The Rise of Market Power
and the Macroeconomic Implications*

Jan De Loecker†
KU Leuven and Princeton University
NBER and CEPR

Jan Eeckhout‡
UPF-ICREA-GSE and UCL

August 4, 2017

Abstract

We document the evolution of markups based on firm-level data for the US economy since 1950. Initially, markups are stable, even slightly decreasing. In 1980, average markups start to rise from 18% above marginal cost to 67% now. There is no strong pattern across industries, though markups tend to be higher, across all sectors of the economy, in smaller firms and most of the increase is due to an increase within industry. We do see a notable change in the distribution of markups with the increase exclusively due to a sharp increase in high markup firms.
We then evaluate the macroeconomic implications of an increase in average market power, which can account for a number of secular trends in the last 3 decades: 1. decrease in labor share; 2. increase in capital share; 3. decrease in low skill wages; 4. decrease in labor force participation; 5. decrease in labor flows; 6. decrease in migration rates; 7. slowdown in aggregate output.

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

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

*We would like to thank Mark Aguiar, Pol Antras, John Asker, Eric Bartelsman, Steve Berry, Emmanuel Farhi, Bob Hall, John Haltiwanger, Xavier Gabaix, Eric Hurst, Loukas Karabarbounis, Patrick Kehoe, Pete Klenow, Esteban Rossi-Hansberg, Chad Syverson, Jo Van Biesebroeck and Frank Verboven for insightful discussions and comments. Shubhdeep Deb, Morgane Guignard and David Puig provided invaluable research assistance. De Loecker gratefully acknowledges support from the FWO Odysseus Grant and Eeckhout from the ERC, Advanced grant 339186, and from ECO2015-67655-P.

†jan.deloecker@kuleuven.be.

‡j.eeckhout@ucl.ac.uk.

1 Introduction

The macroeconomy is experiencing a fundamental long term change in the last decades at a lower frequency than the business cycle. This manifests itself in a number of *secular* trends: some are related to labor market outcomes such as a declining labor share, declining wages and declining labor force participation; others to the slowdown in labor market dynamism with decreased job mobility and lower migration rates; and yet others are related to the capital share and the growth of output. While many explanations have been proposed for each of these secular trends, in this paper we argue that all these trends are consistent with one common cause that hitherto has remained undocumented, the rise in market power since 1980.[1]

The presence of market power has implications for welfare and resource allocation. Firms that can command a price above marginal cost produce less output. In addition to lowering consumer welfare, this has implications for factor demand, for the distribution of economic rents, and for business dynamics such as entry and exit, and resource allocation. In this paper we aim to achieve two goals. First, we document the evolution of markups for the US economy since 1950. Based on firm-level data, we find that while market power was more or less constant between 1950 and 1980, there has been a steady rise in market power since 1980, from 18% above cost to 67% above cost. Over a 35 year period, that is an increase in the price level relative to cost of 1% per year. Second, we investigate the macroeconomic implications of this rise in market power and the general equilibrium effects it has. We show that the rise in market power is consistent with seven secular trends in the last three decades.

While there is ample evidence that market power exists for particular markets and that it is pervasive across sectors, there is surprisingly little systematic evidence of the patterns of market power across the aggregate economy, and over time. The evidence on market power we do have comes from case studies of specific industries for which researchers have access to detailed data and for which they can estimate markups as a proxy for market power. 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.

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. Optimal pricing therefore relates observed price data to estimates of substitution elasticities to uncover the marginal cost of production, and consequently the markup. The combination of requiring data on consumer demand (containing prices, quantities, characteristics, consumer attributes, etc.) and the need for specifying a model of conduct, have limited the use of the so-called demand approach to particular markets.

In this paper we follow a radically different approach to estimate markups, following recent advances in the literature on markup estimation by De Loecker and Warzynski (2012) that builds on Hall (1988), the so-called production-approach. This approach relies on individual firm output and input data, covering a panel of producers over time, and in contrast to the

[1]Outside of the academic literature it has received significant attention, e.g. A lapse in concentration (*The Economist*, September 2016), and the CEA Issue Brief, Benefits of Competition and Indicators of Market Power (May 2016).

demand approach described above, 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, and this for many markets over a long period of time. Second, we can rely on publicly available sources providing us with production data. While there still exist many measurement issues and associated econometric challenges, to our knowledge there is no viable alternative to make progress.[2] This method allows us to uncover the basic pattern of market power over a long period of time and across the entire economy.

This paper starts by documenting the main patterns of markups in the US economy over the last six and a half decades, and in doing so we provide new stylized facts on the cross-section and time-series of markups. This analysis is interesting and important in its own right as increasingly economic models allow for meaningful markup variation across producers and time; and allowing for this firm/time variation has substantially different implications for a variety of questions.[3]

We observe all publicly traded firms covering all sectors of the US economy over the period 1950-2014, provided by Compustat. Because there is no official filing requirement for the privately held firms, our data does not include any privately held firms.[4] And while publicly traded firms are relatively few relative to the total number of firms, because the public firms tend to be the largest firms in the economy, they account for one third of total US employment (Davis, Haltiwanger, Jarmin, and Miranda (2007)) and about 41% sales (Asker, Farre-Mensa, and Ljungqvist (2014)). Moreover, publicly trade firms cover all sectors and industries Our main finding is that while average markups were fairly constant between 1960 and 1980 at around 1.2 (slightly decreasing from 1.27 to 1.18), there was a sharp increase starting in 1980 with average markups reaching 1.67 in 2014. In 2014, on average a firm charges prices 67% over marginal cost compared to only 18% in 1980. Over a period of 35 years, that is an increase markup rate by a factor of 3.6.[5] Moreover, the increase is becoming more pronounced as time goes on, especially after the 2000 and 2008 recessions. Second, there are marked changes in the distribution of markups over time. The increase occurs mainly in the top of the markup distribution. Markups in the top percentiles of the distribution go up most (from 1.4 in 1980 to 2.6 in 2014 for the 90th percentile), whereas those in the lower percentiles are flat or even decreasing. The median markup goes up, but only slightly so and substantially less than the

[2]De Loecker and Scott (2016) apply both methods – the demand approach and the production approach – to estimate markups in the US beer industry, and find that they yield very similar markup estimates.

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

[4]Until the US Inland Revenue Service gives researchers access to the information on tax declaration data of all privately held firms, this remains impossible.

[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.

average markup. Third, there is no strong compositional pattern across industries and the increase occurs mainly within industry. Finally, across industries it is mainly the smaller firms that have higher markups, though this effect disappears at a narrow industry specification, indicating that there is systematic size differences across industries. Within narrowly-defined (here 4 digit) industries firm size and markups are positively correlated.

After we establish these main facts, we take these as given and we discuss the implications of the rise in market power for recent debates in the macro/labor literature. We take the rise in markups as given, and we do not engage in the analysis of how this happened, which lies beyond the scope of this paper, though we provide the reader with a few promising candidates in the concluding remarks. In particular, we show how the rise in markups naturally gives rise to a decrease in the labor share, a decrease in the capital share, a decrease in low skilled wages, a decrease in labor market participation, and decrease in job flows, a decrease in interstate migration, and finally, we show that properly accounting for the rise in markups, there is no productivity slowdown but instead an increase in productivity despite the fact that output growth has slowed down.

At a general level, our findings warrant a more systematic analysis of the aggregate impact, if any, of market power on a variety of outcomes of interest. While there was a tradition to investigate the potential impact of market power on resource allocation, the analysis of Harberger (1954) concluded that profit rates across US (manufacturing) industries during the 1920s were not sufficiently dispersed to generate any meaningful aggregate outcome. This analysis, and its conclusion that market power barely impacts economy-wide outcomes, became the default view held by many economists and policy makers ever since.[6]

The paper is organized as follows. In Section 2 the data and empirical framework to recover markups is presented. The main facts on markups, in both the cross-section and the time series, are presented in Section 3. We derive the macroeconomic implication in Section 4, taking the facts from the previous section as given, building on a stylized model of market-level competition and an aggregate labor market. We conclude in Section 5.

2 Empirical Framework

In this section we discuss the empirical framework and the main data source we rely on to describe the patterns of markups in the US economy. In particular, we observe firm-level output and input data for firms across the US economy. This data is sufficient to measure firm-level markups using minimal assumptions on producer behavior, but without assumptions on product market competition and consumer demand. We apply the method proposed by De Loecker and Warzynski (2012), and we discuss the implementation.

[6]See Asker, Collard-Wexler, and De Loecker (2017) for a discussion, and an application to the oil industry.

2.1 Data

The choice of data is driven entirely by the ability to cover the longest possible period of time, *and* to have a wide coverage of economic activity. This for example rules out using the census of manufacturing establishments, given the decreasing share of manufacturing in the overall employment of US economy, from about 25% in 1960 to 9% in 2014 (Baily and Bosworth (2014)). To our knowledge, Compustat is the only data source that provides substantial coverage of firms in the private sector over a substantial period of time, covering the period 1950 to 2014. While publicly traded firms are relatively few relative to the total number of firms, because the public firms tend to be the largest firms in the economy, they account for one third of total US employment (Davis, Haltiwanger, Jarmin, and Miranda (2007)) and about 41% sales (Asker, Farre-Mensa, and Ljungqvist (2014)).

There are two alternatives to using our publicly trade firms data. A first alternative data source is the Census of Manufacturing, but it accounts for only 8.8% of US employment in 2013. Moreover, the Census of Manufacturing is – by construction – highly selective in the sectors and industries covered. We establish in Appendix B.1 that for manufacturing only, the same result holds. A second alternative data source is to use aggregate industry-level data rather than micro-level data, as originally done by Hall (1988). We perform this robustness analysis in Appendix B.4 both on aggregate IRS data and an aggregation of our own data of public firms. The aggregate patterns across the (aggregate) Compustat dataset and the economy-wide private sector (using IRS) are very similar.

The Compustat data contains firm-level balance sheet information, which allows us to rely on the so-called production approach to measuring markups, and market power. In particular we observe measures of sales, input expenditure, capital stock information, as well as detailed industry activity classifications.[7] 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, are at all correlated to the overall evaluation of the market. Table A.1 in the Appendix provides basic summary statistics of the firm-level panel data used throughout the empirical analysis.

2.2 Markup estimation

Measuring markups is notoriously hard as marginal cost data is not readily available, let alone prices, for a large representative sample of firms. The standard approach in modern Industrial Organization is to specify a particular demand system that delivers price-elasticities of demand, which combined with assumptions on how firms compete, in turn deliver measures of markups through the first order condition associated with optimal pricing. This approach, while powerful in other settings, is not useful here for two distinct reasons. First, we do not want to impose a specific model of how firms compete across a dataset of large firms, or com-

[7]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).

mit to a particular demand system for *all* the products under consideration. Second, even if we wanted to make all these assumptions, there is simply no information on prices and quantities at the product level for a large set of sectors of the economy, over a long period of time, to successfully estimate price elasticities of demand, and specify particular models of price competition for all sectors.

We rely on a recently proposed framework by De Loecker and Warzynski (2012), based on the insight of Hall (1988) to estimate (firm-level) markups using standard balance sheet data on firms, which does not require to make assumptions on demand and how firms compete. Instead markups are obtained by leveraging cost minimization of a variable input of production. This approach requires an explicit treatment of the production function.

2.2.1 Producer behavior

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

$$Q(\Omega_{it}, \boldsymbol{V}_{it}, K_{it}) = \Omega_{it} F_t(\boldsymbol{V}_{it}, K_{it}), \tag{1}$$

where $\boldsymbol{V} = (V^1, ..., V^J)$ captures the set of variable inputs of production (including labor, intermediate inputs, materials,...), K_{it} is the capital stock and Ω_{it} is the Hicks-neutral productivity term that is firm-specific. Because in the implementation we will use information on a bundle of variable inputs, and not the individual inputs, in the exposition we treat the vector $\boldsymbol{V}$ as a scalar V. Following De Loecker and Warzynski (2012) we consider the associated Lagrangian objective function:

$$\mathcal{L}(V_{it}, K_{it}, \Lambda_{it}) = P^V_{it} V_{it} + r_{it} K_{it} - \Lambda_{it}(Q(\cdot) - Q_{it}), \tag{2}$$

where P^V is the price of the variable input, r is the user cost of capital,[8] $Q(\cdot)$ is the technology (1), Q_{it} is a scalar and Λ_{it} is the Lagrangian multiplier. We consider the first order condition with respect to the variable input V, and this is given by:

$$\frac{\partial \mathcal{L}_{it}}{\partial V_{it}} = P^V_{it} - \Lambda_{it} \frac{\partial Q(\cdot)}{\partial V_{it}} = 0. \tag{3}$$

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

$$\theta^V_{it} \equiv \frac{\partial Q(\cdot)}{\partial V_{it}} \frac{V_{it}}{Q_{it}} = \frac{1}{\Lambda_{it}} \frac{P^V_{it} V_{it}}{Q_{it}}. \tag{4}$$

[8] For our purpose here, we maintain the assumption that input markets are competitive and hence P^V and r are equal to marginal revenue product. If they were not, then this will affect the marginal cost and hence the estimate of the markup. The are two opposing effects of market power in the inputs market: 1. Double marginalization, by which the marginal cost of inputs would be higher than under competition; 2. Monopsony power, where the firms squeezes the upstream provider and obtains lower input prices. The measured markup will be biased but it is not clear in which direction. For an analysis with non-competitive inputs, see De Loecker, Goldberg, Khandelwal, and Pavcnik (2016).

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

$$\mu_{it} = \theta_{it}^V \frac{P_{it}Q_{it}}{P_{it}^{V^j} V_{it}}. \tag{5}$$

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

2.2.2 Implementation

We directly observe sales, $S_{it} = P_{it}Q_{it}$ and total variable cost of production, $C_{it} = \sum_j P_{it}^{V^j} V_{it}^j$, measured by the cost of goods sold. The Compustat data 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.[9]

In order to recover markups an estimate of the output elasticity of this input bundle is required. We follow standard practice and rely on a panel of firms, for which we estimate production functions by industry. In particular we consider various specifications of the production function, both at the level of the economy and the industry. For the main results we consider industry-specific Cobb-Douglas production functions, with variable inputs and capital.[10]

For a given industry we consider the production function:

$$q_{it} = \beta_v v_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it}, \tag{6}$$

where lower cases denote logs and $\omega_{it} = \ln \Omega_{it}$, and where q_{it} is measured as the log of deflated firm level sales.[11] We follow the literature and control for the simultaneity and selection bias, inherently present in the estimation of the above equation, and rely on a control function approach, paired with an AR(1) process for productivity to estimate the output elasticity of the

[9] We verify the robustness of our main findings to obtaining markup estimates using intermediate inputs only as the variable input. The latter, however, requires additional assumptions on how to derive a measure of intermediate input use from operating income before depreciations, and the total wage bill, where the latter is imputed from multiplying the reported total number of employees with industry-wide wage data, as done in e.g. Keller and Yeaple (2009).

[10] We also consider more flexible translog production functions, and Cobb-Douglas production functions with time-varying coefficients. Appendix B discusses the main findings using these alternative specifications.

[11] See De Loecker and Warzynski (2012) for a discussion of the measurement of sales and quantities.

variable input, here β_v.[12] The attractive feature of this approach, in our context, is that the control function approach rests on an optimal input demand equation, which is immediate in the cost minimization framework used to recover an expression for the markup. In particular, the insight from Olley and Pakes (1996) is that the (unobserved) productivity term ω_{it} is given by a function of the firm's inputs and a control variable, in our case the variable input bundle; such that $\omega_{it} = h(v_{it}, k_{it})$.

This approach relies on a so-called two-stage approach where in the first stage, the measurement error and unanticipated shocks to sales are purged using:

$$q_{it} = \phi_t(v_{it}, k_{it}) + \epsilon_{it}, \tag{7}$$

where $\phi = \beta_v v_{it} + \beta_k k_{it} + h(v_{it}, k_{it})$. The productivity process is given by $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$, and this gives rise to the following moment condition to obtain the industry-specific output elasticity:

$$\mathbb{E}(\xi_{it}(\beta_v)v_{it-1}) = 0, \tag{8}$$

where $\xi_{it}(\beta_v)$ is obtained, given β_v, by projecting productivity $\omega_{it}(\beta_v)$ on its lag $\omega_{it-1}(\beta_v)$, where productivity is in turn obtained using $\phi_{it} - \beta_v v_{it} - \beta_k k_{it}$, using the estimate ϕ from the first-stage regression of sales on a non-parametric function in the variable input and capital, and year dummies. This approach identifies the output elasticity of a variable input under the assumption that the variable input use responds to productivity shocks, but that the lagged values do not, and more importantly, that lagged variable input use is correlated with current variable input use, and this is guaranteed through the persistence in productivity.

We measure firm-level markups using the estimate of the output elasticity:

$$\mu_{it} = \beta_v \frac{S_{it}}{C_{it}}. \tag{9}$$

As discussed in De Loecker and Warzynski (2012) we correct the markup estimates for the presence of measurement error in sales, ϵ_{it}, which we obtain in the first stage regression (equation (7)).

3 The Evolution of Markups

In this section we present the key stylized facts of markups across the US economy and over time. We document the steady increase in overall markups, and decompose this trend across and within sectors. The drastic increase in markups is further decomposed across the percentiles of the markup distribution, and the time-series properties of markups are discussed. Finally, the link from rising markups to market power is made through auxiliary data on profits, dividends and market valuations.

[12] 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 acquisition, 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.

3.1 A Secular Trend since 1980

Figure 1 presents the weighted average markup, across the economy, over time where weights are based on firm-level sales. Average markups have gone up since the 1980s. In the beginning of the sample period markups were stable and slightly decreasing from 1.27 in the 1960 to 1.18 in 1980. Since 1980 there has been a steady increase to 1.67. In 2014, the average firm charges 67% over marginal cost, compared to 18% in 1980. In Appendix B.5 we report some individual firms' markups.

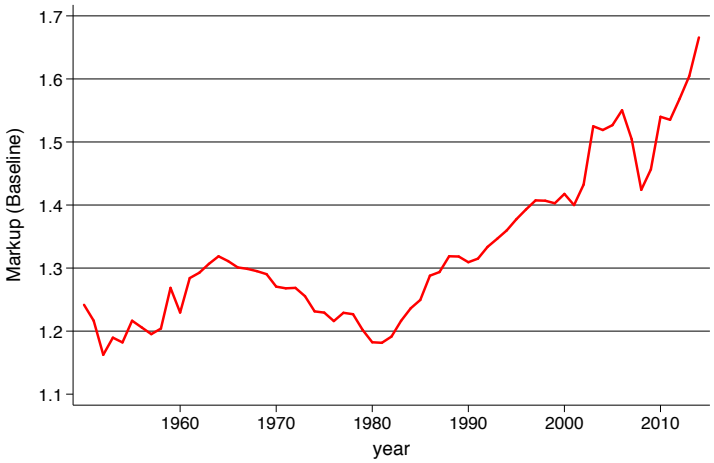


Figure 1: The Evolution of Average Markups (1960 - 2014). Average Markup is weighted by marketshare of sales in the sample.

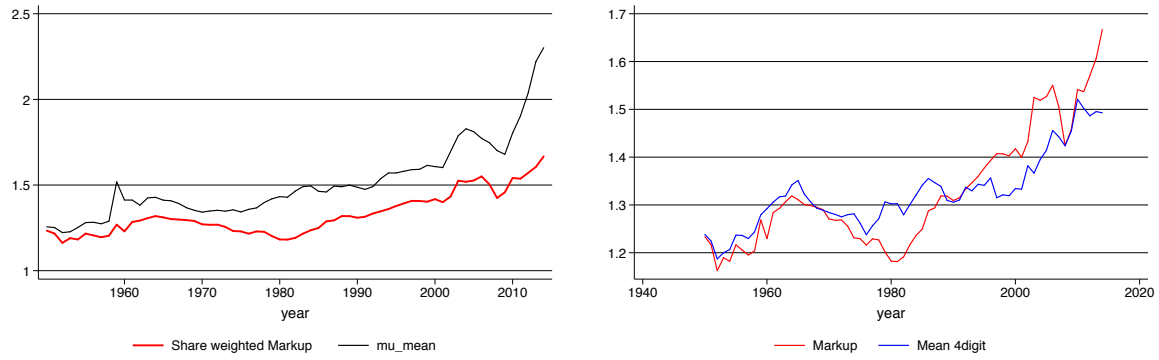
It is well known that the Compustat data is a selected sample and we therefore compare our result to aggregate data from the IRS, and further compare it to the aggregated version of our data (Compustat). The figures in the Appendix B.4 confirm that these same pattern holds, at the aggregate level, although there is a more modest increase in market power for aggregate data. The patterns, however, across the (aggregate) Compustat dataset and the economy-wide private sector (using IRS) are very similar. This highlights the importance of using micro-level data and that industry-level data cannot fully capture the increase in market power. We turn to the micro dimension next.

3.2 Decomposition: Markups and Firm Size

The construction of our measure of markup uses weights given by the sales share of the firm in the economy.[13] When we compare our measure with the unweighted average of markup

[13]The pattern of aggregate markup depends on the aggregation weight used. The share of Sales is the most common, see for example the HHI index. So far we made no assumption on demand and hence there is no welfare measure that guides us which aggregation weight to use. In the Appendix we report aggregate markups for alternative aggregation weights employment and the value of variable inputs. We also report the joint distribution of the individual markup and the corresponding weight variable. The pattern of the increase in markups starting in

in Figure 2a, we observe that the unweighted average is even higher and also increasing more than our benchmark measure: the unweighted markup goes from 1.4 in 1980 to 2.3 in 2014 (and was fairly constant throughout the 1960s and 1970s), compared to an increase from 1.18 to 1.67 for the unweighted measure. The fact that the unweighted measure is always higher indicates that larger firms (as measured by their sales) tend to have a lower markup. This appears at first to contradict the prediction from models of imperfect competition where firms with larger market shares have higher markups. Of course, the previous results hold across all firms across all sectors, while within narrowly defined industries, however, we confirm the positive relationship between markups and firm size.



(a) Unweighted versus Weighted Markup.

(b) Industry disaggregation

Figure 2: Decomposition by market share.

We seek to decompose the weighted mean to get at the source of this discrepancy in the level with the unweighted measure and to get at what has changed over time that can explain the rise in markup. If in the total population, large firms have smaller markups, is this also the case at more disaggregated levels of industry? Is the change over time due to an overall increase in markup (the markup tide has been lifting across all industries) or is there a change in the composition of firms of different sizes?

We therefore investigate whether this negative relation between markup and sales is due to the composition of industry or due to the negative relation between size and markup, and use a decomposition method as in Olley and Pakes (1996).

THE CROSS-SECTIONAL DECOMPOSITION. Denote the share s_{it} of firm i's sales $S_{it} = P_{it}Q_{it}$ in the total sales S_t in the economy (the universe of firms in our sample) by $s_{it} = \frac{S_{it}}{S_t}$. Then we can write the weighted markup, denoted by M_t, as

$$\begin{aligned} M_t &= \sum_i s_{it}\mu_{it} \\ &= \overline{\mu}_t + \sum_i (s_{it} - \overline{s}_t)(\mu_{it} - \overline{\mu}_t) \end{aligned} \tag{10}$$

1980 is robust across the three weighting measures used: sales, employment and the value of variable inputs.

where $\overline{\mu}_t$ is the unweighted average of markups. In Figure 2a, M_t is the red line, $\overline{\mu}_t$ is the black line and the difference is given by N times the covariance between s_{it} and μ_{it}. The covariance is negative and increasing over time, indicating that across the full sample, larger firms have smaller markups.[14]

To evaluate whether markups have gone up within an industry, we make the following decomposition:

$$M_t = \sum_i \left[\frac{S_{it}}{S_{st}}\frac{S_{st}}{S_t}\mu_{it}\right] = \sum_s s_{st}\left[\sum_{i\in s}\frac{S_{it}}{S_{st}}\mu_{it}\right] = \sum_s s_{st}M_{st}, \tag{11}$$

where $S_{st} = \sum_{i\in s} S_{it}$, $s_{st} = \frac{S_{st}}{S_t}$ and M_{st} is the weighted markup within industry s – i.e., $M_{st} = \sum_{i\in s} s_{it}\mu_{it}$, then M_t can further be decomposed as:

$$M_t = \overline{M}_t + \sum_s (s_{st} - \overline{s}_{st})(M_{st} - \overline{M}_t). \tag{12}$$

where $\overline{M}_t = S_t^{-1}\sum_s M_{st}$ is the unweighted average of (weighted) markup across industries, and the last term is the covariance between market share and markups at the industry level.

In Figure 2b we plot, together with the benchmark measure (in red), the term $\overline{M}_t$ (in blue). This term is hovering around 1.35 in between 1960 and 1980 and then increases to 1.50 in 2014. This indicates that average markups across industries have increased since the 1980s, but they account for less than the increase of individual level average markup. The difference between the two measures is proportional to the covariance between the industry share of sales in the entire sample and the average weighted markup in that industry.

In 1980 that covariance term is negative (the red line is below the blue line) which indicates that markups are higher in smaller industries. By 2014 this has changed: the term proportional to the covariance is positive (red is below blue), which means that markups are higher in large industries. This is evidence that there is a change in the composition across industries. Markups have gone up in all industries, but they have gone up more in large industries than they have in small industries.

In sum, economy-wide markups are higher in small firms across the entire economy, though not within narrowly defined industries where larger firms indeed have higher markups as predicted by most standard models of competition. This pattern persists in time. The negative correlation economy-wide indicates that the characteristics of industries vary substantially. Across industries, there is a changing pattern: industries with a larger share of sales in the entire sample used to have smaller markups until the 1980s, but this has reversed by 2014.

[14] In Figure B.8 in the Appendix, we show that especially towards the end to the time period, firms in our sample have become bigger. This is consistent with Autor, Dorn, Katz, Patterson, and Van Reenen (2017) who find that firms size has increased. It is also consistent with the fact that the amount of output in a given market decreases despite a larger firm size, since there are fewer of them (see Lemma 1 below). However, given the negative covariance between market share and markup and as is evident in Figure 2a above, we do not find that it is the large firms that have the largest market power in the cross section. To the contrary.

THE TIME-SERIES DECOMPOSITION. We now decompose the *change over time* in markup by firm size as measured by the share of sales. 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 reallocation). This can be expressed in the following formula:

$$\Delta M_t = \underbrace{\sum_s s_{s,t-1}\Delta M_{st}}_{\Delta \text{ within}} + \underbrace{\sum_s M_{s,t-1}\Delta s_{s,t}}_{\Delta \text{ between}} + \underbrace{\sum_s \Delta M_{s,t}\Delta s_{s,t}}_{\Delta \text{ reallocation}}. \tag{13}$$

We consider the change over 10 year periods starting in 1954. The decomposition for the 4-digit industry classification is reported in Table 1.[15] As we already observed in Figure 1, our baseline measure of markup has slightly decreased in the run up to the 1980s, and has since then increased at an increasing rate. The decomposition shows that since the 1980s, the change in markup is mainly driven by the change within industry. 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.

	Markup	Δ Markup	Δ Within	Δ Between	Δ Realloc.	Mean($\Delta\theta$)	Median($\Delta\theta$)
1964	1.319	0.135	0.067	-0.011	0.079	0.018	0.015
1974	1.231	-0.088	-0.084	0.042	-0.046	0.025	0.013
1984	1.236	0.004	-0.008	0.025	-0.012	0.005	0.001
1994	1.360	0.124	0.126	0.004	-0.007	0.009	0.002
2004	1.519	0.159	0.116	0.031	0.012	0.015	0.005
2014	1.667	0.151	0.187	-0.018	-0.020	0.000	0.000

Table 1: Decomposition of 10 year change in Markup and Output Elasticity

The estimate of firm-level markups is a product of the output elasticity (θ^V_{it}) and the sales-to-variable input expenditure. This allows us to isolate a technology component, given the specified production function in equation (1). The last two columns of Table 1 present the mean and median 10-year change in the output elasticity, using the translog production function. These changes are informative to evaluate the relative importance of the (unweighted) aggregate markup rise, as presented in Figure 2a, and it suggests that during the periods of the strongest markup rise, both median and average output elasticities changed very little.

3.3 The Dispersion of Mark Ups

The evolution of markups is very different for different moments of the distribution. To get an idea for what is going on in the micro data, we plot different moments of the distribution of markups over time (Figure 3). We rank the firms by markup, and to obtain the percentiles

[15]The results for the 2 and 3-digit industry classifications are similar. Those are reported in Table B.1 in the Appendix.

we weigh each firm by its market share. This makes the percentiles directly comparable to our share weighted average.

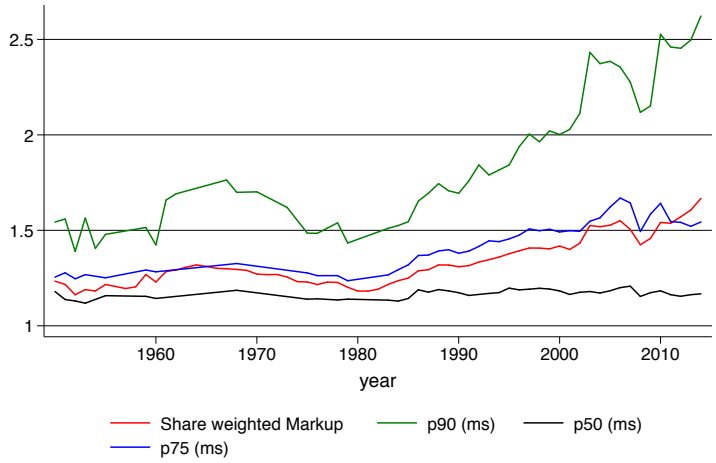


Figure 3: The Evolution of the Distribution (Percentiles) of Markups (1950 - 2014). (The percentiles of the Markup distribution are weighted by marketshare of sales in the sample.)

The increase in the average markup comes entirely from the firms with markups in the top half of the markup distribution. The median and lower percentiles are largely invariant over time. For percentiles above the median, markups increase. Between 1980 and 2014, the 75th percentile increases from 1.3 to 1.52, and the 90th percentile increases from 1.46 to 2.6.

3.4 The Process of Markups

Inspecting the time-series pattern of output, input and markup is directly relevant for studying the extent to which market power has changed over time. Figure 4 shows the evolution of the cross-sectional standard deviation of the shocks in the markup, sales and employment process, assuming this process is autoregressive:

$$x_{it} = \rho x_{it-1} + \varepsilon_{it}, \quad x \in \{\log \mu, \log S, \log L\}. \tag{14}$$

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. The increase in the standard deviation in markup 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).[16]

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

[16]We have also included higher order terms of the persistence and find that those are not important.

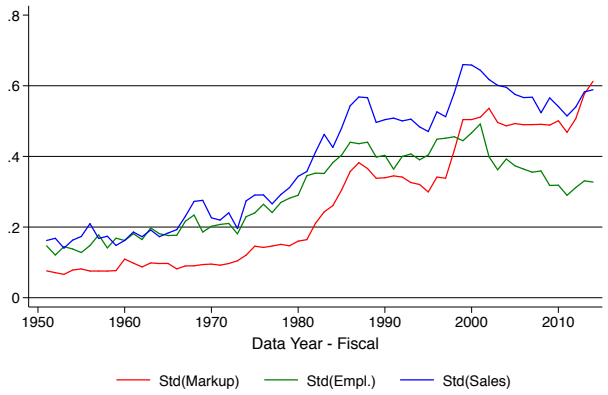


Figure 4: The Evolution of the Standard Deviation of Markups, Sales and Employment (1950 - 2014). (AR(1) in logs on their lag with year and industry fixed effects)

of the shocks to inputs (labor) has. We revisit this lowering of the response of employment to shocks in Section 4, and relate this to the fundamental change in pass-through of shocks to factor demand, and output, consistent with an increase in overall market power.

3.5 Do Markups imply Market Power?

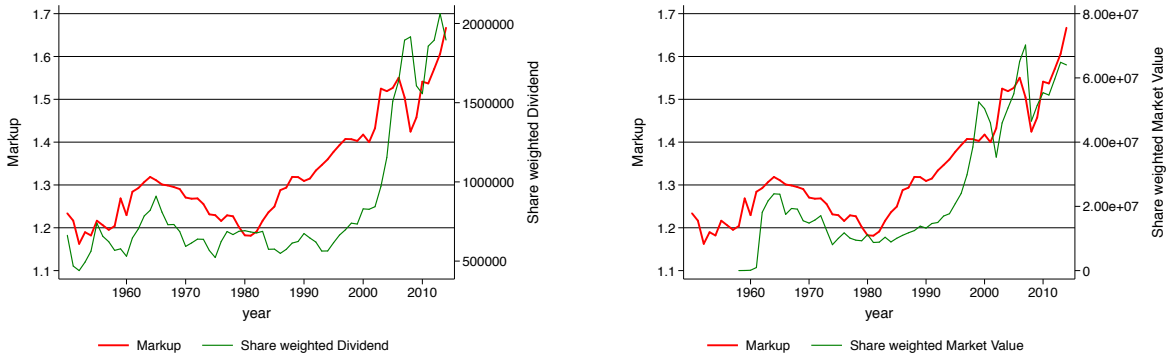
Markups tell us that the margin of revenue over variable costs has increased. That does not necessarily imply that firms are making higher profits. If for example the source of the increase in markups is technological change that reduces variable costs, and the same technological change increases the fixed costs. Consider for example high tech firms that produce software products that need one big up front investment and can be scaled nearly without any additional cost. Such technological change will lead to higher markups (due to lower variable costs), but prices will not drop because firms need to generate revenue to cover fixed costs. As a result, profits will continue to be low and higher markups do not imply higher market power.

So the question is whether higher markups simultaneously lead to higher profits.[17] In order to address this issue, we need to have a measure of profits. Because the measure for profits in Compustat is based on the cost of goods sold, this is not informative. The real value of profits corresponds to what firms pay their shareholders. For that we have two measures: 1. dividends and 2. the market value (or market capitalization). Dividends are 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

[17]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.

profits. The second 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 not necessarily a good measure of contemporaneous profits since it takes into account all future dividends. In addition, it is highly susceptible to changes in expectations about future profitability of the firm. Nonetheless, in a cross section over a large number of firms, this should give us some indication of the profitability of the firm.

Figure 5a clearly illustrates that the evolution of the weighted average of dividends closely tracks that of markups, with sharp reversal in the trend around 1980. The same is true for market value (Figure 5b).



(a) Average Dividends (weighted)

(b) Average Market Value (weighted)

Figure 5: The Evolution of Average Dividends and Average Market Value (as a share of Sales) and Markup (1960 - 2014).

This is not just an artifact of the aggregate data. At the individual firm level, firms with higher markups also have higher dividend margins. Figure B.7 in the Appendix shows that relation. For example, as the dividend margin increases from 0.2 to 0.7, the markup increases from 2 to 3. This establishes that at the individual firm level, higher markups are not exclusively driven by technological change in a competitive market (basically higher fixed costs and/or lower variable costs). Higher markups result in higher dividends and therefore higher profits.

The variable profit rate (variable profits divided by sales) is directly related to the markup $\mu = \frac{P}{c}$, where c is the marginal cost. This follows from the definition of the profit rate:

$$\frac{\Pi^V}{PQ} = \frac{(P-c)Q}{PQ} = 1 - \frac{1}{\mu}, \tag{15}$$

where we assume constant marginal costs and we do not subtract investment in capital. That is why this is a measure of variable profits.[18] In an attempt to obtain external validation, we compare the evolution of this measure of variable profits to the evolution of total profits. We have the data for the evolution of total profits for the economy as a whole from the national

[18]Of course, this is a poor measure of total profits since firms also have capital investment.

accounts data (which includes all firms, not just the publicly traded firms), depicted in Figure 6, below.

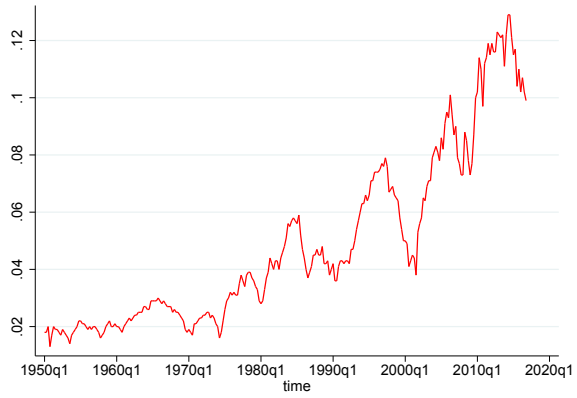


Figure 6: Profit rates. Data from FRED, based on national accounts. Quarterly.

We compare the change in the variable profit rate using our measure of markup in equation (15) to the change in the aggregate profit rate. Since the national profit rate is calculated as a share of GDP and our measure is calculated as a share of total sales, we need to adjust our measure for the variable profit rate for Gross National Income (GNI):

$$\frac{\Pi^V}{GDP} = \frac{\Pi^V}{PQ}\frac{GNI}{GDP} = \left(1 - \frac{1}{\mu}\right)\frac{GNI}{GDP}. \tag{16}$$

Using equation (16), and the fact that $\mu_{1980} = 1.18$ and $\mu_{2014} = 1.67$, we obtain an increase in the profit share of GDP by a factor 2.34:[19]

$$\frac{\frac{\Pi^V_{2014}}{GDP_{2014}}}{\frac{\Pi^V_{1980}}{GDP_{1980}}} = 2.34. \tag{17}$$

The economy-wide profit rate relative to GDP, calculated from the national accounts, has increased fourfold between 1980 and 2014. This is an even bigger increase than the 2.34-fold increase that we obtain from our markup measure. Of course, the two measures are not exactly comparable. First, ours is a measure of variable profits and the national accounts measure is one of total profits which includes capital investments and fixed costs. This seems to indicate that capital investment has increased less than variable costs. Second, the sample of firms differs. Since ours is for publicly traded firms only, this may indicate that markups are even higher for those firms that are not publicly traded. The important conclusion is that there is an enormous increase in profits however it is measured, which is consistent with the increase in market power.

[19]This ratio of GNI/GDP is set at 0.502 in 1980 and 0.565 in 2014, as reported in the national accounts.

4 Macroeconomic Implications

We analyze the macroeconomic implications of the rise in market power. We establish 7 implications that follow from the observed increase in market power. For some implications, this involves further analyzing the data that we have used to calculate the markup, and for other implications we develop a simple static model of industry competition, to illustrate how market power can generate the observed outcomes. In particular we will abstract from some of the heterogeneity that we see in the data.

Of course, given the simplicity of the model and the absence of an estimation, we do not pretend to evaluate the quantitative importance of market power for these implications, nor do we claim that market power is the only force that contributes to these outcomes. We simply take the observed markup trajectory as given and provide a simple framework to interpret various aggregate labor and product market outcomes, without invoking any causal link between market power and the outcomes of interest. The latter lies beyond the scope of the current paper.

Implication 1. 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. 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).[20] The decline since the 1980s occurs in the large majority of industries and across countries (see Karabarbounis and Neiman (2014) and Gollin (2002)).

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

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

[20] 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.

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

property products and this leads to a lower expenditure on labor.[22] 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, the fourfold increase in the total profit rate indicates that if intangibles play a role, it must allow firms to exert more market power, which is the central thesis of our paper.[23] Finally, Elsby, Hobijn, and Şahin (2013) find little support for capital-labor substitution, nor for the role of a decline in unionization. They do find some support for off-shoring labor-intensive work as a potential explanation.

In the context of our setup, the change in the markup has an immediate implication for the labor share. While we have calculated the markup from all variable inputs, we could do so as well for labor alone. Then rewriting the First Order Condition (5), we obtain that at the firm level, the labor share $\frac{wL}{PQ}$ satisfies

$$\frac{wL}{PQ} = \frac{\theta_L}{\mu}, \tag{18}$$

where θ_L is the output elasticity of labor. Profit maximization by individual firms implies an inversely proportional change in the labor share. As markup increases and provided the technology parameter θ remains unchanged over time, we expect to see an decrease in the labor share.

Unfortunately Compustat does not have good data for the wage bill.[24] We can therefore not *directly* verify condition (18) at the firm level.[25] Instead, we rely on aggregate data from the BLS. They report total compensation of employees (expenditure on wages and salaries) as a share of gross domestic income. We compare this BLS measure to a population average of the right hand side of (18), where we assume that the technology parameter has remained unchanged. We use as a measure for markups the one that we calculated above based on total variable costs (COGS) and not the cost of labor because we have no reliable data on the wage bill. We plot this in Figure 7 where we normalize these measures to 100 at the start of our data in 1950.

Our main finding is that the labor share (in green) tracks the inverse markup quite closely, especially since 1980. It seems that the labor share is decreasing somewhat more slowly than the inverse of the average markup, but overall the trend is similar. The fact that this aggregate measure of labor share tracks the inverse of markups is quite remarkable since we are implicitly making the following assumptions, due to absence of good firm-level wage data. We

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

[23] For valuation purposes, intangible assets are valued using different methods including the market approach (how much is another party willing to pay to obtain the rights to those assets), the income approach (the stream of income the assets generate) and the cost approach (how much it would cost to replace the existing asset). If the company has laid out expenses in the past (labor, advertising, R&D,...) to generate the intangibles then they are already measured either in the variable inputs or in the capital inputs. If Coca Cola has invested in building its brand or a pharmaceutical in a drug, then those expenditures are included in variable inputs or capital at some point. If the current value of the brand or the drug is higher than that intangible, then the value assigned to intangibles includes the expected stream of future profits net of the invested capital. Intangibles are therefore an alternative measure of the discounted stream of profits.

[24] There is good data on employment (EMP) but the data on the average compensation (XLR) in the firm is heavily underreported.

[25] Although with additional assumptions on the structure of wages a firm level analysis is feasible.

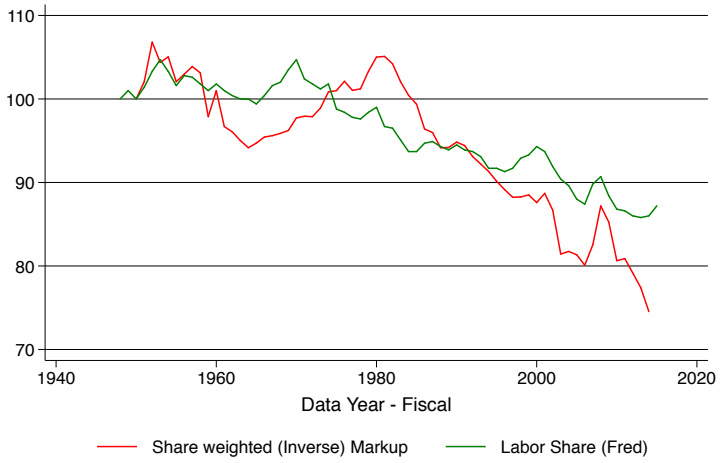


Figure 7: The Evolution of the labor share (BLS), and inverse of the markup (1960-2014) Notes: Labor Share data from BLS. Share of gross domestic income: Compensation of employees, paid: Wages and salary accruals. Disbursements, to persons. 1950=100.

basically treat every worker to be identical in each firm since we only have one markup measure per firm and our model assumes that labor adjusts only in the number workers, not in the composition. Moreover, the markup μ_i in each firm is calculated based on all variable inputs (labor L and material inputs M), whereas the first order condition (18) is supposed to hold for a markup calculated based on labor alone. The association between the declining labor share and the rising (aggregate) market power is robust to the specific weights used to construct the aggregate markup index; Figure B.4 in the Appendix presents the share-weighted markup using alternative weights.

Implication 2. The Secular Decline in the Capital Share

The same logic for the decline in the labor share also applies to materials M, i.e. variable inputs that are used in production. Those are included in our variable cost measure COGS. Now if we consider the evolution of capital investment, which is not included our measure of variable cost and which adjusts at a lower and more long run frequency, then in the increase in markup has implications for the capital share.[26] While the decline in the labor share is widely discussed, the decline in the capital share has received much less attention.[27]

Using a simple accounting rule, we can write the firm's sales as: $PQ = P^V V + rK + \Pi$ where r is the user cost of fixed capital and Π is total profits. Even if capital does not adjust at

[26] 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, adjust at a frequency higher than one year, our unit of time, and capital does not. This assumption allows us to calculate the markup.

[27] 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.

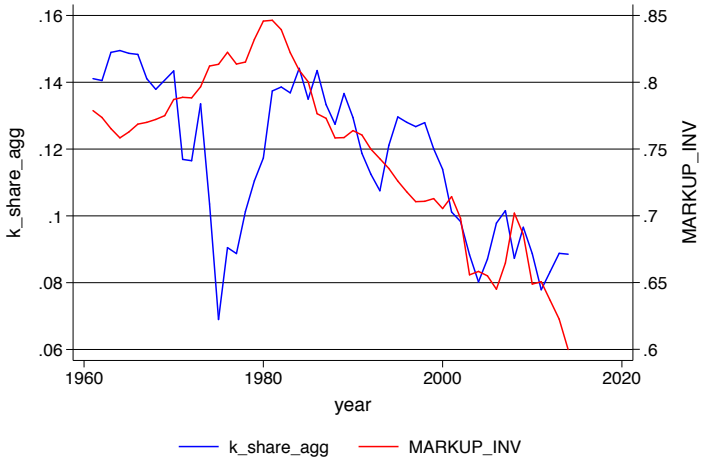


Figure 8: The Evolution of the capital share (our computations), and inverse of the markup (weighted by Sales Share) (1980 - 2014). Notes: Capital Share data, own calculations (from Compustat). Gross Capital (PPEGT) adjusted by the input price deflator (PIRIC from FRED), the federal funds rate, and an exogenous depreciation rate of 12%. 1980=100.

a yearly frequency, on average over a long enough time horizon, the user cost of capital rK as a share of output will be evolving. If market power and profits go up, firms not only lower the labor share, over a long enough time horizon they will eventually also lower the capital share $\frac{rK}{PQ}$. Whenever capital can be adjusted – possibly only over a horizon of several years –, it will be done so satisfying the first order condition $\frac{rK}{PQ} = \frac{\theta_K}{\mu}$. Then the accounting identity can be written as:

$$\frac{P^V V}{PQ} + \frac{rK}{PQ} + \frac{\Pi}{PQ} = 1 \tag{19}$$

$$(\theta_V + \theta_K)\frac{1}{\mu} = 1 - \pi, \tag{20}$$

where we collapse all variable inputs into V and given our Cobb-Douglas technology, θ_V and θ_K are the cost shares of expenditures and $\theta_V + \theta_K = 1$, and where $\pi = \frac{\Pi}{PQ}$ is the total profit rate.[28]

To get an idea of how the capital share has evolved, we construct from our data a measure of capital. We us a measure for Gross Capital that we adjust for the industry-level input price deflator, for the federal funds rate and for an exogenous depreciation rate of 12%. This measure of capital is divided by sales to obtain the capital share. Figure 8 shows the plot of the capital share. 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

[28]The notation is used to distinguish the measure of the total profit rate, π, from the variable profit rate π^V used above. Equation (20) represents the long run relationship where all factors are variable.

times: inflation was high and financial frictions were considered higher. What we learn from the figure is that since 1980, the capital share has moved in lockstep with the inverse of our markup measure. With a long enough horizon, also capital investment adjusts to reflect the same first order condition as labor, and hence there will be a reduction in capital investment as markups increase.

Implication 3. The Secular Decline of Low Skill Wages

For the next four implications – the decline in low skill wages, the decline in labor force participation as well as the decrease in labor and migration flows – we further elaborate on the model that we have used to estimate markups. Or put differently the estimation of markups, and the associated key facts and implications, only rest on cost minimization, and the addition of assumptions of market structure, conduct and demand, preserves this.

Let there be a measure 1 of markets, each with N firms. Each market is indexed by Ω, which also indicates the level of productivity of each firm in the market. Within each market, firms are equally productive. Across the economy, the distribution of Ω is $F(\Omega)$. In the whole economy, there is a measure M of workers who supply unskilled labor. Workers are indexed by skill z expressed in efficiency units and distributed according to $G(z)$. All workers have a common outside option U. The labor market is competitive and the equilibrium wage is w.

Due to the finite number of firms in a given market, each firm has some market power, and in line with Bresnahan (1982), we express the marginal cost pricing behavior in terms of the "conduct parameter".[29] Given a marginal cost c, the pricing decision in market Ω is:

$$P(Q) = c + \lambda h(Q) \tag{21}$$

where Q_i is firm i's output, $Q = \sum_i Q_i$ is the market output, and $h(Q) = -\frac{\partial P(Q)}{\partial Q} Q$ captures some of the features of the demand elasticity. In the case of linear demand $P(Q) = a - bQ$ for example, $h(Q) = bQ$. The term λ is the *conduct parameter* and measures the market structure. In a Cournot model for example, λ is exactly equal to the inverse of the number of firms N in the market. Under perfect competition as N goes to infinity, λ goes to zero, whereas under monopoly with $N = 1$, $\lambda = 1$. It could also measure the extent to which customers are captive, as in Burdett and Judd (1983) for example.

Our measure for markup $\mu = \frac{P}{c}$ can then be written as:

$$\mu = 1 + \lambda \frac{h(Q)}{c}. \tag{22}$$

This makes very transparent the reason why markups can go up. First, if a firm increases its productivity (decreases it's marginal cost c) and other firms do not, then markups increase due to higher efficiency. This cannot be a long run outcome in a competitive market since other firms will adopt that technology. Second, if the elasticity of demand decreases, expressed by

[29] See also Bresnahan (1989) for an overview and Genesove and Mullin (1998) for the estimation of the conduct parameter applied to a particular industry.

an increase in $h(Q)$. And third, if the conduct parameter λ increases, the firm's market power increases and hence so does the markup.

While each of these three reasons can be behind the increase in markup, for the remainder we will focus on a change in λ to reflect the observed increase in μ. A decrease in c of any given firm eventually should lead to a decrease of other firms as well, either because they adopt the new, low cost technology as well, or because they are driven out of the market. And of course, preferences (and therefore h) can and do change, but we find it hard to imagine preferences to change so dramatically to match the increase in market power over such a period of a few decades.

The firms objective is to choose the amount of labor L_i in order to maximize profits:

$$\max_{L_i} P(Q)Q_i - wL_i, \tag{23}$$

where $Q_i = \Omega_i L_i^\theta$. Then the First-order Condition reflects the market structure, and consistent with the conduct rule it satisfies $P(Q) - \lambda h(Q) = c$ where $c = w\frac{\partial L_i}{\partial Q_i} = \frac{w}{\Omega\theta}L_i^{1-\theta}$ is the marginal cost. Then

$$P = \frac{w}{\Omega\theta}L_i^{1-\theta} + \lambda h(Q), \tag{24}$$

and where $h(Q) = -\frac{\partial P}{\partial Q}Q$. In the case of Cournot with linear demand $P = a - bQ$ and with $\frac{1}{\lambda}$ identical firms, the FOC (24) can be written as:

$$a - (1+\lambda)bQ = \frac{w}{\Omega\theta}L_i^{1-\theta}. \tag{25}$$

Consider the case where $\theta = 1$. Then $Q_i = \Omega L_i$ and $Q = \frac{\Omega L_i}{\lambda}$, and the equilibrium condition is

$$a - (1+\lambda)b\Omega L = \frac{w}{\Omega} \;\Rightarrow\; L = \frac{a - \frac{w}{\Omega}}{(1+\lambda)b\Omega},\;\; L_i = \frac{\lambda}{1+\lambda}\frac{a - \frac{w}{\Omega}}{b\Omega} \tag{26}$$

It immediately follows that L is decreasing in λ and L_i is increasing in λ:

$$\frac{\partial L}{\partial \lambda} = \frac{a - \frac{w}{\Omega}}{b\Omega}\frac{-1}{(1+\lambda)^2} < 0 \;\text{ and }\; \frac{\partial L_i}{\partial \lambda} = \frac{a - \frac{w}{\Omega}}{b\Omega}\frac{1}{(1+\lambda)^2} > 0. \tag{27}$$

This establishes the following Lemma:

Lemma 1 *For a given wage and a given market Ω, then the market labor demand L is decreasing in market power λ; an individual firm's labor demand L_i is increasing in market power λ.*

This results states that labor demand shifts inwards as λ increases. Due to market power, total labor demand L is lower as firms restrict the quantity produced. However, the origin of the market power is the fact that there are fewer firms, $N = \frac{1}{\lambda}$. Therefore, while the market labor demand (and output) is smaller, each individual firm has a larger market share and produces more output and demands more labor.[30]

[30]Figure B.8 in the Appendix shows the evolution of the number of firms and their size for the Compustat sample. Even though the sample has shown a substantial increase, at least since the 1990s, there has been an increase in the average size of the firm and a decrease in the number of firms. Related, there has been an increase in the number, and value, of mergers and acquisition, which is of course related to the consolidation of corporate ownership. Figure B.9 plots the number and value of mergers and acquisitions in the US.

Total labor demand L^D is given by the labor demand in each market Ω:

$$L^D(w;\lambda) = \int_{\underline{\Omega}}^{\overline{\Omega}} L(\Omega;w;\lambda)dF(\Omega) = \frac{1}{(1+\lambda)b}\int_{\underline{\Omega}}^{\overline{\Omega}} \left(\frac{a}{\Omega} - \frac{w}{\Omega^2}\right) dF(\Omega), \tag{28}$$

where labor demand is downward sloping $\frac{\partial L^D}{\partial w} < 0$. It follows immediately from Lemma 1 that an increase in λ leads to an inward shift of the labor demand L^D.

We now turn to labor supply. Because the outside option is fixed at U, any worker of skill z will want to work as long as $zw > U$. Therefore, the marginal low skilled worker $z^\star$ is indifferent between work and remaining out of the labor force: $z^\star = \frac{U}{w}$. The total labor supply $L^S(w)$ is then

$$L^S = \int_{\frac{U}{w}}^{1} z dG(z) \overset{G(z)=z}{=} \frac{1-\left(\frac{U}{w}\right)^2}{2}, \tag{29}$$

where the last equality follows for the case where G is uniform. Observe that labor supply is upward slowing in w as long as G is non-degenerate:

$$\frac{\partial L^S}{\partial w} = g\left(\frac{U}{w}\right)\frac{U^2}{w^3} \geq 0, \tag{30}$$

where $g(z)$ is the density of $G(z)$.

Equilibrium in this competitive labor market equates labor supply and labor demand:

$$\int_{\underline{\Omega}}^{\overline{\Omega}} L(\Omega)dF(\Omega) = \int_{\frac{U}{w}}^{1} z dG(z) \tag{31}$$

With upwards sloping labor supply, and a downward shift in the labor demand as λ increases, we can establish the following result.

Proposition 1 *Consider an economy with constant returns to scale. Then equilibrium (nominal and real) wages $w^\star$ and equilibrium labor $L^\star$ are decreasing in market power λ.*

We illustrate this result graphically in Figure 9 for the case with $\lambda = 0$ and $\lambda = 1$. The general case is reported in the Appendix.

Observe that nominal wage $w^\star$ are lower for higher market power λ. But real wages are even lower since real wages are given by $\frac{w^\star}{P^\star}$. Under perfect competition real wages are 1 ($w^\star = P^\star$) whereas under market power $\frac{w^\star}{P^\star} < 1$. Even under perfectly elastic labor supply where nominal wages are constant, real wages are decreasing in market power. To see this, consider the case of perfectly elastic labor supply. Then $w = U$, and with $\theta = 1$ and $\Omega = 1$, real wages are given by:

$$\frac{w}{P} = \frac{U}{\frac{1}{1+\lambda}\left(\frac{U}{\Omega} + \lambda a\right)}. \tag{32}$$

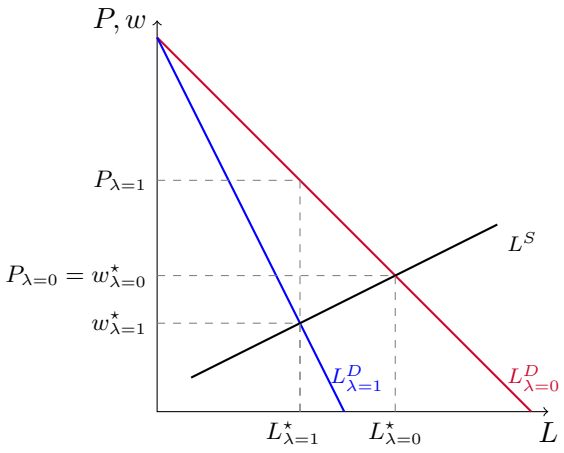


Figure 9: Labor Demand, Labor Supply and Equilibrium for $\lambda = 0$ (perfect competition) and $\lambda = 1$ (monopoly). Parameters $\Omega = 1, \theta = 1$ so that $Q = L$.

Trivially, real wages are decreasing in market power λ. The wage is constant and prices are increasing:

$$\frac{\partial}{\partial \lambda} \frac{w}{P} = \frac{U\left(\frac{U}{\Omega} - a\right)}{\frac{U}{\Omega} + \lambda a} < 0. \tag{33}$$

We can now relate these two theoretical findings to the empirical evidence of the following two implications of the rise in market power. There is ample evidence of low wage stagnation in the last few decades that is consistent with the finding in Result 1. Figure 10a plots the weekly median wage, in constant 1982 prices. There has been little or no change in the median wage since the 1980s. In the presence of economic growth, that means that the share of median wages of GDP per capita has decreased. Figure 10b plots the ratio of median wages to GDP per capita. There has been a secular decline in this ratio from 1.3 in 1980 to 0.75 now.

This is not simply a artifact of how the distribution of earnings is spread over the life cycle.[31]

Implication 4. The Secular Decline in Labor Force Participation

Figure 11 plots the labor force participation rate for the US economy since 1950. There is a sharp decline since the mid 1990s. It appears that this does not fully coincide with the the timing of the rise in market power. However, the sharp increase in the labor force since the 1960s is driven by the increased female participation. This has led to an outward shift of the labor supply. This trend has leveled off in the mid 1990s, which is consistent with the fact that we start to see the impact of a decrease in the total labor force participation due to the rise in market power.

[31] Say for example that due to fact that the age-earnings profile has become steeper and with more variance, the median wage is lower. Guvenen, Kaplan, Song, and Weidner (2017) show that the same is true for life time earnings.



(a) Median Wage (weekly, 1982 prices)

(b) Ratio Median Wage to GDP per Capita.

Figure 10: The Evolution of Median Wages (1980 - 2016). Data from FRED, CPS.

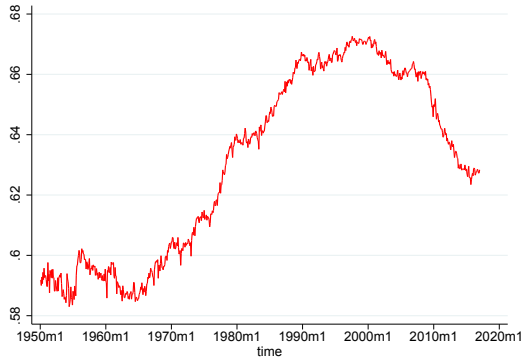


Figure 11: Labor Force Participation (all) (1950 - 2016). Data from FRED, CPS.

Implication 5. The Secular Decline in Labor Market Flows

In order to address further implications, we now extend the model to allow for shocks to the firm's productivity. We will make the minimal assumptions in order to illustrate this theoretically. Let the market productivity take two types $\underline{\Omega}$ and $\overline{\Omega}$. Let there be i.i.d. shocks to a market's productivity where the symmetric Markov transition matrix takes the values $1-p$ on the diagonal and p off the diagonal. That implies that a firm (market) with productivity Ω_k switches to $\Omega_{l\neq k}$ with probability p. There are no frictions and because shocks are i.i.d., there is no aggregate uncertainty.

For a given λ and w and with $\theta = 1$, labor demand in this economy is given by $L^D =$

$\frac{1}{2}\overline{L} + \frac{1}{2}\underline{L}$. In every period, the amount of labor reallocation ΔL is[32]

$$\Delta L = p\left|\frac{1}{2}\overline{L} - \frac{1}{2}\underline{L}\right| = p\frac{1}{2(1+\lambda)b}\left|\frac{a}{\overline{\Omega}} - \frac{w}{\overline{\Omega}^2} - \frac{a}{\underline{\Omega}} + \frac{w}{\underline{\Omega}^2}\right|. \tag{34}$$

It follows immediately that labor reallocation is decreasing in λ:

$$\frac{\partial \Delta L}{\partial \lambda} = -p\frac{1}{2(1+\lambda)^2 b}\left|\frac{a}{\overline{\Omega}} - \frac{w}{\overline{\Omega}^2} - \frac{a}{\underline{\Omega}} + \frac{w}{\underline{\Omega}^2}\right| < 0. \tag{35}$$

This is true for constant w (elastic labor supply). Given linear demand, this result holds more generally if also supply is linear (or concave), and under those conditions we can establish the following result.

Proposition 2 *Consider an economy with constant returns to scale and linear demand for output and linear labor supply. Higher market power leads to a lower responsiveness of labor inputs.*

This is illustrated in Figure 12. High productivity firms $\overline{\Omega}$ face a lower marginal cost than low productivity firms $\underline{\Omega}$. Under perfect competition, the total industry response is along the demand curve and results in a change $\Delta L_{\lambda=0}$. Instead, under monopoly, the response $\Delta L_{\lambda=1}$ is along the marginal revenue curve, the slope of which is steeper by factor two. As a result, the same vertical shift in marginal cost results in a smaller change in labor.

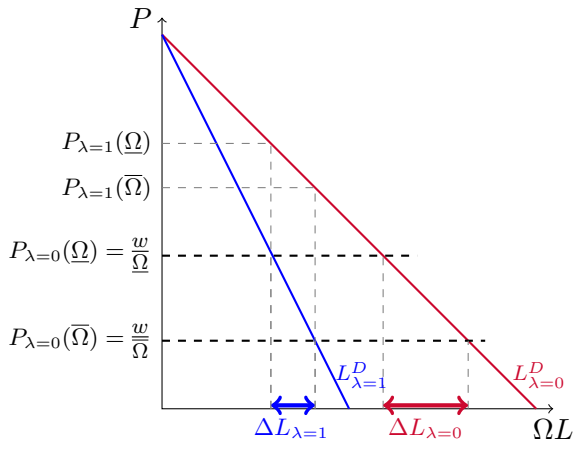


Figure 12: Labor Reallocation: Perfect Competition ($\lambda = 0$) and Monopoly ($\lambda = 1$).

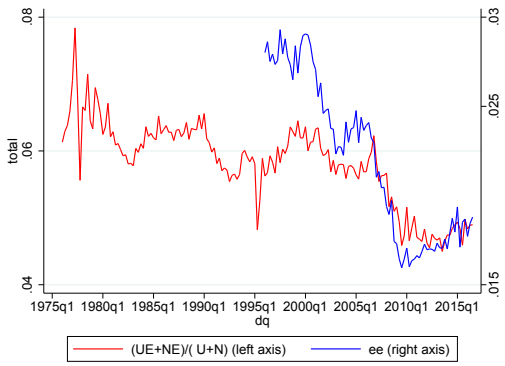
The magnitude of labor reallocation in response to a change in market power depends on the elasticity of the (residual) demand curve. This argument is similar to the one in the debate on pass-through: the extent to which firms with market power passes on a change in the marginal cost to the consumer. It is well-known that with linear demand, there is imperfect

[32] As is well-known, output and therefore labor demand is not necessarily monotonic in productivity Ω and therefore there are parameter values such that $\overline{L} < \underline{L}$. This however is irrelevant for the reallocation argument we make here, where only the absolute value of the difference matters.

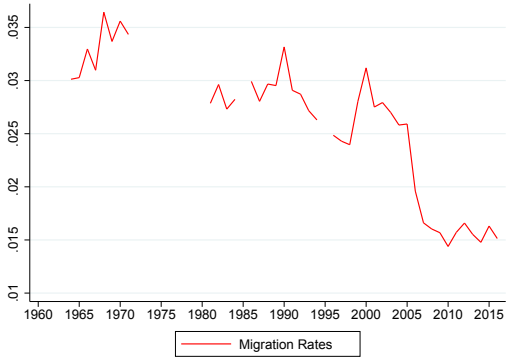
passthrough (see for example Melitz and Ottaviano (2008)), but not with constant elasticity demand.[33] Eventually it is an empirical question what the elasticity of demand and supply is, and therefore whether and how much pass through there is. There is quite a bit of evidence of incomplete passthrough, most of which comes from the effect of changes in the exchange rate or reduction in tariffs, see Campa and Goldberg (2005) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016).

We have treated wages w as exogenous, which is equivalent to perfectly elastic labor supply. Whenever labor supply is upward sloping (as in Figure 9), a change in λ also affects wages. It is easily verified that an increase in market power λ still leads to a decrease in labor reallocation as long as labor supply is not too convex.[34]

The decrease in labor reallocation as a result of the increase in market power (Result 2) forms the basis for two macroeconomic implications. There has been a secular decline in business dynamics broadly defined. This is manifest in several observed outcomes, most notably in the evolution of labor market flows and in the evolution of migration rates. The likelihood of switching jobs and the flows in and out of unemployment have dropped (and as a result the duration of employment has increased).



(a) Labor Market Flows (1976-2016).



(b) Migration, between MSA's (1962 - 2016).

Figure 13: Labor Reallocation (1976-2016). CPS data.

Figure 13a shows the evolution of labor market flows based on CPS data. The sharpest decline is in the Employment-to-Employment (EE) flows, though we only have data on EE flows since 1996. The probability for an employed worker to switch to another job has gone down from 2.9% to 1.8%. There is also a decrease in the flow rate of the unemployed and those not in the labor force from 6.5% in 1980 to 4.7% now.

Our model in conjunction with the rise in market power provides a mechanism for the

[33] Even with CES demand, passthrough can be incomplete depending on the preference aggregation. Yeh (2017) for example analyzes the responsiveness of large firms to i.i.d. shocks and he uses a Kimball aggregator of the CES demand.

[34] Or when wages are perfectly inelastic. In that case there is zero labor reallocation.

decline in job flows.[35] The simplified static model is meant to illustrate the role of market power in a setting with full blown firm dynamics as in Jovanovic (1982) and Hopenhayn (1992). Firms that receive positive productivity shocks augment output by augmenting inputs such as their labor force, and they reduce output by reducing inputs when a negative shock hits.[36]

An obvious reason why job flows could have decreased is that the volatility of productivity shocks (p in our stylized model) has decreased over time. Then even with constant market power, we would see lower job flows. However, as Decker, Haltiwanger, Jarmin, and Miranda (2014) find for the US economy, the volatility of shocks has not decreased in the last few decades. If anything, it has increased. But as they point out, it is not the volatility of the shocks that has decreased, but rather the *responsiveness* of firm's output and labor force decisions to the existing shocks.[37]

If the distribution of idiosyncratic shocks remains unchanged, then the responsiveness of the labor input and therefore the transmission of shocks will lead to a decrease in job flows as market power increases. This can illustrate the pattern of decreased job flows observed in the US labor market and plotted in Figure 13a.

Observe that this is consistent with what we find on the evolution of the cross-sectional standard deviation of the noise term reported in Figure 4. While there is an increase in the standard deviation of sales, the standard deviation of employment is much more moderate, and even declines while that of sales increases. This confirms the finding by Decker, Haltiwanger, Jarmin, and Miranda (2014) that the transmission of the shocks to labor has become more attenuated over time.

Implication 6. The Secular Decline in Migration Rates

Finally, for the same reason as the decrease in the job flow rates, the transmission of shocks has implications on the job relocation choices across different local labor markets. There has been a marked decrease in interstate migration as well as the migration between metropolitan areas. Figure 13b shows that relocation rate has nearly halved, from around 3% in 1980 to 1.5% now.[38] If firms are based in different local labor markets and a fixed fraction of all job relocation decisions are between local labor markets, then lower job flow rates will automatically give rise to lower migration rates.

[35]There are several potential alternative explanations for the decline in job flows: demographic change (aging workforce, Fallick, Fleischman, and Pingle (2010)), 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.

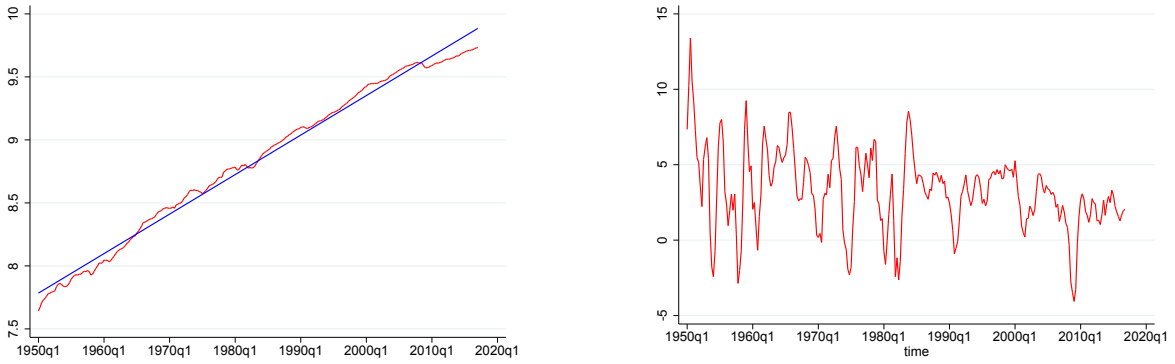
[36]Recent work by Baqaee and Farhi (2017) draws attention to the fact that firm-level productivity shocks can give rise to a nonlinear impact on macroeconomic outcomes. For example, models with network linkages such as Gabaix (2011) give rise to such non-linearities. So does market power in the presence of incomplete passthrough as we show in this section.

[37]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.

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

Implication 7. The Slowdown in Aggregate Output

In recent decades, people have pointed out a slowdown in GDP. As can be seen from Figure 14, GDP growth since the great recession has not recovered and it appears the economy is on a path below trend. That decrease did not necessarily coincide with the rise of market power in 1980.[39] Growth was high during much of the 1980s and 1990s. But the slowdown in GDP does coincide with the sharper increase of market power after the great recession.



(a) GDP with linear trend (log scale)

(b) GDP growth. Percent Change Last Year.

Figure 14: Growth of Real GDP per capita (1950 - 2016). Data from FRED, Quarterly.

The question is what the cause is of this slowdown. There are advocates who argue that the slowdown in aggregate output is due to a slowdown in factor augmenting productivity, TFP (see for example Gordon (2016)). The productivity of workers and firms is growing at slower rates because there is less innovation. This is often referred to as the "productivity paradox" of information systems.[40] Despite the enormous technological advances in IT, very little of that technological revolution shows up in aggregate productivity statistics.[41]

Now in the light of the rise in market power, we ask what the implications are for productivity growth. For a given level of total factor productivity, an increase in market power leads to a reduction in the quantity produced and an increase in prices. Since GDP measure quantities (and price changes are factored out via CPI adjustments), we would expect to see a decline in GDP if productivity stays constant. And if productivity does increase, we expect to see an adjustment downwards of GDP, a decrease in quantities.

By its very nature, productivity at the aggregate level of the economy is unobservable. So we need a model to derive productivity as the residual after accounting for observable inputs. We use the firm's first-order condition to back out TFP. Consider our simple model of production with with labor as the only input. Then the first order condition (again) can be written

[39] If anything, the big difference between the period before and after 1980 is the higher volatility in GDP before, the great moderation.

[40] Sometimes also called the Solow computer paradox after Solow (1987) who writes "You can see the computer age everywhere but in the productivity statistics."

[41] For a review of the facts and the potential causes, see Brynjolfsson (1993)

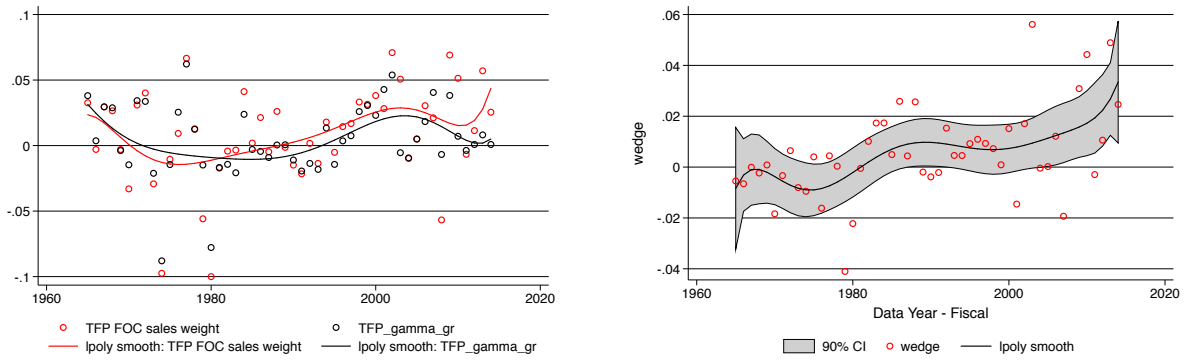
as:[42]

$$P\Omega_i L_i^{\theta-1}\theta\mu_i^{-1} = w \;\;\Rightarrow\;\; \Omega_i = \mu_i \frac{w}{P}\frac{1}{\theta}L_i^{1-\theta}. \tag{36}$$

To get a better insight, consider the case where $\theta = 1$ and the marginal product of labor constant. Then we can write productivity as:

$$\Omega_i = \mu_i \frac{w}{P}. \tag{37}$$

Productivity in the expression for Ω is equal to the real wage $\frac{w}{P}$ multiplied by the markup μ.[43] This is the adjustment mentioned before. Without market power ($\mu = 1$), productivity (Ω) equals the real wage. Now the gradual rise of market power (an increase in μ) increases productivity growth directly because μ increases. Higher market power means that we measure higher productivity than what is simply observed in patterns of real wages. The reason: the quantity produced in a market decreases as firms have market power, and as a result, for a given technology, GDP will be lower. Vice versa, if we see a level of technology, the implied productivity must be scaled up.



(a) Productivity Growth of Ω_t: data and smooth local polynomial. $\mu = 1$ (black); μ_t (red).

(b) Difference in Productivity Growth of $\Omega_t(\mu_t)$ and $\Omega_t(1)$.

Figure 15: Productivity Growth (1964 - 2014). Data: FRED, Compustat and own calculations.

In Figure 15, we use the real labor cost from aggregate data, as well as the firm's markup μ_i and its employment L_i to calculate labor productivity Ω_i at the individual firm level. We then make two aggregations. The first is the one where we use our measure for M_t as calculated above. In Figure 15a we report the growth rate of this productivity measure (in red), both the data points as well as a smooth local polynomial. The second (in black) is the measure where we set $\mu_t = 1$ to capture the notion of productivity obtained when assuming that markets are competitive (or equivalently, since we are looking at growth rates, when market power remains

[42] If in our technology we would include capital, say $Q = \Omega L^{\theta}K^{1-\theta}$, we would obtain a of productivity $\Omega = \mu\frac{w}{P}\frac{1}{\theta}L^{1-\theta}K^{\theta-1}$.

[43] This is similar as the observation of Hall (1988) using aggregate data, and De Loecker (2011) using micro data.

unchanged). We believe that productivity growth should capture the impact of markups and we find that productivity growth is higher since 1980 than when it does not incorporate the influence of changing markups. From Figure 15b it is also evident that the difference between the two measures widens.

More importantly, we find that except for a dip around the great recession, productivity actually increases and hovers around 3 to 4% after the 2000. Without accounting for the increase in market power, we find a decrease in measured productivity. This indicates that properly accounting for market power, there is no productivity slowdown but instead an increase in productivity.

5 Concluding Remarks

Using micro data on the accounts of publicly traded firms in the US starting in 1950, we find that markups have been relatively constant between 1950 and 1980 at around 20% above marginal cost. From 1980 onwards, there has been marked change in this pattern with markups steadily rising from 18% to nearly 67% in 2014, a three and a half fold increase.

We have documented the properties of this rise, which is mainly due to a change within industry, and which can be attributed mostly to the increase of markup of the firms with the highest markups already. Markups tend to be larger in smaller firms, but this seems to be due mainly to a composition effects across industries as this disappears when we decompose the markup at the fine industry level.

We use this increase to investigate the implications this has for secular macroeconomic trends in the last decades: 1. decrease in labor share; 2. decrease in capital share; 3. decrease in low skill wages; 4. decrease in labor force participation; 5. decrease in labor flows; 6. decrease in migration rates; 7. slowdown in output and GDP.

There are of course other secular trends that happen to coincide with the increase in market power: 1. decrease in startup rate of new firms[44] due to higher barriers erected by incumbents; 2. decrease in the long term interest rate[45] due to the fall in demand for capital (due to quantity reduction of firms with market power) and an increase in the supply of capital (due to higher profits); 3. the increase in wage inequality[46] from to the decrease in low skill wages as documented above together with an increase in skilled wages due to profit sharing by managers; 4. the great moderation, or the idea that output fluctuations have decreased since 1980,[47] has lost appeal since the great recession, but which is still very much observed in the data, especially in the last decade, with of course, the notable exception of 2008.

As we have shown, with incomplete pass-through the same individual firm level productiv-

[44]See amongst others Davis, Haltiwanger, Jarmin, and Miranda (2007) and Karahan, Pugsley, and Şahin (2016)

[45]See for example Caballero, Farhi, and Gourinchas (2017). This is often also linked to Greenspan's conundrum, the inability of the Federal Reserve to increase the long term Treasury yield despite sizable increases in the federal funds rate.

[46]See amongst many others Autor, Katz, and Kearney (2008).

[47]See for example Bernanke (2004) and Davis and Kahn (2008). To get a visual idea of the difference in output volatility before and after 1980, simply look at Figure 14b.

ity shocks lead to lower output adjustments. There is evidence however that the transmission of aggregate shocks is also incomplete. As a result, higher market power would translate in smaller aggregate fluctuations GDP in response to aggregate shocks as well. While we believe that each of these four secular trends are as important as, if not more important than, the ones we have analyzed, we leave them as a priority for future research.

There are two profound policy consequences that result from the rise in market power. The first relates to inflation. Relative to a scenario without a rise market power and with the same evolution of technology, prices would have gone up by 1% per year (42% over 35 years, from 1.18 to 1.67). This implies that inflation has been higher than it would have been without the rise in market power. This is surprising given the low inflation rates in the last decades – and particularly low ones since the great recession. Of course, monetary policy is not the appropriate policy tool to remedy that source of inflation. That would be the prerogative of anti-trust policy.

The second consequence pertains to the value of the stock market. Stocks prices reflect the discounted flow of dividends and thus profits. The stock market under market power is therefore overvalued relative to a competitive economy. If investors believe that the current profit level that is four times higher than in 1980 is permanent, then we would expect that the stock market capitalization would be four time higher than under perfect competition. That would put the Dow Jones at 5,500 instead of its nearly 22,000 level today. Of course, that is a naive calculation since market capitalization also reflects the extensive margin, the total value of goods sold, and that decreases with market power. How big the effect of market power is on the stock market value depends on the elasticity of demand. Note also that the stock market can go up if output deceases, for example when productivity growth is zero and market power and hence profits increase. In the presence of market power, the stock market is therefore not a good gauge of an economy's output.

While we have focused on the consequences of the rise in market power, in this paper we have remained silent on the causes. Especially when it comes to finding remedies, it is essential to understand the causes. We limit ourselves to listing a number of candidate explanations: the rise of merger & acquisition activity (Figure B.9 in the Appendix), deregulation, the higher share of network goods, the increase in wholesale transaction (Ganapati (2016)), private equity, improved product differentiation, increased vertical and financial integration of competitors. A common factor in each of these is technology. Rapid technological change allows firms to better create and preserve situations of market power.

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

Appendix A Data

Table A.1: Summary Statistics (1950-2014)

Variable	Mean	Median	Nr Obs
Sales	1,844	122	371,118
Cost of Good Sold	1,274	71	367,870
Capital stock	216	6	336,933
Cost of Good Sold/Sales	0.74	0.66	346,939

Notes: Million USD deflated using the GDP Deflator with base year 2010. The industry classification is based on the North American Industry Classification System.

Appendix B Further Evidence and Robustness

B.1 Alternative Technology Specifications

We consider measures of markups different from our benchmark reported in Figure 1. The main results are obtained by relying on a Cobb-Douglas production function with time-invariant production function coefficients. In this section we present the main facts of the paper relying on two alternative production technologies: industry-specific Cobb-Douglas productions and industry-specific production functions allowing for time varying output elasticities.

The Figure below plots both series, and we find that the pattern is very similar to that in our benchmark specification. If anything, the increase since 1980 is sharper and even more dramatic, starting at a lower level and catching up to the level in the benchmark.

Both these approaches, however, keep the output elasticities fixed over time, and given that we consider a rather unusually long sample period (1955-2014), this might be overly restrictive. Therefore we explore the sensitivity of our results and consider an industry-specific translog production function. As discussed in De Loecker and Warzynski (2012), this functional form permits time-varying output elasticities while preserving the identification results, and it constitutes a parsimonious way to generate firm/time-specific output elasticity differences, which ultimately impact the estimate of markups. The moment conditions are as before and exploit the static optimization of the variable inputs:

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

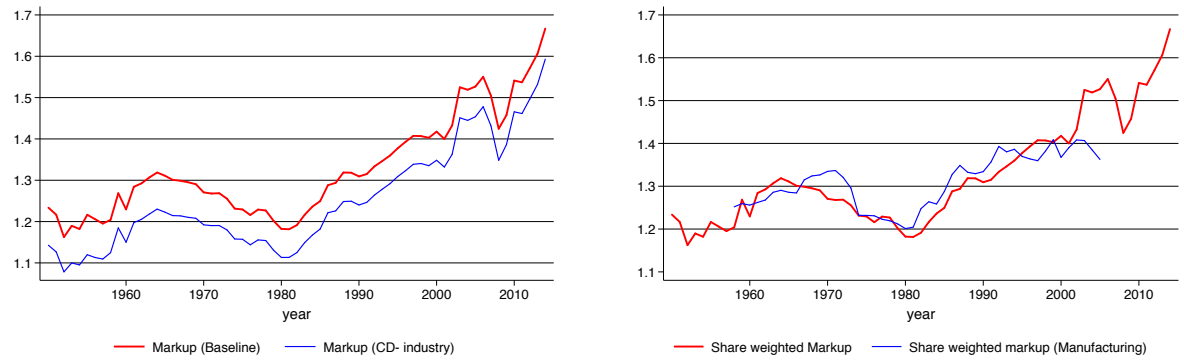
where now the vector $\boldsymbol{\beta}$ contains two additional parameters.

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

$$q_{it} = \beta_{v1} v_{it} + \beta_{k1} k_{it} + \beta_{v2} v_{it}^2 + \beta_{k2} k_{it}^2 + \omega_{it} + \epsilon_{it} \tag{B.2}$$

This yields and estimate for the output elasticity of the composite variable input: $\theta_{it}^{v} = \beta_v 1 + 2\beta_{v2} v_{it}$, and we calculate the markup as before. We do not consider the interaction term between v and k to minimize the potential impact of measurement error in capital to contaminate the parameter of most interest – i.e. the output elasticity. See Collard-Wexler and De Loecker (2016) for a discussion and implications of the presence of measurement error in capital. The results are, however, not sensitive to this omission.

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



(a) Average Markup weighted by market share - Industry Cobb-Douglas (1950 - 2014). Compustat.

(b) Average Markup in Manufacturing. NBER CES Data.

Figure B.1: Markup by Industry (Compustat Data) and for Manufacturing only (NBER CES Data).

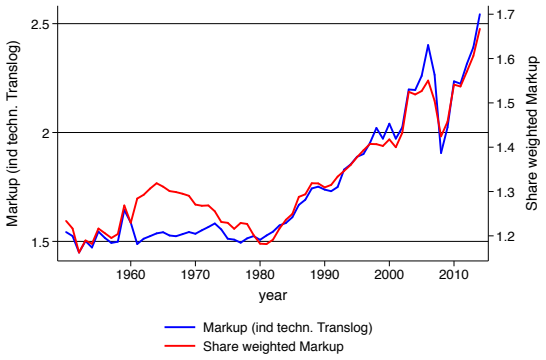
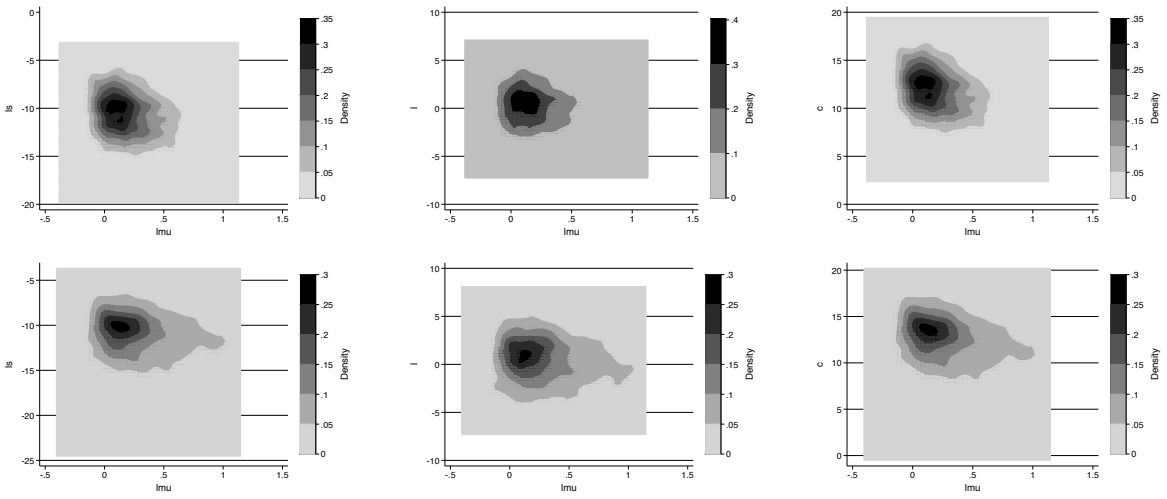


Figure B.2: Average Markup weighted by market share - Industry Translog Production Function (1950 - 2014). Data from CPS.

B.2 Alternative Aggregation Weights

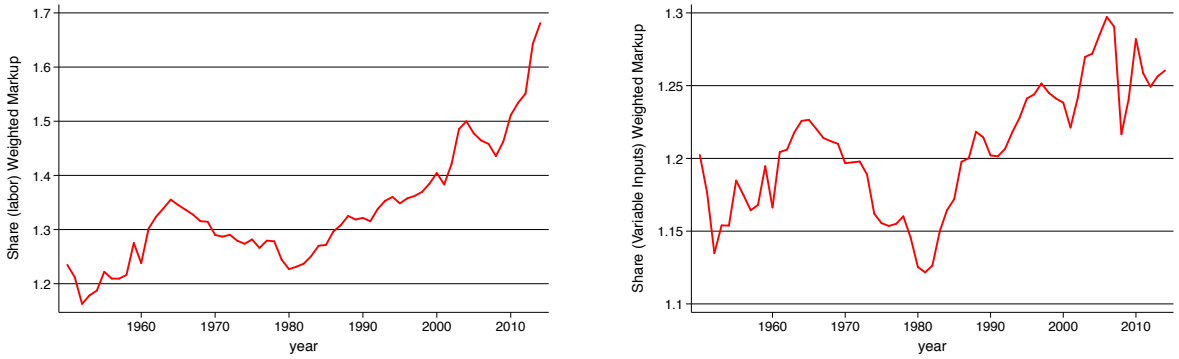
As is standard in the literature, we have calculated the average markup using the sales share of the firm in the entire sample. Different aggregation weights will affect the average. We consider employment L_{it} and the value of the variable input $P_{it}^V V_{it}$ as alternative variables to sales to calculate the weights. First in Figure B.3 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 2014. This allows us to inspect the raw data. The three variables confirm the same story: 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; markups are negatively correlated with size of sales, employment and variable inputs in the entire data set, which is evident from the fact that the ridge of the contour is negatively downward sloping.



(a) Sales. (b) Employment. (c) Value of variable input.

Figure B.3: Joint Distribution of the firm-level Markup and the Share of Sales PQ, Employment L and the value of the variable input $P^V V$. The year 1980 is on the first row, and the year 2014 on the second row.

Based on these alternative weights, in Figure B.4 we calculate the aggregate markup for each year as an alternative for our benchmark measure in Figure 1. The evolution of the average markup under these different weights follows our benchmark closely with an initial hump shape between 1950 and 1980 and a steady rise from 1980 to 2014. The level of the employment weighted markup is very similar, while the level of the value-of-input weighted markup is lower. This is as expected since the gap between sales and input value widens as the markup increases. After all, this is how markup is defined. Therefore less weight is put on high markups with input value than in with sales. Because we have seen a shift towards extreme markups, this explains the lower average with input weights.



(a) Employment.

(b) Value of variable input.

Figure B.4: Share weighted Markup: alternative weights: Employment and the Value of the variable input.

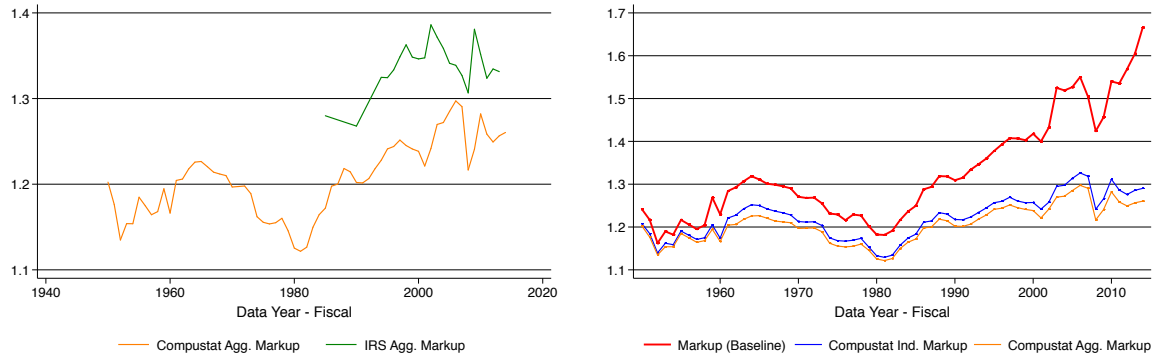
B.3 Decomposition of change in markup at different industry level aggregation

	Markup	Δ Markup	Δ Within	Δ Between	Δ Realloc.
2-digit industry					
1971	1.268	-0.028	-0.025	0.041	-0.032
1981	1.181	-0.087	-0.085	0.016	-0.017
1991	1.315	0.134	0.123	0.016	-0.006
2001	1.400	0.085	0.089	0.006	-0.010
3-digit industry					
1971	1.268	-0.028	-0.026	0.034	-0.027
1981	1.181	-0.087	-0.084	0.022	-0.025
1991	1.315	0.134	0.104	0.033	-0.003
2001	1.400	0.085	0.089	-0.003	-0.004
4-digit industry					
1971	1.268	-0.028	-0.028	0.035	-0.035
1981	1.181	-0.087	-0.089	0.026	-0.024
1991	1.315	0.134	0.090	0.045	-0.002
2001	1.400	0.085	0.061	0.011	0.010

Table B.1: Decomposition of 10 year change in Markup at different industry levels.

B.4 Markups using aggregate-level data

In the literature (first and foremost, see Hall (1988)) the analysis of markups has been performed using aggregate data. In this section we aim to achieve two goals. First, we want to perform the analysis on different data from ours that covers the entire economy, not just the



(a) IRS Economy-wide vs Compustat

(b) Aggregation within Compustat.

Figure B.5: Panel (a): Markups based on aggregates: IRS (Total Business Receipts/Total Cost of Goods sold and Compustat (Total Sales/ Total Cost of Goods Sold) and $\theta = 0.85$. Panel (b): repeat Compustat aggregate, and compare to base-line measure, M_t firm-level sales-share weighted markup, and Compustat aggregate (Compustat Ind. Markup) based on aggregate industry-wide markups, using industry sales weights.

publicly traded firms. To that end, we use IRS aggregate data and we compare the findings of the IRS data to the aggregate version of our data – i.e., summing sales and cost of goods sold over the entire Compustat data.

First, Figure B.5a plots the aggregate markup time-series for the aggregate US economy for the period (1985-2013) for which there is publicly available IRS data on Total Business Receipts and Total Cost of goods sold.[48] We compare this to the analogue in our firm-level data (Compustat), and for the period where we can compare both series, we find a very similar pattern. In fact, the level difference is almost trivial in nature since we simply calibrated the output elasticity at 0.85 for both samples.

Second, we aggregate our data to the industry level averages. We have found that most of the increase in markups can be attributed to within industry increases (see Table 1). As a result, we expect to find a more moderate increase in the markup calculated from industry averages because those within industry differences are averaged out. Figure B.5b reports two markup measures as well as our benchmark. One calculates the (industry) weighted markup based on industry average sales and industry average variable inputs in the Compustat sample. The other calculates the markup based on total sales and total inputs in the entire sample, as if we calculated one economy-wide markup measure. Those two measures show a steady increase starting in 1980, but it is less pronounced than the increase in our benchmark model where markups are calculated based on micro-level data.

[48] Own calculations based on the source: `https://www.irs.gov/uac/`, and in particular we rely on document *14a.–Corporation Income Tax Returns: Selected Balance Sheet, Income Statement, and Tax*, where we compute the aggregate markup using $0.85\times$(Total Business Receipts/Total Cost of Goods Sold).

B.5 Examples of Firms

We present the markup, sales and employment for the years 1980 and 2014, for 5 selected firms.

	Markup μ_i			Sales S_i millions (2010 $)			Empl. L_i thousands		
	1980	1990	2014	1980	1990	2014	1980	1990	2014
Google (Alphabet)			2.71			60,600			53
WalMart	1.17	1.10	1.15	3,702	48,800	444,000	27	328	2,200
Mylan	1.05	1.49	1.87	49.9	136	7,093	0.23	0.51	30
Apple	1.50	1.97	1.49	263	8,324	168,000	1	14	97
General Electric	1.19	1.45	1.71	56,200	86,500	134,000	402	298	305

Table B.2: Examples of firms: individual Markup, Sales and Employment (1980, 1990, 2014).

B.6 Industry-specific trends

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.

B.7 Additional Figure Dividend Margin

We plot a local polynomial regression of firm-level average markup (across all years) and the share of total dividends in total sales (both over the firm lifespan)

B.8 Size and Number of Firms (Compustat)

Our sample has entry and exit that varies over time. Nonetheless, we can look at the sample of firms and see how the size of firms has evolved. Consistent with the interpretation that the increase in market power stems from firms increasing market share, we find that firm size in terms of employment L is increasing, at least since the 1990s. The number of firms in the sample of publicly traded firms has steadily increased through the mid 1990s, and then has steadily fallen.

B.9 Evidence on Mergers and Acquisitions

We report the number and the average value of mergers and acquisitions as reported by Thomson Reuters - Mergers and Acquisitions. We overlay the share-weighted aggregate markup trajectory (in red).

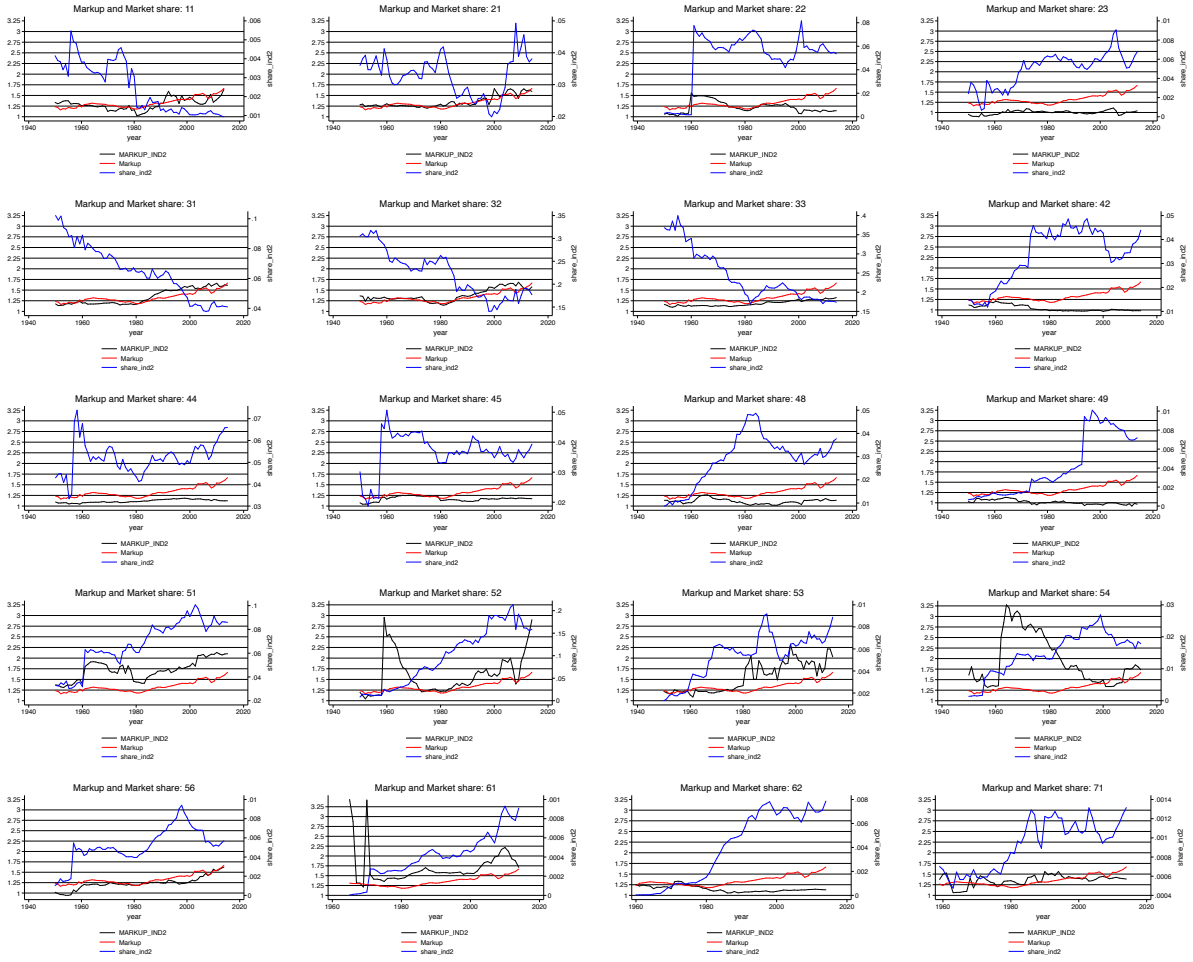


Figure B.6: Industry Specific Average Markups.

Notes. Industry weighted average markup (black) and the aggregate weighted markup (red) on the left axis, in addition to the NAICS 2digit's share in total aggregate sales (blue). NAICS code: 11 Agriculture, Forestry, Fishing and Hunting, 21 Mining, Quarrying, and Oil and Gas Extraction, 22 Utilities, 23 Construction, 31-33 Manufacturing, 42 Wholesale Trade 44-45 Retail Trade, 48-49 Transportation and Warehousing, 51 Information, 52 Finance and Insurance, 53 Real Estate and Rental and Leasing, 54 Professional, Scientific, and Technical Services, 55 Management of Companies and Enterprises 56 Administrative and Support and Waste Management and Remediation Services, 61 Educational Services, 62 Health Care and Social Assistance, 71 Arts, Entertainment, and Recreation, 72 Accommodation and Food Services.

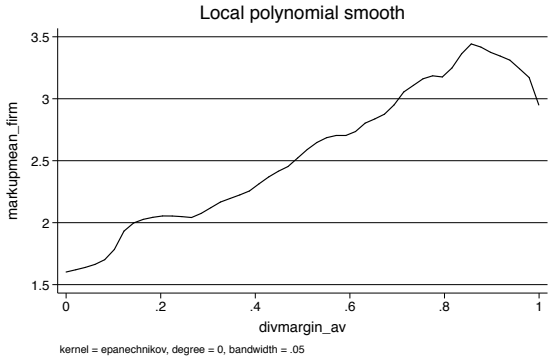


Figure B.7: The Firm-level Relation between Markup and Dividend Margin. Notes: . All firms in the sample, all years.

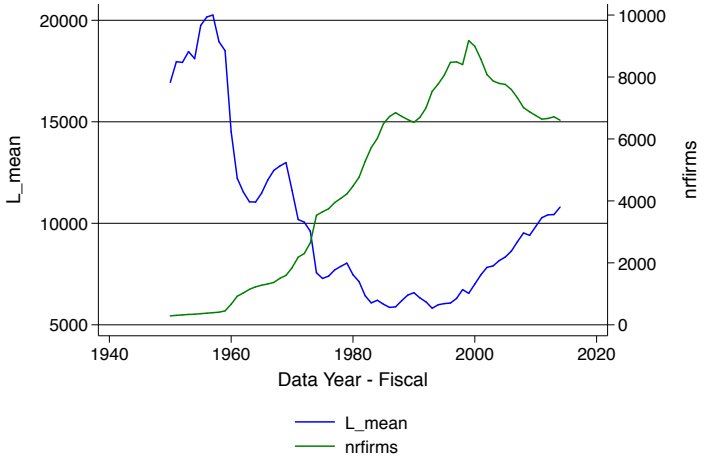


Figure B.8: Average Firm size L and Number of Firms. (1950 - 2014). Compustat data.

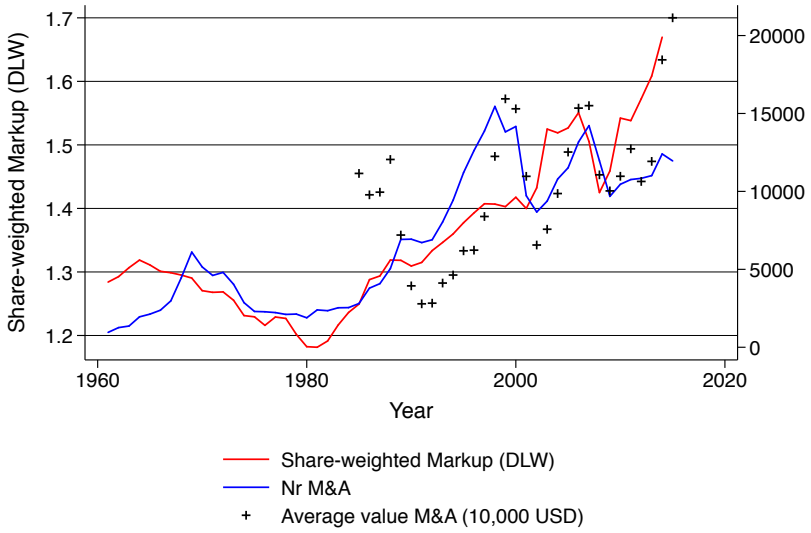


Figure B.9: Number and average value of M&A in the Compustat sample. (1950 - 2014). Notes. Compustat data.

References

ANTRAS, P. (2004): "Is the US aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution," *Contributions in Macroeconomics*, 4(1).

ARMENTER, R. (2015): "A bit of a miracle no more: the decline of the labor share," *Business Review*, Q3, 1–9.

ASKER, J., A. COLLARD-WEXLER, AND J. DE LOECKER (2017): "Market Power and Production (Mis)Allocation. A Study of the World Oil Market," mimeo.

ASKER, J., J. FARRE-MENSA, AND A. LJUNGQVIST (2014): "Corporate investment and stock market listing: a puzzle?," *Review of Financial Studies*.

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.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): "Trends in US wage inequality: Revising the revisionists," *The Review of economics and statistics*, 90(2), 300–323.

BAILY, M. N., AND B. P. BOSWORTH (2014): "US manufacturing: Understanding its past and its potential future," *The Journal of Economic Perspectives*, 28(1), 3–25.

BAQAEE, D. R., AND E. FARHI (2017): "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," Discussion paper, National Bureau of Economic Research.

BARKAI, S. (2017): "Declining Labor and Capital Shares," Chicago Booth mimeo.

BASU, S., AND J. G. FERNALD (1997): "Returns to scale in US production: Estimates and implications," *Journal of political economy*, 105(2), 249–283.

BERGER, D., AND J. VAVRA (2017): "Shocks vs. Responsiveness: What Drives Time-Varying Dispersion?," Discussion paper, National Bureau of Economic Research.

BERNANKE, B. (2004): "The great moderation," *Washington, DC*.

BRESNAHAN, T. F. (1982): "The oligopoly solution concept is identified," *Economics Letters*, 10(1-2), 87–92.

——— (1989): "Empirical studies of industries with market power," *Handbook of industrial organization*, 2, 1011–1057.

BRYNJOLFSSON, E. (1993): "The productivity paradox of information technology," *Communications of the ACM*, 36(12), 66–77.

BURDETT, K., AND K. L. JUDD (1983): "Equilibrium price dispersion," *Econometrica: Journal of the Econometric Society*, pp. 955–969.

BURNSIDE, C. (1996): "Production function regressions, returns to scale, and externalities," *Journal of monetary Economics*, 37(2), 177–201.

CABALLERO, R., E. FARHI, AND P.-O. GOURINCHAS (2017): "Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares," MIT mimeo.

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.

DAVIS, S. J., AND J. A. KAHN (2008): "Interpreting the great moderation: Changes in the volatility of economic activity at the macro and micro levels," *The Journal of Economic Perspectives*, 22(4), 155–180.

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., 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.

EDMONDS, C., V. MIDRIGAN, AND Y. XU (2015): "Competition, Markups and the Gains from International Trade," *The American Economic Review*, 105, 3183–3221.

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.

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.

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. (2016): "The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies," Yale mimeo.

GENESOVE, D., AND W. P. MULLIN (1998): "Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914," *The RAND Journal of Economics*, pp. 355–377.

GOLLIN, D. (2002): "Getting income shares right," *Journal of political Economy*, 110(2), 458–474.

GORDON, R. J. (2016): *The rise and fall of American growth: The US standard of living since the civil war*. Princeton University Press.

GUVENEN, F., G. KAPLAN, J. SONG, AND J. WEIDNER (2017): "Lifetime Incomes in the United States over Six Decades," Discussion paper, National Bureau of Economic Research.

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

HARBERGER, A. C. (1954): "Monopoly and Resource Allocation," *The American Economic Review*, 44(2), 77–87.

HOPENHAYN, H. A. (1992): "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 1127–1150.

HYATT, H. R., AND J. R. SPLETZER (2013): "The recent decline in employment dynamics," *IZA Journal of Labor Economics*, 2(1), 5.

JOVANOVIC, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, pp. 649–670.

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).

KARAHAN, F., B. PUGSLEY, AND A. ŞAHIN (2016): "Demographic Origins of the Startup Deficit," NY Fed 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.

MELITZ, M. J., AND G. I. OTTAVIANO (2008): "Market size, trade, and productivity," *The review of economic studies*, 75(1), 295–316.

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.

SOLOW, R. M. (1987): "We'd better watch out," *New York Times Book Review*, 36.

YEH, C. (2017): "Are firm-level idiosyncratic shocks important for U.S. aggregate volatility?," University of Chicago.