlines, department stores, electronic shopping and mail order. For the companies in our data set that fall in the same industry classification, we construct an average markup and plot them jointly with the markups obtained in the literature (Figure 7.1).

Whenever there is overlap, the patterns of markups obtained with the demand approach closely follow those obtained with our cost-based approach. This is remarkable because not only are the methods different, they rely on different data. This is testament to the fact that the estimates we obtain are robust across different methods and data sources.

We perform further robustness exercises in the Appendix and in the Online Appendix.

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 study the evolution of market power. For each firm, we estimate both markups and profitability, and we document the properties of their distribution. We find that from 1980 onwards, markups have risen from 21% to nearly 61% in 2014, an increase of 40 points. For the same period, average profit rates have increased from 1 percent of sales to 8 percent.

We attribute this rise in market power nearly exclusively to the increase for 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 revenue 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. We find that reallocation accounts for two thirds of the rise.

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. 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 Estimating Output Elasticities

A crucial component to measure markups is to obtain an estimate of the output elasticity of a variable input of production (θ^V). While the production approach to markup estimation, described in De Loecker and Warzynski (2012) 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 identify and estimate the elasticity in the data. We use two distinct methods to estimate the output elasticity of the production function. First, we estimate a parametric production function for each sector-year using recent techniques that take into account the well-known potential biases discussed in the literature. Second, we non-parametrically estimate the output elasticity using (constructed) cost shares. Both approaches have their advantages and disadvantages, which we discuss below.

A.1 Production Function Estimation

We follow standard practice and rely on a panel of firms, for which we estimate production functions for each (2-digit) industry. For the benchmark specification, we consider a sector-year-specific Cobb-Douglas production function, with a variable input bundle and capital as inputs. For *each* industry s we consider the production function (PF1 in the main text):

$$y_{it} = \theta_t^V v_{it} + \theta_t^K k_{it} + \omega_{it} + \epsilon_{it}, \tag{A.1}$$

where lower cases denote logs and $\omega_{it} = \ln \Omega_{it}$, where y_{it} is a measure of realized firm's output, and ϵ_{it} captures measurement error in output – i.e., $y_{it} = \ln(Q_{it} \exp(\epsilon_{it}))$.

We depart from the standard specification in 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).[58] 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 (7), 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.[59] Allowing the production function coefficients to vary over time is also a parsimonious way to account for factor-biased technological change.

When we consider the production function with overhead (PF2 in the main text), the specification is given by:

$$y_{it} = \theta_t^V v_{it} + \theta_t^K k_{it} + \omega_{it} + \theta^X x_{it} + \epsilon_{it}, \tag{A.2}$$

with $x = ln(X)$, and X captures (deflated) SG&A.

[58] In 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.

[59] Because of data scarcity, we use a five year rolling window around the year where we estimate the technology. In Section 6 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.

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: dealing with *unobserved productivity* shocks (ω_{it}); and extracting *units* of output and inputs from revenue and expenditure data (i.e. the omitted price variable bias) . 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.

We follow the literature and control for the simultaneity and selection bias, inherently present in the estimation of equation (A.1), and rely on a control function approach, paired with a law of motion for productivity, to estimate the output elasticity of the variable input.[60] 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 react to contemporaneous shocks to productivity, but it is correlated with the persistent productivity term. This requires us to restrict the production function to a particular class to guarantee that the coefficients of interest – which determines the output elasticity – are identified.

A.1.1 Control function

We build on the insight from Olley and Pakes (1996) that (unobserved) productivity ω_{it} can be expressed as an (unknown) function of the firm's state variables and observables. This is obtained by considering input (or investment) demand, and inverting out for productivity to yield:

$$\omega_{it} = h_t(d_{it}, k_{it}, z_{it}), \tag{A.3}$$

where d_{it} is the control variable. We consider two cases: a variable input in production (in our case COGS, v), and investment (i) and Ackerberg, Benkard, Berry, and Pakes (2007) provide an excellent treatment of the two types of control variables, z_{it} captures output and input market factors that generate variation in factor demand (for input d) across firms, conditional on the level of productivity and capital. The latter is critical to allow for imperfectly competitive product markets when estimating production functions. Standard approaches in the literature on production function estimation are restricted to either perfect competition or models of common markups (monopolistic competition paired with CES demand). Instead, we rely on De Loecker and Warzynski (2012) andDe Loecker, Goldberg, Khandelwal, and Pavcnik (2016) to allow for imperfect competition in product markets, and thus markup heterogeneity across firms. In practice this amounts to allowing for input demand shifters that move around the optimal amount of a variable input, conditional on a firm's productivity and capital stock.

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.

In the case of a static control, $d_{it} = v_{it}$ this is given by:

$$y_{it} = \phi_t(v_{it}, k_{it}, z_{it}) + \epsilon_{it}. \tag{A.4}$$

[60] We 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.

The output elasticity is obtained by constructing moments on the productivity shock, which is obtained by considering a productivity process given by $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$. It gives rise to the following moment condition to obtain the industry-year-specific output elasticity:

$$\mathbb{E}\left(\xi_{it}(\theta_t)\begin{bmatrix} v_{it-1} \\ k_{it} \end{bmatrix}\right) = 0, \tag{A.5}$$

where $\xi_{it}(\theta_t)$ is obtained by projecting productivity $\omega_{it}(\theta_t)$ on its lag $\omega_{it-1}(\theta_t)$, with $\theta_t = \{\theta_t^V, \theta_t^K\}$, where productivity is in turn obtained from $\phi_{it} - \theta_t^V v_{it} - \theta_t^K k_{it}$, 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 that lagged variable input use is correlated with current variable input use, through serially correlated input and output market conditions, captured in z_{it}. In the case of PF2, an additional moment identifies the output elasticity of the overhead (SG&A) input, $\mathbb{E}(\xi_{it}(\theta_t)x_{it}) = 0$.

In the case of the OP approach, we can in fact identify and estimate the output elasticity using a simple non-linear regression:

$$y_{it} = \theta_t^V v_{it} + \phi_t(i_{it}, k_{it}, z_{it}) + \epsilon_{it}, \tag{A.6}$$

and rely on the identification arguments made in Ackerberg, Caves, and Frazer (2015) – i.e., if the variable input bundle v (COGS) is non-dynamic and chosen at t, after the investment decision, made at $t-1$, while allowing for productivity shocks to hit the firm in between these two periods. This approach has the advantage that it is simple to implement, and does not require to consider the subsequent second stage. Compared to the static control, discussed above, the investment policy function needs to be increasing in productivity (conditional on capital and variables captured by z), and the specification adopted here limits the scope of strategic interaction among firms.

We consider both controls (COGS and investment) and we find very similar results for the estimated output elasticities. Below we plot the two series, using the static and dynamic control variable, and we aggregate the industry-year specific output elasticities of COGS using industry sales.

A.2 Units

As pointed out in De Loecker and Goldberg (2014) standard production data, whether it is Compustat or census data, records revenue and expenditures, rather than physical production and input use (with the exception of a few manufacturing industries). In the presence of product differentiation (be it through physical attributes or location) an additional source of endogeneity presents itself, through unobserved output and input prices. This has been the topic of recent research, for a recent treatment see De Loecker, Goldberg, Khandelwal, and Pavcnik (2016). A first observation is that the error term, ϵ_{it}, will in general contain output and input prices (scaled by the relevant technology parameters). De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) show that the correlation of input expenditures to this error yield biased estimates of the output elasticity. However, in their setting physical output quantities are observed, and the unobserved input prices, reflecting differentiation, are the only source

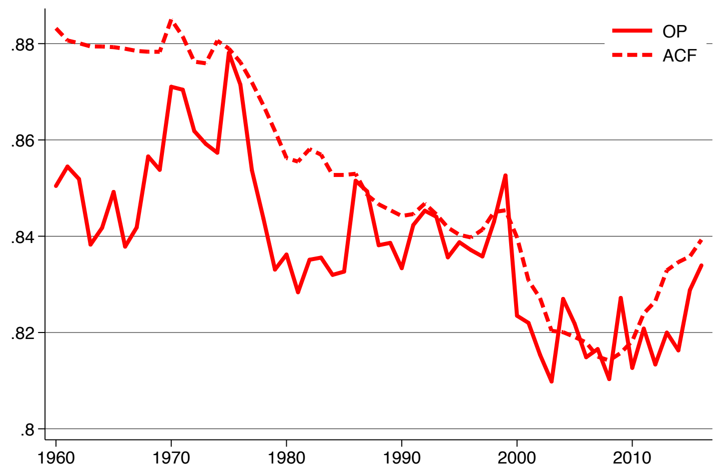


Figure A.1: Output Elasticities under Alternative Data Generating Processes: Olley and Pakes (1996) and Ackerberg, Caves, and Frazer (2015)

of the price error. We do not observe output price variation, and we are therefore left with the following structural error term:

$$\omega_{it} + p_{it} - \theta_t^V p_{it}^v - \theta_t^K p_t^K, \tag{A.7}$$

where we let the user cost of capital be industry-time specific, but input prices potential vary across firms reflecting variation in quality, location and other exogenous factors. We follow De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and let the wedge between the output and input price (scaled by the output elasticity) be a function of the demand shifters and productivity difference.[61] In the case of OP, the inclusion of the variable z should therefore capture the relevant output and input market forces that generate differences in output and input price. Of course, productivity differences that influence the wedge between output and input prices are automatically captured by the inclusion of the control function. Note that not observing output prices has, the perhaps unexpected, benefit that output price variation absorbs input price variation, therefore eliminating part of the variation in the error term. In the extreme case we are left with just the productivity unobservable, and this puts us back in the standard framework introduced above.

Under the alternative DGP, where the static control (COGS) is used, and where the output elasticity is identified in the second stage, we follow De Loecker, Goldberg, Khandelwal, and Pavcnik (2016). The main difference lies in the fact that we cannot rely on observed output prices, and we therefore have to rely on constructed measures of market share (at various levels of aggregation) to eliminate the variation in the price error wedge.

In practice, we consider market share, measured at various level of aggregation (2,3, and 4 digit), to take into account additional variation in output and input markets. As discussed

[61]See De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a micro-foundation for this, and application. An alternative is De Loecker (2011) and specify a specific demand system. This, however, limits the scope of markup heterogeneity, across firms and time, and we are precisely interested in describing markups in a flexible fashion.

in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) this is an exact control when output prices, conditional on productivity, reflect input price variation, and when demand is of the (nested) logit form. We have subjected our analysis to a host of different specifications of the production function (such as the translog production function), and we find very similar results for the estimated output elasticities. As discussed in the main text, the main findings on aggregate markups, are furthermore not sensitive to the use of a common time-invariant calibrated output elasticity of 0.85. We also considered an alternative specification, including SG&A as a factor of production, and document a comparable rise in aggregate markups.

Finally, there are a host of possible measurement error and endogeneity concerns with any single specification one could consider. We do not attempt to provide the one final set of output elasticities for all sectors of the US economy, using the Compustat data. Rather we consider a variety of specifications, and show that the main facts we are interested (under the maintained year-sector specific Cobb-Douglas production function) are not sensitive to these. The aggregate markup can be expressed in terms of the potential bias $\psi_{st} = \hat{\theta}_{st} - \theta_{st}$ in the production function coefficient:

$$
\begin{aligned}
\mu_t &= \sum_i m_{it}\hat{\mu}_{it} - \sum_i m_{it}\psi_{st} \\
&= \hat{\mu}_t - \sum_s m_{st}\psi_{st}
\end{aligned}
$$

We have no prior to belief that there is a particular correlation between the weight of an industry in the economy, or in the sample, and the bias introduced by either the simultaneity, selection or omitted price variable bias.

Appendix B Data: Summary Statistics

B.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. 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).[62] 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.

[62] Rather 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 (**??**).

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

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

Notes: Thousands 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.

B.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 about 420,000 establishments. These establishments aggregate into about 650,000 retail firms, about 314,00 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.[63]

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.

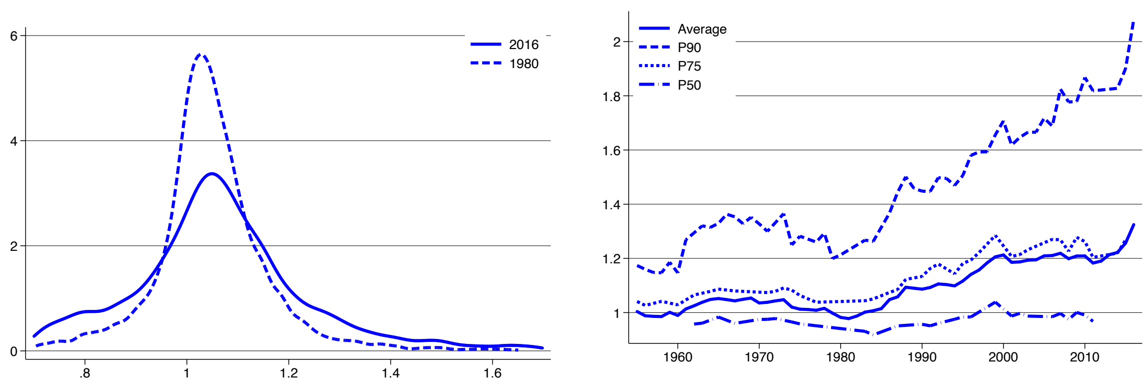
Perhaps the most common way of defining "firm" in the recent firm heterogeneity literature is to say that all of a firm's establishments in a given 4-digit SIC industry (roughly equivalent to a 6-digit NAICS industry) are a distinct firm (this is the approach taken by Hsieh and Klenow (2009), Autor, 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

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

sector Census (e.g. all of Walmart's firms in all of retail, NAICS codes 44-45, are a single firm).

Notice that even though we do not have information on overhead directly, one can obtain multiple sources of information about overhead costs in the Census data: 1) Census flags the auxiliary establishments of multi-unit firms and links them to other establishments of the same firm. Auxiliary establishments include headquarters, other facilities mainly engaged in general management functions, and facilities that mainly engage in R&D. Census has taken a systematic approach to identifying and flagging auxiliary establishments across most sectors of the economy since 1997. 2) Census conducts various business surveys that elicit information about various types of overhead. For example, the 2012 Annual Survey of Manufactures includes questions about software expenses, the cost of purchased communication services, advertising and promotional expenses, and the cost of purchased professional and technical services. As a second example, the Survey of Industrial R&D collects data on research and development expenses for all large firms and a sample of smaller ones. While the data are scattered across a variety of sources and databases, there is great potential in constructing firm-level measures of overhead costs using the Census data, and combine it with the markup analysis. This is left for future work and lies beyond the scope of this paper.[64]

Appendix C The Distribution of Markups with Technology PF2



(a) Kernel Density of Unweighted Markups

(b) Percentiles of the Markup Distribution (revenue weighted)

Figure C.1: Distribution of Markups μ_{it}: Kernel Density Plots (unweighted)

Appendix D 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

[64]We would like to thank an anonymous referee for pointing this out to us, and for providing a careful and detailed discussion on the various datasources at census to measure overhead costs.

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. We consider a fixed-proportion (Leontief) technology in the intermediate variable. This is the case considered in De Loecker and Scott (2016), Ackerberg, Caves, and Frazer (2015) and Gandhi, Navarro, and Rivers (2011), and avoids the potential identification issues surrounding intermediate inputs in the classical setting. Consider:

$$Q_{it} = \min\{\theta^M M_{it}, L_{it}^{\theta_L} K_{it}^{K} \Omega_{it}\} \tag{D.1}$$

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 (D.1) requires input choices in fixed proportions:

$$\mu_{it}(L) = \frac{1}{\mu_{it}^{-1} + \frac{P_t^M M_{it}}{P_{it} Q_{it}}} \tag{D.2}$$

where μ_{it} is obtained with the standard formula, and where the output elasticity is the estimated labor coefficient – i.e., $\mu_{it} = \theta^L \frac{S_{it}}{wL_{it}}$. P_t^M We compute the material expenditure by netting the wage bill from COGS.

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

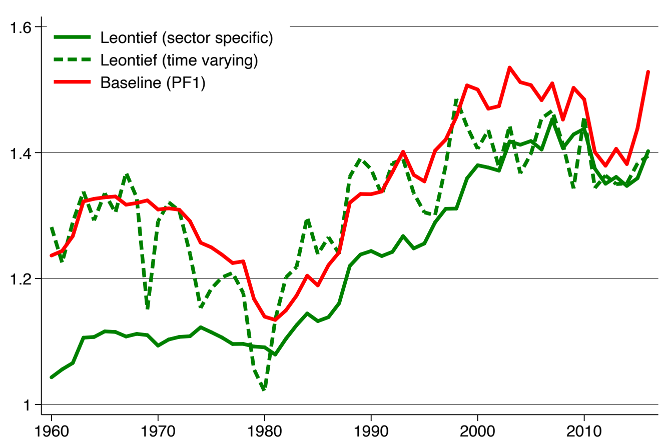


Figure D.1: Industry Specific Average Markups

Appendix E Markup Distribution: Autoregressive Process

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:

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

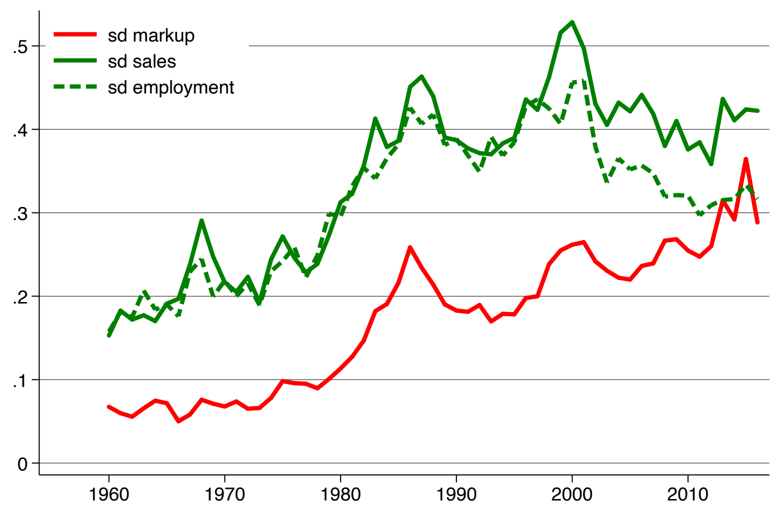


Figure E.1: The Evolution of the Standard Deviation of Markups, Sales and Employment (1960 - 2016). (AR(1) in logs on their lag with year and industry fixed effects; The estimated persistence is 0.84).

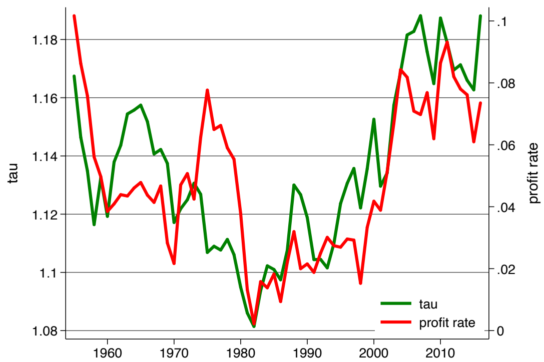
Figure E.1 shows the evolution of the cross-sectional standard deviation of the shocks in the markup, sales and employment processes. Starting in 1980, there is clearly a sharp rise in the standard deviation of the markup μ and a more moderate increase in that of sales S. Interestingly, there is much 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).[65]

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.

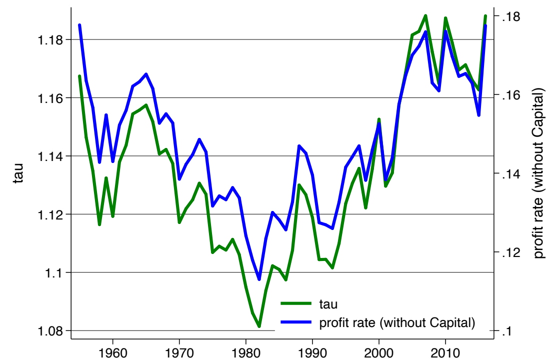
[65] As robustness checks, we have also included higher order terms of the persistence and find that those are not important.

Appendix F Aggregate Profits

We compare aggregate profits with the markup measure τ proposed in Traina (2018).



(a) π_t



(b) π_t^{OPX}

Figure F.1: τ versus Profit Rates

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).

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.

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

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.

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., AND J. EECKHOUT (2018a): "Global Market Power," NBER working paper.

——— (2018b): "Some Thoughts on the Debate about (Aggregate) Markup Measurement," Discussion paper, mimeo.

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. PAVCNIK (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. WARZYNSKI (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.

——— (2019): "How costly are markups?," Discussion paper, National Bureau of Economic Research.

EECKHOUT, J., AND X. WENG (2017): "The Technological Origins of the Decline in Labor Market Dynamism," mimeo.

ELSBY, M. W., B. HOBIJN, AND A. ŞAHIN (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.

——— (2018): "How EU Markets Became More Competitive Than US Markets: A Study of Institutional Drift," Working Paper 24700, 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. ŞAHIN (2016): "Demographic Origins of the Startup Deficit," NY Fed mimeo.

KEHRIG, M. (2011): "The cyclicality 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.

KOH, D., R. SANTAEULALIA-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.

MERTENS, M. (2019): "Micro-Mechanisms behind Declining Labor Shares: Market Power, Production Processes, and Global Competition," Halle Institute for Economic Research mimeo.

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.

RUBENS, M. (2019): "Monopsony power and factor-biased technology adoption," University of Leuven mimeo.

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.

——— (2019): "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, 33(3), 23–43.

TRAINA, J. (2018): "Is Aggregate Market Power Increasing? Production Trends using Financial Statements," Chicago Booth mimeo.

Online Appendix: Not For Publication

The Rise of Market Power and the Macroecononomic Implications

JAN DE LOECKER, JAN EECKHOUT AND GABRIEL UNGER

November 15, 2019

Appendix 1 Further Robustness of the Technology

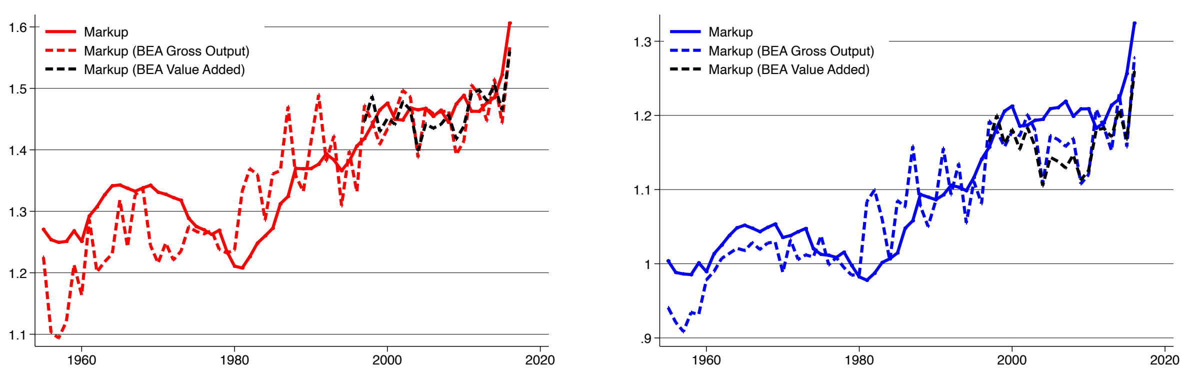
1.1 Comparison to alternative measures

Unfortunately we cannot compare the markups from the production approach to demand approach, since we do not have the data (see De Loecker and Scott (2016) for such an analysis). We can, however, contrast our finding to the accounting approach. This approach relies, as mentioned above, on the premise that average and marginal costs are equated (which is consistent with assuming constant returns to scale in production), and moreover that all cost items are considered variable inputs in production. Interestingly enough, if we impose these assumptions the markup in the DLW framework reduces to the ratio of sales-to-total costs. The aggregate version (sales-weighted) of the latter is in fact identical to the aggregate profit rate, which highlights that the accounting approach makes no distinction between profits and markups.

We can further decompose the importance of the two assumptions: average equals marginal costs (which coincides with constant returns to scale, and no presence of fixed cost), and all cost items are variable. For example, we can keep the first assumption but simply only consider COGS to be the only variable input in production. This case happens to coincide with the calibrated version (see Figure 2a), with the only difference that the output elasticity is now calibrated to one, and the entire aggregate markup series is shifted up. Finally, we can move items from the cost structure, like SG&A, into the variable cost factors. The framework we rely on then asks the important question whether the various variable input factors of production are perfect substitutes or not. In the more plausible case where they are not, we would obtain two first-order conditions to compute the markup (one for each variable input). If we take one variable input to be the benchmark (say cost of goods sold), then the (aggregate) markup series obtain from the second variable input (say SG&A) should be identical. When we apply this procedure to the data, we find very distinct patterns for SG&A indicating that this cost item contains a sufficient amount of fixed and quasi-fixed factors of production, or fixed costs (overhead) all together. The pattern is stark, and even when we insert estimated output elasticities for SG&A (from estimating the adjusted production function), we find a drastic different series. We interpret this evidence strongly in favor of the fact that SG&A contains many cost items that are not perfect substitutes with the cost items features in COGS, and in addition fixed in nature (i.e. overhead labor, brand value, royalties, innovative activities, advertising, etc.) that should not be subsumed in the variable cost of production, and hence determine the level of marginal costs, and the associated markup. Of course more work, that lies beyond the scope of this paper, is needed to dig into the various cost items, and decompose the the bundle SG&A itself into production and non-production components (much in the spirit of what we did with employment, with the wage bill being, at least partly, a part of the COGS input), and consider the appropriate value of the stock, and the associated user cost.

1.2 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.[66] 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.



(a) PF1

(b) PF2

Figure 1.1: 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).[67] Once

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

[67] In Figure 2.1 in the Online Appendix, we plot the scatter plot of the Compustat weights against the BEA sam-

we adjust for the economy-wide sample weights, we obtain a measure for markups (both for PF1 in Figure 1.1a and PF2 in Figure 1.1b) 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 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.

1.3 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 3.1a 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 markups, one weighted by the revenue shares of the entire sample of firms, the other weighted only by the share of firms in

ple 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.

the subsample $\mathcal{G}$.[68] Both measures give very similar ratios, that are very close to one (Figure 3.1b). 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 3.2 in the Online Appendix. We find that the pattern of the average markup is very 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.

1.4 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.

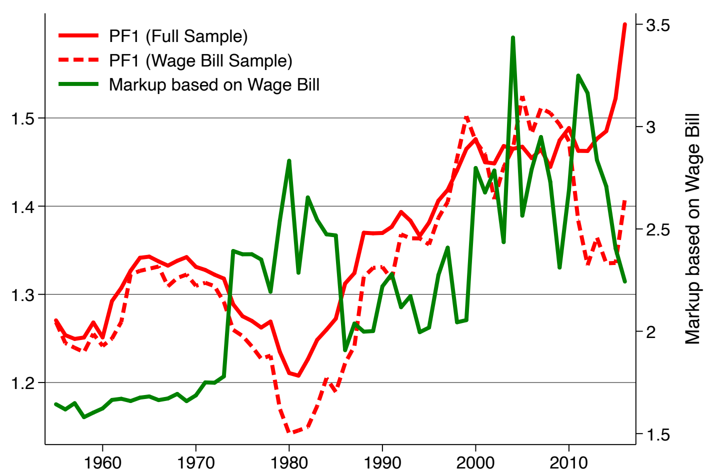


Figure 1.2: Average Markup derived from Wage Bill: PF1 (full sample); PF1 (report XLR);

[68]Formally, these ratio are obtained as:

$$\frac{\sum_{i\in\mathcal{G}} \frac{S_{it}}{S_t}\mu_{it}}{\sum_{i\in\mathcal{G}} \frac{S_{it}}{S_t}\mu_{it}^D} \quad \text{and} \quad \frac{\sum_{i\in\mathcal{G}} \frac{S_{it}}{S_t(\mathcal{G})}\mu_{it}}{\sum_{i\in\mathcal{G}} \frac{S_{it}}{S_t(\mathcal{G})}\mu_{it}^D}. \tag{1.1}$$

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.[69] 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 revenue weighted average markup based on employment rather than COGS. In Figure 1.2 we report the markup estimate for different specifications.

1.5 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.[70] We have not a 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 (7) to K and X instead of to V would give us an expression for the markup μ_{it}^K and μ_{it}^X:

$$\mu_{it}^K = \theta_{it}^K \frac{P_{it}Q_{it}}{P_{it}^K K_{it}} \quad \text{and} \quad \mu_{it}^X = \theta_{it}^X \frac{P_{it}Q_{it}}{P_{it}^X X_{it}}. \tag{1.2}$$

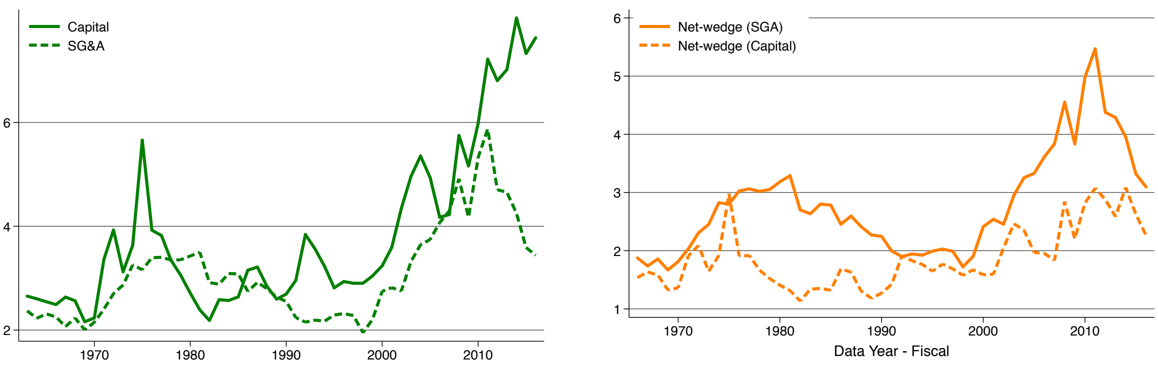
If K and X are variable inputs that adjust within one period, then the markups in (1.2) are identical and equal to the markup μ_{it} that we obtained for V (provided V is variable and fully adjusts within one period). This offers an over-identifying restriction of the model, and a simple way to evaluate whether K and X are variable. In Figure 1.3a we report the markups μ_{it}^K and μ_{it}^X, if these inputs were indeed variable and statically optimized. [71]

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

[69] Reporting the wage bill is not an SEC requirement.

[70] It 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 17 we consider the census of manufacturing plants and consider material inputs separately.

[71] The 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 and *all* factors of production to be variable.



(a) Wedge Capital and SG&A

(b) Net-Wedge Capital and SG&A

Figure 1.3: Fixed Factors and Markup Measurement

increase in the benchmark measure μ_{it}^{V} slows down. The delay appears to indicates 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:

$$\psi_{it} = \frac{\hat{\mu}_{it}}{\mu_{it}} = \frac{\theta^K P_{it}^K K_{it}}{\theta^V P_{it}^V V_{it}}, \tag{1.3}$$

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 1.3b 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 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 all components booked under SG&A (X) constitute variable inputs in production, in addition to the assumptions that they are perfect substitutes to all components captured by cost-of-goods-sold (V).

Appendix 2 Sample weights BEA vs. Compustat

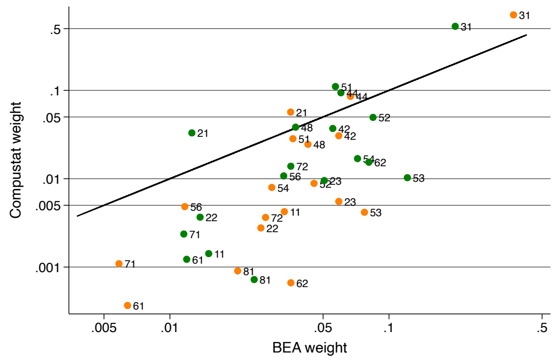
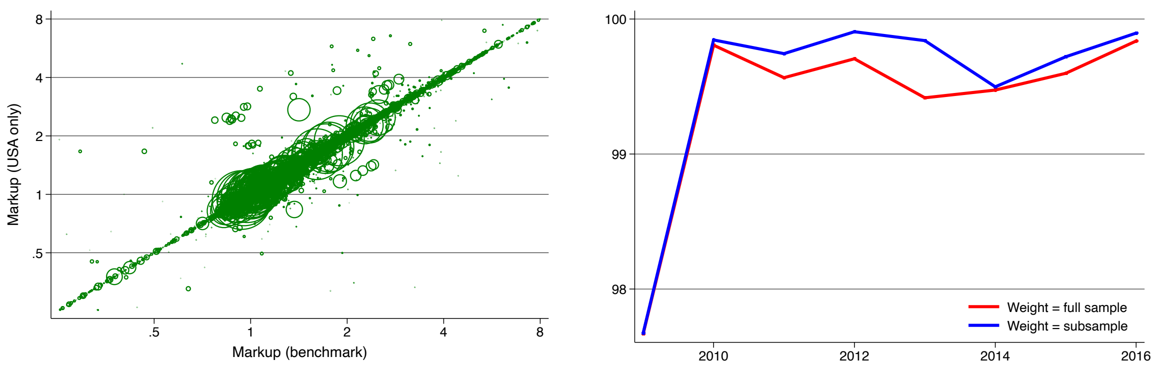


Figure 2.1: Sample weights: Compustat versus Economy wide (BEA) – log scale, 1980 (orange) and 2016 (green). For the list of NAICS codes, see Table 15.1.

Appendix 3 Consolidated Accounts: Geographical Decomposition



(a) Scatter plot of μ_{it} and μ_{it}^D (size of scatter point is weight of the yearly sales share)

(b) Ratio of average consolidated to domestic (USA only) markup

Figure 3.1: 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.

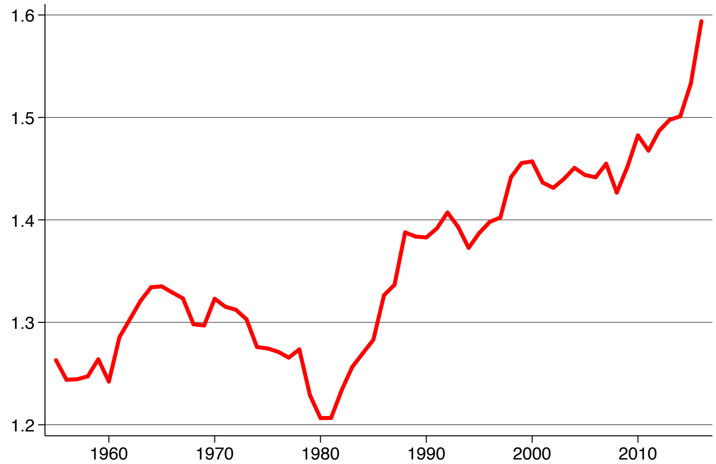


Figure 3.2: Average Markup (PF1) using only those firms in the sample that are incorporated in the United States (84.3% of all firm-year observations).

Appendix 4 Decomposition: details (equation (9))

year	Markup	Δ Markup	ΔWithin	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 4.1: Decomposition of aggregate markups at the firm level (equation (9)).

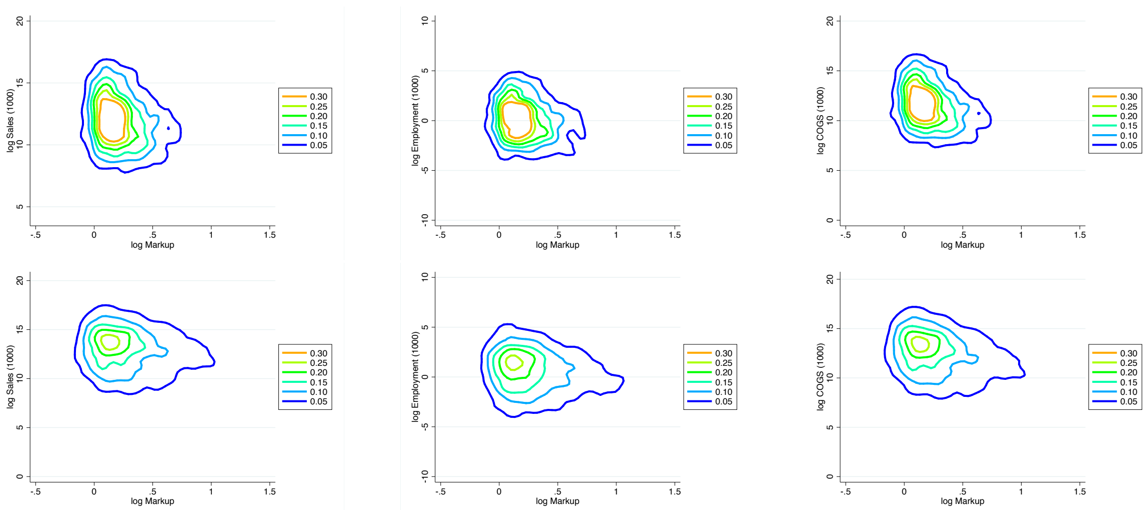
Appendix 5 A Selection of Firms' Individual Markups

	Markup μ_i					Empl. L_i thousands	Sales S_i millions (2010 $)
	1980	1990	2000	2010	2016	2016	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 5.1: Individual Firms' Markup (baseline measure PF1)

Appendix 6 Contour Plots of the Markup Distribution with Revenue and Inputs

In Figure 6.1 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 Revenue, 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, 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.



(a) Sales. (b) Employment. (c) COGS.

Figure 6.1: Joint Distribution of the firm-level Markup and the Share of Revenue 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.

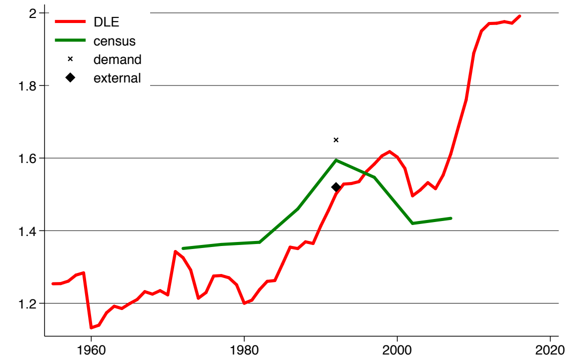
Appendix 7 Comparing markup measures: 7 case studies

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).

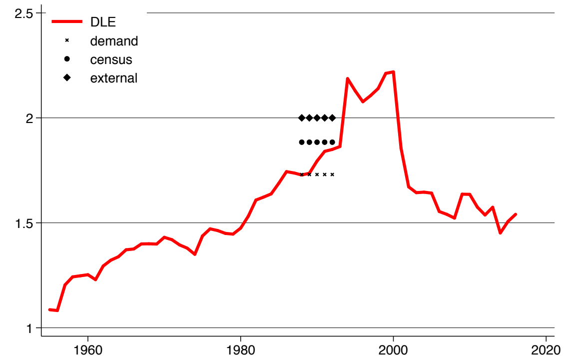
An important observation made in De Loecker and Scott (2016) is that a strict comparison of various approaches is difficult for at least two separate reasons. First, except for a handful of sectors, there is no overlap (in either time or geography) between the data required for applying the production 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.

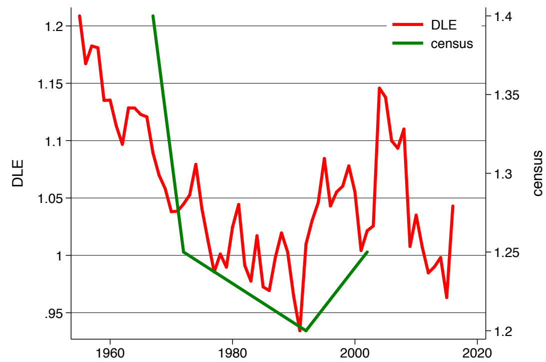
The underlying sources are: *Brewing*: Compustat: share-weighted aggregate markup, demand and census based on analysis in De Loecker and Scott (2016), and external evidence from the 1996 Consumer Reports; *RTE Breakfast Cereal*: Compustat: share-weighted aggregate markup for Kellogs and Quaker), for 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.



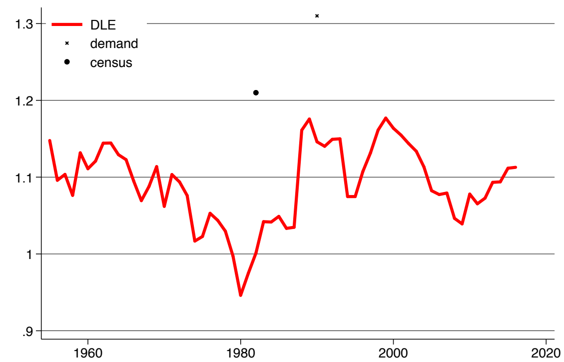
(a) Markups: US Beer Industry



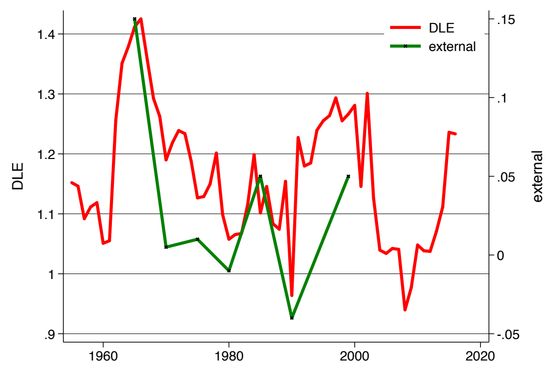
(b) Markups: US RTE Breakfast Cereal Industry



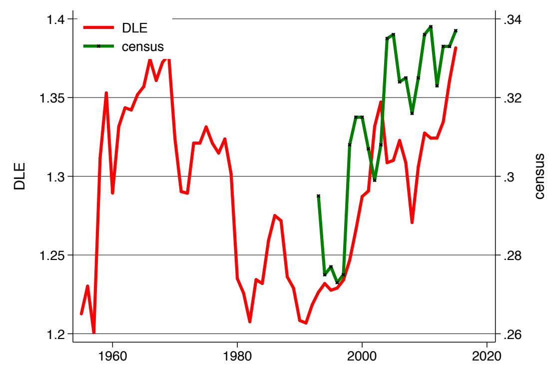
(c) Markups: US Steel Industry



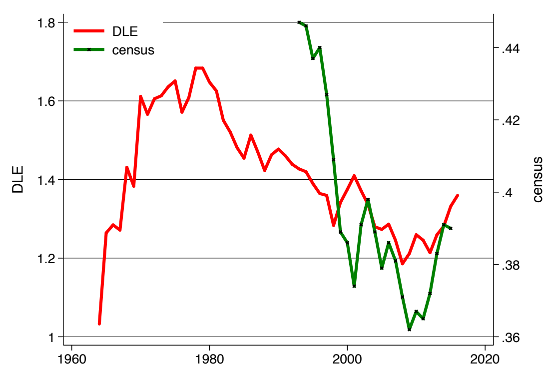
(d) Markups: US Automobile Industry



(e) Markups: US Airlines Industry



(f) Markups: US Department Store Industry



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

Figure 7.1: Markups: DLE and alternative sources

Appendix 8 Profit Rate without Capital

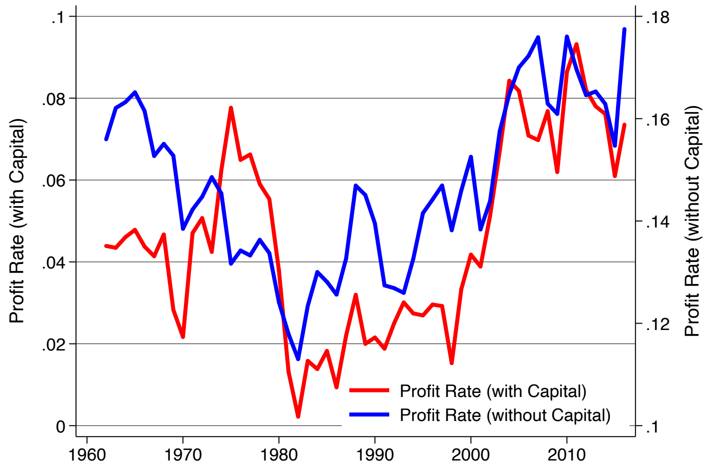
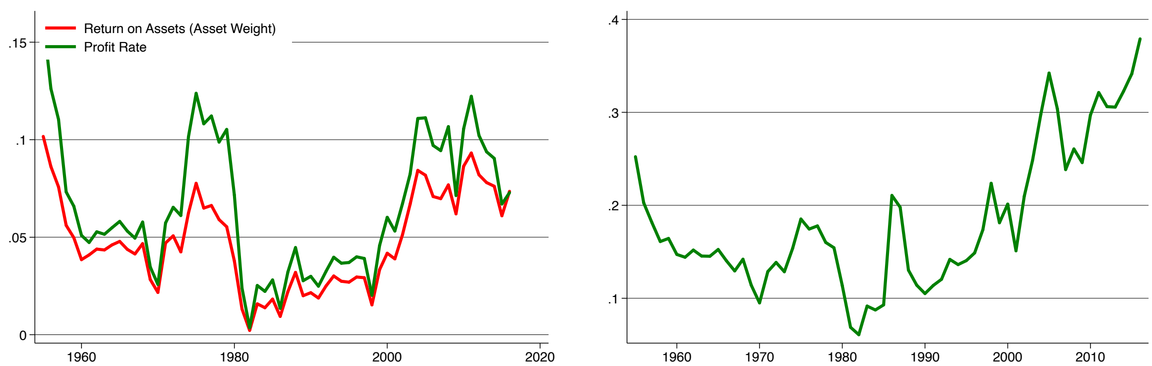


Figure 8.1: Profit Rate: with and without Captial.

Appendix 9 Return on Assets



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

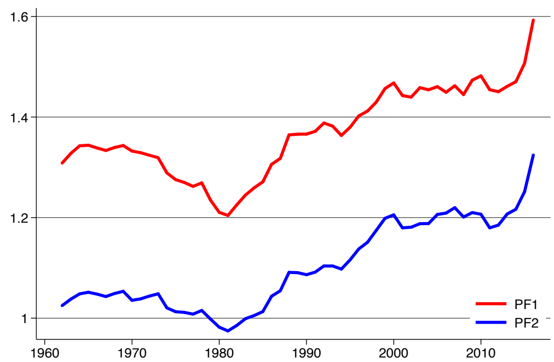
(b) Return on Assets (revenue weight).

Figure 9.1: Average Return on Assets.

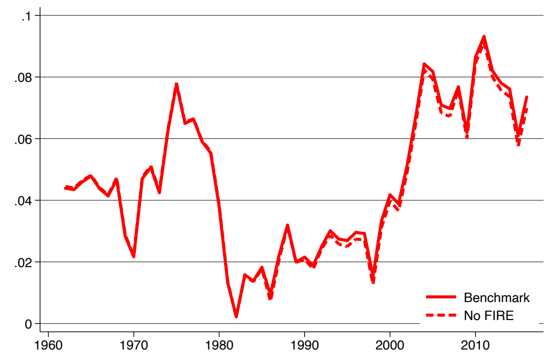
Appendix 10 Excluding FIRE

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 10.1. We find little difference in the pattern of the average markups for both specifications PF1 and PF2 (Figure 10.1a) compared to the sample with FIRE. Likewise for the profit rate (Figure 10.1b).



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



(b) Profit Rate without FIRE

Figure 10.1: Markup and profit rate without Finance, Insurance and Real Estate (FIRE).

Appendix 11 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 11.1: 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 11.1 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 (10) 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 12 Industry-specific trends

12.1 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.

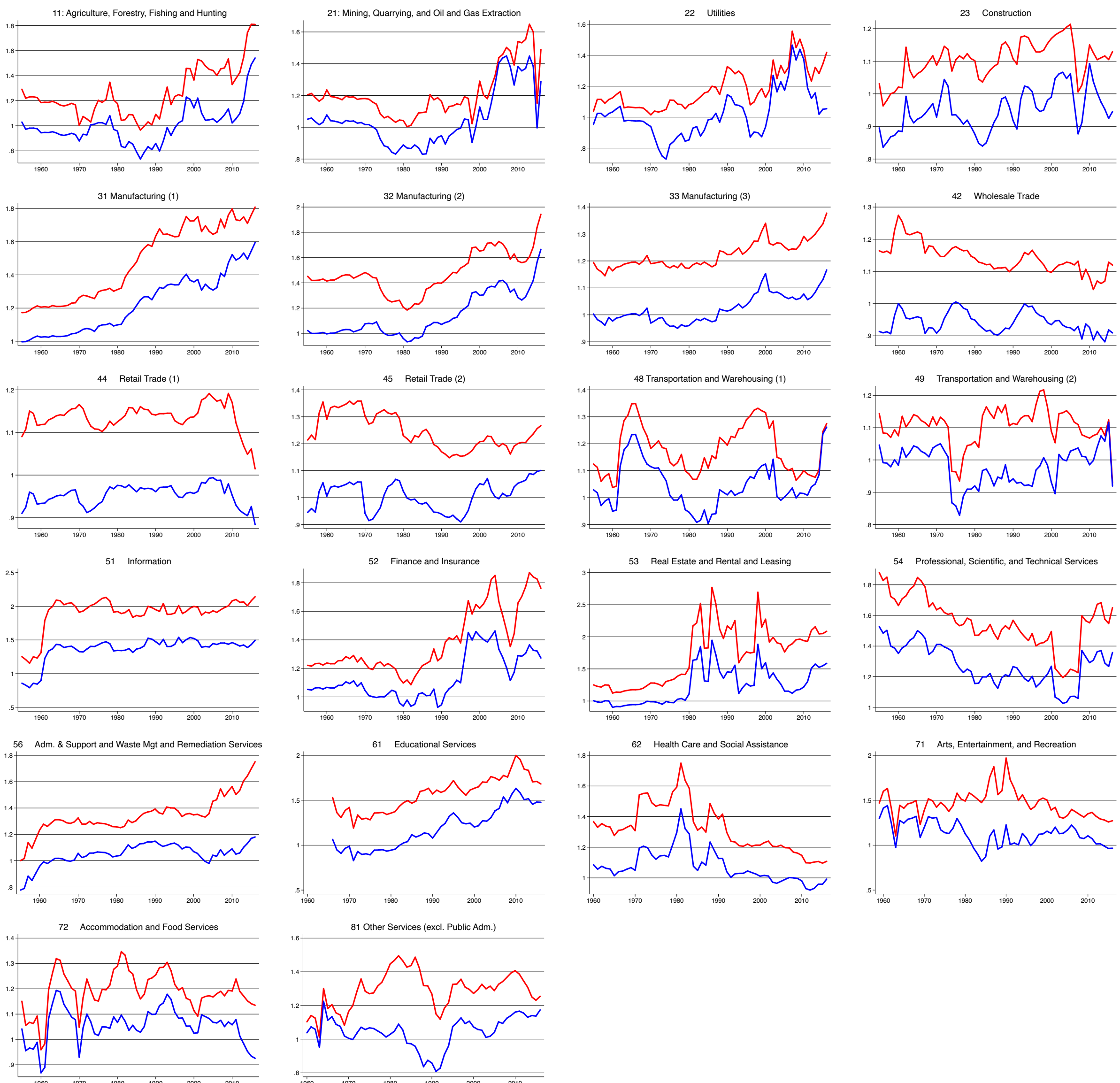


Figure 12.1: Industry Specific Markups.

12.2 Estimated Output Elasticities by 2-digit Sector

Figure 12.2: Industry Specific Output Elasticities.

12.3 Sales Shares by 2-digit Industry

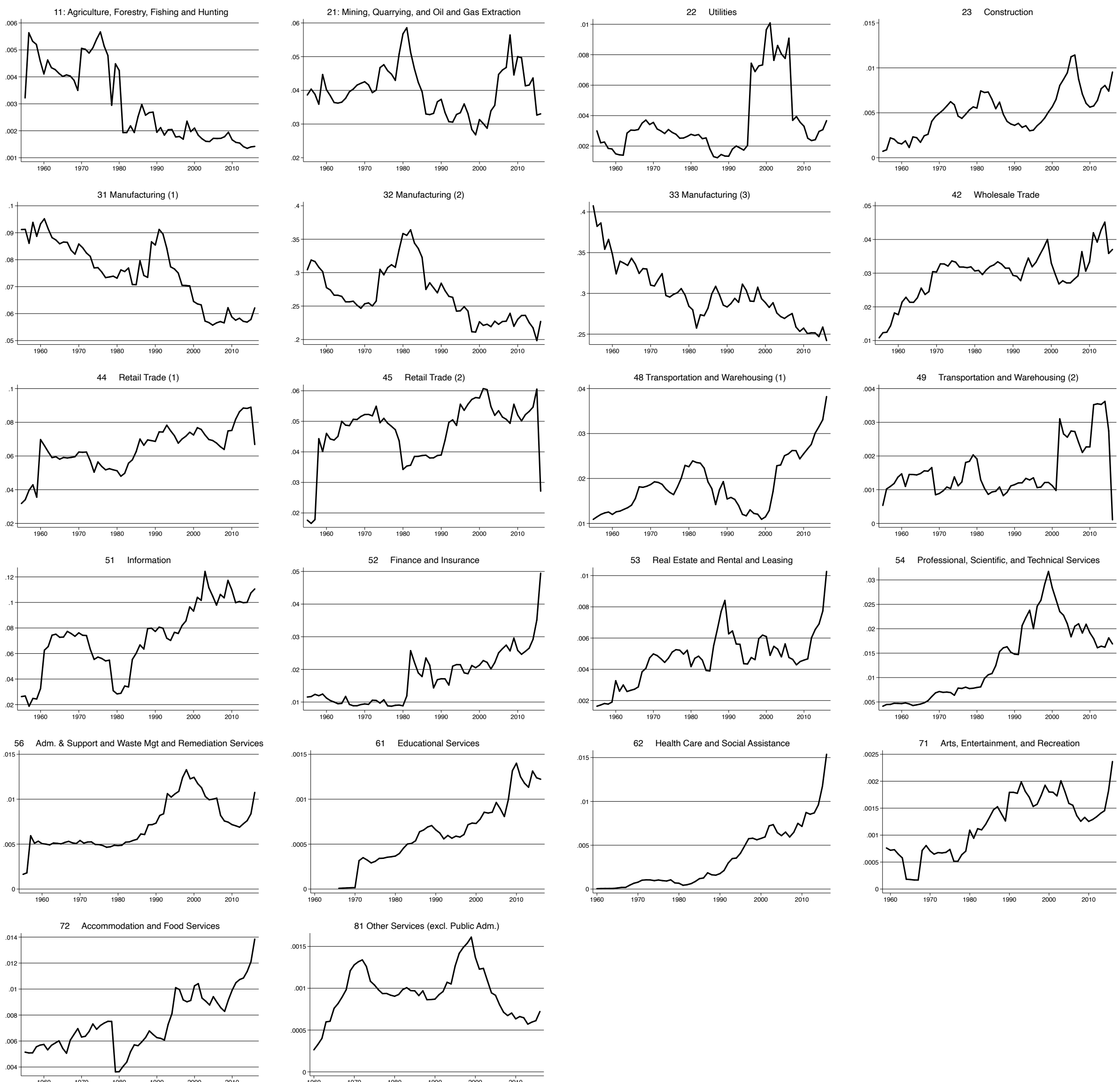


Figure 12.3: Industry Specific Market Shares.

12.4 Decomposition Within Firm by 2-digit Industry

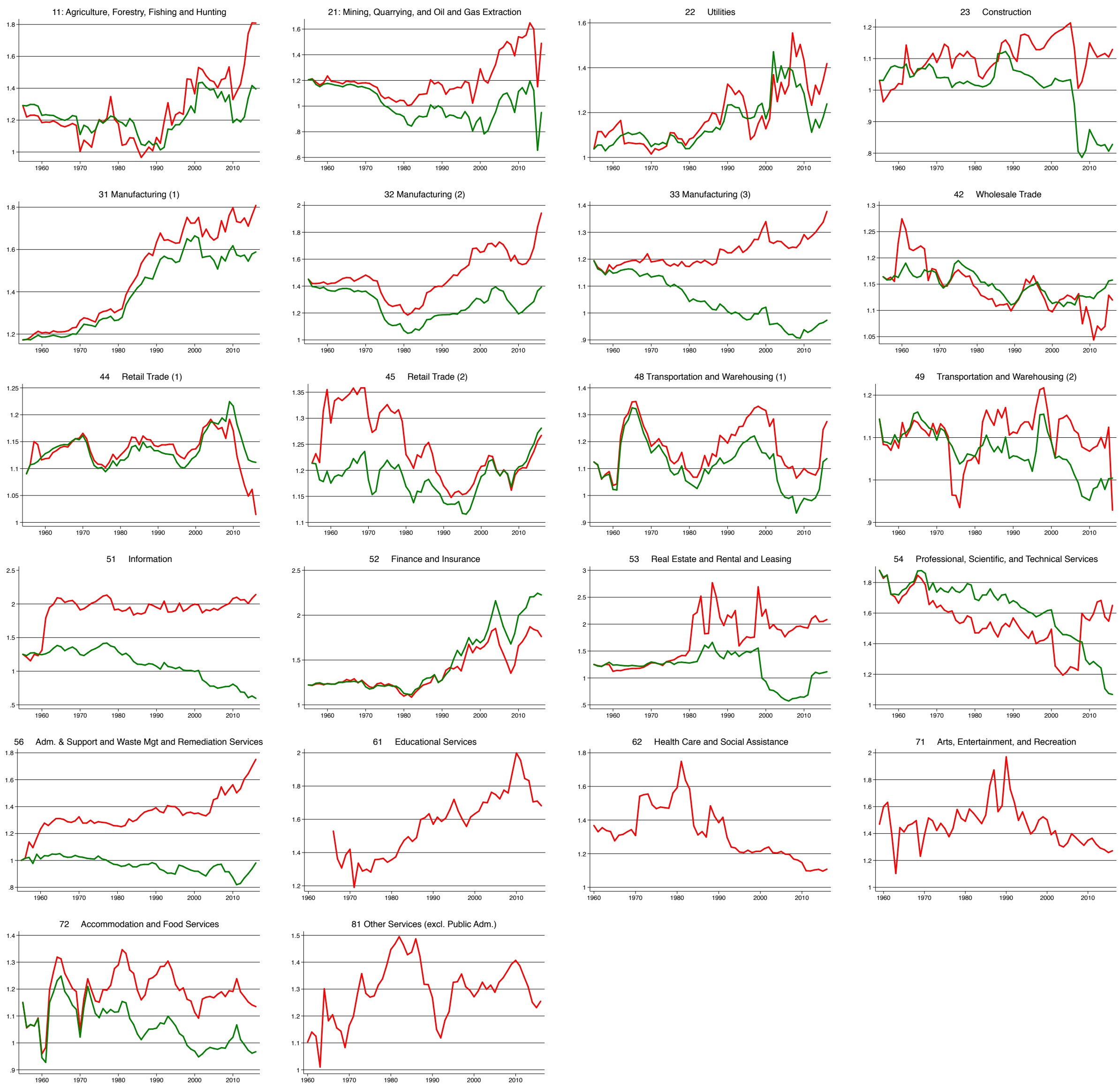
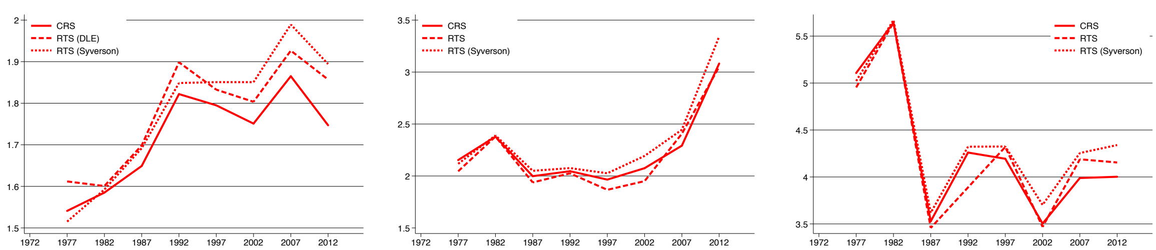


Figure 12.4: Decomposition Within Firm by 2-digit Industry.
For sectors 61, 62, 71, and 81 there is not enough data for at least one year to calculate the time difference. Recall that to calculate the elasticities, for that reason we use a five year rolling windows.

Appendix 13 Returns to scale in the Censuses.

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 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 6.3), 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 variable 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 (22)). 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.



(a) Manufacturing (b) Retail (c) Wholesale

Figure 13.1: 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 13.1. 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.

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).

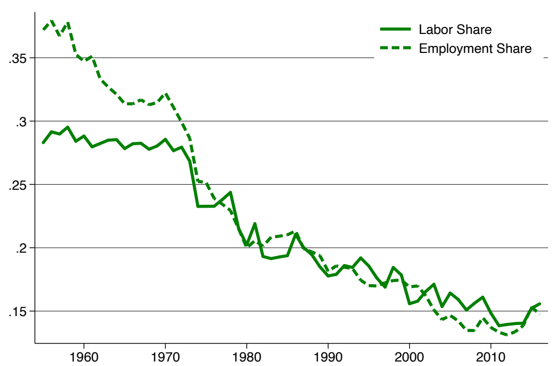
The biggest restriction is that one has to rely on labor as a variable input in production. While we were able to formally check materials as a variable input for the manufacturing sector, and found very similar markup patterns, this is not expected to hold outside manufacturing; especially not for the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

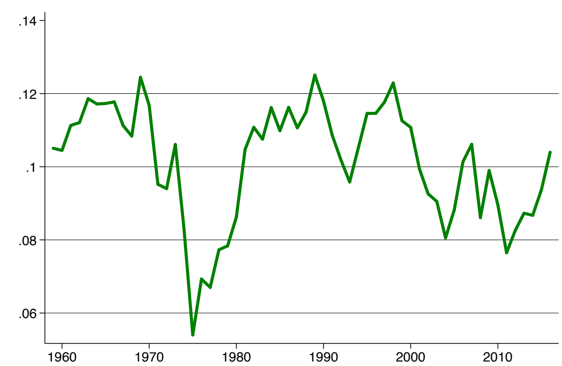
With those caveats in mind, we do, however find a robust set of facts: aggregate markups increase, and this increase is mostly due to the rise of margins at the top of the weighted markup distribution. There are, however, sectoral differences, and the rise of markups happens at different points in time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

Appendix 14 The Labor and Capital Shares in the Micro Data

In Figure 14.1a we report the revenue 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.



(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 14.1: Labor and Capital Shares.

The magnitude would roughly be double, though the 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%).

Instead, the capital share (Figure 14.1b) is much more stable.

Appendix 15 Results by sector

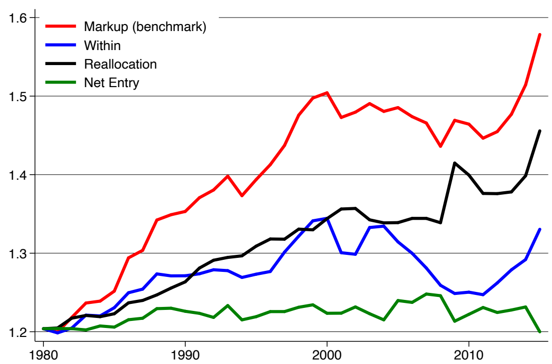
We present additional sector-level results on the time-series of markups, and the underlying components using the decomposition introduced in the main text. We present the aggregate markup by sector (2 digit) for the last year in our data (2016), alongside the change between 1980 and 2016, and the sector's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

Table 15.1: Markups by Sector, 1980-2016: PF1.

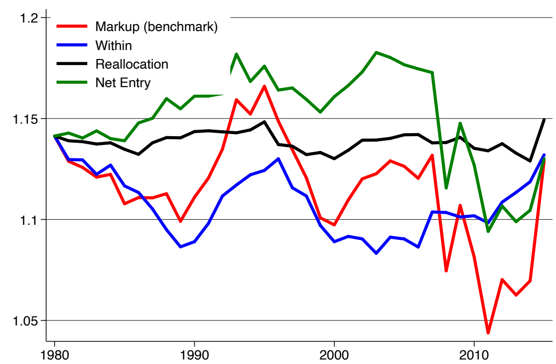
NAICS	Sector	Markup		Sales Share
			Change	
		2016	1980-2016	2016
11	Agriculture, Forestry, Fishing and Hunting	1.81	0.00	0.00
21	Mining, Quarrying, and Oil and Gas Extraction	1.49	0.34	0.03
22	Utilities	1.41	0.07	0.00
23	Construction	1.16	0.03	0.01
31	Manufacturing (1)	1.81	0.05	0.06
32	Manufacturing (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

Appendix 16 Reallocation Decomposition: by Sector

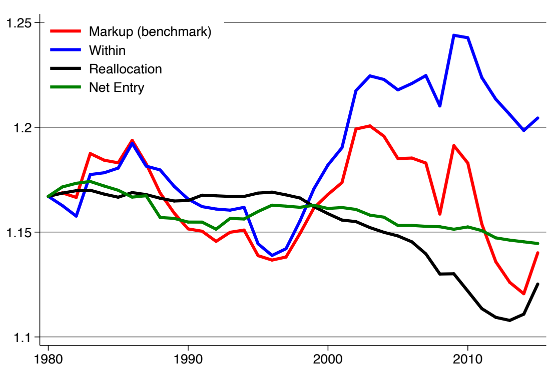
In Figures 16.1 and 16.2 we repeat the decomposition exercise with the counterfactuals for a few leading (broad) sectors separately – i.e., Manufacturing, Wholesale, Retail and Services. In the manufacturing sector both the within and the reallocation component contributed positively to the sharp rise 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.



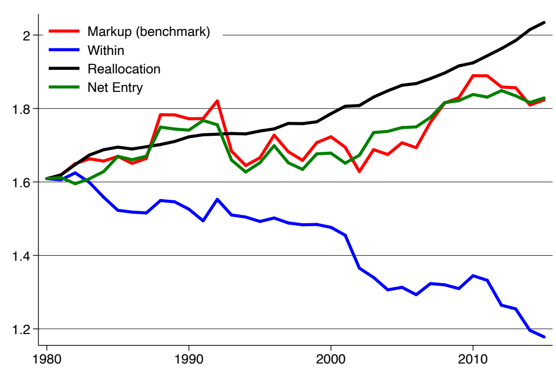
(a) Manufacturing (31-32-33)



(b) Wholesale (42)



(c) Retail (44-45)



(d) Services (54-56-61-62-71-72-81)

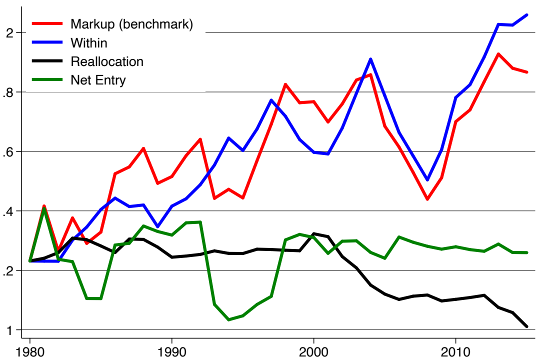
Figure 16.1: Firm Level Decomposition by Sectors

The Wholesale sector looks markedly different, where the aggregate markup is almost uniquely driven by the pattern of within markup growth.[72] 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.[73]

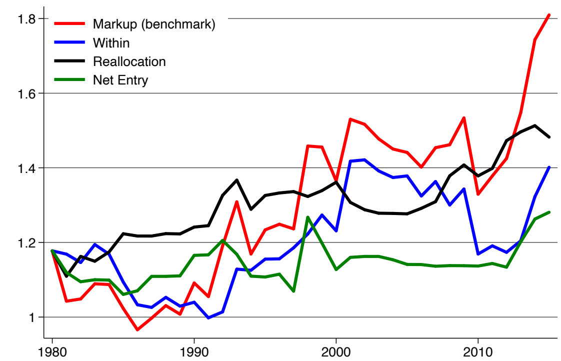
The Retail sector's aggregate markup, on the other hand, is uniquely driven by overall markup growth leading up to the late nineties. From there onward, the negative reallocation

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

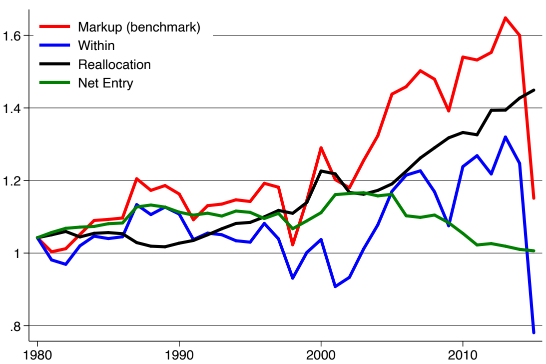
[73] We present the decompositions for the remaining sectors (Finance, insurance and real estate; Agriculture, Mining and Utilities) in Figure 16.1.



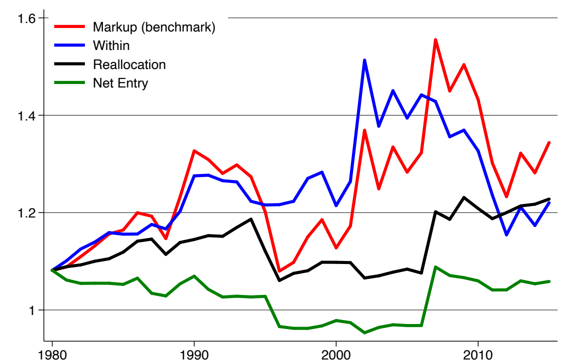
(a) FIRE: Finance, Insurance, Real Estate (52-53)



(b) Agriculture (11)



(c) Mining (21)



(d) Utilities (22)

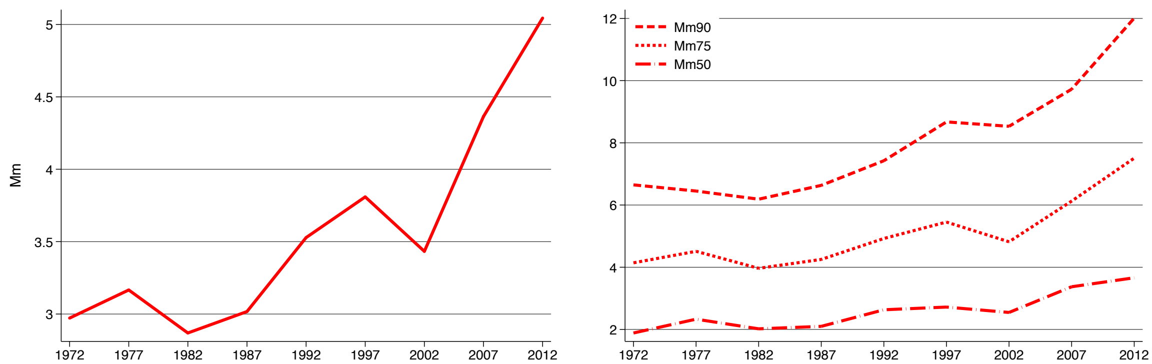
Figure 16.2: Firm Level Decomposition by Sectors

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.

Appendix 17 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 17.1 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.



(a) Manufacturing: average

(b) Manufacturing: percentiles

Figure 17.1: Markups in the US Census of Manufacturers. The variable input is materials. Averages and percentiles are revenue 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 17.2 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.

Appendix 18 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

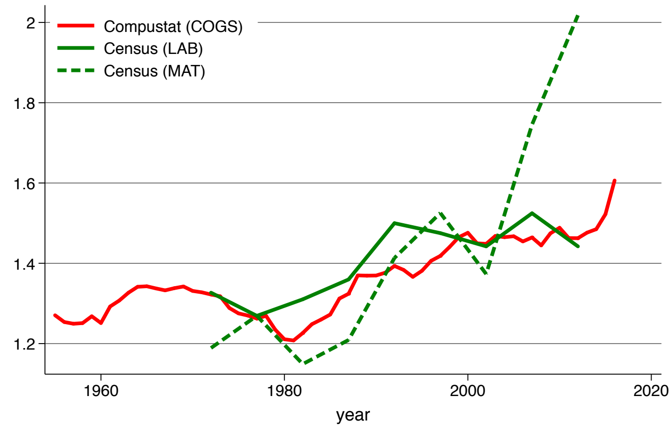


Figure 17.2: Aggregate Markup in Manufacturing: Variable inputs

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.

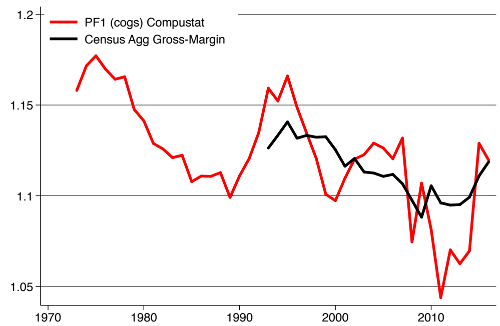
Taking these set of limitations as given, we perform the following comparison. In Figure 18.1 we compare the aggregate markup series for wholesale as obtained by applying our approach to the set of Compustat firms, relying on the variable input cost of goods sold (red line), to the publicly reported gross margins of the entire wholesale sector (black line).[74] Both series paint a very similar picture of the aggregate variable profit margins in the total sector.

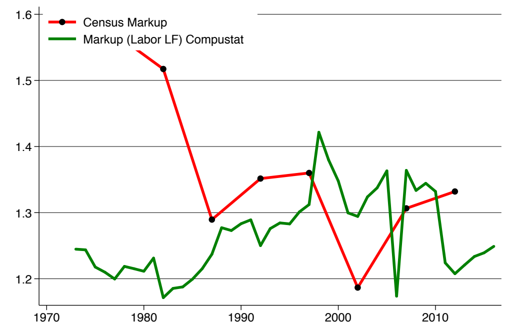
Finally, we consider the sector-aggregate markup using the first order condition of employment. The green line in the second panel in the figure below denotes the series obtained with the set of Compustat firms.[75] We overlay the aggregate markup obtained using the micro-census data (using a calibrated 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 stronger drop from the 1982 to the 1987 census year, but the trajectories are in fact the same.

[74]The 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 ($\mu = P/c$).

[75]The 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.

Figure 18.1: Wholesale trade: margins in census and Compustat





Appendix 19 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(Market Value)		
ln(Markup PF2)	0.71	0.64	0.56	0.18	0.73	0.65	0.59	0.29
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
ln(Sales)					0.81	0.81	0.83	0.68
					(0.00)	(0.00)	(0.00)	(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(Dividends)		
ln(Markup PF2)	1.03	0.93	0.79	0.23	1.01	0.90	0.77	0.25
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
ln(Sales)					0.94	0.92	0.93	0.76
					(0.01)	(0.01)	(0.01)	(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 19.1: firm-level Regressions: market values and dividends on markups (PF2) (clustered standard errors by firm in brackets).